ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Framework for improving the prediction rate with respect to outdoor thermal comfort using machine learning

Jaemin Jeong[1], Jaewook Jeong[2]*, Minsu Lee[3], Jaehyun Lee[4]

[1] *Department of safety engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: ss96011@seoultech.ac.kr
[2] *Department of safety engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: jaewook.jeong@seoultech.ac.kr
[3] *Department of safety engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: alstn9881@seoultech.ac.kr
[4] *Department of safety engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: archi0528@seoultech.ac.kr

Abstract: Most of the construction works are conducted outdoors, so the construction workers are affected by weather conditions such as temperature, humidity, and wind velocity which can be evaluated the thermal comfort as environmental factors. In our previous researches, it was found that construction accidents are usually occurred in the discomfort ranges. The safety management, therefore, should be planned in consideration of the thermal comfort and measured by a specialized simulation tool. However, it is very complex, time-consuming, and difficult to model. To address this issue, this study is aimed to develop a framework of a prediction model for improving the prediction accuracy about outdoor thermal comfort considering environmental factors using machine learning algorithms with hyperparameter tuning. This study is done in four steps: i) Establishment of database, ii) Selection of variables to develop prediction model, iii) Development of prediction model; iv) Conducting of hyperparameter tuning. The tree type algorithm is used to develop the prediction model. The results of this study are as follows. First, considering three variables related to environmental factor, the prediction accuracy was 85.74%. Second, the prediction accuracy was 86.55% when considering four environmental factors. Third, after conducting hyperparameter tuning, the prediction accuracy was increased up to 87.28%. This study has several contributions. First, using this prediction model, the thermal comfort can be calculated easily and quickly. Second, using this prediction model, the safety management can be utilized to manage the construction accident considering weather conditions.

Key words: Outdoor thermal comfort, physiological equivalent temperature, prediction model, machine learning, bayesian optimizer

1. INTRODUCTION

The construction workers mainly do their jobs at the outdoor environmental conditions. So, they are affected by various environmental factors such as temperature, humidity, wind, and air quality

[1]. When the construction workers are exposed to a poor climate conditions, performance of worker is significantly affected [1-4]. For example, the hot temperature can be caused by various heat diseases, and cold temperature can affect negative cognitive [1]. In this regard, the poor weather condition can be connected with the construction workers' accidents. Since the number of fatalities and injuries in construction are bigger than those of other industries, the construction industry is known as the most dangerous one [1].

Refer to the previous research, construction accidents are presumed to be related to workers' thermal uncomfortable [1]. Lee et al. (2021) presented a result of the frequency of fatality and injury accident considering the thermal comfort for over 10 years. Over 80% construction accidents were occurred in the uncomfortable range which has been defined by 18ºC or over 23ºC ('ºC' is called by celcious as PET unit) [1]. Therefore, it is also necessary to consider safety management according to thermal comfort.

In the previous researches, the predicted mean vote (PMV) is used to evaluate the thermal sensation of humans. PMV, however, is focused on the indoor condition, which can be controlled by HVAC. Therefore, it is not suitable for use in uncontrollable outdoor conditions [1-4]. To address this issue, the Physiological Equivalent Temperature (PET) is used as one of the thermal comfort index to evaluate outdoor conditions [1,3,5-6]. However, since it is difficult to calculate PET manually, specialized simulation tool is required [1,7-8]. So, it can be a hurdle to utilize the PET for the safety manager at the construction site. To address this issue, previous researches suggested several prediction methods for the PET [4-6]. First, PET under specific conditions was predicted through experiments [6]. Second, the PET is predicted using various machine learning algorithms such as regression analysis, decision tree, and Support Vector Machine [4-5]. However, there have some limitations as below. First, it can only be used under specific conditions as these researches are analyzed with a small amount of data [4-6]. Second, the prediction accuracy is not very high [4].

In this regards. This study aims to suggest the framework of a prediction model for Physiological Equivalent Temperature considering database. This framework can provide high prediction accuracy for PET considering only the environmental factors which can be easily collected at the construction site.

2. MATERIALS AND METHODS

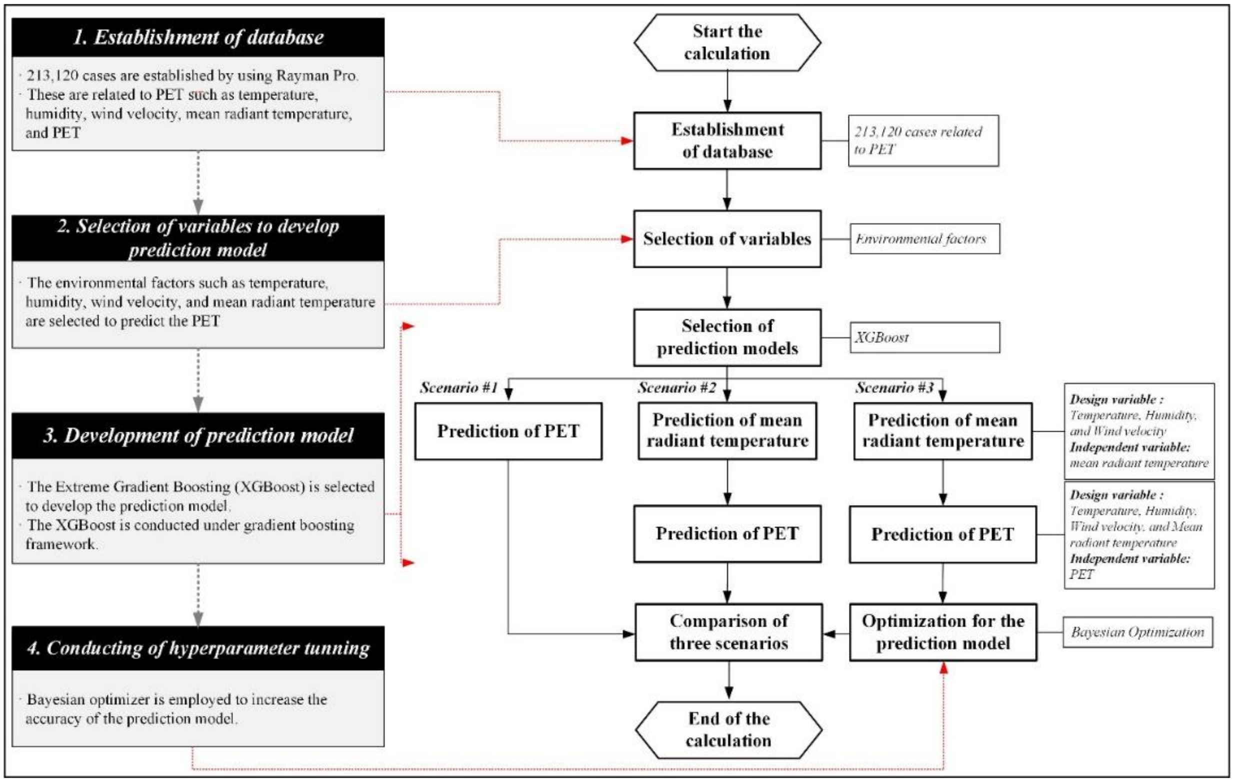

Figure 1. Research and decision making process

As shown in Figure 1, This study is done in four steps: i) Establishment of database, ii) Selection of variables to develop prediction model, iii) Development of prediction model; iv) Conducting of hyperparameter tuning.

2.1. Establishment of database

In this study, the database is established to develop the prediction model for PET. The actual accident cases in construction and climate factors among 2006 to 2019 were collected. Lee et al. (2021) calculated 213,120 cases for PET using RaymanPro, a specialized simulation tool in consideration of fatality and injury accident in the construction industry for 10 years [1].

This study utilizes 213,120 cases to develop the prediction model for PET. The collected data are included with regard to air temperature (t_a), relative humidity (rh), wind velocity (v), mean radiant temperature (t_{mrt}), and PET.

2.2. Selection of variables to develop prediction model

There are two major variable groups in the thermal comfort calculation as environmental factors and personal factors. First, environmental factors which are comprised of air temperature (t_a), relative humidity (rh), wind velocity (v), mean radiant temperature (t_{mrt}). t_a, rh, and v can be collected from the Korea Meteorological Administration (KMA) in South Korea, day by day. However, t_{mrt} is hard to measure and so a globe thermometer is needed to measure radiant heat directly from a warm object. Environmental factors are important to measure the PET [1-2,4-5]. Second, personal factors are made up of metabolic rate and clothing. Metabolic rate is the activity level of the body surface area. And clothing is associated with the thermal performance of an individual's clothing [1,9]. However, refer to the previous researches, when calculating the PET using RaymanPro, the metabolic rate and the clothing are fixed as 80 w and 0.9 clo. So, PET is not affected by metabolic rate and clothing [1]. By contrast the personal factors, the PET is mainly affected by four environmental factors (t_a, rh, v, and t_{mrt}) [1,5]. So, environmental factors are selected as variables to develop the prediction model.

2.3. Development of prediction model

This study is aimed to develop a framework for the prediction model which can be easily used by site safety managers. As mentioned above, the PET is mainly affected by the environmental factors such as t_a, rh, v, and t_{mrt}. These, therefore, can be selected as design variables. However, except for t_{mrt}, because others can easily be collected from the KMA in South Korea, the site manager can collect these data every single day. However, t_{mrt} is hard to measure and so a globe thermometer is needed to measure radiant heat directly from a warm object. It is not easy for site managers to collect. However, t_{mrt} is also one of the important variables that have great influence on the prediction model for PET. It can be guessed that if four environmental factors including t_{mrt} are considered, the prediction accuracy would be increased compared to three environmental factors.

In this study, three scenarios are considered to suggest the prediction model with the highest prediction accuracy. As shown in Figure 1, three scenarios in this paper are as follows.

Scenario #1 : The PET is predicted by considering three environmental factors such as t_a, rh, and v as design variables which are easily collected from KMA at a construction site.

Scenario #2 : Before prediciting the PET, the t_{mrt} is predicted in consideration with t_a, rh, and v. And then, the PET is predicted by considering four environmental factors such as t_a, rh, v, and t_{mrt} as design variables.

Scenario #3: Before prediciting the PET, the t_{mrt} is predicted in consideration with t_a, rh, and v. And then, the PET is predicted by considering four environmental factors such as t_a, rh, v, and t_{mrt} as design variables. Hereafter, the hyperparameter tuning is employed for guaranteeing the highest prediction accuracy.

The Extreme Gradient Boosting (XGBoost) is an integrated algorithm based on gradient boosting, which is widely used in this area currently, is used to develop the prediction model [10-11]. Because cause-based decision trees and gradient boosting machine are integrated, XGBoost has a rapid process speed, excellent precision levels, and few overfitting problems. With such abilities, XGBoost can be used efficiently to develop a prediction model for database [10-12].

For this reason, this study selects XGBoost as machine learning algorithms to develop the prediction model for the PET with the highest prediction accuracy. And the collected data is classified by 70% and 30% as training set and test set.

Finally, This study carries out the k-fold cross validation to identify the performance evaluation and avoid overfitting problems for the developed prediction model. Before utilizing a machine learning, the dataset should be divided into training data set and test data set. After dividing the dataset, the training dataset is used for training the prediction model. After developing the prediction model using the training data set, the test data set is utilized to evaluate the ability of the prediction model [13]. The k-fold cross validation is performed by conducting the procedure of k times for training and testing. In particular, the k-fold cross validation can be evaluated accurately for a dataset [13]. In previous research, the k value was set to 5 to maximize the performance of CV [14]. So, refer to the previous researches, this study also sets k to 5.

2.4. Conducting of hyperparameter tuning

In this study, the hyperparameter tuning is employed to increase the prediction accuracy of machine learning algorithms. Prediction model can be greatly affected by hyperparameters. Therefore, hyperparameter tuning is very important to enhance the prediction accuracy. In previous studies, various methods were considered to conduct the hyperparameter tuning such as grid search, random search, and Bayesian optimizer which have been mainly applied [15]. The grid and random searches are a substantial iteration of hyperparameters tunings. Thus, it is difficult to find the optimal value of the prediction model [15]. As an alternative to these problems, Bayesian optimization, which can present the optimal prediction accuracy [15-17].

Four hyperparameters of the XGBoost are selected by considering the previous researches [12,18] (refer to Table 1). The maximum number of iterations is set to 1,000 iterations.

Table 1. Descriptions of Hyperparameters of XGBoost for Bayesian Optimization

Machine learning algorithm	**Hyper parameter**	**Description**	**Minimum**	**Maximum**
XGBoost	Learning rate (Float)	· Step size shrinkage used in update to prevent overfitting.	0	1
	Colsample bytree (Float)	· The subsample ratio of columns for constructing each tree	0	1
	Reg lambda (Float)	· L2 regularization term on weights	0	1
	Max depth (Int)	· The maximum depth of a tree	5	100

3. RESULTS AND DISCUSSION

3.1. Comparison of three prediction models for outdoor thermal comfort

As shown in Figure 2, the results of each scenario were presented. The detailed explanation is as follows.

For scenario #1, the prediction accuracy and 5-fold validation were 85.74% and 85.32% Temperature (t_a), Relative humidity (rh), and wind velocity (v) were considered as design variables and PET was considered as independent variables in the scenario #1. The prediction accuracy was very low compared to others.

For scenario #2, the prediction accuracy and 5-fold validation were 86.55% and 86.16. The prediction accuracy is slightly higher compared to scenarios #1. The prediction process was conducted by two stages. First, considering t_a, rh, and v, t_{mrt} was predicted first. At that time, the prediction accuracy of t_{mrt} was 64.66%. And then, through t_a, rh, v, and t_{mrt} (predicted value) as design variables, the PET was predicted. The prediction accuracy of PET was 86.55%. Although the prediction accuracy of t_{mrt} was low, the prediction accuracy of PET was increased compared to scenario #1.

For scenario #3, as the prediction accuracy and 5-fold validation were 87.28% and 86.00%. The prediction accuracy has the highest prediction accuracy compared to others. For scenario #3, the hyperparameter tuning was performed in scenario #3. Through the hyperparameter tuning with Bayesian optimization, the prediction accuracy was improved. After conducting hyperparameter tuning, ‘colsample_bytree’, ‘learning_rate’, ‘max_depth’, and ‘reg_lambda’ were fixed at 0.77, 0.13, 12, and 0.51 to suggest the optimal values. Based on the results of the three scenarios, scenario #3 was suggested as the best prediction model.

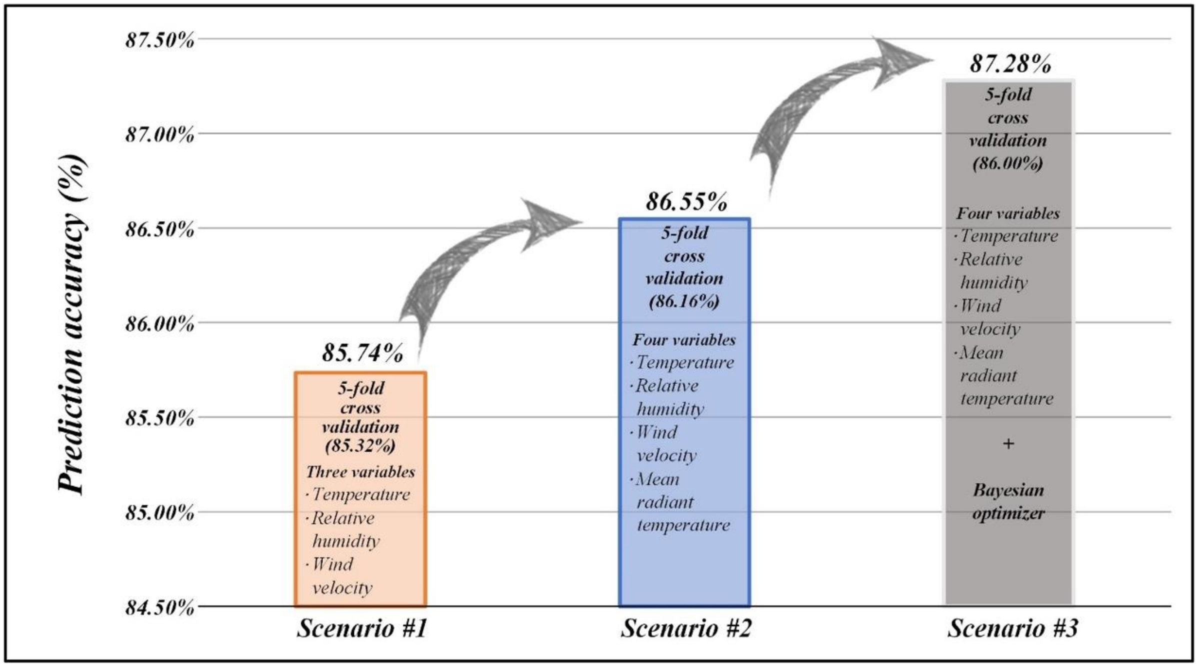

Figure 2. Results of the prediction model considering several environmental factors

3.2. Discussion

In this study, using the established data on 213,120 cases, the framework of a prediction model for PET as the thermal comfort index was presented considering environmental factors to easily use safety management at the construction site. Several previous researches were efforted to predict the PET through the experiment or prediction model. In the previous researches, due to a small amount of data, the prediction accuracy was low or can be used for specific conditions. To address this issue, this study suggested a framework of the prediction model with guaranteeing high prediction accuracy based on 213,120 cases. The prediction accuracy of the prediction model was maximized through the two-step prediction process and hyperparameter tuning in consideration of the three environmental factors. So, the results of this study suggested prediction accuracy up to 87.28%. Furthermore, through the 5-fold cross validation, the prediction model was proved and identified about overfitting problem. Since there is no significant difference between prediction accuracy and 5-fold cross validation, the prediction model allows it to ensure accurate estimation as well as solve the overfitting problem.

The result of this study can be utilized as follows. When starting daily work, the site safety manager measures PET using a prediction model by collecting three environmental factors such as t_a, rh, and v from KMA using simple internet access. The expected PET level of that day can be obtained through the proposed prediction model using three factors, Then the site safety manager the site manager conduct the risk assessment of tasks considering the calculated PET level at comfort or discomfort range.

4. CONCLUSION

The PET is divided by the comfort range (between 18ºC and 23ºC) and discomfort range (under 18ºC or over 23ºC). In the our previous researches, the number of construction accidents is mainly occurred at the discomfort ranges. Therefore, the PET needs to be considered to reduce or eliminate the construction accidents. However, not only is it hard to calculate the PET, but the specialized simulation is required. To solve this issue, this study aimed to develop a framework of the prediction model with the highest prediction accuracy using machine learning and hyperparameter tuning. This study is processed by four steps. i) Establishment of database, ii) Selection of variables to develop prediction model, iii) Development of prediction model; iv) Conducting of hyperparameter tuning.

(i) Establishment of database: 213,120 cases related to PET are established in consideration of accident occurred. (ii) Selection of variables to develop a prediction model: The design variables and independent variables are selected to develop the prediction model for PET. (iii) Development of prediction model: The XGBoost is selected as a machine learning algorithm which has high efficiency and accuracy. Three scenarios are presented to ensure the highest prediction accuracy for the prediction model. (iv) Conducting of hyperparameter tuning: the methodology of optimization is introduced to improve the prediction accuracy.

The results of this study are as follows. First, the prediction accuracy of scenario #1 (Considering three environmental factors) is very low (85.74%) compared to others. Second, the prediction accuracy of scenario #2 (Considering four environmental factors) is improved (86.55%) compared to scenario #1. Third, the prediction accuracy of scenario #3 (Considering four environmental factors and conducting the hyperparameter tuning) has the highest value (87.28%).

Based on the results, when the more variables are considered, the higher the prediction accuracy and the hyperparameter tuning can be ensured about high prediction accuracy.

This study has several contributions. First, in terms of technical aspects, the framework can be suggested, which have the possibility with ensuring the highest prediction accuracy based on the 213,120 cases. Second, in terms of the practical aspect, the site manager can plan the safety management considering PET comfort ranges. In the past, it was hard to calculate the PET. However, using this framework, the site manager can easily collect environmental factors from KMA and predict the PET. Using predicted PET, the site manager can utilize the safety management at the construction site.

The limitations of this study are as follows. First, this study presented the framework for prediction of PET. While various machine learning algorithms should be considered for suggesting the prediction model with high accuracy, the XGBoost was used in this study. Second, As the prediction accuracy of t_{mrt} was 64.66%, It is considered to be very low. Third, the environmental factors and personal factors are considered to calculate the thermal comfort. However, Since the Rayman Pro is used to calculate the outdoor thermal comfort taking into account environmental factor, personal factor is not considered in this study.

The future research will conduct following several considerations. First, the developed prediction model will be validated with other statistical analyses or new data set. Second, the prediction accuracy of the prediction model will be maximized through the various machine learning algorithms. Third, the hyperparameter tuning will be employed to overcome low prediction accuracy of t_{mrt}. Fourth, the prediction model will be developed by considering various design variables not only the environmental factor but also the personal factors. Fifth, for the next research, the performance of a developed prediction model will be evaluated from real practice to calculate the outdoor thermal comfort of workers.

ACKNOWLEGEMENTS

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 22CTAP-C163805).

REFERENCES

[1] M. Lee, J. Jeong, J. Jeong, J. Lee, J. "Exploring fatalities and injuries in construction by considering thermal comfort using uncertainty and relative importance analysis", International journal of environmental research and public health, vol. 18, no. 11, pp. 5573-5602, 2021.

[2] J. M. Kim, M. J. Kong, T. H. Hong, K. B. Jeong, M. H. Lee, "Physiological response of building occupants based on their activity and the indoor environmental quality condition changes", Building and Environment, vol. 145, pp. 96-103, 2018.

[3] R. De Dear, J. Xiong, J. Kim, B. Cao, "A review of adaptive thermal comfort research since 1998", Energy and Buildings, vol 214, pp 109893, 2020.

[4] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, G. Brager, "Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning", Building and Environment, vol 129, pp. 96-106, 2018.

[5] K. Liu, T. Nie, W. Liu, Y. Liu, D. Lai, "A machine learning approach to predict outdoor thermal comfort using local skin temperatures." Sustainable Cities and Society, vol. 59, pp. 102216, 2020.

[6] D. Lai, X. Zhou, Q. Chen, "Modelling dynamic thermal sensation of human subjects in outdoor environments", Energy and Buildings, vol. 149, pp. 16-25, 2017.

[7] S. Coccolo, J. Kämpf, J. L. Scartezzini, D. Pearlmutter, D. "Outdoor human comfort and thermal stress: A comprehensive review on models and standards", Urban Climate, vol. 18, pp. 33-57, 2016.

[8] S. Shooshtarian, C.K.C. Lam, I. Kenawy, I. "Outdoor thermal comfort assessment: A review on thermal comfort research in Australia", Building and Environment, vol. 177, pp. 106917, 2020.

[9] J. Kim, T. Hong, J. Jeong, C. Koo, K. Jeong, "An optimization model for selecting the optimal green systems by considering the thermal comfort and energy consumption", Applied Energy, vol. 169, pp. 682-695. 2016.

[10] S. Bai, M. Li, R. Kong, S. Han, H. Li, L.Qin, "Data mining approach to construction productivity prediction for cutter suction dredgers", Automation in Construction, vol. 105, pp. 102833, 2019.

[11] A. Shehadeh, O. Alshboul, R.E. Al Mamlook, O. Hamedat, "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression", Automation in Construction, vol. 129 pp.103827, 2021.

[12] S. Lim, S. Chi, "Xgboost application on bridge management systems for proactive damage estimation:, Advanced Engineering Informatics, vol 41, pp.100922, 2019.

[13] N. Baldo, E. Manthos, M. Miani, “Stiffness modulus and marshall parameters of hot mix asphalts: Laboratory data modeling by artificial neural networks characterized by cross-validation”, Applied Sciences, vol 9(17), pp.3502, 2019.

[14] S, Lim, S, Chi, “Xgboost application on bridge management systems for proactive damage estimation”, Advanced Engineering Informatics, vol 41, [[.100922, 2019.

[15] Y. Xia, C. Liu, Y. Li, N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring", Expert Systems with Applications, vol 78, pp.225-241, 2017.

[16] Pandas, https://pandas.pydata.org/ (Accessed on 10.08.21).

[17] XGBoost, https://xgboost.ai/ (Accessed on 10.08.21).

[18] J. Choi, B. Gu, S. Chin, J.S. Lee, "Machine learning predictive model based on national data for fatal accidents of construction workers", Automation in Construction, vol. 110, pp. 102974, 2020.