

Research Article

Application of Stochastic Gradient Boosting Approach to Early Prediction of Safety Accidents at Construction Site

Yoonseok Shin 

Department of Architectural Engineering, Kyonggi University, Suwon 16227, Republic of Korea

Correspondence should be addressed to Yoonseok Shin; shinys@kyonggi.ac.kr

Received 28 July 2019; Revised 9 October 2019; Accepted 8 November 2019; Published 20 December 2019

Academic Editor: Dong Zhao

Copyright © 2019 Yoonseok Shin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The construction industry is one of the deadliest industries in the United States and Korea. The number of accidents at a construction site has been recently increasing despite institutional supports and managerial efforts. A proactive prediction of safety accidents is the best way to prevent them, but a dynamic change in particular conditions of a construction project makes the prediction very tricky and complicated. Moreover, preventive work for any safety accident at a construction site mainly depends on the intuitive and subjective opinions of practitioners with limited experience. The stochastic gradient boosting (SGB) approach may be an attractive alternative to conventional methods for predicting safety accidents because of its superior predictive performance. Therefore, SGB is applied to an early prediction of safety accidents at a construction site in order to examine its applicability to the construction safety domain. The prediction result of the proposed model is compared to an artificial neural network model and a decision tree model. The proposed model shows a slightly better result compared to the ANN and DT models. Moreover, the result of the proposed model also demonstrates the advantages of a simple parameter set in constructing a model and a comprehensible decision-making procedure for safety management.

1. Introduction

The construction industry is one of the largest but deadliest industries in the United States. According to a current state of fatal work injuries by industrial accidents surveyed in the US as of 2017, out of the total 5,147 workers, the fatal work injuries from the construction industry consist of 971, the transportation and warehouse sector of 882, the agriculture, forestry, fishing, and hunting of 581, and the professional and business services and others of 532 [1]. According to a current state of the death toll caused by industrial accidents by industry surveyed in Korea as of 2016, out of the total 1,777 workers, the death toll from the construction industry accounts for 554 people, the manufacturing industry for 408, the mining industry for 364, others for 293, and transport warehouse and communication for 129 [2]. The death toll reported from the construction field is the highest. Despite many managerial efforts made out of this poor performance, the accident statistics got worse in the past decade [3]. This might imply that the construction industry has still

employed ineffective, traditional approaches to safety management and that there is an urgent need for innovations [4].

That is, in order to prevent construction safety accidents (hereinafter construction accidents), proactive safety management that can predict risk in advance is needed rather than reactive safety management that analyzes the construction accident cases that already happened. Previous studies mainly focused on clarification of the relationship between the causes and construction accidents by analyzing the construction accident cases that had occurred [5–7]. However, due to its peculiar characteristics of a construction project, i.e., differences in scale, region, workers, contractor, and construction period, it is believed that each project hardly has identical risk. As a result, it is very complicated for a safety manager to accurately judge and predict how risky a possible safety accident is in the current project based on his/her own intuition and experience.

Various data mining approaches have been applied to solve this problem. To be more specific, a decision tree (DT)

[3, 8–10], an artificial neural network (ANN) [11–13], and a support vector machine (SVM) [14, 15] have been used to perform an intricate prediction of construction accidents, which is affected by various factors. DT is a very powerful and humanly comprehensible approach because of its intuitive binary “if-then” rule-based structure [16]. However, the instability of DT is known to researchers. In other words, a small change in the training samples can bring about a large difference not only in a tree but also in the analytical result [17]. ANN usually provides better results than the traditional simple-cut technique. However, this approach is a black-box technique [18], so it lacks human interpretability. SVM has attracted much attention because of its capacity for self-learning and high performance in generalization [19]. However, it needs a considerable time of trials and errors to determine a suitable kernel function [18]. Additionally, it has a high level of algorithmic complexity and needs extensive memory [16].

In the recent machine learning approaches, the boosting approach, which was culminated in the research by Freund and Schapire who introduced Adaboost M1 (also known as the AdaBoost algorithm), has drawn much attention because of its robustness and easy application [20]. The boosting approach was originally proposed to combine several weak classifiers to improve classification performance. Previous studies found that in terms of predictive performance, a boosting method is superior to the other competing methods including DT, ANN, and SVM, even when the dataset is defective and small [21–23]. In addition, the applicability of the boosting approach is much higher than that of competing techniques because of its single parameter [22]. Furthermore, the instability of DTs can be overcome by the boosting approach, i.e., growing a forest of DTs and classifying the instances of a majority vote for the classifications given by individual trees [24, 25].

In 2002, Friedman proposed stochastic gradient boosting (SGB) as a relatively new tree-based regression and classification method to optimize predictive performance [26, 27]. SGB is a hybrid that combines the advantages of both boosting and bagging approaches [26]. SGB provides several advantages such as a limited number of user-defined parameters and the ability to model nonlinear relationships, manage qualitative and quantitative variables, and remain robust despite missing values and outliers of data [28]. Furthermore, the stochastic component of SGB is powerful not only to provide superior predictive performance for parametric approaches but also to reduce the occurrence of overfitting by excluding a certain fraction of input data through a randomized approach, which is one of the representative characteristics of a bagging algorithm [26, 29].

These advantages allowed SGB actively utilized in the various domains, such as image classification [30, 31], mobile phone customer type discrimination [32], software reliability prediction [33], software maintainability prediction [34], genomic selection [35], financial fraud detection [36], airfoil self-noise prediction [37], rock burst damage prediction [38], drug combination prediction [39], vulnerable web component prediction [40], tree canopy cover estimation [29], pillar stability prediction [41], and

credit risk prediction [42]. In the construction domain, Shin firstly applied SGB to the preliminary construction cost estimation [43], and Tixier et al. secondly applied it to the construction injury prediction [44]. Unfortunately, there have been rarely tries to apply SGB to safety management, which recently needs innovative challenges in the construction domain. For this reason, in this study, SGB is applied to an early prediction of accidents at a construction site to examine its applicability to construction safety management.

In the next section, preventive efforts for construction accidents made in Korea will be discussed. The stochastic gradient boosting approach is going to be reviewed theoretically in the third section. An early prediction model based on SGB is applied to a dataset from the actual case data of construction accidents in Korea and is compared to that of ANN and DT. Finally, conclusions and suggestions for further study will be made.

2. Construction Accidents and Preventive Efforts in Korea

In Korea, the construction industry has experienced a serious labor shortage due to workers' avoidance of 3D, i.e., dirty, difficult, dangerous, jobs. In addition, the construction workers who are currently engaged are aging and are concerned that they would gradually become vulnerable to construction accidents. According to the statistical data of the Korean Statistical Information Service [2], fatal injuries by construction accidents in Korea are total of 2,596 for the past five years (2012–2016), as shown in Table 1. The fatal injuries of the construction industry were annually increasing to 554 in 2016 from 496 in 2012, even though the total industrial fatal injuries were annually decreasing for the same years. Fatal injuries are highly likely to occur, if any, in a construction project compared to other industries, due to its unique features including outdoor production, high place work, and heavy equipment work. Moreover, given that construction projects have been increasingly higher and scaler and that the number of foreign workers with different cultural backgrounds and languages has been increasing, safety accidents may not decrease in the construction industry [45].

The construction industry in Korea has made practical and institutional efforts to prevent construction accidents for a long time. For a reactive prediction of any construction accidents, many efforts have been made by analyzing accident data that had occurred to prepare comprehensive guidelines for construction accidents based on the statistical results by type in terms of cause, time, and the mechanism of the accidents. However, these efforts have a limitation in that various dynamically changing construction projects have not been considered. Furthermore, under Article 41 (2) of the Industrial Safety and Health Law, it is imperative to make efforts to reduce accidents by evaluating the degree of risk assessment in advance for each work at each construction site [46]. Risk assessment is a series of processes that identify hazards and risk factors of a workplace, estimate and determine the probability and severity of injuries or illnesses

TABLE 1: Fatal injuries in the Korean construction industry and overall industry.

Year	Construction (persons) (A)	Overall (persons) (B)	Remark (%) ((A/B) * 100)
2016	554	1,777	31.18
2015	493	1,810	27.24
2014	486	1,850	26.27
2013	567	1,929	29.39
2012	496	1,864	26.61
Sum	2,596	9,230	—

caused by hazards and risk factors, and establish and implement reduction measures. Under the law, each site conducts a five-phase evaluation of preparation of risk assessment, hazards identification using checklist, risk estimation, risk evaluation, and risk control and implementation. However, this method also has a limitation in that safety managers or practitioners make decisions based on subjective and limited experience and judgment. Recently, state-of-the-art technology, i.e., Big data analysis, Internet of Things, and image processing, applied to construction safety management [47–49]. And data mining techniques, i.e., decision tree analysis, association rule analysis, and machine learning, also applied to predict construction accidents proactively [3, 5, 10, 12, 13]. Although these efforts are not enough for practical application in the construction project yet, scientific and systematic decision support tools to help the safety managers proactively predict how risky a possible safety accident are needed in the long term.

3. Stochastic Gradient Boosting Approach

SGB proposed by Friedman is a hybrid of bagging and boosting approaches [26]. SGB is an ensemble learning algorithm combined with boosting and decision tree and makes a prediction by weighting the ensemble members of all trees as shown in Figure 1 [50].

In the iteration, the new model is built along the gradient descent direction of the loss function of the previous tree. The essence of SGB is to minimize this loss function between the classification function and real function by training the classification function $F^*(x)$. The construction accident prediction refers to the classification problem. For multiple classifications, the surrogate loss function can be considered as the following equations [26].

The loss function can be expressed as shown in the following equation:

$$\begin{aligned}
 \psi(\{y_k, F_k(x)_1^K\}) &= -\sum_{k=1}^K y_k \log[p_k(x)] \\
 &= -\sum_{k=1}^K y_k \log[P_r(y_k = 1 | x)] \quad (1) \\
 &= -\sum_{k=1}^K y_k \log\left[\frac{\exp(F_k(x))}{\sum_{l=1}^K \exp(F_l(x))}\right],
 \end{aligned}$$

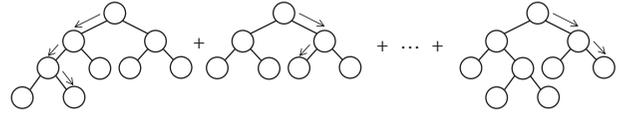


FIGURE 1: Ensemble of boosting the decision trees.

where $X = \{x_1, x_2, \dots, x_n\}$ is the input variables, k is the number of classes, y is the output variable, and $p_k(x)$ is the probability.

Then, the following equation can be obtained:

$$\begin{aligned}
 \tilde{y}_i^k &= -\left[\frac{\partial \psi \{y_{ij}, F_j(x_i)\}_{j=1}^K}{\partial F_k(x_i)} \right]_{\{F_j(x)=F_{j,m-1}(x)\}_1^K} \quad (2) \\
 &= y_i^k - p_k(x_i),
 \end{aligned}$$

where K -trees are induced, each of which to predict the corresponding current residuals $\{y_i^k - p_k(x_i)\}_{i=1}^N$. This produces K -trees each with L -terminal nodes at iteration m , Rklm. As mentioned above, a separate line search is performed in each terminal node l of each tree k , as shown in the following equation:

$$Y_{klm} = \arg \min_{x_i \in Y_{klm}} \phi(y_i^k, F_{k,m-1}(x_i) + Y), \quad (3)$$

where $\phi_k = -y_k \log[p_k(x)]$, and each of the functions is updated and then SGB is established.

4. Application

4.1. Data Collection and Analysis. Data from 22,935 construction accident cases were collected from construction projects executed by general contractors in Korea. These were overall data to be reported to the Korea Occupational Safety and Health Agency (KOSHA) in 2014 [51]. Then, valid 16,248 cases were used in this study by excluding the invalid data which had been ambiguously classified or had many missing values. The collected data of 16,248 cases were divided randomly into 30 test datasets and 16,218 training data.

Generally, the prediction accuracy of prediction of construction accidents on-site correlates with the amount of accident information available concerning for project scale, project location, project type, season, and others. In this study, the factors for predicting the construction accidents were determined in two steps. Firstly, a list of factors affecting a construction accident prediction was made by

TABLE 2: Factors of construction accidents on-site.

Variable	Type	Feature		
Dependent variable (output)	Type of accident	Nominal	Fall (1), wipeout (2), hit (3), cut (4), narrowness (5), crash (6)	
	Gender		Male (1), female (2)	
	Day of the week		Weekdays (1), weekend (2)	
	Location		Seoul (1), capital area (except for Seoul) (2), etc. (3)	
	Nationality		Korean (1), foreigner (2)	
	Season		Spring (1), summer (2), Autumn (3), winter (4)	
	Time	Nominal	06–12 hour (1), 12–18 hour (2), 18–22 hour (3), 22–06 hour (4)	
	Type of employment		Regular (1), nonregular (2)	
	Independent variable (input)	Type of occupation		Management (1), engineering (2), labor (3), operation (4)
		Type of project		Residential and commercial (1), public and healthy (2), etc. (3)
Age			Integer (17–83)	
Project duration			Integer (1–13,752)	
Project scale			Integer (0.1–2,000)	
Rate of process		Numerical	Real number (5–90)	
Workforce			Integer (5–3,000)	
Year of service			Real number (0.8–30)	

reviewing the construction accident case database given by KOSHA. Lastly, appropriate factors were selected from the list through interviews with construction practitioners who were very experienced in safety management in Korea. Consequently, 15 factors, i.e., input variables, were selected in this study, as indicated in Table 2. And the construction accident prediction results were representatively classified into 6 categories: fall, wipeout, hit, cut, narrowness, and crash, which account for over 87% of accidents occurring at construction sites in 2014 [13].

4.2. Applying SGB to Predicting the Construction Accident.

In this study, the SGB-based prediction model for construction accidents is evaluated by applying to actual cases of construction accidents. A construction accident was predicted using SGB as follows: (1) The classification function $F^*(x)$ was trained using training data. In the dataset, each “ x_i ” of the training set has project scale, project duration, rate of process, and others. Each result, i.e., the accident type, was allocated to “ y_i ”. (2) After training was completed using parameters such as the number of additive trees, learning rate (shrinkage coefficient), and a maximum and minimum number of levels, the series of trees $F^*(x)$ which maps “ x ” to “ y ” of the training dataset (y_i, x_i) with a minimized loss function $(\psi(\{y_k, F_k(x)_1^k\}))$ was found. (3) The expected value of $F^*(x)$, i.e., expected accident, was calculated for a new case of the test dataset (y_j, x_j) .

The construction accident prediction model of this study was developed using STATISTICA software. STATISTICA employs an implementation of the method usually referred to as SGB by Friedman [26]. To conduct an SGB training procedure, the parameters, i.e., the learning rate, the number of additive trees, and subsampling proportion, have to be set, as shown in Figure 2. At first, the learning rate was set at “0.1” as it leads to better results in terms of prediction error

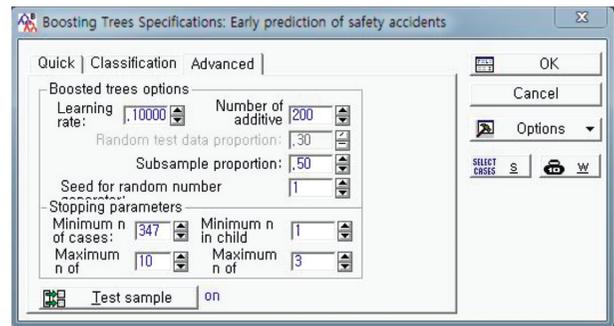


FIGURE 2: Parameter setting for the SGB model.

according to previous empirical studies [26, 52]. The subsample proportion and the seed for random number generators were also set to 0.5 and 1, respectively, according to a previous study [40]. The number of additive terms was set to 200. Figures 3 and 4 show how consecutive trees in the boosting step improve the quality of the SGB model which has randomly selected training data and testing data in each dataset. As additive terms are added to the model, it is observed that the prediction error of the training data continued decreasing. And a certain point that the error estimation for the testing data started to increase means the optimal number of trees to avoid overfitting. As a result of training, the SGB model in this study showed that the optimal number of additive trees was 60, and the maximum tree size was 3, as shown in Figure 4.

For the verification of the SGB model, the same cases were applied to the ANN and DT models to compare the results. It is because ANN and DT approaches showed superior performance with prediction accuracy in previous studies [3, 13]. Statistical Package for the Social Sciences (SPSS) Statics 19 was used to construct ANN and DT models

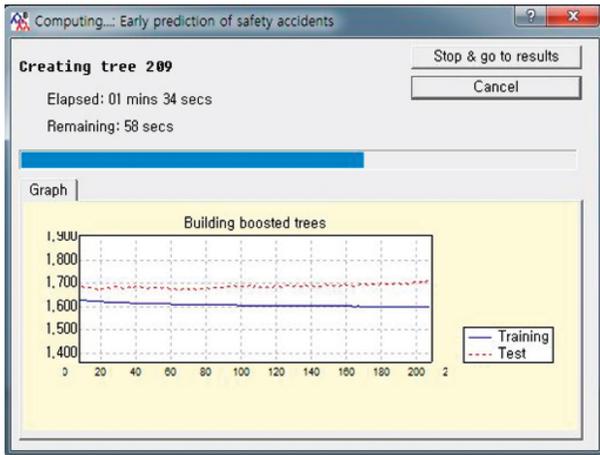


FIGURE 3: Training process of the SGB model.

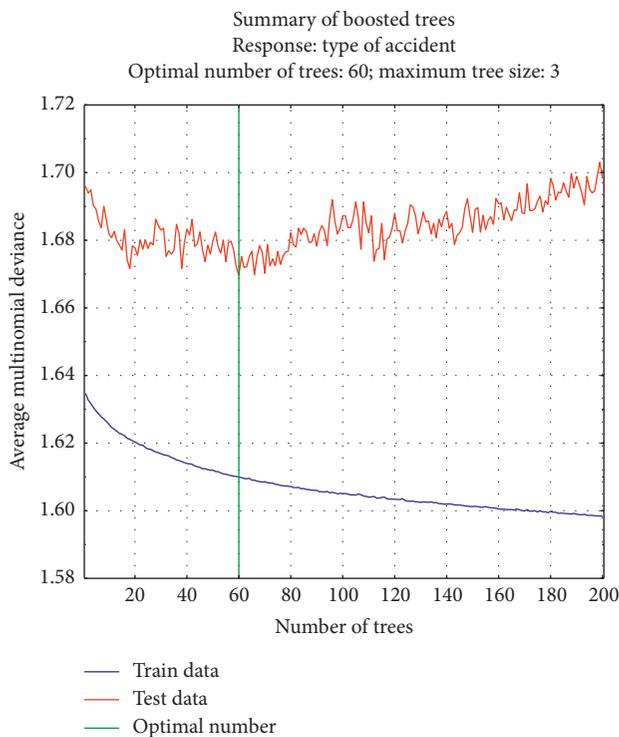


FIGURE 4: Training results of the SGB model.

in this study. To construct model using ANN and DT, optimal parameters have to be selected beforehand: a learning rate, the number of hidden neurons, the momentum for ANNs and the maximum tree depth, the number of minimum cases for high and low-rank node, and level of significance of chi-square automatic interaction detection (CHAID) method, etc. for DT. We determine the values from repeated experiments.

5. Results and Discussion

5.1. Results of Evaluation. In the evaluation of the three prediction models, this study includes the second and the

third most likely construction accidents along with the first most likely construction accident recommended. The output results of this study are based on discrete variables, not on continuous variables. For this reason, it is believed more reasonable to predict a construction accident by considering the three most likely accidents (from the 1st to 3rd highly possible accidents) rather than the most likely accident in each model. Table 3 indicates a summary of the results from the 30 test data using SGB, ANN, and DT. This table consists of the actual values (target) and the predicted results of SGB, ANN, and DT models. However, there are the first, second, and third prediction values based on probabilities of six construction accidents and the cumulative prediction accuracy from the order of the first place arranged rather than presenting the first prediction value only.

The SGB model provided comprehensible information about new cases for prediction, which is an advantage inherent in the decision tree. Firstly, the importance of each dependent variable in predicting construction accident was estimated, as shown in Figure 5. These values indicate the relative importance of each variable by assigning 100% to the highest value and then by scaling the others accordingly. Lastly, the tree structures of the model are given as shown in Figure 6. This shows the prediction rules, such as the applied variables, and their influence in the proposed model. So, an intuitive understanding of the whole structure of the model is possible.

5.2. Discussion of Results. This study was conducted using the same test set and the training set of construction accident data. In terms of prediction accuracy of the most likely accident, SGB showed a slightly higher result of 36.6% compared to ANN of 33.3% and to DT of 33.3%. When the second and the third most likely accidents were included, ANN showed the best result of 80.0%, which is higher compared to SGB of 76.6% and to DT of 70%. In construction accident prediction, it is difficult to conclude that in terms of performance, the SGB model is superior to both of the ANN and DT models because the gap of accuracies among the models is very slight. However, the even similar performance of the SGB model is noticeable because ANN and DT models have been proven the superior performance with prediction accuracy of construction accidents in previous studies. Additionally, the SGB model provided additional information, i.e., importance plot and structure model, which helps the safety managers and practitioners comprehend the decision-making process intuitively. Moreover, SGB was a simple and easy to construct model compared to ANN and DT because it has only three parameters.

Consequently, these results reveal that SGB, which is a relatively new AI approach to construction safety management, has the potential to apply to the construction prediction. It will help safety managers and practitioners solve the difficult problem of an early prediction based on the accident cases occurring in the past and their own limited experience due to the dynamic changes in some conditions of construction projects. In addition, SGB is possibly utilized in many areas since the boosting approach

TABLE 3: Comparison of prediction results of SGB, ANN, and DT models.

No.	Target	SGB model prediction			ANN model prediction			DT model prediction		
		1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
1	5	5	6	3	5	2	4	2	5	1
2	6	5	6	3	5	2	4	1	6	2
3	2	2	3	1	1	2	4	2	3	1
4	1	3	1	6	1	6	4	1	2	3
5	1	5	1	3	1	5	4	1	3	2
6	3	3	6	2	4	1	2	2	3	1
7	3	2	1	3	1	4	2	2	3	1
8	4	4	2	3	1	4	2	2	3	1
9	5	4	5	1	5	2	4	1	6	2
10	4	3	4	2	1	4	2	2	3	1
11	2	4	3	1	4	1	2	4	3	1
12	2	2	3	1	2	1	4	2	3	1
13	5	5	6	1	1	5	4	1	5	2
14	1	4	1	2	4	2	1	2	3	1
15	1	2	1	3	1	2	4	1	2	3
16	3	4	6	5	4	1	2	2	3	1
17	4	2	1	4	1	4	2	1	2	3
18	3	2	1	3	1	2	3	2	3	1
19	2	1	5	6	5	2	4	2	1	3
20	5	1	5	4	2	1	4	1	6	2
21	1	5	1	3	2	4	1	2	1	3
22	2	5	6	4	2	5	4	2	1	3
23	1	1	2	3	2	1	4	2	3	1
24	1	6	2	5	1	6	4	1	2	3
25	6	2	1	3	2	4	1	2	3	1
26	4	4	3	2	1	2	4	1	4	3
27	1	1	2	3	1	2	5	1	2	3
28	4	4	1	6	1	2	4	2	3	1
29	4	4	1	2	1	4	6	1	4	3
30	1	4	3	1	1	6	2	1	4	3
Accumulated accuracy		1st 36.6% (11/30)	1st-2nd 63.3% (19/30)	1st-2nd-3rd 76.6% (23/30)	1st 33.3% (10/30)	1st-2nd 60.0% (18/30)	1st-2nd-3rd 80.0% (24/30)	1st 33.3% (10/30)	1st-2nd 63.3% (19/30)	1st-2nd-3rd 70.0% (21/30)

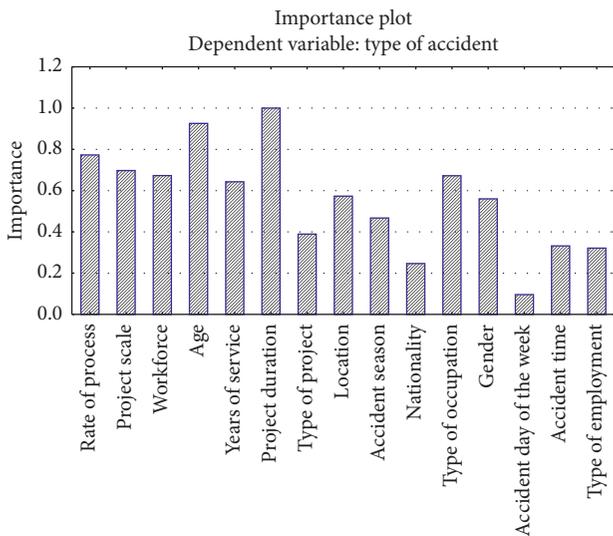


FIGURE 5: Importance of dependent variables of the SGB model.

can use the existing AI techniques such as ANN and SVM as well as DT as a base learner in the boosting algorithms. The construction accident prediction method proposed in this

Tree graph for type of accident
Num. of nonterminal nodes: 1, num. of terminal nodes: 2
Tree number: 3; category: 4

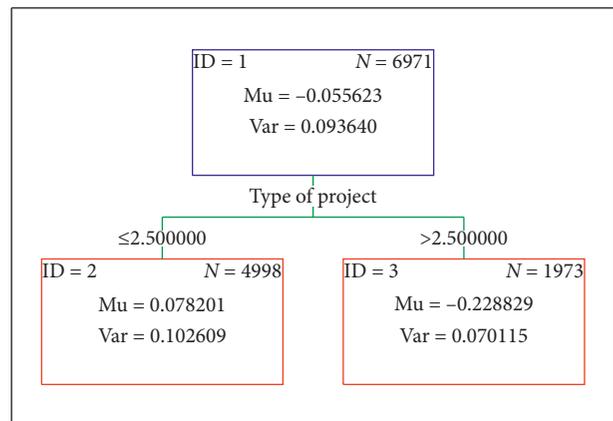


FIGURE 6: A fragment of ensemble trees.

study is ultimately expected to improve the effectiveness of safety management by helping safety managers and practitioners review possible construction accidents from a wider perspective.

6. Conclusion

This study proposed an early prediction model for construction accidents to assist the decision-making of safety managers in the construction field. To validate the applicability of the proposed model to construction safety management, its results were compared to those of ANN and DT models, which have proven to have a high predictive performance of construction accident. In an experiment with three models, the SGB model showed similar results using actual case data for a construction accident in Korea. Moreover, the SGB model can provide additional information about the factors for predicting the construction accidents to support safety managers and practitioners to comprehend the decision-making procedure. Lastly, the SGB model is easier to construct than the ANN and DT models because it needs fewer parameters. These results demonstrate that SGB has the dual advantages of boosting and decision trees. SGB has the considerable potential to be a leading technique for the next generation of construction safety management system.

In this study, an examination was carried out on the application of SGB in the early prediction of a construction accident. Although the prediction result from SGB is better than that of ANN or DT, it is difficult to conclude the performance of the model is enough for practical applications. Therefore, further detailed analysis for the quality of the collected data and statistical verification for factors in construction accident occurrence will be necessary to utilize the proposed model for an actual new project.

Data Availability

The construction accident data used to support the findings of this study have been deposited in the website depository of the Korea Occupational Safety and Health Agency (<http://www.kosha.or.kr/kosha/data/construction.do>). The depository provides representative cases of construction accidents occurred in Korea. The access to overall data needs official requests to an agent in charge.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by Kyonggi University Research Grant 2017.

References

- [1] Bureau of Labor Statistics (BLS), *Number and Rate of Fatal Work Injuries, by Industry*, Bureau of Labor Statistics (BLS), Washington, DC, USA, 2019, <https://www.bls.gov/charts/census-of-fatal-occupational-injuries/number-and-rate-of-fatal-work-injuries-by-industry.htm>.
- [2] The Korean Statistical Information Service (KOSIS), "Status of deaths by occurrence," The Korean Statistical Information Service (KOSIS), Daejeon, Republic of Korea, 2019, http://kosis.kr/statHtml/statHtml.do?orgId=118&tblId=DT_11806_N022&vw_cd=MT_ZTITLE&list_id=118_11806_ciek6458&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE.
- [3] Y. Cho, Y. C. Kim, and Y. Shin, "Prediction model of construction safety accidents using decision tree technique," *Journal of the Korea Institute of Building Construction*, vol. 17, no. 3, pp. 294–303, 2017.
- [4] B. Esmaeili and M. R. Hallowell, "Diffusion of safety innovations in the construction industry," *Journal of Construction Engineering and Management*, vol. 138, no. 8, pp. 955–963, 2011.
- [5] D. G. Lee and D. E. Lee, "Data mining in construction fatal accident: association rule mining," in *Proceeding of the Association of Architectural Institute of Korea*, pp. 426–430, Busan, Republic of Korea, December 2010.
- [6] J. H. Kim, "The improvement of providing protective equipment preventing a fall in small scale construction site," Master's thesis, Kyunghee University, Suwon, Republic of Korea, 2013.
- [7] A. Karimi and E. Kashi, "Investigating the effect of geometric parameters influencing safety promotion and accident reduction (case study: Bojnurd-Golestan National Park road)," *Cogent Engineering*, vol. 5, no. 1, pp. 1–14, 2018.
- [8] T. Rivas, M. Paz, J. E. Martín, J. M. Matías, J. F. García, and J. Taboada, "Explaining and predicting workplace accidents using data-mining techniques," *Reliability Engineering & System Safety*, vol. 96, no. 7, pp. 739–747, 2011.
- [9] G. Mistikoglu, I. H. Gerek, E. Erdis, P. E. Mumtaz Usman, H. Cakan, and E. E. Kazan, "Decision tree analysis of construction fall accidents involving roofers," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2256–2263, 2015.
- [10] S. Sarkar, A. Patel, S. Madaan, and J. Maiti, "Prediction of occupational accidents using decision tree approach," in *Proceeding of the 2016 IEEE Annual India Conference (INDICON)*, pp. 1–6, Bangalore, India, December 2016.
- [11] Y. M. Goh and D. Chua, "Neural network analysis of construction safety management systems: a case study in Singapore," *Construction Management and Economics*, vol. 31, no. 5, pp. 460–470, 2013.
- [12] D. A. Patel and K. N. Jha, "Neural network model for the prediction of safe work behavior in construction projects," *Journal of Construction Engineering and Management*, vol. 141, no. 1, Article ID 04014066, 2015.
- [13] Y.-C. Kim, W. S. Yoo, and Y. Shin, "Application of artificial neural networks to prediction of construction safety accidents," *Korean Society of Hazard Mitigation*, vol. 17, no. 1, pp. 7–14, 2017.
- [14] F. Yajuan and C. Jia, "Study on prediction model of building construction safety accidents based on GA-SVM," in *Proceedings of the 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering*, pp. 460–462, Xi'an, China, November 2013.
- [15] Y. Zhou, W. Su, L. Ding, and H. Luo, "Predicting safety risks in deep foundation pits in subway infrastructure projects: support vector machine approach," *Journal of Computing in Civil Engineering*, vol. 31, no. 5, Article ID 04017052, 2017.
- [16] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review," *European Journal of Operational Research*, vol. 180, no. 1, pp. 1–28, 2007.
- [17] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, "Boosted decision trees as an alternative to

- artificial neural networks for particle identification,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2-3, pp. 577–584, 2005.
- [18] S. H. An, U. Y. Park, K. I. Kang, M. Y. Cho, and H. H. Cho, “Application of support vector machines in assessing conceptual cost estimates,” *Journal of Computing in Civil Engineering*, vol. 21, no. 4, pp. 259–264, 2007.
- [19] M.-Y. Cheng, H.-S. Peng, Y.-W. Wu, and T.-L. Chen, “Estimate at completion for construction projects using evolutionary support vector machine inference model,” *Automation in Construction*, vol. 19, no. 5, pp. 619–629, 2010.
- [20] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *Machine Learning: Proceedings of the Thirteenth International Conference*, vol. 13, pp. 148–156, Morgan Kaufmann, San Francisco, CA, USA, 1996.
- [21] D. Arditi and T. Pulket, “Predicting the outcome of construction litigation using boosted decision trees,” *Journal of Computing in Civil Engineering*, vol. 19, no. 4, pp. 387–393, 2005.
- [22] Y. Shin, D.-W. Kim, J.-Y. Kim, K.-I. Kang, M.-Y. Cho, and H.-H. Cho, “Application of AdaBoost to the retaining wall method selection in construction,” *Journal of Computing in Civil Engineering*, vol. 23, no. 3, pp. 188–192, 2009.
- [23] Y. Shin, T. Kim, H. Cho, and K.-I. Kang, “A formwork method selection model based on boosted decision trees in tall building construction,” *Automation in Construction*, vol. 23, pp. 47–54, 2012.
- [24] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 161–168, Pittsburgh, PA, USA, June 2006.
- [25] J. Bastos, “Credit scoring with boosted decision trees. MPRA paper,” *Munich Personal RePEc Archive*, vol. 8156, pp. 1–13, 2008.
- [26] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [27] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, New York, NY, USA, 2013.
- [28] J. Zhou, X. Li, and H. S. Mitri, “Comparative performance of six supervised learning methods for the development of models of hard rock pillar stability prediction,” *Natural Hazards*, vol. 79, no. 1, pp. 291–316, 2015.
- [29] S. Godinho, N. Guiomar, and A. Gil, “Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm,” *International Journal of Remote Sensing*, vol. 39, no. 14, pp. 4640–4662, 2018.
- [30] R. Lawrence, A. Bunn, S. Powell, and M. Zambon, “Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis,” *Remote Sensing of Environment*, vol. 90, no. 3, pp. 331–336, 2004.
- [31] L. Li, Y. Wu, and M. Ye, “Multi-class image classification based on fast stochastic gradient boosting,” *Informatica*, vol. 38, no. 3, pp. 145–153, 2014.
- [32] D. Steinberg, M. Golovnya, and N. S. Cardell, “Mobile phone customer type discrimination via stochastic gradient boosting,” *International Journal of Data Warehousing and Mining*, vol. 3, no. 2, pp. 32–53, 2007.
- [33] N. Kiran and V. Ravi, “Software reliability prediction by soft computing techniques,” *Journal of Systems and Software*, vol. 81, no. 4, pp. 576–583, 2008.
- [34] M. Elish and K. Elish, “Application of treenet in predicting object-oriented software maintainability: a comparative study,” in *Proceedings of the 13th European Conference on Software Maintenance and Reengineering (CSMR ’09)*, pp. 69–78, Kaiserslautern, Germany, March 2009.
- [35] J. O. Ogutu, H. P. Piepho, and T. Schulz-Streeck, “A comparison of random forests, boosting and support vector machines for genomic selection,” *BMC Proceedings*, vol. 5, no. S3, p. 11, 2011.
- [36] D. G. Whiting, J. V. Hansen, J. B. McDonald, C. Albrecht, and W. S. Albrecht, “Machine learning methods for detecting patterns of management fraud,” *Computational Intelligence*, vol. 28, no. 4, pp. 505–527, 2012.
- [37] A. Patri and Y. Patnaik, “Random forest and stochastic gradient tree boosting based approach for the prediction of airfoil self-noise,” *Procedia Computer Science*, vol. 46, pp. 109–121, 2015.
- [38] J. Zhou, X.-Z. Shi, R.-D. Huang, X.-Y. Qiu, and C. Chen, “Feasibility of stochastic gradient boosting approach for predicting rockburst damage in burst-prone mines,” *Transactions of Nonferrous Metals Society of China*, vol. 26, no. 7, pp. 1938–1945, 2016.
- [39] Q. Xu, Y. Xiong, H. Dai et al., “PDC-SGB: prediction of effective drug combinations using a stochastic gradient boosting algorithm,” *Journal of Theoretical Biology*, vol. 417, pp. 1–7, 2017.
- [40] M. Elish, “Enhance prediction of vulnerable web components using stochastic gradient boosting trees,” *International Journal of Web Information System*, vol. 15, no. 2, pp. 201–214, 2018.
- [41] H. Ding, G. Li, X. Dong, and Y. Lin, “Prediction of pillar stability for underground mines using the stochastic gradient boosting technique,” *IEEE Access*, vol. 6, pp. 69253–69264, 2018.
- [42] K. Halteh, K. Kumar, and A. Gepp, “Using cutting-edge tree-based stochastic models to predict credit risk,” *Risks*, vol. 6, no. 2, p. 55, 2018.
- [43] Y. Shin, “Application of boosting regression trees to preliminary cost estimation in building construction projects,” *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 149702, 9 pages, 2015.
- [44] A. J.-P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, “Application of machine learning to construction injury prediction,” *Automation in Construction*, vol. 69, pp. 102–114, 2016.
- [45] S.-H. An, H.-W. Ryu, and M.-S. Park, “Selecting key accident risk factors in building construction projects,” *Journal of the Korea Institute of Building Construction*, vol. 11, no. 1, pp. 45–50, 2011.
- [46] Korea Ministry of Government Legislation, *Industrial Safety and Health Act*, Korea Ministry of Government Legislation, Sejong City, Republic of Korea, 2019, <http://www.law.go.kr/lsInfoP.do?lsiSeq=203199&efYd=20181018#0000>.
- [47] M. Y. Hong, J. Y. Park, H. R. Lee, C. E. Im, H. J. Jang, and J. W. Son, “A study on the development of smart construction safety management framework using big data analysis based on the internet of things,” in *Proceeding of 2017 Conference of Architectural Institute of Korea*, pp. 1170–1175, Gyeongju, Republic of Korea, 2017.
- [48] H. Ryu, “Safety management method for construction workers based on IoT(internet of things),” in *Proceeding of 2017 Conference of Architectural Institute of Korea*, pp. 873–874, Jeju, Republic of Korea, 2017.

- [49] S.-M. Hong, B.-C. Kim, T.-W. Kwon, J.-H. Kim, and J.-J. Kim, "A study on prevention of construction opening fall accidents introducing image processing," *Journal of KIBIM*, vol. 6, no. 2, pp. 39–46, 2016.
- [50] J. Ye, J. H. Chow, J. Chen, and Z. Zheng, "Stochastic gradient boosted distributed decision trees," in *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pp. 2–6, Hong Kong, China, November 2009.
- [51] Korea Occupational Safety and Health Agency (KOSHA), *The Accident Cases of Construction Industry*, Korea Occupational Safety and Health Agency (KOSHA), Seoul, Republic of Korea, 2019, <http://www.kosha.or.kr/kosha/data/construction.do>.
- [52] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *Statistics in Medicine*, vol. 22, no. 9, pp. 1365–1381, 2003.

