Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identifying of the latent causes underlying preventable major construction accidents using stacking hybrid machine learning algorithms

Tolga BARIŞIK

tolga.barisik@yeniyuzyil.edu.tr

Istanbul Yeni Yüzyıl University

Fatih YILMAZ

Istanbul Yeni Yüzyıl University

Article

Keywords: Construction accidents, Hybrid machine learning, Logistic regression, Occupational safety, Stacking learning, Stochastic gradient descent

Posted Date: May 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9349406/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Backround:

This study aims to introduce an optimized hybrid modeling approach to the causality of major construction site accidents. To examine the factors influencing the transformation of minor injuries into major injuries/fatalities using machine learning methods, and to uncover the potential latent causes underlying major construction site accidents.

Methods

A dataset consisting of 3157 accident records from different construction projects was modeled using logistic regression and stochastic gradient descent algorithms. For more reliable results, it was optimized using a stacking optimization algorithm.

Results

The dominant causes and impact levels in the transformation of minor accidents into major accidents were identified. The rate at which minor incidents transform into major accidents on construction sites is 14%. Approximately 22.5% of major accidents were avoidable or could have been resolved with minimal damage. In construction sites, the most significant types of major accidents are vehicle accidents and falls from height.

Conclusions

Machine learning methods have shown that many factors, beyond the use of personel protective equipment (PPE) or worker negligence, can have a higher impact on the probability of major accidents occurring. This study contributes to the literature by integrating stacking-based hybrid machine learning with accident causality analysis to identify preventable major accidents.

1. Introduction

The construction industry is one of the riskiest sectors in the world in terms of worker safety. Many workplace accidents resulting in serious injury or death are preventable. However, construction work involves complex processes carried out in both open and closed areas, using a wide variety of materials and equipment. Most importantly, despite advancements in technology, it is still a sector that heavily relies on human labor, and therefore, personal factors play a significant role in accidents. For this reason, as has been the subject of numerous studies in the past, research into the causality of these accidents continues today.

Construction site accidents often have multiple interconnected causes. Vivek et al. (2012) categorize accident causality theories into four generations: accident tendency theory, domino theories, injury epidemiology models, and systems theories [1]. Sequential models began with Heinrich's "domino theory" (Heinrich, 1931), which defines an accident as a series of events occurring in a specific order [2]. This model was revised and updated by Bird and Loftus (1976), establishing causality among managerial factors [3]. Other studies add human information processing models to the sequential accident approach, defining accidents in terms of human behavior and actions [4, 5].

In systemic accident causality models, accidents occur due to a complex interplay of latent conditions and active failures [6]. One of the sequential accident theories that still retains its importance today is Petersen's "accident-event-causality" theory [7]. Various hazardous situations and human behaviors in the work environment lead to near misses, followed by minor accidents, and then, in a sequential order, to major occupational accidents.

Since many studies have shown that the causes of accidents can vary between occupations, accident causality models specific to the construction sector have also been developed in recent years. These models include a range of factors, from project safety performance, control, organizational and technological characteristics, and building design, to causal links between organizational or system events and occupational events [8, 9]. However, some studies have analyzed the causes in the construction industry and reported significantly different causes and conditions for accidents with non-fatal consequences compared to those with fatal consequences [10, 11].

The new literature on accident causality in the construction industry includes some different approaches appropriate to the nature of the sector. Within the framework of developing data processing and methodological approaches in the construction sector, it is observed that integrated methodological approaches that address human and behavioral factors (age, experience, education level, vocational training, etc.) have developed due to the hazardous nature of the sector [12–18].

Data mining and machine learning approaches with large datasets have been receiving increasing attention in recent years. Shao et al. (2023) examined collapse accidents in construction using correlation rules and stated that they originate from human factors, material and machinery conditions, and management factors, and that there is a close relationship between these and the construction organization [19]. Tong et al. (2021) aimed to establish a causality model based on 2–4 correlation rules for highway construction accidents and determined the accident causality factors from four perspectives: human, material, environment, and management [20].

Halabi et al. (2022) modeled 23,057 fall accidents in the US construction industry using logistic regression with a range of variables including project type, end-use of construction, work activity, worker's age and occupation, and location of the fall [21]. A similar machine learning modeling methodology was used to predict the types of construction site accidents in Korea [22].

Tixier et al. (2016), Koc et al. (2023a), Koc et al. (2023b), and Xia et al. (2024) investigated construction work accidents using big data and Machine Learning Modeling, Random Forest Algorithm, and Particle Swarm Optimization methods [23–26].

Statistical studies aimed at optimizing various machine learning algorithms have also gained importance in recent years. Reducing the unique methodological handicaps of different algorithms, obtaining results closest to reality, and making the best choices in solving decision-making problems in various fields are crucial. Stacking is one of the most useful algorithms in this sense and has recently been used in many different fields such as agriculture, health, education, and disaster management. For example; Xuan et al. (2025) in maximizing agricultural production, Granata et al. (2023) in soil moisture detection, Zhang et al. (2024) in brain cancer research, Kalagotla et al. (2021) in diabetes prediction, Ghasemieh et al. (2023) in predicting heart disease, Ma et al. (2018) in dosage prediction, Zhou and Jiao (2022) in detecting cheating in exams [27, 33]. It has been used in water flow estimation by Lu et al. (2023), Li et al. (2023), in optimizing the anchor pulling capacity of disaster relief tents by Dou et al. (2020), in landslide risk estimation by Rahman et al. (2021), and in determining flood risks [34–37]. In addition, it has been applied in the construction field by Meharie et al. (2022) in cost estimation of road construction works [37]. Looking at the literature, it is seen that batch machine learning is most used in the health and agriculture fields, there are some applications in

disaster risk management, it is very limited in the construction field, and there is no study on occupational safety in construction.

Research confirms that workplace accidents are caused by multiple factors. Accident factors are also influenced by spatial differences. Therefore, there is still a strong need for accident modeling studies in construction. The literature lacks a large-scale data study focusing on the interaction between near misses and major accidents in construction, and the role of human-behavioral factors in relation to environmental-managerial factors within this interaction.

In this study, we aim to introduce an optimized hybrid modeling approach to the causality of major construction site accidents resulting in serious injury or death. Using various personal characteristics of employees, the work environment, and a range of managerial factors, we aim to examine the factors influencing the transformation of easily averted minor injuries into major injuries or deaths using machine learning methods. We also aim to uncover potential latent causes underlying major construction site accidents, different from what is currently known, with the lowest possible error rate.

This approach has several key objectives. Firstly, we aimed to model a dataset of 3157 records of near misses, minor injuries, serious injuries, or fatalities using two different algorithms-one linear (logistic regression) and the other nonlinear (stochastic gradient descent)-and then optimize the outputs of the previous two algorithms using a third algorithm (stacking) to ultimately create a more robust and reliable accident prediction, classification, and causality model.

Secondly, based on the information in the existing dataset, we wanted to classify all records as major accidents and minor incidents with both models, thus predicting major accidents that were actually major but predicted as minor, i.e., preventable major accidents, and to compare the results with the real dataset by determining the characteristics related to these predictions.

Thirdly, using algorithms that are quite suitable for use with categorical variables, we aimed to obtain quantitative data on the potential impact levels of variables such as age, gender, education level, professional experience, accident time, location, day, season, and accident causes on the probability of workers experiencing major accidents.

Finally, by examining the effects of various human-behavioral, environmental, and managerial factors on the severity of accidents, we aimed to gain insight into the hidden clues related to major accidents and the interactions between risk factors.

2. Methods

2.1. Data collection and design

After combining data from different projects, data refinement studies were conducted. Categories with low observed frequency but which could be included in another category were combined into a single group (e.g., in the occupation variable: technician-mechanic; in the education variable: vocational high school, vocational college). The month of the accident was converted into 3-month seasonal groups. Similarly, some categories related to the cause of the accident were combined and grouped into 13 categories. Thus, the frequencies of the subcategories of each variable were made as homogeneous and balanced as possible. Twelve variables were used in the accident model, one dependent and eleven independent. Most variables in the accident model are categorical. The age and experience duration variables, which were numerical, were also converted into categorical variables. This data refinement study aimed to facilitate the interpretation and comprehensibility of the analysis results.

As a result, a dataset consisting of a total of 3157 accident records was created. Of these, 2716 were near misses or minor accidents, and 441 were major injury and fatal accidents (3).

While the dependent variable has two categories, the independent variables have more than two categories. In the logistic regression analysis, the dependent variable, "accident outcome," was grouped into two categories: "near miss or minor injury" and "major injury and death." According to the records in the dataset, accidents resulting in the worker being unable to work for 3 days or more were categorized as major accidents; accidents resulting in no injury or injuries that rendered the worker unable to work for a maximum of 2 days were categorized as minor incidents. The independent variables included in the model are: the worker's occupation, construction site experience, education level, age, marital status, type of accident, working hours, month of accident, day of accident, cause of accident, and location of accident.

The data used in the analysis is real data created by OH&S professionals using follow-up records of accident victims who applied to the OH&S unit. The records are uniform, contain the same variables, and are maintained with the least possible error. The decision regarding the accident cause variable was determined within the framework of the reports and experiences of the OSH professionals who reported, investigated, or treated the accident.

2.2. Ethical Considerations

The dataset used in this study is publicly available and fully anonymized, containing no personally identifiable information. The data were obtained from occupational health and safety records collected for administrative and reporting purposes.

All methods were carried out in accordance with relevant guidelines and regulations. Since the study is based on secondary anonymized data and does not involve direct human participation, ethical approval and informed consent were not required.

2.3. Accident causality models and optimization

In machine learning, various algorithms are trained to find patterns and relationships between various variables in large datasets and to make the best decisions and predictions based on this analysis. Logistic Regression (LR) and Stochastic Gradient Descent (SGD) are important algorithms used for this purpose. Both algorithms can be used when the dependent variable has a categorical structure and provide reliable and interpretable results. LR and SGD also provide important information (coefficients) about the direction and strength of the influence of each subcategory of the independent variables on the dependent variable.

Stacking Optimization Algorithm (SOA), also known as Stacked Generalization, is an ensemble method in machine learning. Essentially, it aims to create a superior final model by combining the predictions of multiple models. SOA can be used to overcome the limitations of a classical model and reduce error rates.

In LR analysis, provided the dependent variable is categorical, the other independent variables in the model can be categorical, numerical, or ordinal. Therefore, logistic regression is a very useful, reliable, and easily interpretable method. SGD is also one of the most frequently used algorithms in large datasets due to its simplicity and ease of application. It is a powerful optimization algorithm frequently used in machine learning applications to find the model parameters that provide the best fit between predicted and actual outputs. Model outputs generated with both algorithms can also be used to make predictions suitable for research purposes. In this research, since the dependent variable has a binary categorical structure, binary LR and SGD algorithms were preferred. In the final stage, SOA was applied to optimize the model.

In this research, these three algorithms are used in a hierarchical (sequential) approach to predict, model, and classify construction site accidents, and to gain insight into the probability and susceptibility of workers to major accidents by comparing the resulting findings.

The following analyses were carried out in stages as part of the research:

First, a large dataset consisting of near misses, minor injuries, serious injuries, or fatalities was modeled with two algorithms—one linear (LR) and the other nonlinear (LLD)—to explore the potential effects and levels of various factors on accidents, and to calculate predictions regarding factors that may influence the transformation of minor events into major accidents based on data obtained from model outputs, and to classify accidents on construction sites.

In the second stage, the outputs of the previous two algorithms were optimized with SOA, and a more robust and reliable accident prediction, classification, and causality model was obtained.

In the third stage, based on the information in the existing dataset, all data were classified as major accidents and minor events using LR and LLD. Which major accidents, actually recorded as major but predicted as minor by both algorithms, could be prevented were predicted.

In the fourth stage, the characteristics related to these preventable accident predictions were determined and evaluated comparatively with the actual dataset.

In the fifth stage, coefficients expressing the direction and strength of the influence of all independent variable categories on the dependent variable were calculated using both LR and SGD algorithms. Thus, quantitative data were obtained regarding the probability of workers experiencing major accidents and the potential impact of variables such as age, gender, education level, professional experience, accident time, location, day, season, and accident causes.

In the final stage, based on the effects of various human-behavioral, environmental, and managerial factors on accident severity, latent patterns and interactions between risk factors related to major accidents were interpreted, and practical implications were presented.

Since including all the tables created for data presentation would be too long, some tables include categories with the highest frequency and percentage or summary information for the variables. This hierarchical hybrid modeling framework enables both predictive accuracy and interpretability, which are often difficult to achieve simultaneously in construction safety studies.

3. Discussion

The findings provide both descriptive and model-based insights into accident severity and preventability. As shown in Table 1, major accidents are predominantly associated with specific occupations, early working hours, and human-related factors.

Table 1
Characteristics of all major accidents

Occupation of worker	**n**	**%**	**Time of the accident/incident**	**n**	**%**
Unskilled worker	105	23.8	3rd hour	70	15.8
Moulder	76	17.2	1st hour	66	14.9
Plumber	44	9.9	6th hour	51	11.5
Ironworker	43	9.7	8th hour	49	11.1
Electrician	42	9.5	7th hour	44	10
Assembler	32	7.2	2nd hour	43	9.7
Plasterer-Mason	31	7	5th hour	40	9
Welder	17	3.8	9 + hour	40	9
Team leader-foreman	10	2.2	4th hour	38	8.6
Administrative staff	10	2.2	**Month of accident/incident**	**n**	**%**
Others	31	7	Winter(December-January-February)	181	41
Construction site experience of worker	**n**	**%**	Autumn(September-October-November)	91	20.6
1–5 month	228	51.7	Summer(June-July-August)	85	19.2
2–3 weeks	81	18.3	Spring(March-April-May)	84	19
6–11 month	69	15.6	**Day of the accident/incident**	**n**	**%**
1–4 years	41	9.3	Monday	84	19
0–1 weeks	19	4.3	Thursday	67	15.2
5–9 years	2	0.04	Wednesday	67	15.2
10–19 years	1	0.01	Friday	64	14.5
Worker's education level	**n**	**%**	Tuesday	64	14.5
Primary	309	70	Saturday	55	12.4
Secondary	106	24	Sunday	40	9
Vocational	14	3.1	**Factor that caused the accident/incident**	**n**	**%**
Higher	12	2.7	Not using PPE or inappropriate PPE	118	26.7
Age	**n**	**%**	Insufficient vocational or occupational safety training	80	18.1
25–39	189	42.8	Failure to comply with occupational safety rules	67	15.2
40–54	124	28.1	Hazardous speed working	43	9.7
18–24	66	14.9	Hazardous transportation, storage and stacking of loads	37	8.4

Occupation of worker	n	%	Time of the accident/incident	n	%
55–66	39	8.8	Failure to comply with work discipline	35	7.9
14–17	23	5.2	Using tools and equipment hazardously	23	5.2
Marital status	**n**	**%**	Messy-disorganized workplace or slippery-rough floor	19	4.3
Single	197	44.6	Equipment without maintenance/periodical inspection	7	1.5
Married with 2 or more children	130	29.4	No machine protection or removed from equipment	8	1.8
Married with 1 child	101	22.9	Others	4	1
Married with no children	13	2.9	**Location of the accident/incident**	**n**	**%**
Type of accident/incident	**n**	**%**	Inside the building	190	43
Falling from height	221	50.1	Outdoor area	149	33.7
Material falling or getting stuck between materials	78	17.6	Workshops	35	7.9
Compression crushing, piercing, cutting by an object	38	8.6	Roads within the construction site	22	4.9
Exposure to electric shock	28	6.3	Building exterior	17	3.8
Machine-related accidents	27	6.1	Social facilities(dining hall, dormitory etc.)	15	3.4
Vehicle accidents	27	6.1	Roof	10	2.2
Heart attack, illness, fight, suicide, etc.	7	1.5	Other	1	1
Others	15	3.4			

The overall performance metrics of the developed models are summarized in Table 2, demonstrating the comparative reliability of LR, SGD, and stacking approaches.

Table 2
Overall model significance scores for all algorithms based on various parameters

Model	AUC	CA	F1	Prec	Recall
Stack	0.984	0.959	0.839	0.914	0.776
Logistic regression	0.984	0,959	0.841	0.904	0.787
Stochastic gradient descent	0.870	0.937	0.775	0.773	0.778

When examining the overall reliability coefficients of the calculated accident prediction model, the LR algorithm is seen to have a higher overall reliability coefficient than SGD. Overall, all models created with the three algorithms show high overall reliability coefficients. Both SOA and LR have the same and quite high reliability scores for the AUC and CA parameters. For the Prec parameter, the Stacking algorithm slightly increases the reliability score. The characteristics of major accidents misclassified as minor (i.e., preventable accidents) by the optimized model are detailed in Table 3.

Table 3
Characteristics of major accidents predicted as minor incident (preventable accidents) by the optimized SGD algorithm

Occupation of worker	**n**	**%**	**Time of the accident/incident**	**n**	**%**
Unskilled worker	19	19.1	6th hour	33	33.3
Moulder	18	18.1	5th hour	24	24.1
Electrician	12	12.1	4th hour	21	21.1
Plumber	9	9	3rd hour	11	11.1
Assembler	8	8	2nd hour	7	7
Ironworker	8	8	7th hour	2	2
Plasterer-Mason	6	6	1st hour	1	1
Administrative staff	6	6	**Month of accident/incident**	**n**	**%**
Team leader-foreman	3	3	Winter(December-January-February)	32	32.3
Welder	3	3	Summer(June-July-August)	27	27.2
Others	7	7	Spring(March-April-May)	24	24.1
Construction site experience of worker	**n**	**%**	Autumn(September-October-November)	16	16.1
1–5 month	53	53.5	**Day of the accident/incident**	**n**	**%**
2–3 weeks	16	16.1	Thursday	18	18.1
6–11 month	12	12.1	Wednesday	18	18.1
1–4 years	12	12.1	Monday	17	17.1
0–1 weeks	5	5	Saturday	16	16.1
5–9 years	1	1	Friday	13	13.1
Worker's education level	**n**	**%**	Tuesday	10	10.1
Primary	72	72.7	Sunday	7	7
Secondary	21	21.2	**Factor that caused the accident/incident**	**n**	**%**
Vocational	3	3	Not using PPE or inappropriate PPE	25	25.5
Higher	3	3	Failure to comply with occupational safety rules	23	23.3
Age	**n**	**%**	Hazardous transportation, storage and stacking of loads	12	12.1
25–39	63	63.6	Failure to comply with work discipline	11	11.1
18–24	19	19.1	Hazardous speed working	9	9
40–54	17	17.1	Insufficient vocational or occupational safety training	8	8
Marital status	**n**	**%**	Using tools and equipment hazardously	7	7

Occupation of worker	n	%	Time of the accident/incident	n	%
Single	38	38.3	Messy-disorganized workplace or slippery-rough floor	2	2
Married with 2 or more children	35	35.3	Equipment without maintenance/periodical inspection	1	1
Married with 1 child	23	23.3	No machine protection or removed from equipment	1	1
Married with no children	3	3	**Location of the accident/incident**	**n**	**%**
Type of accident/incident	**n**	**%**	Inside the building	50	50.5
Falling from height	31	31.3	Outdoor area	30	30.3
Material falling or getting stuck between materials	31	31.3	Building exterior	6	6
Compression crushing, piercing, cutting by an object	21	21.1	Social facilities(dining hall, dormitory etc.)	4	4
Machine-related accidents	9	9	Roads within the construction site	3	3
Exposure to electric shock	3	3	Workshops	3	3
Other (heart attack, illness, fight, suicide, etc.)	3	3	Roof	1	1
Poisoning	1	1	Other	1	1

The rate at construction sites where minor incidents turn into major accidents is 14%. LR predicted 94 out of 441 major accidents (21.3%) as minor, while SGD predicted 98 out of 441 as minor (22.2%). Characteristics of the major accidents that occurred are given in Table 1; data on the major accidents predicted as minor using SOA optimization are given in Table 3.

According to SOA, 99% of major accidents are predicted to be minor. This means that approximately 22.5%, or one-quarter, of major accidents were preventable or could have been resolved with minimal harm. Furthermore, SOA predicts 32 incidents (7.2%) recorded as minor injuries as major injuries. In other words, these incidents could have resulted in major injury but were near missed.

In the distribution of all major accidents by occupation, unskilled laborers and molders experience the most accidents. They are followed by plumbers, blacksmiths, and electricians (Table 1). According to SOA, in the distribution of preventable major accidents by occupation, unskilled laborers and molders are again in first place. However, there is a significant change in the subsequent ranking. Electricians are in third place, followed by plumbers, assemblers, and ironworkers. It is particularly noticeable that electricians are moving up in the ranking of preventable accidents, while ironworkers are moving down.

The distribution by work experience duration shows that lack of experience is a significant factor, both generally and in preventable major accidents. Construction sites have a large number of inexperienced and low-educated workers. Therefore, the proportion of primary school graduates is naturally high in the distribution by education level. However, the results show that middle school graduates have a higher numerical impact on major accidents occurring on construction sites. Their share is over 20%, both generally and in preventable major accidents. Those classified as having secondary education include not only middle school graduates but also high school graduates. At least some

of them perform relatively less hazardous jobs than primary school graduates. These workers also have a relatively higher level of perception. Perhaps their awareness is expected to be higher. Some are mid-level managers or office workers on construction sites, but because construction sites are dynamic environments, they may also need to be in hazardous environments. Some employees who generally do not work in hazardous field jobs and perform administrative tasks such as management and supervision (such as administrative staff, team leader-foreman), despite having received training, may have lower awareness of hazards when they go out into the field. We can also say that middle-level and managerial employees are more vulnerable to accidents resulting in serious injury or death, and that occupational safety training and monitoring processes related to them are inadequate.

Similarly, in construction sites, which are heavy and dangerous jobs, workers are predominantly in the 25–39 age range, resulting in a high percentage for this group. However, there is a striking finding in the distribution of preventable major accidents compared to the overall distribution of major accidents. The 40–54 age group ranked second in major accidents. However, in preventable or avoidable accidents, the 18–24 age group ranked second, while the 40–54 age group dropped to third place. The 55–66 age group is not included in this ranking at all. Although older workers in the 40–66 age range are more vulnerable to major accidents due to factors such as physical and emotional limitations, as well as work pressure and work environment factors, we can say that workers in this age group are more successful than younger workers in avoiding major workplace accidents. This is most likely due to the experience, expertise, and safety knowledge that comes with age. Another striking conclusion drawn from this data is that 19 of the 66 major accidents that occurred in the 18–24 age group were preventable but were not. Young people are more affected and harmed by preventable accidents due to their age and inexperience. There is a difference in the distribution of preventable accidents according to marital status compared to the general distribution. While single people account for 44.6% of all major accidents, and those married with 2 or more children account for 29.4%, the share of single people in preventable accidents decreases by approximately 6 percentage points, while the share of those married with 2 or more children increases by the same amount.

When examining the distribution of preventable major accidents by accident type, there is a striking difference compared to the distribution by cause. In the general distribution, falls from height were the leading cause by a large margin. However, in preventable major accidents, falling materials and entrapment between materials have become equally important as falls from height. The proportion of accidents involving being pinched, crushed, or cut by an object, as well as machine-related accidents, has also increased from 8.6% to 21.1%. The proportion of machine-related accidents is also increasing slightly. This tells us something important. Accidents involving falling or pinching materials constitute 31% of all preventable accidents, while accidents involving being punctured, cut, or pierced by objects constitute 21%, and it is possible to prevent or overcome a large portion of these types of accidents with minimal damage. Here, managerial interventions such as a well-organized work environment, maintenance and inspection of machinery and equipment, and field supervision become crucial.

In terms of distribution by working hours, major accidents were generally more frequent during the first few hours of work. An increase was also observed towards the later hours, but the ratios were similar during the intermediate hours. The opposite was true for avoidable major accidents. The first hour had the lowest number of accidents, followed by a steadily increasing trend up to the 6th hour. It then decreased again in subsequent hours. The last working hour (7th or 8th hour depending on the situation) is usually spent on finishing (collecting) tasks, so there are fewer dangerous processes that could lead to accidents. However, each additional working hour after the first hour results in fatigue and a lack of attention and motivation. Or, as the hours progress, workers may not use certain safety precautions and equipment and may rush to finish the job. An interesting point in the data is that the hours with the fewest near misses or minor incidents were also the hours with the highest number of preventable accidents. Furthermore, the first working hour ranked second in the distribution of major accidents, but last in the distribution of

preventable major accidents. This means that the first hour of work is when the risk of a major accident related to the adaptation process is highest. However, each hour after that results in more preventable accidents, but also more injuries or damage.

Winter months account for a significant portion of the overall distribution of major accidents (41%). The percentages are similar in other seasons, around 20%. However, the impact of winter and autumn decreases in preventable major accidents, while the proportion increases in spring and summer. Although the workload is relatively lower in winter months, more serious accidents occur due to challenging climatic conditions such as wind and rain.

However, considering that construction sites operate in open climates, spring and summer months offer the most comfortable thermal conditions. Because it is neither too hot nor too cold, it is the season with the lowest risk of accidents. However, according to our results, the number of preventable major accidents also increases significantly during these periods. In fact, it appears that a significant portion of major accidents occurring during the spring and summer months, which offer the most suitable climatic conditions for construction sites, could be reduced. Major accidents occur most frequently on Mondays. The rates on other weekdays are similar, around 15%. Preventable major accidents are slightly more common on Thursdays and Wednesdays.

These variations in working hours and seasonality also demonstrate the importance of motivation and strict adherence to safety procedures in terms of being prepared to recognize potential hazards. It also shows that managers and HSE teams need to make the necessary organizational adjustments, considering the risks that these time-dependent variations may create, and to improve work discipline and the effectiveness of training and monitoring processes.

In the distribution of accident causes, lack of personal protective equipment (PPE) ranks first, both in the overall distribution and in preventable accidents. Behavioral errors such as workers' failure to comply with occupational safety rules and work discipline become more dominant. However, faulty handling, storage, or stacking of loads ranks third in preventable major accidents. Here, managerial factors such as the failure to provide appropriate handling equipment or the lack of proper maintenance of such equipment can be as influential as the workers' faulty behavior. Failure to comply with work discipline and working at dangerous speeds have also been identified as significant factors in preventable major accidents.

In the overall distribution of major accidents, both managerial and behavioral factors were found to be significant in their occurrence. In preventable major accidents, however, deficiencies in employee behavior played a dominant role. Of course, it is useful to evaluate this result together with other factors such as the distribution of accidents/incidents according to working hours, days of the week, and seasons. Because, although employee behavior seems to be the primary factor, there may be direct and indirect triggers such as the work environment, mechanics, climate, or working hours and work schedule that trigger these negative behaviors. Another important finding here is that one out of seven major accidents (14.2%) caused by poorly maintained or periodically untested equipment, seven out of 23 accidents (30.4%) caused by dangerous use of vehicles and equipment, and twelve out of 37 accidents (32.4%) caused by dangerous handling and storage operations were avoidable or could have been prevented with minor damage. As with the distribution of major accidents, indoor and outdoor areas rank first and second in terms of the locations where preventable accidents occurred. However, subsequently, facades and social facilities rank 3rd and 4th in preventable major accidents. The coefficients obtained from LR and SGD models, indicating the direction and magnitude of factor effects, are presented in Table 4.

Table 4
Logistic regression and stochastic gradient descent coefficients (ref.variable=major injury)

Logistic regression				**Stochastic gradient descent**			
Factors	**Coef (+)**	**Factors**	**Coef (-)**	**Factors**	**Coef (+)**	**Factors**	**Coef (-)**
Vehicle accidents	4.111	Object collision	-3.907	8th hour	13.426	Object collision	-9.636
Other (heart attack, illness, fight, suicide, etc.)	3.311	Falling from the same level, sliding, or friction.	-3.573	Falling from height	13.399	Falling from the same level, sliding, or friction.	-8.166
8th hour	3.002	Flying or scattered material	-3.437	Vehicle accidents	10.686	Flying or scattered material	-5.753
14–17	3.044	Hitting something	-2.972	9th hour	8.823	5th hour	-5.396
Exposure to electric shock	2.959	5th hour	-2.850	Explosion accidents	7.464	Concrete worker	-4.42
10 + hour	2.917	6th hour	-2.821	Foreign objects entering the body's natural cavities	6.703	Hitting something	-4.165
Falling from height	2.834	4th hour	-2.764	Exposure to electric shock	6.033	Exposure to welding light	-3.921
Explosion accidents	2.427	18–24	-2.511	Material falling or getting stuck between materials	6.026	Fatigue, insomnia, or low mood	-3.911
9th hour	2.330	Strain	-2.309	10 + hour	5.557	Strain	-3.628
Foreign objects entering the body's natural cavities	1.685	Concrete worker	-2.042	14–17	5.550	Carpenter or Insulation Installer	-3.411
55–66	1.591	Exposure to welding light	-1.998	55–66	4.061	Operator-Driver	-2.841
Poisoning	1.426	25–39	-1.715	Compression crushing, piercing, cutting by an object	3.931	18–24	-2.837
Roof	1.205	Fatigue, insomnia, or low mood	-1.641	7th hour	3.392	Administrative staff	-2.815

Logistic regression				Stochastic gradient descent			
Factors	**Coef (+)**	**Factors**	**Coef (-)**	**Factors**	**Coef (+)**	**Factors**	**Coef (-)**
Plumber	1.192	Exposure to harmful chemicals or radiation	-1.480	3rd hour	3.078	Roofer	-2.8
Machine-related accidents	1.183	Failure to adhere to work discipline	-1.322	6–11 month	2.942	6th hour	-2.664
Ironworker	1.144	Operator-Driver	-1.286	Plumber	2.927	5–9 month	-2.533
Winter (December-January-February)	0.861	Fire	-0.974	Ironworker	2.637	Married and childless	-2.4
Material falling or getting stuck between materials	0.859	Summer (June-July-August)	-0.813	1st hour	2.574	Poisoning	-2.331
Electrician	0.806	Primary education	-0.803	40–54	2.519	Vocational education	-2.211
Workshops	0.720	Concrete Batching Plant	-0.746	Insufficient lighting or ventilation	2.472	Exterior Facade	-2.126
1st hour	0.759	Technician-Repairman	-0.709	Lack of machine protective equipment	2.236	4th hour	-1.939
Machinery and equipment that is not maintained or has not undergone periodic inspections	0.682	3rd hour	-0.623	Machinery and equipment that is not maintained or has not undergone periodic inspections	2.117	Concrete Batching Plant	-1.765
Unskilled worker	0.621	Welder	-0.615	Machine-related accidents	2.112	Exposure to harmful chemicals or radiation	-1.646
Higher Education	0.585	İlkBahar(Mart-Nisan-Mayıs)	-0.058	Electrician	2.075	Failure to adhere to work discipline	-1.477
Carpenter or Insulation Installer	0.577	Tuesday	-0.574	Winter (December-January-February)	1.988	Summer (June-July-August)	-1.418
7th hour	0.563	2nd hour	-0.512	Friday	1.599	Wednesday	-1.33
Administrative staff	0.565	0–1 week	-0.441	10–19 years	1.392	1–5 months	-1.258

Logistic regression				Stochastic gradient descent			
Factors	**Coef (+)**	**Factors**	**Coef (-)**	**Factors**	**Coef (+)**	**Factors**	**Coef (-)**
6–11 months	0.520	40–54	-0.407	Roof	1.794	1–4 years	-1.231
2–3 weeks	0.4	Painter	-0.349	Failure to use PPE or PPE not being suitable for the job.	1.163	Insufficient vocational or occupational safety training	-1.095
Moulder	0.388	1–4 years	-0.309	Painter	1.161	25–39	-0.961
Outdoor area	0.387	Team leader-foreman	-0.291	Using hazardous tools and equipment	1.155	Welder	-0.941
Failure to use PPE or PPE not being suitable for the job.	0.374	Married and childless.	-0.276	Inside the building	1.153	Other (heart attack, illness, fight, suicide, etc.)	-0.872
Hazardous handling, storage and stacking of goods.	0.372	5–9 years	-0.227	Unskilled worker	1.087	Married with one child	-0.787
Working at hazardous speeds.	0.360	Wednesday	-0.187	Married with two or more children.	1.014	Spring (March-April-May)	-0.754
Cluttered/disorganized workplace or slippery/rough floor	0.353	Assembler	-0.167	Construction Site Roads	0.945	Social Facilities	-0.751
Insufficient vocational or occupational safety training	0.339	Compression crushing, piercing, cutting by an object	-0.143	Single	0.888	Team leader-foreman	-0.679
Construction Site Roads	0.339	Plasterer-Mason	-0.138	Secondary education (middle school-high school)	0.775	Plasterer-Mason	-0.497
Vocational education	0.324	Secondary education (middle school-high school)	-0.104	Workshops	0.726	0–1 week	-0.492
Friday	0.300	Failure to comply with workplace safety rules	-0.044	Outdoor area	0.676	Thursday	-0.467
Thursday	0.279	Married with two or more children	-0.041	2–3 weeks	0.625	Working at hazardous speeds	-0.418

Logistic regression				Stochastic gradient descent			
Factors	**Coef (+)**	**Factors**	**Coef (-)**	**Factors**	**Coef (+)**	**Factors**	**Coef (-)**
Using hazardous tools and equipment	0.250	1–5 month	-0.016	Higher Education	0.616	Hazardous handling, storage and stacking of goods.	-0.377
Inside the building	0.197	Saturday	-0.009	Failure to comply with workplace safety rules	0.541	Fire	-0.36
Monday	0.195	Sunday	-0.004	2nd hour	0.516	Moulder	-0.329
Single	0.188	Exterior Facade	-0.211	Monday	0.477	Technician-Repairman	-0.28
Insufficient lighting or ventilation	0.179			Autumn (September-October-November)	0.432	Primary education	-0.274

Both algorithms identify numerous factors that have a strong positive influence on the likelihood of major workplace accidents, meaning workers' susceptibility to serious injury or even death. Most have a coefficient of 1+. The 8th, 9th, and 10th hours of the workday are particularly sensitive to major accidents. In the overall distribution of major accidents, falls from heights and falling materials were the most common types (70%). However, according to the LR and SGD results, vehicle accidents, electrocution, explosions, and drowning are equally, or even more, significant major accident factors.

The very young (14–17 age group) and very old (55–66) age groups are more susceptible. In the overall distribution, the 25–39 and 40–54 age groups were in the top two. According to their occupation, plumbers, blacksmiths, and electricians are more prone to major accidents. In the overall distribution, unskilled laborers and molders were in the top two. Regarding the location of the accident, according to both algorithms, roofs, workshops, and construction site roads stand out as prominent locations for major accidents. In the overall distribution, 77% of major accidents occurred indoors and in open areas.

In terms of work experience, 6–11 months, 2–3 weeks, and 10–19 years are significant. However, there's a different finding here. In the overall distribution, the group with the most major accidents was the one with 1–5 months of experience, at 51.7%. Regarding education level, although they had lower coefficient values, secondary and higher education graduates were found to be significantly at higher risk.

The algorithm outputs, based on accident causes, give us significantly different results compared to general statistics. According to LR, the most important major accident factors are, in order: poorly maintained or untested equipment, lack of PPE, and hazardous load handling and storage. According to SGD, based on the coefficient value, they are ranked as follows: inadequate lighting/ventilation, lack of machine guards, and poorly maintained/uninspected machinery and equipment. The coefficient values for these are around 2.2. Lack of PPE comes in fifth place. In the causal distribution of major accidents, lack of PPE was by a large margin the most significant factor. This was followed by insufficient training, failure to comply with occupational safety rules, and working at dangerous speeds.

Results

Research into the causality of construction site accidents using deep learning or machine learning approaches has gained attention in recent years. Traditional approaches are insufficient for optimizing errors in output data. The construction sector, due to its unique and labor-intensive nature, is a high-risk sector where human factors play a significant role in workplace accidents. Determining the level of influence of various factors on accident causality in construction sites is not easy. Therefore, studies on accident causality will continue to be important until the most reliable results are obtained.

In this study, we hierarchically and integratedly examined major accidents resulting in serious injuries at construction sites using three different optimized machine learning approaches. We predicted the causes of major accidents and identified those that could have been avoided using our algorithms. We also investigated the causes of avoidable accidents and determined the factors that may influence workers' susceptibility to major accidents.

The rate at which minor incidents escalate into major accidents on construction sites is 14%. According to SOA results, 99 out of 441 major accidents recorded as major were predicted to be minor. This means that approximately 22.5%, or one-quarter, of major accidents were preventable or could have been resolved with minimal damage. We found that electricians, those with only a secondary school education, mid-level managers, those aged 18-24, and married individuals with two or more children were at higher risk of preventable major accidents.

In the overall distribution by accident type, falls from height were by a large margin in first place. However, in preventable major accidents, falling materials and entrapment between materials were found to be as significant as falls from height. The proportion of accidents caused by being pinched, crushed, cut, or machine-related accidents also increased from 8.6 to 21.1. Accidents involving falling or pinching materials accounted for 31% of all preventable accidents, while accidents involving being punctured, cut, or pierced by objects accounted for 21%.

While major accidents occur more frequently during the first and last hours of work, preventable major accidents are more common during average hours. In preventable major accidents, faulty handling, storage, or stacking is as significant a cause of accidents as lack of personal protective equipment (PPE) and non-compliance with occupational safety regulations. 14.2% of accidents caused by poorly maintained/uninspected equipment, 30.4% of accidents caused by dangerous equipment use, and 32.4% of accidents caused by dangerous handling and storage are preventable or averted major accidents. In terms of accident locations, exterior facades and social facilities are significant factors in preventable major accidents.

In general, it is observed that the coefficient values of factors influencing major accidents, as calculated by SGD, are higher than those calculated by LR. Both algorithms calculate the highest coefficients for the factors that are considered influential and significant in major accidents, and these factors are largely the same. However, when ranked from highest to lowest according to coefficient values, significant differences are observed in the ranking of the two algorithms. Furthermore, the number of factors with more than one coefficient value calculated according to the SGD algorithm is 34, while this number is 16 in LR. In our research, SGD is a stronger algorithm in terms of factor discrimination and classification.

According to official data and numerous studies, falls from height are the leading cause of major injuries in construction accidents. Our study also found it to be a significant cause, but not the primary one. According to LR (Learning Lottery), vehicle accidents are the most common cause. This is followed by other causes such as heart attacks, strokes, and suicides at the workplace, the time of day the accident occurs (8th hour), the worker's age (14-17 age group), and then electrocution. As can be seen, six different accident cause factors were calculated to be

more important than falls from height. In SGD (Student Selection and Placement), the time of day the accident occurs (8th hour) is again the most common cause, followed by falls from height, and then vehicle accidents.

Some major accident factors stand out in both algorithms, although their ranking differs. These include vehicle accidents, electric shocks, explosions, the 14-17 age group, and the 8th hour of work. On the other hand, heart attacks and fights are very strong major accident causes in LR, while in SGD they are negatively significant, meaning they are strongly considered minor accident causes. Poisoning is also a factor. These are rare occurrences in construction sites, but when they do happen, they can result in serious injury, loss of work days, or death.

Lack of PPE is not the primary cause of accidents. Lack of machine guarding, poor maintenance of machinery and equipment, inadequate lighting and ventilation, and dangerous equipment use are also equally significant factors in major accidents as a lack of PPE. According to both algorithms, employees with higher education degrees are more likely to be involved in major accidents.

Furthermore, according to both algorithms, accident time, accident type, worker's age and occupation are more significant factors in the occurrence of major accidents than other accident factors (accident cause, accident location, accident season, accident day, worker's experience, education level and marital status). This shows us that in accident prevention efforts, accident time, worker's age and occupation should be given at least as much importance as accident cause and education.

We found that accident times are significant in minor-major accident transitions and need to be carefully evaluated. This means that it is important to distribute the effectiveness of managerial control activities evenly throughout the shift.

Construction workers in management positions and those with high levels of education are at a higher risk than expected of major accidents. The effectiveness of safety training needs to be improved to adapt to changing work environments and technological conditions. Managers' risk awareness must be increased.

Vehicle accidents were the dominant major accident type with a high coefficient in both algorithms. However, operator-drivers were more likely to be involved in minor accidents. In serious and fatal vehicle accidents on construction sites, workers in the surrounding area are affected more than drivers and operators.

The causality of major accidents and minor incidents may differ, but they interact. Therefore, it can be beneficial to examine them separately using different methods and consider these outcomes in construction site risk management.

The impact of personal-behavioral barriers on accident causation is relatively decreasing but still persists; conversely, the positive impact of proactive managerial interventions, particularly in preventing major accidents, is becoming more visible than before. Therefore, appropriate organizational arrangements prioritizing occupational safety on construction sites remain important.

In conclusion, we see that the most significant causes of major or preventable major workplace accidents on construction sites are not limited to falls from height, failure to use PPE, or worker negligence. Machine learning methods show that many other factors may have a much greater impact on the probability of major accidents than is currently known. The causality of construction site accidents may involve interactions far different from what current statistics reveal. There may be hidden patterns within accident causality that are not visible. Machine learning methods can reveal the unseen dimensions of major construction site accidents. Instead of prescriptive approaches, identifying specific problem sources can yield more effective results. Large datasets and machine learning

approaches can provide important information about the interaction patterns between various environmental, mechanical, and human factors in the construction sector and the strength of these interactions.

The study has unique aspects as well as limitations. First, the data is limited to variables recorded at construction sites. Second, the data covers only one country except for a few projects abroad. Third, the observed frequencies in some variable categories are low. However, we tried to reduce these problems to some extent by using a large data set. The proposed stacking-based framework offers a data-driven decision support tool for proactive safety management in construction environments.

Declarations

Ethical approval and consent to participate

Not applicable. This study is a non-intervention research. Research data is based on anonymous and publicly available statistical records determined according to criteria required for reporting to official institutions under local legislation in countries where construction projects are carried out. Data was not obtained from human experiments or individual surveys, and ethical committee approval was not required.

Data availability statement

The dataset supporting the findings of this study is publicly available in the Mendeley Data repository:

Yilmaz, Fatih; Barisik, Tolga (2025), “Data of Transformation of Minor Incidents into Major Accidents in Construction”, Mendeley Data, V1.

https://doi.org/10.17632/xvr6cgg34y.1

The dataset includes anonymized demographic and occupational characteristics of workers, as well as information on accident type, location, time, and causes. All data are fully anonymized and publicly accessible without restriction.

Declaration of competing interests

The authors have nothing to declare

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions:

Auhor 1: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft

Author 2: Conceptualization, Data curation, Methodology, Project administration, Software, Supervision, Writing – review and editing

References

1. Vivek V. Khanzode, J. Maiti, P.K. Ray. Occupational injury and accident research: A comprehensive review. Safety Science, Volume 50, Issue 5, June 2012, Pages 1355-1367. https://doi.org/10.1016/j.ssci.2011.12.015
2. Heinrich, H. (1931) Industrial Accident Prevention: A Scientific Approach, McGraw-Hill, New York.
3. Bird, F.E. and Loftus, R.G. (1976). Loss Control Management, Institute Press, Loganville, GA.
4. Gordon, R., Flin, R., Mearns, K. (2005). Designing and evaluating a human factors investigation tool (HFIT) for accident analysis, Safety Science, Volume 43, Issue 3, Pages 147-171, https://doi.org/10.1016/j.ssci.2005.02.002.
5. Katsakiori, P., Sakellaropoulos, G. and Manatakis, E. (2009). Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. Safety Science, 47(7), 1007–15.
6. Reason, J., Hollnagel, E., Paries J. (2006). Revisiting the Swiss cheese model of accidents. Journal of Clinical Engineering, 27 pp. 110-115.
7. Petersen, Dan., (1984). Human-Error Reduction and Safety Management, Aloray, New York.
8. Huang, X. and Hinze, J. (2006). Owner's role in construction safety. Journal of Construction Engineering and Management,132(2), 164–73.
9. Behm, M. and Schneller, A. (2013). Application of theLoughborough Construction Accident Causation model: a framework for organizational learning. Construction Management and Economics, 31(6), 580–95.
10. Li, Y. and Bai, Y. (2008). Comparison of characteristics between fatal and injury accidents in the highway construction zones. Safety Science, 46(4), 646–60.
11. Gibb, A., Lingard, H., Behm, M., & Cooke, T. (2014). Construction accident causality: learning from different countries and differing consequences. Construction Management and Economics, 32(5), 446–459. https://doi.org/10.1080/01446193.2014.907498
12. Mohammadi, A., Tavakolan, M and Khosravi, Y. (2018). Factors influencing safety performance on construction projects: A review, Safety Science, Volume 109, Pages 382-397, https://doi.org/10.1016/j.ssci.2018.06.017.
13. Yılmaz, F. (2023). "Analysis of the interaction between safety measures and workers' safety awareness from the construction workers' perspective", Engineering, Construction and Architectural Management, Vol. 30 No. 1, pp. 41-55. https://doi.org/10.1108/ECAM-07-2021-0564.
14. Gokler, C.A., Kilic, H. S., Gao, S., & Deliveli, N. (2024). Axiomatic design for safe construction considering lean and ergonomic principles: an application in Turkey. International Journal of Occupational Safety and Ergonomics, 1–18. https://doi.org/10.1080/10803548.2024.2371191
15. Mohandes, S.R., Sadeghi, H., Fazeli, A., Mahdiyar, A. Hosseini, M.R., Arashpour, M. and Zayed, T. (2022). Causal analysis of accidents on construction sites: A hybrid fuzzy Delphi and DEMATEL approach, Safety Science, Volume 151, 105730, https://doi.org/10.1016/j.ssci.2022.105730.
16. Cal, M. and Yilmaz, F. (2025). Evaluation of the factors determining thermal comfort and occupational health conditions of employees in multi-storey business buildings in the light of expert experiences, Bulletin of the Polish Academy of Sciences Technical Sciences, 73, 1 e152215, DOI: 10.24425/bpasts.2024.152215
17. Abbasianjahromi, H., & Etemadi, A. (2019). Applying social network analysis to identify the most effective persons according to their potential in causing accidents in construction projects. International Journal of Construction Management, 22(6), 1065–1078. https://doi.org/10.1080/15623599.2019.1683688
18. Elsebaei, M., Elnawawy, O., Othman, A. A. E., & Badawy, M. (2020). Causes and impacts of site accidents in the Egyptian construction industry. International Journal of Construction Management, 22(14), 2659–2670. https://doi.org/10.1080/15623599.2020.1819523

19. Shao, L., Guo, S., Dong, Y., Niu, H. and Zhang, P. (2023), "Cause analysis of construction collapse accidents using association rule mining", Engineering, Construction and Architectural Management, Vol. 30 No. 9, pp. 4120-4142. https://doi.org/10.1108/ECAM-11-2021-0991
20. Tong, R., Zhao, H., Zhang, N., Li, H., Wang, X. and Yang, H. (2021), "Modified accident causation model for highway construction accidents (ACM-HC)", Engineering, Construction and Architectural Management, Vol. 28 No. 9, pp. 2592-2609. https://doi.org/10.1108/ECAM-07-2020-0530
21. Halabi, Y., Xu, H., Long, D., Chen, Y. Yu, Z., Alhaek, F. and Alhaddad, W. (2022). Causal factors and risk assessment of fall accidents in the U.S. construction industry: A comprehensive data analysis (2000–2020), Safety Science, Volume 146, 105537, https://doi.org/10.1016/j.ssci.2021.105537.
22. Kang, K. and Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model, Safety Science, Volume 120, Pages 226-236, https://doi.org/10.1016/j.ssci.2019.06.034.
23. Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B. and Bowman, D. (2016). Application of machine learning to construction injury prediction, Automation in Construction, Volume 69, Pages 102-114, https://doi.org/10.1016/j.autcon.2016.05.016.
24. Koc, K., Budayan, C., Ekmekcioğlu, Ö. and Tokdemir, A.B. (2023a). Predicting Cost Impacts of Nonconformances in Construction Projects Using Interpretable Machine Learning, Journal of Construction Engineering and Management, Volume 150, Issue 1, https://doi.org/10.1061/JCEMD4.COENG-13857
25. Koc, K., Ekmekcioğlu, Ö. and Gurgun, A.P. (2023b). Developing a National Data-Driven Construction Safety Management Framework with Interpretable Fatal Accident Prediction, Journal of Construction Engineering and Management, Volume 149, Issue 4 https://doi.org/10.1061/JCEMD4.COENG-12848
26. Xia, X., Xiang, P., Khanmohammadi, S., Gao, T. and Arashpour, M. (2024). Predicting Safety Accident Costs in Construction Projects Using Ensemble Data-Driven Models, Journal of Construction Engineering and Management, Volume 150, Issue 7, https://doi.org/10.1061/JCEMD4.COENG-14397
27. Fu Xuan, Wei Su, Zhen Chen, Xianda Huang, Weiguang Zhai, Xuecao Li, Yelu Zeng, Zhi Li, Jingsuo Li, Jianxi Huang, (2025). Performance of stacking machine learning and volume model for improving corn above ground biomass prediction, Plant Phenomics,Volume 7, Issue 3, 100068, https://doi.org/10.1016/j.plaphe.2025.100068.
28. Granata, F.; Di Nunno, F.; Najafzadeh, M.; Demir, I. A. (2023). Stacked Machine Learning Algorithm for Multi-Step Ahead Prediction of Soil Moisture. *Hydrology*, *10*, 1. https://doi.org/10.3390/hydrology10010001
29. Zhang, Hw., Wang, Yr., Hu, B. *et al.* (2024). Using machine learning to develop a stacking ensemble learning model for the CT radiomics classification of brain metastases. *Sci Rep* 14, 28575 https://doi.org/10.1038/s41598-024-80210-x
30. Satish Kumar Kalagotla, Suryakanth V. Gangashetty, Kanuri Giridhar. (2021). A novel stacking technique for prediction of diabetes, Computers in Biology and Medicine, Volume 135, 104554, https://doi.org/10.1016/j.compbiomed.2021.104554.
31. Alireza Ghasemieh, Alston Lloyed, Parsa Bahrami, Pooyan Vajar, Rasha Kashef. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients, Decision Analytics Journal, Volume 7, 100242, https://doi.org/10.1016/j.dajour.2023.100242.
32. Ma Z, Wang P, Gao Z, Wang R, Khalighi K. (2018). Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. PLOS ONE 13(10): e0205872. https://doi.org/10.1371/journal.pone.0205872
33. Zhou T, Jiao H. Exploration of the Stacking Ensemble Machine Learning Algorithm for Cheating Detection in Large-Scale Assessment. Educ Psychol Meas. 2023 Aug;83(4):831-854. doi: 10.1177/00131644221117193. Epub 2022 Aug 13. PMID: 37398846; PMCID: PMC10311957.

34. Lu, M.; Hou, Q.; Qin, S.; Zhou, L.; Hua, D.; Wang, X.; Cheng, L. A. (2023). Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. Water, 15, 1265. https://doi.org/10.3390/w15071265
35. Li, L., Zuo, L., Wei, G. et al. A stacking machine learning model for predicting pullout capacity of small ground anchors. AI Civ. Eng. 3, 11 (2024). https://doi.org/10.1007/s43503-024-00032-8
36. Dou, J., Yunus, A.P., Bui, D.T. et al. (2020). Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides 17, 641–658 https://doi.org/10.1007/s10346-019-01286-5
37. Mahfuzur Rahman, Ningsheng Chen, Ahmed Elbeltagi, Md Monirul Islam, Mehtab Alam, Hamid Reza Pourghasemi, Wang Tao, Jun Zhang, Tian Shufeng, Hamid Faiz, Muhammad Aslam Baig, Ashraf Dewan. (2021). Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh, Journal of Environmental Management, Volume 295, 113086, https://doi.org/10.1016/j.jenvman.2021.113086.
38. Meharie MG, Mengesha WJ, Gariy ZA, Mutuku RN. (2022), "Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects". Engineering, Construction and Architectural Management, Vol. 29 No. 7 pp. 2836–2853, doi: https://doi.org/10.1108/ECAM-02-2020-0128