Heliyon 10 (2024) e32215

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A deep neural network algorithm-based approach for predicting recovery period of accidents according to construction scale

Ji-Myong Kim [a], Manik Das Adhikari [b], Junseo Bae [c], Sang-Guk Yum [b,*]

[a] Department of Architectural Engineering, Mokpo National University, Mokpo, 58554, South Korea
[b] Department of Civil and Environmental Engineering, Gangneung-Wonju National University, Gangneung, 25457, South Korea
[c] Division of Smart Cities, Korea University, Sejong, 30019, South Korea

ARTICLE INFO

Keywords:
Construction accident
Accident recovery period
Construction scale
Deep learning algorithm
Small-to-medium sized construction site
Large construction site

ABSTRACT

Despite ongoing safety efforts, construction sites experience a concerningly high accident rate. Notwithstanding that policies and research to reduce the risk of accidents in the construction industry have been active for a long time, the accident rate in the construction industry is considerably higher than in other industries. This trend may likely be further exacerbated by the rapid growth of large-scale construction projects driven by urban population expansion. Consequently, accurately predicting recovery periods of accidents at construction sites in advance and proactively investing in measures to mitigate them is critical for efficiently managing construction projects. Therefore, the purpose of this study is to propose a framework for developing accident prediction models based on the Deep Neural Network (DNN) algorithm according to the scale of the construction site. This study suggests DNN models and applies the DNN for each construction site scale to predict accident recovery periods. The model performance and accuracy were evaluated using mean absolute error (MAE) and root-mean-square error (RMSE) and compared with the widely used multiple regression analysis model. As a result of model comparison, the DNN models showed a lower prediction error rate than the regression analysis models for both small-to-medium and large construction sites. The findings and framework of this study can be applied as the opening stage of accident risk assessment using deep learning techniques, and the introduction of deep learning technology to safety management according to the scale of the construction site is provided as a guideline.

1. Introduction

As a 3D industry, the construction industry is often associated with a perception that it experiences a higher frequency of accidents compared to other industries. According to the status of occupational accidents in 2021 reported by the Korea Labor Administration,[1] the number of injured (injured, dead, or sick due to occupational accidents or diseases) occurred in the order: other businesses (37 %) > manufacturing (25.8 %) > the construction industries (24.4 %). However, the number of deaths occurred in the order: the construction industry (26.5 %) > manufacturing industry (24.6 %) > other businesses (22.2 %), demonstrating that the construction

* Corresponding author. Sang-Guk Yum
E-mail addresses: jimy6180@gmail.com (J.-M. Kim), rsgis.manik@gmail.com (M.D. Adhikari), junseobae@korea.ac.kr (J. Bae), skyeom0401@gwnu.ac.kr (S.-G. Yum).

[1] https://www.moel.go.kr/policy/policydata/view.do?bbs_seq = 20220300882 (accessed July 20th, 2022).

https://doi.org/10.1016/j.heliyon.2024.e32215
Received 23 October 2023; Received in revised form 25 May 2024; Accepted 29 May 2024
Available online 31 May 2024

industry shows a higher severity of disasters than other industries. The main causes of accidents in the construction industry were various, such as collapse, jamming, crushing, being hit by objects, overturning, and bumping. In particular, it was reported that the highest number of deaths in the construction industry occurred in workplaces with fewer than five employees (231 people). Analysis of the number of accident victims (the total number of injuries and deaths caused by occupational accidents) shows a similar trend to the number of casualties. The number of accident victims was high in other businesses (38.9 %), construction (26.3 %), and manufacturing industries (23.7 %). However, the accident fatalities were found in construction (50.4 %), manufacturing (22.2 %), and other businesses (14.9 %). The death toll in the construction industry was 2.3 times higher than that in the manufacturing industry, which had the second highest rate possible.

As shown in the previous statistics, the construction industry is exposed to more safety accidents than other industries due to the nature of the industry [1–5,[1]]. To summarize some of these characteristics, first, in the work environment, the construction industry is directly exposed to the external environment due to the specificity of the work environment and thus is greatly affected by outdoor work, weather, and topography. In addition, the work has many risks, such as the repetitive assembly and dismantling of temporary structures and the handling of many construction machines and heavy objects [2]. Second, in the internal environment of the construction industry, the inequity and unilateral nature of the construction contract, such as the low construction price according to the lowest bid schedule, the construction order period, and the lack of a construction period, make the construction industry vulnerable to safety accidents [3,4]. Third, there is a change in the technological environment due to the rapid influx of new technologies and new construction methods introduced in the construction industry due to the high-rise complex of buildings, construction in the city center, etc. Thus, insufficient safety measures, cost-cutting, shortened construction periods, and neglect of safety considerations contribute to increased accidents in the construction industry [5]. The transient nature of construction work, insecure employment, and limited safety education opportunities lead to a lack of safety awareness among workers [3,6]. Factors such as long, unpredictable working hours and fatigue from continuous work further exacerbate safety concerns. Additionally, challenges such as an aging workforce and the influx of foreign workers hinder the development of a safety culture in the construction industry. These factors collectively elevate the risk of accidents in construction compared to other industries. Consequently, this study addresses the heightened risk of accidents in the construction industry, particularly emphasizing the significance of construction site scale. Recent data from the Korea Labor Administration highlights the high incidence of accidents in the construction industry, revealing an alarming situation where the number of casualties is disproportionately higher in small- and medium-sized construction sites compared to large construction sites. These small- and medium-sized construction sites often lack adequate safety management structures, leading to an increased accident rate. This alarming situation of construction sites at different scales necessitates a focused investigation into the contributing factors. While existing research has made progress in safety management, there remains a critical gap in tailored models that account for the unique challenges faced by construction sites of varying sizes. Therefore, there is an urgent requirement to study the prevention of accidents at construction sites by analyzing problems in the construction industry and deriving improvement plans to reduce accidents in that industry. The present study aims to fill this gap by proposing a model leveraging advanced deep learning techniques to predict and analyze recovery periods (i.e., the total number of treatment days due to accidents) of accident occurrences, considering the distinct characteristics of the construction site scale. The ultimate goal is not only to advance our understanding of these risks but to significantly contribute to reducing accidents within the construction sector, specifically emphasizing enhancing safety measures for small- and medium-sized construction projects.

2. Background of the study and literature review

2.1. Background of the study

In many studies, small-to-medium-sized construction sites are considered as a major cause of accidents at construction sites. Table 1 shows the number of accidents by construction site size that occurred between 2017 and 2020, according to the Korean National Statistical Office in 2020.[2] Although the number of accident victims by year is slightly different, it can be seen that 85–90 % of the number of casualties occur in small-to-medium-sized sites, while the rest occur in large-scale sites. Thus, most casualties occur in small-to-medium-sized construction sites.

Furthermore, Table 2 shows that the casualty rate by construction size was 6.02 % for small-to-medium-scale construction sites and 2.2 % for large construction sites, indicating that the casualty rate in small-to-medium-sized construction sites is about 3.82 times higher. This clearly confirms that the number of accidents and the accident rate are significantly greater in small-to-medium-sized construction sites than in large construction sites.

As the previous statistical data shows, small-to-medium-sized construction sites are more vulnerable to accidents than large construction sites. The reasons can be summarized as follows. First, compared to large-scale construction sites (construction costs more than 12 billion KRW), small-to-medium-sized construction sites (construction costs less than 12 billion KRW) are relatively lacking in safety management organization and technical systems. In particular, under the current safety management system, small- and medium-sized construction sites are classified as worksites that are subject to technical guidance by an accident prevention expert guidance organization, while large construction sites are classified as worksites that are subject to the appointment of safety managers. In other words, small- and medium-sized construction sites with a construction cost of less than 12 billion KRW are not obliged to

[2] https://kosis.kr/statHtml/statHtml.do?orgId=118&tblId=DT_118006_001&vw_cd=MT_ZTITLE&list_id=C_14_001&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE (accessed July 20th, 2022).

Table 1
Number of accidents by construction scale over the period from 2017 to 2020.

Construction Scale	2017	2018	2019	2020
	Number of casualties (%)	Number of casualties (%)	Number of casualties (%)	Number of casualties (%)
Medium & small	46,817 (89.4 %)	52,114 (87.6 %)	57,523 (85.7 %)	72,063 (85.5 %)
Large	4947 (10.6 %)	7364 (12.4 %)	9624 (14.3 %)	10,447 (14.5 %)

Table 2
Casualties rate by construction scale.

Size of Company	Total number of employees[a]	Number of casualties[b]	Casualty rate[c]
Medium & small	1,197,564 (71.7 %)	72,063 (85.5 %)	6.02 %
Large	473,834 (28.3 %)	10,447 (14.5 %)	2.2 %

[a] Total number of construction workers in 2020, according to Statistics Korea.
[b] Number of casualties in the construction industry as of 2020, according to Statistics Korea.
[c] Total number of casualties/number of employees.

appoint a safety manager. This shows that the absence of a safety management organization has a significant impact, as the number of accidents occurring at construction sites with the cost of less than 12 billion KRW, which sites are subject to technical guidance by an accident prevention expert guidance organization, is much higher than that at sites subject to the appointment of safety managers. The absence of an on-site safety organization leads to non-implementation of safety training, non-compliance with basic safety rules, and prior safety inspection, which raise the incidence of accidents [7–9]. Second, small- and medium-sized construction sites are fairly economically coarse compared to large-scale construction sites. Small- and medium-sized construction sites often have shorter construction periods and lower construction costs, prioritizing efficiency and speed, which can contribute to a higher incidence of accidents. Moreover, this economic inferiority leads to insufficient, or avoidance of, investment in safety, deteriorating the work environment of small- and medium-sized construction sites. In addition, safety management for accident prevention is insufficient since the investment necessary for safety, such as specialized technology, equipment, and functions related to safety, is not well made. Furthermore, the short construction period of small- and medium-sized construction sites makes it difficult to understand the construction status, making it challenging to invest in improving the construction environment [10,11]. Third, these sites lie in the blind spot of safety management through policy support and institutional regulation of small- and medium-sized construction sites. For example, the Severe Accident Punishment Act, which was enforced in 2022 to broaden the scope and subject of safety and health obligations and significantly strengthen the severity of punishment, excludes small-scale construction sites (i.e., less than five full-time workers). In addition, small-scale construction companies (outside the top 1000 in construction capacity) have no disadvantage in bidding even in the event of an accident, so the safety and health awareness of business owners and other managers is unsatisfactory, and they are working formally on the use of occupational safety and health management expenses [7,11,12].

2.2. *Review of construction site disaster risk analysis and construction risk quantification model*

Despite various efforts to prevent accidents in the construction industry, accidents are becoming more numerous and are centered on small- and medium-sized sites. Thus, it is impossible to moderate the total number of accidents in the construction industry without reducing safety accidents at small- and medium-sized construction sites [10]. To minimize accidents effectively, proactive investments in risk reduction are crucial. This necessitates the use of advanced predictive models to identify and mitigate potential accident risks, making them an essential component of safety measures. This accident risk prediction can be utilized to prevent and reduce accident risk by identifying the types and amounts of accidents that can occur, eventually ensuring the continuity of the construction project and affecting the increase in profits. These measures can further underwrite the reduction of accidents in the construction industry. Nevertheless, as discussed above, although small and large construction sites have marked differences in characteristics, studies and models for quantifying construction accident risk considering these differences are deficient.

The accident risk analysis of a construction project manages potential risks at the construction site by preventing and dropping risks in advance and finally decreases the risk of accidents. This can eventually lead to a successful project by reducing the risk of accidents. Additionally, prediction through accident risk analysis can be adopted as a yardstick for strategically allocating and executing limited resources in a construction project commendably [13,14]. Consequently, sophisticated accident risk analysis is crucial for fruitful and sustainable construction projects. The reason is that, in accident risk assessment, the opinions or experiences of experts, engineers, and clients are often utilized, and subjective indicators that are problematic to index, other than opinions or experiences, are included in accident risk assessment [15,16]. For example, qualitative methodologies, such as surveys or checklists based on construction-related experts' or engineers' knowledge and experience, are repeatedly used for accident risk [17]. The reason is that to use a quantitative methodology, securing reliable data is the most significant factor, but due to the limitation of collecting reliable data, a qualitative methodology is sometimes forced in the construction industry [18,19]. Furthermore, there are hitches in quantitative analysis due to the influence of various indicators due to the specificity and complexity of the construction site [3]. This is because the construction industry has numerous external work processes, and since a number of specialized companies for each process work within a short construction period within a fixed place, the construction industry is unprotected from countless and compound risk indicators, resulting in extraordinary uncertainty [20,21]. Thus, reliable data collection and aggregation are essential to advance and refine the accident risk analysis of construction sites, and dependable and advanced analysis techniques are prerequisites for scientific and quantitative data analysis. Although active investment, policies, and systems have been enhanced and introduced to reduce accidents in the construction industry for decades, the number of accidents in the construction industry is still extraordinarily associated with other industries; every year, a lot of the budget is spent on accident treatment [22,23].

In the previous study, for scientific and quantitative analysis of accidents at construction sites, there were many studies to advance accident analysis and increase efficiency through research, such as identifying, analyzing, and evaluating accident risk factors. For example, Cabello et al. [23] analyzed construction accident data through data mining to analyze patterns and factors for each construction process. They analyzed the accident data accumulated in the database and subdivided the accident investigation for each stage of construction. Their study delivered a framework for construction workers to advance accident prevention measures, accident action plans, and risk management. Martinez-Rojas et al. [24] built an IoT infrastructure and an open-source library to support accident-related decision-making by construction personnel. IoT can be used at all stages of the life cycle of a construction project, not just on the construction site, to improve accident-related data collection and information management. These measures contributed to real-time decision-making ability for efficient and active disaster risk management. In addition, an integrated safety environment framework for gathering and analyzing quantitative and qualitative data was proposed to improve accident safety performance in the construction industry [25]. In this study, quantitative data, as well as qualitative data (unsafe communication, work conditions, cognitive conflict between logic, cost tradeoff, etc.), were combined to advance the assessment of the accident risk and safety environment at the construction site. Allison et al. [22] presented a method of calculating the cost of an accident through an *ex-post* approach to governing the average cost of an accident for the quantitative assessment of an accident. When the volatility and

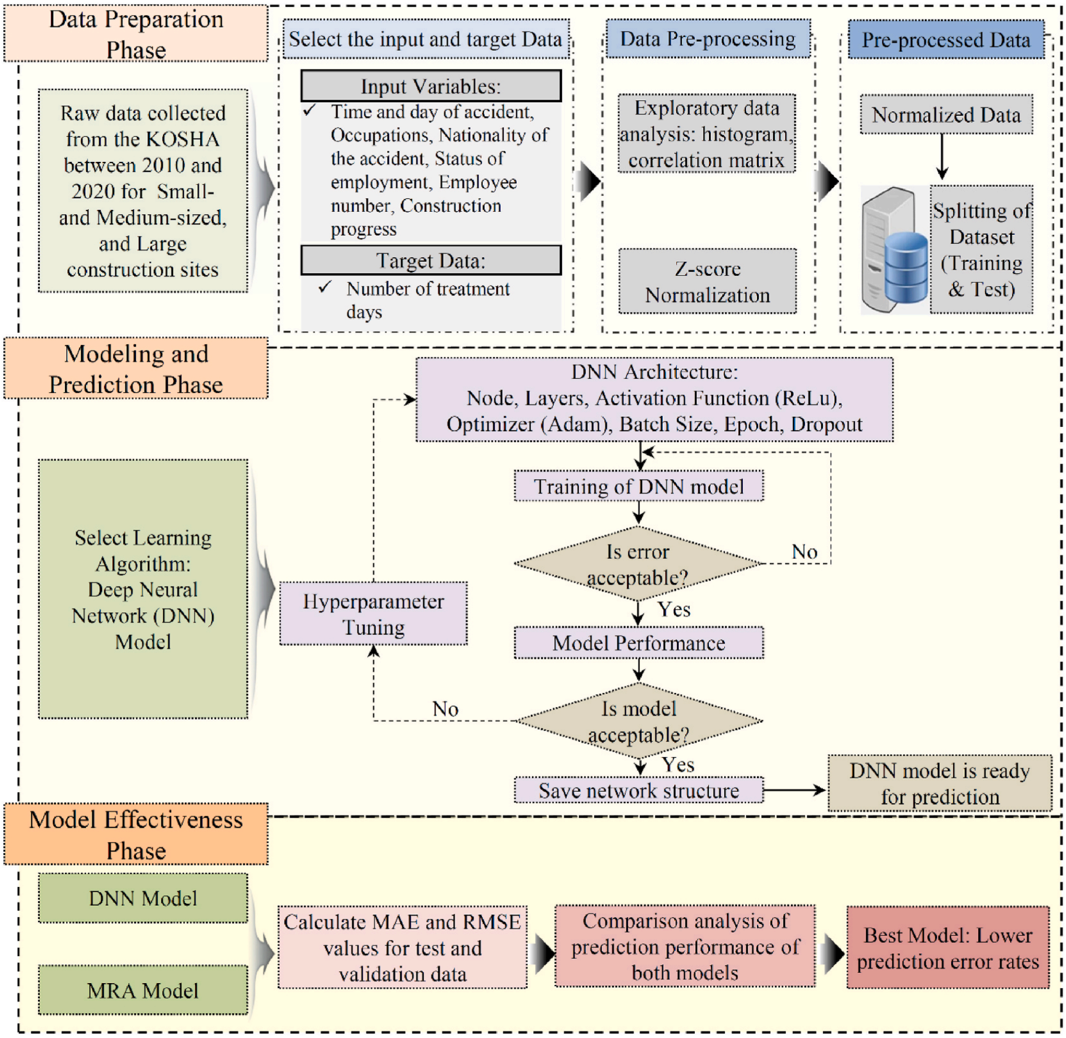

Fig. 1. Flow chart of the proposed DNN model with its different phases.

uncertainty of accidents are considered, the cost estimate and variability are enormous, depending on the nature of the accident [22]. Betsis et al. [26] created a database of accidents occurring in construction work and analyzed potential accident trends, stipulating correlations between various parameters. Through this study, the associations between various accidents were identified; consequently, an improvement plan for accident risk reduction was suggested.

As seen in the previous studies, analysis methods and frameworks for effective safety management have been proposed by identifying and evaluating risk indicators through the analysis of construction accident data in many existing studies. Nevertheless, despite these many studies, the accident rate in the construction industry is still great compared to that in other industries, and continuous research and efforts to improve that rate are immediately compulsory. As an extension of these efforts, this study recommends a model for predicting accident recovery periods (i.e., number of treatment days) at construction sites according to the scale of the construction site by applying deep learning techniques. Deep learning techniques can have more complex structures than other artificial intelligence techniques, such as fuzzy logic and machine learning, so they can quickly and precisely process more multifaceted and outsized amounts of data [27]. This technique can be actively adopted to analyze the massive quantity of big data produced and accumulated by unmanned aerial vehicles, IoT, sensors, CCTVs, etc., which are promptly being implemented at construction sites in recent years, and is expected to contribute to reducing the risk of accidents at construction sites [28–30]. Hence, this study proposes a method for accident analysis by predicting recovery periods of accidents at construction sites that consider the characteristics of construction site scale using deep learning algorithms.

3. Data and methodology

The ultimate goal of this study is to present a framework for predicting accident recovery periods by construction site size using deep learning algorithms. The detailed aims are as follows: First, accident cases and indicators occurring at the construction site are collected and classified by the scale of the site. Second, deep learning algorithm models are generated grounded on the collected data. The deep learning models were developed using Python 3.8. Third, to verify the deep learning models, the prediction results are compared with the multiple regression analysis models. The multiple regression model is a technique mainly used in the development of existing predictive models, and it was introduced for error and validation of deep learning models. Multiple regression models were developed using IBM Statistical Package for the Social Sciences (SPSS) V23. The prediction results (i.e., mean absolute error (MAE) and root mean squared error (RMSE)) of the deep learning algorithm models and the multiple regression analysis models were calculated and paralleled individually. Fig. 1 depicts the overall framework of the proposed DNN model with its different phases.

Table 3
Variable description and numerical assignment of variables.

Variable		Descriptions	Unit
Output	Accident recovery periods	Total number of treatment days due to accident	Numeric
Input	Time of accident	Time zone at the time of the accident Morning (6–12), afternoon (13–18), evening and night (18–24), dawn (0–6)	Nominal 1. Dawn 2. Evening and night 3. Afternoon 4. Morning
	Day of the accident	Day of the week when the accident occurred	Nominal 1. Monday 2. Tuesday 3. Sunday 4. Wednesday 5. Friday 6. Saturday 7. Thursday
	Classification of Occupations	Classification according to the Korean Standard Occupational Classification	Nominal 1. Equipment, machine operator and assembly worker 2. Professionals and related workers 3. Craft and related trades workers 4. Manager 5. Elementary workers
	Nationality of the accident	Classification of migrant workers and non-migrant workers by nationality of the accident	Nominal 0. non-migrant worker 1. Migrant worker
	Status of employment	Classified into regular and non-regular workers according to the employment status of the accident	Nominal 0. irregular worker 1. regular worker
	Employee number	Total number of employees at the accident site	Numeric
	Construction progress	Construction progress at the time of the accident (%)	Numeric

3.1. Data collection

In this study, deep learning models were developed based on accident cases that occurred at Korean construction sites from 2010 to 2020. Accident cases were collected from the Korea Occupational Safety and Health Agency (KOSHA). The reason this study selected KOSHA data is the high consistency of the data. KOSHA is a government institution established in 1987 in accordance with the Korea Occupational Safety and Health Agency Act to support employers in preventing accidents and to help workers work safely and healthily. Also, among many of KOSHA's representative tasks, such as disaster-related R&D, safety education, and prevention technology dissemination, the data is extremely dependable because it contains electronic data conversion through accident counting by industrialists [31].

The present study incorporated several input variables (i.e., time of the accident, employee number, occupational classification, nationality of the individuals involved in the accident, employment status, day of the accident, and the progress of construction at the site) identified through previous research as risk indicators associated with construction site accidents [1, 3, 5, 29–31]. Additionally, accident type classification, accident details, accident date and time, and total number of hospitalization days due to the accident were involved. Among them, the number of treatment days (i.e., the total number of treatment days due to accidents) was used as an output variable to compute the severity of accidents and fatal accidents were excluded for uniformity of cases. The input and output variable was obtained from the KOSHA dataset. A log transformation was applied to ensure a more normal distribution of the data. Here, we focus solely on non-fatal injuries, despite the significant occurrence of fatal accidents across different construction scales, including small and medium-sized industries, and aim to ensure uniformity and consistency in risk assessment. In general, small- and medium-sized construction sites are more unprotected from complex issues and several risk indicators, such as contract system, safety education, and the difficulty and complexity of work, indicating that uncertainty about the rate of fatal accidents is comparatively more

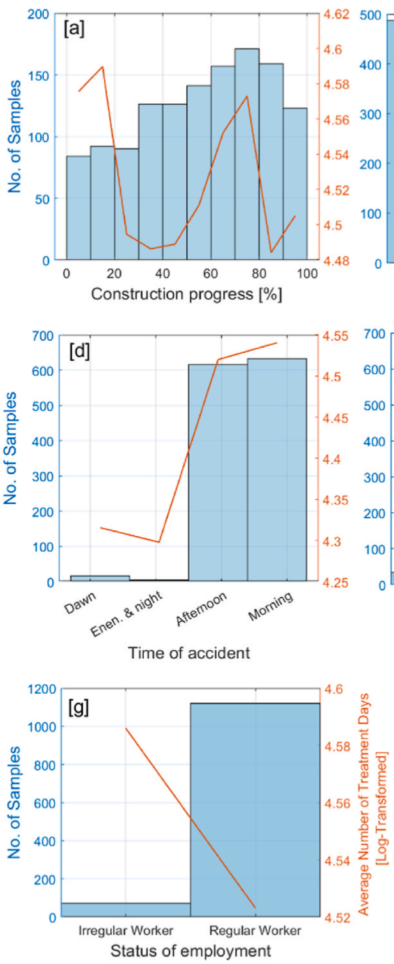

Fig. 2. Data distribution and the relationship between the output variable and input variables of the small-to-medium construction sites: (a) construction progress, (b) employee number, (c) day of the accident, (d) time of the accident, (e) classification of occupations, (f) nationality of the accident, and (g) status of employment (Note: Even. & night – evening and night, PW- professionals and related workers, CTW- craft and related trades workers, EW- elementary workers).

significant than large construction sites. The list of input variables is illustrated in Table 3. The time of the accident and the day of the accident are extensively used indicators for accident risk measurement and evaluation, as they are affected by the concentration of workers, the difficulty of construction and the occurrence and severity of accidents. Each was entered as a nominal variable [3,32]. Classification of occupations, the accident's nationality, and employment status are also statistically significantly related to accidents and are frequently applied as accident-related indicators [4,33]. For example, the nationality of the accident distinguished non-migrant workers from migrant workers. Non-migrant workers have longer careers and more opportunities for safety education than migrant workers; thus, they are less vulnerable to accidents. In addition, the employment status was divided into regular and non-regular workers since regular workers are less vulnerable to accidents than non-regular workers. Classification of occupations, the accident's nationality, and employment status were all accepted as nominal variables. Employee number is often used to indicate the risk of accidents since the larger the construction site, the less vulnerable workers are to accidents because workers have opportunities for safety education and high-quality safety management by safety officers [4,34]. The variable was entered as a numeric variable. As for the construction progress, as the construction project progresses and various processes unfold, the construction site becomes more composite, and as many specialized construction companies are involved on-site, the risk of accidents increases. For this reason, the variable is broadly adopted as a risk index for accidents, and in this study, it was entered as a numeric variable [6]. After selecting input and target variables, data cleaning was done, which involved feature selection for predictive variables and outlier detection within instance values to ensure data integrity for accurate predictions. Additionally, data transformations are essential for maintaining consistency in deep learning models, including converting categorical variables to numerical form and ensuring that input data is numeric and normalized [35,36]. Chollet [37] also highlights that deep learning frameworks natively handle numerical data, making the numerical assignment of categorical variables a common preprocessing step for seamless integration into the model. Accordingly, the numerical assignment of categorical variables has been performed (Table 3). Further, the collected input and output variables were divided according to the size of the construction site into small- and medium-sized construction sites and large construction sites. The

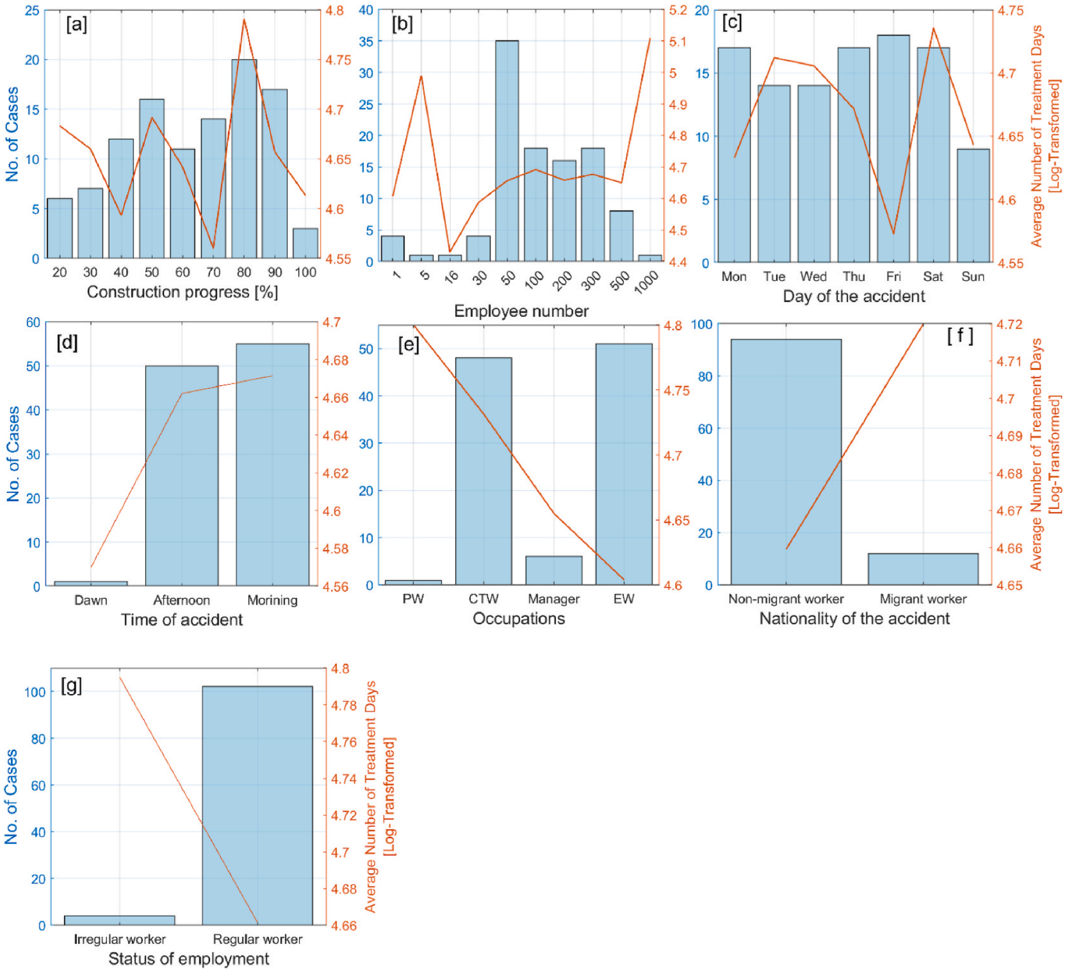

Fig. 3. Data distribution and the relationship between the output variable and input variables of the large construction sites: (a) construction progress, (b) employee number, (c) day of the accident, (d) time of the accident, (e) classification of occupations, (f) nationality of the accident, and (g) status of employment (Note: PW- professionals and related workers, CTW- craft and related trades workers, EW- elementary workers).

classification of construction site size was according to the Korea Occupational Safety and Health Act. This act classifies construction sites with a total construction cost of less than 2 billion won as small-sized, construction sites with a total construction cost of (2−12) billion won as medium, and construction sites with a total construction cost of 12 billion won or more as large [7]. Through this, small- and medium-sized construction sites (construction sites with a total construction cost of less than KRW 12 billion) and large construction sites (construction sites with a total construction cost of more than KRW 12 billion) were separated. In the present study, a total of 3954 accident cases were gathered, and among them, 3647 small- and medium-sized sites and 307 large construction sites were classified according to size.

Exploratory data analysis visually explores the link between accident severity and independent variables. Fig. 2 illustrates data distributions and the relationship between the output (i.e., average treatment days/recovery period) and input variables for small-to-medium construction sites. The relationship between construction progress and the average number of treatment days due to accidents is depicted in Fig. 2a. It reveals that early construction stages (0–20 % progress) result in higher average treatment days, suggesting heightened severity or extended recovery. Conversely, accidents during the intermediate construction stage (20–60 %) require fewer treatment days, indicating potential safety improvements as construction progresses. Similarly, accidents lead to shorter treatment durations in the final construction stages (80–100 % progress), possibly due to increased safety awareness as the project wraps up. In the case of employee numbers, the construction sites with less than 30 workers tend to have a higher average treatment duration following accidents, while construction sites with more than 50 workers exhibit shorter treatment periods, as depicted in Fig. 2b. It could be due to smaller construction teams encountering specific challenges or conditions leading to longer treatment periods, whereas larger construction teams may benefit from improved safety measures and quicker response, resulting in shorter treatments. Furthermore, accidents on weekends (Saturdays and Sundays) tend to result in longer treatment durations, while those at the beginning of weekdays (Mondays and Tuesdays) are associated with shorter treatment periods, as indicated in Fig. 2c. There is a slight increase in treatment times later in the week. This underscores the substantial influence of the accident day on severity and recovery times, possibly due to varying response dynamics and conditions during weekends versus weekdays. Moreover, morning and afternoon accidents lead to longer treatment days, whereas evening, night, and dawn accidents are linked to shorter durations (Fig. 2d). Fewer workers during dawn and evening-night shifts may contribute to these patterns, potentially affecting accident outcomes and response efficiency. Fig. 2e demonstrates that elementary workers undergo extended treatment durations following accidents, while professionals and related workers recover more quickly. Craft and related trades workers and managers fall in between. These disparities suggest that professionals and related workers are typically more safety-conscious and experienced, often in less physically demanding roles. In the case of the nationality of the accident, the migrant workers experience longer treatment durations after accidents, while non-migrant workers achieve shorter recoveries (Fig. 2f). Similarly, Fig. 2g shows irregular workers experience longer treatment durations after accidents, while regular workers recover more quickly. This may be because irregular workers often lack a sense of

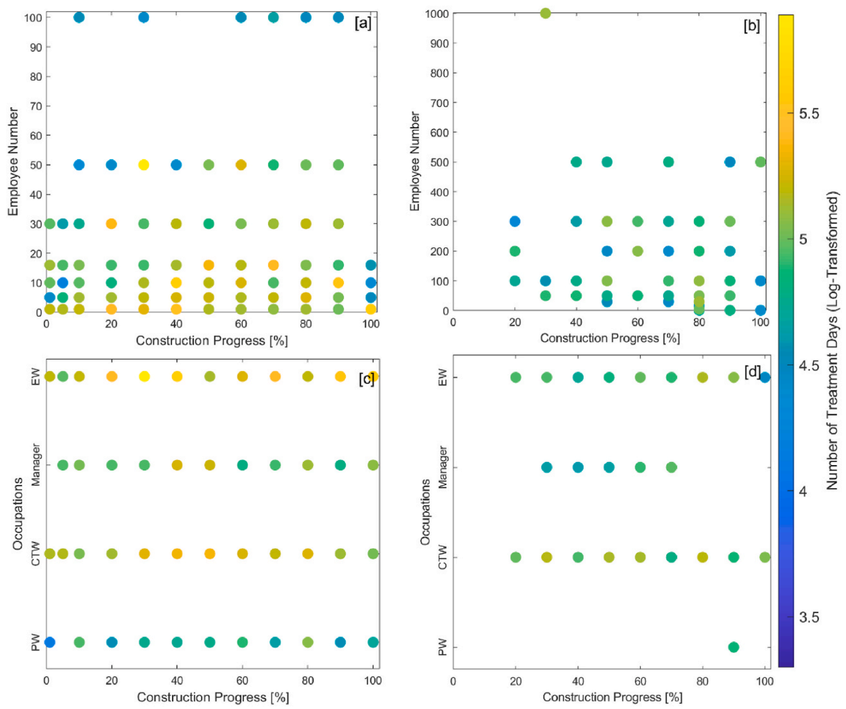

Fig. 4. The inter-relationship of the number of treatment days with employee number & construction progress for (a) small-medium construction sites and (b) large construction sites, while (c) & (d) represents the inter-relation of number of treatment days with classification of occupations & construction progress, respectively (note: PW– Professionals and related workers, CTW– Craft and related trades workers, EW– Elementary workers).

belonging to their current construction site and opportunities for safety education.

Fig. 3 depicts the relationship between input and output variables for large construction sites. It was observed that the accidents tend to result in longer treatment durations during 80–90 % of the construction stages, with the highest number of accident cases observed (Fig. 3a). In terms of employee numbers, sites with fewer than 5 workers or more than 1000 workers exhibit longer average treatment durations following accidents, despite having fewer accident cases, as shown in Fig. 3b. Additionally, accidents occurring on Saturdays typically lead to longer treatment durations, while those on Mondays and Fridays are associated with shorter recovery periods, as illustrated in Fig. 3c. Moreover, accidents during morning and afternoon hours are linked to longer treatment days, whereas dawn accidents are associated with shorter durations (Fig. 3d). Notably, no accident cases are reported during the evening and night times for large construction sites. Furthermore, Fig. 3e highlights that professionals, craft workers, and related trade workers tend to undergo prolonged treatment durations following accidents, while elementary workers and managers recover more quickly. In terms of nationality, migrant workers experience longer treatment durations compared to non-migrant workers (Fig. 3f). Similarly, irregular workers experience longer treatment durations after accidents compared to regular workers, as depicted in Fig. 3g. The exploratory data analysis revealed that factors like time of accident, employee number, occupational classification, and construction progress all significantly impact recovery times for both small and large construction sites. Interestingly, nationality and employment status follow similar recovery patterns in both cases. Thus, understanding these trends is essential for optimizing safety protocols across all construction phases and ensuring worker well-being throughout the project lifecycle for both the small-to-medium and large construction sites.

In the case of an accident that wasn't life-threatening, the injured worker received medical care and took time off work. Therefore, the severity of the accident could affect the duration of workdays lost. Consequently, Fig. 4(a and b) exhibits the distribution of recovery period (i.e., treatment days) due to accidents, taking into account the number of employees and construction progress for both the small-to-medium and large construction sites. The analysis revealed that small-to-medium construction sites with fewer than 50 workers, especially during the construction progress phase between 20 % and 90 %, experienced longer treatment periods. Conversely, sites with more than 50 workers revealed fewer workdays lost throughout the entire construction process. On the other hand, it was observed that the recovery period throughout the construction process in large construction sites is less than in small-to-medium construction sites, irrespective of employee number. Fig. 4(c and d) highlights the variations in treatment days influenced by occupational classifications and construction progress in both construction sites, reflecting the nature of work and associated risk levels. Notably, in both cases, elementary workers and craft and related trades workers typically had longer treatment durations, while professionals and related workers consistently had shorter recovery periods throughout the construction process. The physical demands of jobs held by elementary and craft workers increased the likelihood of accidents with more severe injuries, resulting in longer treatment times. Interestingly, managers experienced shorter treatment periods throughout the construction progress compared to the other occupational classes. This interconnection is vital for refining safety protocols throughout construction projects.

Furthermore, exploring the correlation between variables is captivating, particularly in identifying highly correlated variables. The degree of correlation among variables is quantified through Pearson's correlation coefficient [35,38]. Upon thorough analysis of the correlation matrix, we observe a noteworthy correlation of 0.143 between work progress and the number of employees for the small-to-medium construction sites (Fig. 5a), while 0.16 between employee number and status of employment for the case of large construction sites (Fig. 5b). However, this correlation does not reach a level where the removal of either variable is warranted. Consequently, none of the variables are considered redundant, enabling us to advance to the next analysis and modeling phase. Additionally, data normalization is crucial for effective predictive deep-learning models [39]. This is particularly important for diverse datasets like construction accident data, where features can have significantly different measurement units. During training, features often span varying scales, leading high-range variables to dominate those with smaller ranges, adversely affecting predictions [35,36,40]. Therefore, data normalization aims to minimize bias, ensuring equitable feature contribution and enhancing pattern recognition by minimizing the influence of dominant features on the model's overall performance. Consequently, the collected data went through a preprocessing process that normalized the data through z-score normalization. The z-score normalization was adopted considering the

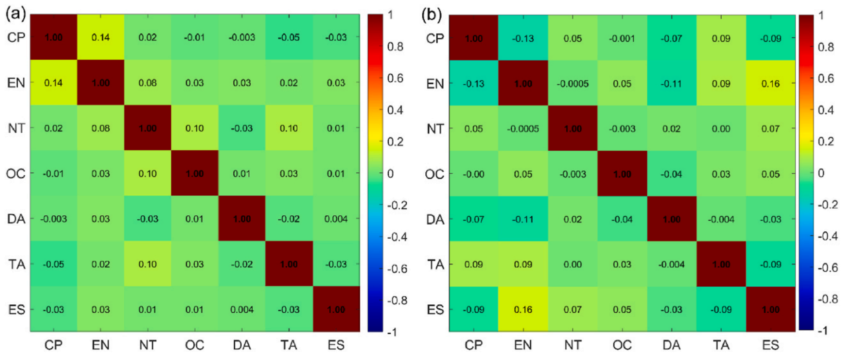

Fig. 5. Heatmap showing Pearson correlation matrix between each variable for (a) small-to-medium construction sites and (b) large construction sites (note: CP– Construction progress, EN– Employee number, NT– Nationality, OC– Occupations Classification, DA– Day of the accident, TA– Time of accident, ES– Status of employment).

diversity of construction accident cases and indicators since it adjusts the range of data using standard deviation and average and corrects for differences in different units and quantities [35,41]. Table 4 shows the descriptive statistics of input and output variables. The normalized data were indiscriminately separated into learning data (70 %) and test data (30 %). Among them, 30 % of the learning data was used as validation data.

3.2. Framework of deep learning algorithm model

This study employed a deep learning algorithm widely utilized for predictive and recognition tasks across diverse industrial domains to develop models. The deep learning algorithm comprises a neural network with several layers; subsequently, it can be applied to various data types and is extensively applied in innumerable industries [42]. For example, there are various models depending on the processing method, such as Deep Neural Network (DNN), Generative Adversarial Network (GAN), Recurrent Neural Network (RNN), Auto Encoder (AE), and Convolutional Neural Network (CNN). Among them, DNN is commonly applied for the prediction and classification of data with complex non-linear relationships, as multiple layers act to identify diverse specific functions [43]. Consequently, this study suggests a framework for developing accident prediction models according to the construction scale using DNN, considering the non-linearity of the accident data of the construction site [35,44,45]. To determine the optimal DNN model, the prediction error was estimated by comparison with the multiple regression analysis (MRA) models. The evaluation indicators of prediction error for artificial neural networks, known as MAE and RMSE, were calculated and contrasted individually [46]. The MAE calculates the difference between the actual and predicted values as an absolute value and averages them, while RMSE shows the residual as one measure. Therefore, the closer the MAE and RMSE are to 0, the higher the predictive power of the model. The detailed workflow and overview of the proposed DNN model with its different phases are depicted in Fig. 1.

The DNN model implements an optimization model through a backpropagation algorithm that changes the weight of each neural network node. Thus, for the optimization model, it is compulsory to find the optimal network structure scenario and hyperparameter through trial-and-error methods [47]. To achieve an optimal model configuration, it is crucial to establish the appropriate number of nodes and layers in the network structure and set hyperparameters such as dropout rate, batch size, epoch count, choice of optimizer, and activation functions. Hyperparameters are parameters that control the learning process [48]. For example, dropout is a regularization penalty to solve the overfitting problem that causes poor performance of deep learning models. The batch size sets the data learning unit of the DNN model and affects the efficient training of the model. Epoch specifies the number of learning in the data learning process of the model, and the optimizer controls the model's stability by governing the model's data learning rate. The activation function specifies how to regulate the least-cost function [49]. To address the limitations of the available data and align with the network structure and hyperparameters used in related prior studies, we configured the model with three layers, and dropout was set to 0 or 0.2. Additionally, we set the batch size to 5 and conducted training for 1000 epochs [44–47]. Optimizer announced an optimization algorithm called Adaptive Moment Estimation (Adam) using the moment theory of the stochastic objective function. The Adam Method has been accepted in numerous areas owing to the convenience of calculation and the diversity of applications [50]. As the activation function, the Rectified Linear Unit (ReLu) function established to supplement the sigmoid function was used [51]. The block diagram of DNN architecture, which includes input variables, hidden layers and output (number of treatment days, i.e., recovery period), is depicted in Fig. 6.

4. Results and discussion

4.1. Results

Table 5 illustrates the variation in the number of nodes for simulation based on network structure and hyperparameters. It also presents the MAE and RMSE values corresponding to different construction scales (small- and medium-sized construction sites, large construction sites) and dropout rates. Among the scenarios, the scenario with the minimum MAE and RMSE values was designated as the final model. In the small- and medium-sized construction site model, the scenario generally had a larger loss function when the dropout was 0.2 than when the dropout was 0. Moreover, as the number of nodes increased, the MAE and RMSE values tended to decline, and when the numbers of nodes were 400–400–400, the MAE and RMSE values displayed the smallest values. On the other

Table 4
Descriptive statistics of variables.

Variables		Small- and Medium-sized construction site					Large construction site				
		N	Min.	Max.	Mean	Std. Devi-ation	N	Min.	Max.	Mean	Std. Devi-ation
Output	[a]Accident recovery periods	3647	3.30	5.89	4.52	0.36	307	4.19	5.12	4.67	0.24
Input	Time of accident	3647	1.00	4.00	3.46	0.58	307	1.00	4.00	3.50	0.55
	Day of the accident	3647	1.00	7.00	3.99	2.05	307	1.00	7.00	4.14	2.07
	Classification of Occupations	3647	2.00	5.00	3.84	1.00	307	2.00	5.00	4.00	0.99
	Nationality of the accident	3647	0.00	1.00	0.05	0.22	307	0.00	1.00	0.11	0.32
	Status of employment	3647	0.00	1.00	0.94	0.23	307	0.00	1.00	0.96	0.19
	Employee number	3647	1.00	100.00	10.69	14.15	307	1.00	1000.00	163.16	159.87
	Construction progress	3647	1.00	100.00	50.61	27.04	307	20.00	100.00	63.02	21.89

[a] Accident recovery periods, i.e., number of treatment days presented in log-transformed form.

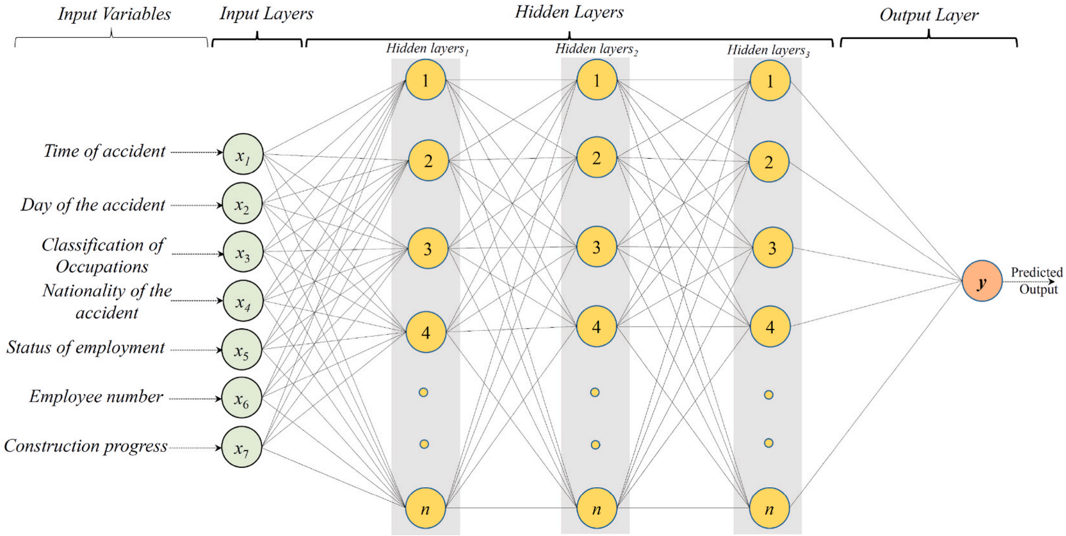

Fig. 6. The block diagram of DNN architecture.

Table 5
Learning results.

Network Structure Scenario	Small- and Medium-sized construction sites				Large construction sites			
	Dropout (0)		Dropout (0.2)		Dropout (0)		Dropout (0.2)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
5–5–5	0.332	0.389	0.342	0.401	0.282	0.335	0.368	0.444
10–10–10	0.324	0.381	0.357	0.402	0.231	0.289	0.270	0.339
25–25–25	0.317	0.374	0.324	0.384	0.257	0.320	0.283	0.349
50–50–50	0.302	0.359	0.322	0.382	0.197	0.254	0.283	0.347
75–75–75	0.291	0.354	0.315	0.374	0.199	0.254	0.254	0.318
100–100–100	0.284	0.346	0.302	0.362	0.182	0.232	0.226	0.279
200–200–200	0.266	0.335	0.297	0.356	0.158	0.204	0.216	0.267
300–300–300	0.264	0.332	0.291	0.352	0.158	0.201	0.186	0.229
400–400–400	**0.263**	**0.331**	0.285	0.345	0.126	0.167	0.193	0.249
500–500–500	0.263	0.333	0.284	0.347	0.141	0.191	0.188	0.233
600–600–600	0.266	0.334	0.289	0.354	0.134	0.181	0.198	0.253
700–700–700	0.261	0.333	0.285	0.349	**0.095**	**0.138**	0.187	0.235
800–800–800	0.268	0.334	0.285	0.349	0.109	0.162	0.152	0.188
900–900–900	0.267	0.333	0.285	0.349	0.129	0.177	0.161	0.205
1000–1000–1000	0.269	0.332	0.284	0.349	0.120	0.161	0.168	0.206

hand, the large construction site model revealed that the loss function was larger when the dropout was 0.2 than when the dropout was 0. Similarly, as the number of nodes grew, the MAE and RMSE values tended to minimize. When the number of nodes was 700–700–700, the MAE and RMSE values showed the smallest values. Table 6 presents the optimized network structure and hyperparameters for the model used in small-to-medium-sized and large construction sites.

To ensure the final model's effectiveness with new data, we calculated MAE and RMSE values for both the test data and verification data. Table 7 presents the MAE and RMSE values for verification and test data in both the small- and medium-sized and large construction site models. Specifically, for the small- and medium-sized construction site model, the verification data yielded MAE and

Table 6
Hyper-parameter and network structure.

Type	Configuration	Small- and Medium-sized construction site model	Large construction site model
Hyper Parameter	Dropout	0	
	Batch Size	5	
	Epoch	1000	
	Optimizer	Adaptive Moment Estimation (Adam) Method	
	Activation Function	Rectified Linear Unit (ReLu) function	
Network structure	Hidden Layers	3	
	Node	400–400–400	700–700–700

Table 7
Comparison of the analysis results of the verification data and test data.

Models	Small- and Medium-sized construction site				Large construction site			
	Validation data		Test data		Validation data		Test data	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MRA			0.276	0.467			0.185	0.082
DNN	0.258	0.317	0.263	0.332	0.227	0.302	0.109	0.057
MRA/DNN (%)			−4.7 %	−28.9 %			−40.9 %	−30.3 %

RMSE values of 0.258 and 0.317, respectively, while the test data produced values of 0.263 and 0.332. Similarly, for the large construction site model, the MAE and RMSE values for the verification data were 0.227 and 0.302, respectively, and for the test data, they were 0.109 and 0.057. Notably, both models showed no significant difference between the MAE and RMSE values of the test data and those of the validation data, suggesting that the overfitting problem of the model was overlooked. Furthermore, additional validation of the prediction error rates of the models was compared with MRA models. The MRA model is a method for estimating the correlation between factors by means of a statistical technique and is commonly used in the field of prediction [52,53]. The MRA model was constructed using the same input and output variables as used in the DNN model. Table 7 presents the results of the comparison between the models. The small- and medium-sized construction site DNN model showed lower prediction error rates than the MRA model at 4.7 % in MAE and 28.9 % in RMSE. The large construction site DNN model also showed lower prediction error rates than the MRA model at 40.9 % in MAE and 30.3 % in RMSE.

4.2. Discussion

This study proposed a framework for developing an accident prediction model according to the construction scale by applying the

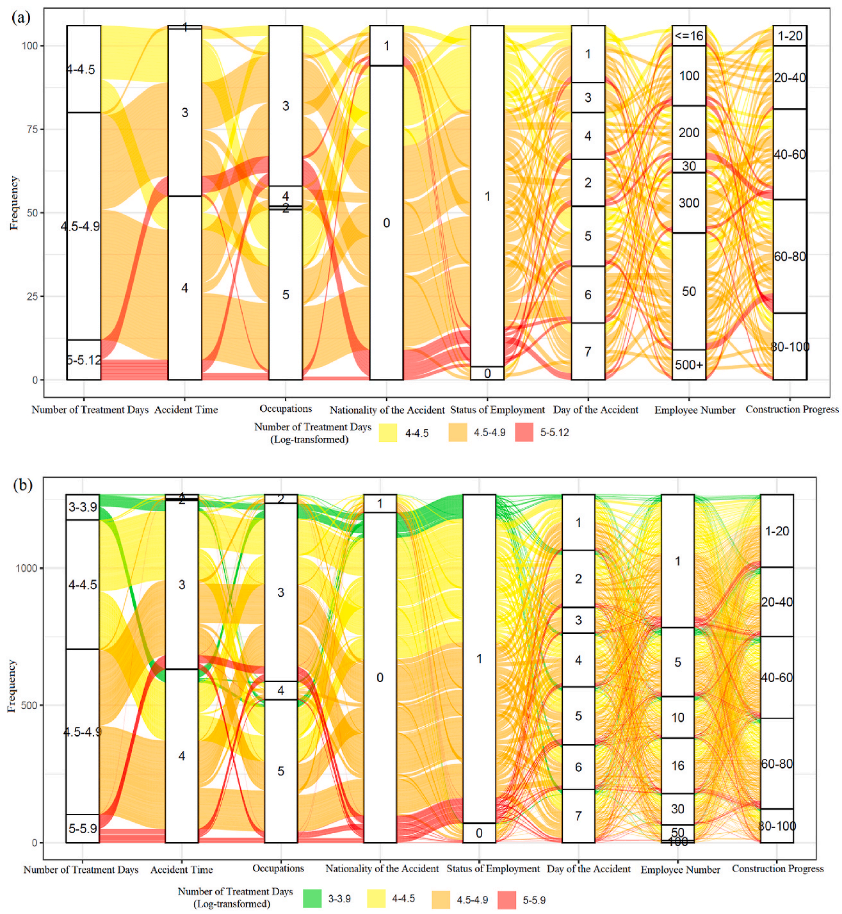

Fig. 7. Alluvial flow diagram exhibiting inter-correlations among the input and output variables for (a) large construction sites and (b) small-to-medium construction sites. The numerical values presented in each bar of the corresponding variable are discussed in Table 3.

DNN algorithm. For data collection, accident cases that occurred at the construction sites were collected from the KOSHA and based on the total construction cost, divided into small- and medium-sized construction sites (construction sites with a total construction cost of less than 12 billion won), and large construction sites (Construction sites with a total construction cost of 12 billion won or more). The alluvium diagram (Fig. 7) provides valuable insights into the intricate relationship between output and input variables that significantly influence the severity of construction accidents for both large and small-to-medium construction sites. Notably, it was observed that accidents during the morning and afternoon required the longest recovery or treatment periods (Fig. 7 a & b), which suggests the scheduling of construction activities to reduce the likelihood of accidents during critical times. Concerning occupation classes, both elementary and craft-related trades workers required longer recovery periods, while professionals and related workers recovered more quickly in both cases, highlighting the need for customized safety measures and training programs. Additionally, it was observed that the migrant and irregular workers had longer recovery periods than regular and non-migrant workers. These differences imply that unique challenges or circumstances faced by migrant workers, including limited job experience due to frequent job changes, contribute to their extended treatment periods, emphasizing the need to address safety disparities and ensure the well-being of all worker groups. An analysis of accidents on specific days of the week, employee numbers, and construction progress revealed that these factors showed minimal variation and oscillated within a narrow range in both cases. There was no significant indication from the data analysis that the day of the week consistently affected the duration of accident recovery. The analysis revealed that accidents occurring during the early stages of construction (0–20 % progress) resulted in an average recovery period of 98 days in the case of small-to-medium construction sites, indicating heightened severity or prolonged recovery periods. Interestingly, these early-stage accidents, which are frequent, often involve construction sites with fewer than five workers (Fig. 7b). On the other hand, in the case of large construction sites, accidents during the advanced construction stage (40–80 % progress) necessitated most treatment days (approximately 97 days) (Fig. 7a). Further investigation is needed to understand this specific trend. In terms of employee numbers, small-to-medium construction sites with fewer than sixteen workers tend to see longer treatment durations following accidents, while sites with more than 50 workers experienced shorter treatment periods. This pattern is also mostly followed in large-scale construction sites. This variation may stem from the unique challenges or conditions encountered by smaller construction teams, leading to extended recovery periods, whereas larger teams could benefit from better safety protocols and quicker responses, resulting in shorter treatment durations. Interestingly, it was noted that large construction sites with more than 1000 workers also experienced longer treatment periods despite having fewer accident cases. This discrepancy could be due to various factors, such as the severity of accidents, the nature of injuries, or the effectiveness of response and treatment protocols. The inter-relationships identified provide a data-driven roadmap for enhancing construction site safety, enabling the precise implementation of measures to reduce accident severity and promote worker well-being. Nonetheless, it is important to note that the relationships between the number of treatment days and input variables, including accident timing, employee numbers, occupational classification, nationality, employment status, day of the accident, and construction progress, predominantly exhibit non-linear patterns. This underscores the necessity for advanced modeling approaches to effectively predict accident recovery periods for planning, resource allocation, and minimizing downtime in the construction industry.

The DNN model was deemed more suitable for predicting the severity of construction accidents, given its capacity to capture the

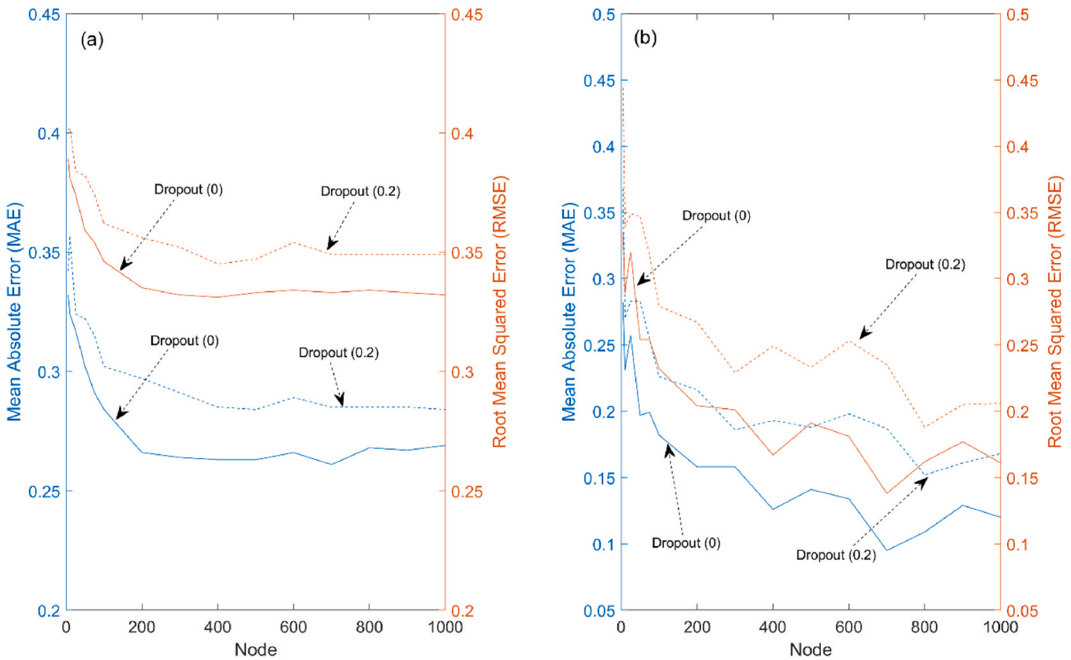

Fig. 8. Performance evaluation of network architecture at various dropout rates for (a) Small- and Medium-sized and (b) large construction sites using MAE and RMSE.

non-linearity and uncertainty inherent in the data. Consequently, we opted for the DNN model to predict the accident recovery period for construction sites following accidents. After that, the final optimal DNN model was established by learning network scenarios and hyper-parameters using input and output variables through trial-and-error methods. The network architecture performance at various dropout rates for both small- and medium-sized and large construction sites is depicted in Fig. 8a and b, respectively. In the prediction results of the final optimal model, it was observed that the small- and medium-sized construction site model exhibited superior predictive capabilities compared to the MRA model, with a 4.7 % lower MAE and a 28.9 % lower RMSE. The large construction site model also presented greater predictive power than the MRA model by 40.9 % in MAE and 30.3 % in RMSE. Additionally, both DNN models demonstrated greater predictive accuracy in the verification results than the MRA models. Consequently, the results and framework of this study are considered to be highly significant. The reason is that the DNN models can better reveal the non-linearities of accidents and various influencing indicators at the construction site than the MRA model [47,48,54]. When comparing the two DNN models, the small- and medium-sized construction site model showed a 36.2 % higher prediction error in MAE and 1.4 % in RMSE. This shows that small- and medium-sized construction sites are more unprotected from complex issues and several risk indicators, such as contract system, safety education, and the difficulty and complexity of work, indicating that uncertainty about prediction is comparatively great for large construction sites. To unravel this problem, it is considered that it can be supplemented with additional research by securing more influencing indicators and acquiring data for small- and medium-sized construction sites.

Hence, according to the framework offered in this study, by building a model for predicting accident recovery periods at construction sites, numerous construction-related players will be able to derive an accurate and objective amount of lost workdays. For example, a small- and medium-sized site contractor can develop an accident prediction model (i.e., recovery period) using the proposed DNN algorithm, quantify accident risk, and manage accidents without needing expert consulting or separate risk assessment. In particular, the ordering party can set the risk probability level in consideration of the asset status and risk appetite through pre-prediction of the amount of accident risk, and based on this, it can be used as a guideline for budgeting, such as accident preparedness reserve and safety management budget. Furthermore, reducing and preventing accidents through appropriate investment parallel to the risk of accidents will be possible. It can also be utilized as a criterion for judging whether the contractor purchases insurance for a construction site or whether the insurance premium purchased is suitable. For additional risks, it will be possible to proactively establish a financial risk transfer strategy in advance through special contracts or expansion of coverage. This will finally enable active accident management at the construction site, thereby contributing to the reduction of the high accident rate and social costs due to accidents at the construction site.

Nonetheless, in this study, a model was established based on the accident cases collected by KOSHA. Thus, additional research is needed to compare and verify the results of this study by securing supplementary data from insurance companies or institutions that can signify the financial loss of construction site accidents and accidents in various countries. In addition, we intend to develop further models to incorporate data and features related to fatal accidents, thereby providing a more comprehensive tool for risk assessment in the construction industry. Further, for the gradual sophistication of the model, it is indispensable to discover additional influencing indicators and secure superfluous data. Moreover, although only the DNN algorithm was utilized in this study, it is crucial to improve the reliability of the model through comparative or cross-validation studies by introducing other machine learning and deep learning algorithms in future studies. Additionally, owing to the nature of the DNN algorithm, the mutual relationship and weight between each node could not be acknowledged. This illustrates the disadvantage that the user cannot access the process and basis of the prediction result by unilaterally presenting the prediction result through the black box of the DNN algorithm. Hence, it is compulsory to define the user variables and weights by introducing XAI (eXplainable AI) to improve the dependability and acceptability of the model in the future and to relieve user anxiety [55].

5. Conclusion

The construction industry is more prone to accidents than other industries. In addition, the accident rate at construction sites is increasing due to the recent increase in complexity and scale of construction projects and the upsurge in urban construction. An advanced and accurate accident risk quantification model is desperately necessary for accident rate reduction and prevention. However, although there are significant differences in the characteristics of small-to-medium and large construction sites in existing accident risk quantification models and studies, they do not replicate this. Therefore, this study proposes a framework for developing a deep learning model for accident analysis by predicting recovery periods of accidents based on the scale of construction projects.

To develop predictive models, accident-related indicators allowed in earlier studies were collected based on accident case data of construction sites, and DNN algorithm models were created. The optimal combination was found to set up the optimal DNN model, and the model was refined through trial and error by referring to the optimal hyper-parameter components and network structure scenarios utilized in former studies. The proposed model significantly aids in conserving workdays within the construction industry by accurately estimating recovery periods for accidents across various construction scales. Employers can leverage these forecasts to strategize temporary work reassignments, modify project timelines, and allocate resources more effectively, thereby mitigating disruptions resulting from injuries. Through the predictive capacity of the proposed model, innumerable construction-related entities will be able to predict accident-related risks easily and quickly in advance.

The DNN model developed in this study showed a low prediction error through comparative verification with the conventional MRA model, so it is considered that it can contribute to the prediction of the accident recovery period at the construction site. The research results and framework offered in this study can also be adopted as reference materials for accident loss reduction and proficient safety management and are projected to eventually contribute to the decline of the accident rate in the construction industry. Furthermore, the outcomes and framework of this study can be of practical assistance to managing construction projects in high-risk industries and other risk studies. In addition, it is judged that the deep learning model of this study will be capable of advancing the accident prediction technique at the construction site and expanding the accuracy of accident prediction. Moreover, it is probable that the model can be further advanced through additional validation of the effectiveness of the model and continuous data collection.

Funding statement

This Research was supported by Research Funds of Mokpo National University in 2023.

Additional information

No additional information is available for this paper.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Ji-Myong Kim: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Manik Das Adhikari:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Data curation. **Junseo Bae:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology. **Sang-Guk Yum:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis.

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.

Acknowledgments

The authors thank the esteemed reviewers for their valuable comments and suggestions that helped improve the manuscript.

References

[1] Kim Y.H., Kim J.H., Kim H.G., Artificial intelligence for safety and disaster management in the construction industry, Korean J. Construct. Eng. Manag. 22 (4) (2021) 4-8. Available at https://www.auric.or.kr/user/rdoc/RdocDirectoryStep3.aspx?dbname=CMAG&organCode2=kicem01&yearmonth=202108&n=4.

[2] J.M. Kim, T. Kim, S. Ahn, Loss assessment for sustainable industrial infrastructure: focusing on bridge construction and financial losses, Sustainability 12 (13) (2020) 5316, https://doi.org/10.3390/su12135316.

[3] J.M. Kim, T. Kim, K. Son, J. Bae, S. Son, A quantitative risk assessment development using risk indicators for predicting economic damages in construction sites of South Korea, J. Asian Architect. Build Eng. 18 (5) (2019) 472–478, https://doi.org/10.1080/13467581.2019.1681274.

[4] J.M. Kim, T. Kim, J. Bae, K. Son, S. Ahn, Analysis of plant construction accidents and loss estimation using insurance loss records, J. Asian Architect. Build Eng. 18 (6) (2019) 507–516, https://doi.org/10.1080/13467581.2019.1687089.

[5] J.M. Kim, K.C. Ha, S. Ahn, S. Son, K. Son, Quantifying the third-party loss in building construction sites utilizing claims payouts: a case study in South Korea, Sustainability 12 (23) (2020) 1015, https://doi.org/10.3390/su122310153.

[6] S.G. Yum, S. Ahn, J. Bae, J.M. Kim, Assessing the risk of natural disaster-induced losses to tunnel-construction projects using empirical financial-loss data from South Korea, Sustainability 12 (19) (2020) 8026, https://doi.org/10.3390/su12198026.

[7] Y.R. Jang, S.S. Go, A study on the priority safety management items in the medium and small sized construction sites, Korean J. Construct. Eng. Manag. 21 (4) (2020) 38–49, https://doi.org/10.6106/KJCEM.2020.21.4.038.

[8] T.W. Kim, S. Park, B. Choi, Y. Kang, K. Park, W. Jeong, C. Koo, Differences Among University Students, Professors, and Practitioners on the Construction Technologies in the Fourth Industrial Revolution, Korean J. Constr. Eng. Manag. 23 (3) (2022) 095–113, https://doi.org/10.6106/KJCEM.2022.23.3.095.

[9] H. Park, C. Koo, Effect of virtual reality-based construction safety education on the learning performance of construction workers Using CAMIL theory, Korean J. Constr. Eng. Manag. 23 (3) (2022) 104–115, https://doi.org/10.6106/KJCEM.2022.23.3.104.

[10] K.S. Bae, J.D. Yoon, H.S. Ahn, K.B. Shim, Industrial Accident Status Analysis and Policy Direction: Focusing on Small and Medium-Sized Construction Sites, Korea Labor Institute, Labor Welfare and Labor Insurance Research, 2013.

[11] J.H. Won, Y.H. Yoon, T.G. Oh, H.G. Park, S.H. Jeong, Measures for Assigning Responsibility to the Client to Prevent Accidents at Small-Scale Construction Sites in the Construction Industry, *Korea Occupational Safety And Health Agency*, 2019.

[12] M. Jeong, Kingpin for the prevention of safety accidents at construction sites and the act on punishment of serious accidents, Construct. Eng. Manag. 23 (2) (2022).

[13] S. Ahn, T. Kim, J.M. Kim, Sustainable risk assessment through the analysis of financial losses from third-party damage in bridge construction, Sustainability 12 (8) (2020) 3435, https://doi.org/10.3390/su12083435.

[14] S. Laryea, Risk pricing practices in finance, insurance and construction, in: Proceedings of the Construction and Building Research Conference of the Royal Institution of Chartered Surveyors, Dublin, Ireland, 4–5 September 2008, University of, Reading, 2008, 2008, pp. 1–16. Avialable at, https://centaur.reading.ac.uk/16292/.

[15] S. Baker, D. Ponniah, S. Smith, Techniques for the analysis of risks in major projects, J. Oper. Res. Soc. 49 (1998) 567–572, https://doi.org/10.1057/palgrave.jors.2600548.

[16] I. Dikmen, M.T. Birgonul, A.E. Arikan, A critical review of risk management support tools, in: *Proceedings Of the 20th Annual Conference Of Association Of Researchers In Construction Management*, 1145–1154 (Association of Researchers in Construction Management, Edinburgh, UK, 1–3 September 2004, Heriot-Watt University, 2004. https://www.arcom.ac.uk/.

[17] G. Wood, R.C.T. Ellis, Risk management practices of leading UK cost consultants, Eng. Construct. Architect. Manag. 10 (2003) 254–262, https://doi.org/10.1108/09699980310489960.
[18] K.R. Molenaar, Programmatic cost risk analysis for highway megaprojects, J. Construct. Eng. Manag. 131 (2005) 343–353, https://doi.org/10.1061/(ASCE)0733-9364(2005)131:3(343).
[19] E. Cagno, F. Caron, M. Mancini, A multi-dimensional analysis of major risks in complex projects, Risk Manag. 9 (2007) 1–18, https://doi.org/10.1057/palgrave.rm.8250014.
[20] P.X. Zou, G. Zhang, J. Wang, Understanding the key risks in construction projects in China, Int. J. Proj. Manag. 25 (2007) 601–614, https://doi.org/10.1016/j.ijproman.2007.03.001.
[21] D. Baloi, A.D. Price, Modeling global risk factors affecting construction cost performance, Int. J. Proj. Manag. 21 (2003) 261–269, https://doi.org/10.1016/S0263-7863(02)00017-0.
[22] W.R. Allison, K.H.C. Hon, B. Xia, Construction accidents in Australia: evaluating the true costs, Saf. Sci. 120 (2019) 886–896, https://doi.org/10.1016/j.ssci.2019.07.037.
[23] T.A. Cabello, M. Martinez-Rojas, A.J. Carrillo-Castrillo, C.J. Rubio-Romero, Occupational accident analysis according to professionals of different construction phases using association rules, Saf. Sci. 144 (2021) 925–7535, https://doi.org/10.1016/j.ssci.2021.105457.
[24] M. Martinez-Rojas, M.J. Gacto, A. Vitiello, G. Acampora, J.M. Soto-Hidalgo, An internet of things and fuzzy markup language based approach to prevent the risk of falling object accidents in the execution phase of construction projects, Sensors 21 (2021) 6461, https://doi.org/10.3390/s21196461, 2021.
[25] F. Lestari, Y.R. Sunindijo, M. Loosemore, Y. Kusminanti, B. Widanarko, A safety climate framework for improving health and safety in the Indonesian construction industry, Int. J. Environ. Res. Publ. Health 17 (2020) 7462, https://doi.org/10.3390/ijerph17207462.
[26] S. Betsis, M. Kalogirou, G. Aretoulis, M. Pertzinidou, Work accidents correlation analysis for construction projects in Northern Greece 2003–2007: a retrospective study, Saf. Now. 5 (2019) 33, https://doi.org/10.3390/safety5020033.
[27] B. Zhong, X. Pan, P.E. Love, L. Ding, W. Fang, Deep learning and network analysis: classifying and visualizing accident narratives in construction, Autom. ConStruct. 113 (2020) 103089, https://doi.org/10.1016/j.autcon.2020.103089.
[28] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inform. 29 (2015) 239–251, https://doi.org/10.1016/j.aei.2015.02.001.
[29] Zdenek, K., Hainan, C., & Xiaowei, L. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Autom. ConStruct. 89, 58–70, https://doi.org/10.1016/j.autcon.2018.01.003.
[30] F. Weili, D. Lieyun, L. Hanbin, E.L. Peter, Falls from heights: a computer vision-based approach for safety harness detection, Autom. ConStruct. 91 (2018) 53–61, https://doi.org/10.1016/j.autcon.2018.02.018.
[31] J.M. Kim, K.K. Lim, S.G. Yum, S. Son, A deep learning model development to predict safety accidents for sustainable construction: a case study of fall accidents in South Korea, Sustainability 14 (3) (2022) 1583, https://doi.org/10.3390/su14031583.
[32] Nevada Department of Transportation, Risk Management and Risk-Based Cost Estimation Guidelines; Nevada Department of Transport: Carson City, NV, USA, 2012.
[33] S. Ahmed, Causes and effects of accident at construction site: a study for the construction industry in Bangladesh, Int. J. Sustain. Constr. Eng. Technol. 10 (2019) 18–40. https://penerbit.uthm.edu.my/ojs/index.php/IJSCET/article/view/4499.
[34] J.M. Kim, K. Son, S.G. Yum, S. Ahn, Analyzing the risk of safety accidents: the relative risks of migrant workers in construction industry, Sustainability 12 (2020) 5430, https://doi.org/10.3390/su12135430.
[35] L. Pérez-Sala, M. Curado, L. Tortosa, J.F. Vicent, Deep learning model of convolutional neural networks powered by a genetic algorithm for prevention of traffic accidents severity, Chaos, Solit. Fractals 169 (2023) 113245, https://doi.org/10.1016/j.chaos.2023.113245.
[36] F. Ettensperger, Comparing supervised learning algorithms and artificial neural networks for conflict prediction: performance and applicability of deep learning in the field, Qual. Quantity 54 (2) (2020) 567–601, https://doi.org/10.1007/s11135-019-00882-w.
[37] F. Chollet, *Deep Learning with Python*. Shelter Island, Manning Publications Company, 2017.
[38] M. Szóstak, Analysis of occupational accidents in the construction industry with regards to selected time parameters, Open Eng. 9 (1) (2019) 312–320, https://doi.org/10.1515/eng-2019-0027.
[39] K.P. Murphy, Machine Learning: a Probabilistic Perspective, MIT press, 2012.
[40] K. Li, H. Xu, X. Liu, Analysis and visualization of accidents severity based on LightGBM-TPE, Chaos, Solit. Fractals 157 (2022) 111987, https://doi.org/10.1016/j.chaos.2022.111987.
[41] S.W. Bae, J.S. Yoo, Apartment price estimation using machine learning: gangnam-gu, Seoul as an example, Real Estate Stud. 24 (2018) 69–85.
[42] Zhong, G., Wang, L.N., Ling, X., & Dong, J. An overview on data representation learning: From traditional feature learning to recent deep learning. J. Financ. Data Sci. 2, 265–278, https://doi.org/10.1016/j.jfds.2017.05.001.
[43] J. Gu, Z. Wang, J. Kuen, Recent advances in convolutional neural networks, Pattern Recogn. 77 (2018) 354–377, https://doi.org/10.1016/j.patcog.2017.10.013.
[44] R. Zhu, X. Hu, J. Hou, X. Li, Application of machine learning techniques for predicting the consequences of construction accidents in China, Process Saf. Environ. Protect. 145 (2021) 293–302, https://doi.org/10.1016/j.psep.2020.08.006.
[45] M. Alkaissy, M. Arashpour, E.M. Golafshani, M.R. Hosseini, S. Khanmohammadi, Y. Bai, H. Feng, Enhancing construction safety: machine learning-based classification of injury types, Saf. Sci. 162 (2023) 106102, https://doi.org/10.1016/j.ssci.2023.106102.
[46] H. Na, B.H. Park, Developing accident models of rotary by accident occurrence location, Int. J. Highw. Eng. 14 (2012) 83–91, https://doi.org/10.7855/IJHE.2012.14.4.083.
[47] J. Kim, S. Yum, S. Son, K. Son, J. Bae, Modeling deep neural networks to learn maintenance and repair costs of educational facilities, Buildings 11 (4) (2021) 165, https://doi.org/10.3390/buildings11040165.
[48] J.M. Kim, J. Bae, S. Son, K. Son, S.G. Yum, Development of model to predict natural disaster-induced financial losses for construction projects using deep learning techniques, Sustainability 13 (9) (2021) 5304, https://doi.org/10.3390/su13095304.
[49] J.-D. Ryu, S.-M. Park, S.-H. Park, C.-W. Kwon, I.-S. Yoon, A study on the development of a model for predicting the number of highway traffic accidents using deep learning, J. Korean Soc. 17 (2018) 14–25.
[50] D.P. Kingma, L.J. Ba, ADAM: a method for stochastic optimization, Int. Conf. Learn. Represent. 9 (2018) 1–15.
[51] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. rocess. Syst. 25 (2012) 1097–1105, https://doi.org/10.1145/3065386.
[52] L.L. Nathans, F.L. Oswald, K. Nimon, Interpreting multiple linear regression: a guidebook of variable importance, Practical Assess. Res. Eval. 17 (2012) n9. Avialable at, https://files.eric.ed.gov/fulltext/EJ977607.pdf.
[53] M. Krzywinski, N. Altman, Multiple linear regression, Nat. Methods 12 (2015) 1103–1104.
[54] J.M. Kim, S.G. Yum, H. Park, J. Bae, A deep learning algorithm-driven approach to predicting repair costs associated with natural disaster indicators: the case of accommodation facilities, J. Build. Eng. 42 (2021) 103098, https://doi.org/10.1016/j.jobe.2021.103098.
[55] A. Dikshit, B. Pradhan, Interpretable and explainable AI (XAI) model for spatial drought prediction, Sci. Total Environ. 801 (2021) 149797, https://doi.org/10.1016/j.scitotenv.2021.149797.