Driver Fatigue Recognition at Construction Site: Using Electrodermal Activity and Supervised Machine learning Algorithm

Huiying Li[1], Jeonghyeun Chae[2]

[1] *Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Korea,* E-mail address: hyeyeong1111@yonsei.ac.kr
[2] *Department of Architecture and Architectural Engineering, Yonsei University, Seoul 03722, Korea,* E-mail address: jh.chae@yonsei.ac.kr

1. INTRODUCTION

Fatality rate of construction industry is higher than other industries or averaged fatality rate of industries. Construction equipment-related accidents are responsible for one out of five deaths in the construction industry in Korean construction industry [1]. To prevent equipment-related accident in construction industry, alarm systems are mandatory to installed in equipment [2]. Proximity alarm and back-up alarm are well-known alarm systems. The goal of proximity alarm is to alert surrounding ground-worker that equipment is near, and danger is imminent, and alert equipment operator about existence of obstacles or ground-workers around the equipment. Back-up alarm alert the ground-workers who locate at the path of the back-up equipment. These alarm systems can alert the ground workers and induce them to avoid collision. Also, alarm systems can alert the equipment operator and induce them to be cautious about their surroundings or react to imminent danger. However, even though alarm system can alert the equipment operator for short amount of time, if the equipment operator experiencing driver fatigue, then alarm effect cannot last long, and much worse, it cannot affect to the equipment operator at all. Driver fatigue is physical/mental fatigue that equipment operator experiences due to prolong driving, sleep deprivation, physical fatigue, constant noise, monotonous task, and more factors. Driver fatigue can induce slower reaction and lowered alertness, i.e., desensitization or numbness, poor decision making, poor driving skills [3]. Driver fatigue is well known equipment-related accident leading indicator. Driver fatigue is prominent in construction site, especially to dump truck drivers. As dump truck driver should constantly drive same route to deliver material numerous times during work hours, it can be perceived as monotonous task, which is a one of the major causes of driver fatigue [4]. Therefore, managing driver fatigue can be one of the solutions for preventing equipment-related accident in construction site.

Driver fatigue recognition is important for managing driver fatigue. If the individual experiences driver fatigue, he/she needs to rest for the recovery. And driver fatigue recognition using quantitative method can help such process by suggesting the objective level of driver fatigue that individual experiencing. For driver fatigue recognition, quantitative measurement of driver fatigue is necessary. Many researchers use physiological signals, such as electrocardiograph (ECG), electroencephalogram (EEG), and electrodermal activity (EDA) to measure the level of driver fatigue [4-6]. As physiological signal can reflect body reaction to stimulus, it's performance of measuring psychological state such as stress, alertness is validated with many research. In this research, EDA signal is used to measure the psychological state (i.e., driver fatigue) of the experiment subject. Every physiological signal's performance has validated by many researchers, but EDA has some advantages compared to other physiological sensors. For example, the form factor of Empatica E4 device, which is a EDA sensor used in this research, is shape like regular smartwatch. The weight of the E4 device is only 25g, it can be considered as un-intrusive physiological sensor. Other physiological signals such as ECG, EMG, EEG are required more intrusive method of measuring. The other advantages of using EDA signal is robustness. EDA signal is robust to noise, compared to other physiological signal [7]. These two advantages of EDA signal make it a great tool for measuring psychological features such as alertness or fatigue of working individual. Detailed explanation of EDA signal is presented in next section.

In this study, we use supervised machine learning algorithm to recognize driver fatigue. There exist some research using statistical method such as student t test to recognize driver fatigue [8]. However, as physiological signal is measurement of body reaction to outer stimulus, recorded physiological signal has large inter-personal difference [9]. Therefore, in order to generalize the recognition performance, there exist the need to use large sample and model that trained with such large sample. Supervised machine learning algorithm can support such role.

Experiment is conducted to 20 subject and collected EDA signal is labeled with the reaction time data. After collecting and filtering the EDA data, the authors use 5 different classification model to recognize driver fatigue. The model that shows highest accuracy is random forest model and the accuracy is 72%. Result of this study shows that EDA sensor and supervised machine learning algorithms can minimize the inter-personal difference problem and the classification model can recognize driver fatigue. This can contribute to current body of knowledge, as there is no study using EDA sensor and supervised machine learning algorithms to recognize driver fatigue, and eventually this result can contribute to safer construction industry.

2. METHODOLOGY

2.1. Electrodermal activity

Electrodermal activity means the value of skin conductance. When person perceived stimulus, it will lead to the activation of sympathetic nervous system (SNS). Activation of SNS will leads to sweat secretion and sweat on the skin will change the skin conductance level. And we can collect the change of skin conductance with EDA sensor, and through this mechanism, we can measure how much the individual is alerted by stimulus [10].

EDA signal can be decomposed into two different types of signal. First is skin conductance response (SCR), and second is tonic skin conductance level (SCL). SCR related to discrete environmental stimuli. Driving is consisted with constant decision making and perceiving external stimuli. Therefore, SCR can reflect the change of alertness or amount of driver fatigue. SCL related to continuous slow changing level of EDA signal. As SCL is related to psychological state, it can reflect the numbness or desensitization of the driver [11].

2.2. Experiment design

In this section, brief explanation of experiment is given. Subject will sit in quiet room and play driving simulation for 10 minutes. During simulation, to provide the subject the experience of construction equipment operator, alarm will be emitted for 10 to 25 second interval. Also, this alarm work with another measure of alertness of the subject, which is reaction time. When the subject perceived alarm sound they are instructed before the experiment to press brake pedal. And by comparing the brake time (i.e., when the pedal is pressed) and alarm time (i.e., when the alarm is emitted) we can obtain reaction time.

Total number of subjects are 20, averaged age of subjects is 27.4. The total number of features used in this experiment is 2, which is SCR, SCL. Total experiment time is 10 minutes, and gender ratio is seven to 3, for men and women respectively.

2.3. Data labeling

For using supervised machine learning algorithm, SCR and SCL dataset should be labeled. Labeling process need validation data. In this study, we use reaction time to label the EDA data. Reaction time for 25 alarms, 20 subjects are collected. After collecting the raw data, each data is scaled and averaged for each alarm repetition. For example, we average the scaled data for 20 people into one average data. After obtain averaging dataset, linear regression is applied to see the trend of reaction time as the time passes. As shown in the figure, the reaction time data show increasing trend, which can be related to decreased alertness. Decreased alertness during driving situation indicates driver fatigue. Therefore, with analysis with reaction time, this data shows that the subjects experiencing driver fatigue during playing driving simulation. For labeling the EDA data, total dataset is divided in half. This means we divide the dataset into early stage and later stage and label the data of early stage as non-driver fatigue class and label later stage as driver fatigue class.

2.4. EDA data processing

As mentioned before, EDA signal is more robust to noise than other physiological signals. However, there still exist noise in the collected raw EDA data. To remove the noise from the EDA signal, filtering process is needed. Highpass filter and rolling filter is used to smoothing the raw EDA signal. This process can remove noise such as popping noise, line noise. After filtering EDA data, it can be decomposed into SCR and SCL data with continuous decomposition analysis, using Matlab program. The sampling frequency of Empatica E4 device is 4Hz, which means it records four EDA data point for every seconds. As mentioned before, total experiment time is 10 minutes. Therefore, from one subject we can extract total 2400 datapoint for each feature. As experiment is conducted to 20 subjects, we can use 48,000 data point for each feature.

2.5. Supervised machine learning algorithms

Logistic regression classification model uses the probability measure to classify the categories in a way. In this study we use linear-logistic regression to classify the “early stage” and “late stage”. The decision boundary is setting as the time standard that can classify the time based on the half time of the whole reaction time. In general, theorical explanation, logistic regression measured categories based on decision boundary. For example, in the two categories model like this study, we can set the whole probability as 100% equals to 1. And odds those the probability with bigger than 50% to 1. In contrast, the probability with smaller than 50% will be measure at 0. In this way, the probability with each category will be correctly classified. The original function of logistic regression is written as

$$P(Y) = 1/1 + e^{-(\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn)}$$

$$= 1/1 + e^{-(\beta T * X)}$$

In this function, the X_1, X_2,…,X_n are the predictor variables, $ß_0$,$ß_1$…$ß_n$ are regression coefficients. So, we have two features phasic EDA data and Tonic EDA data, in this function, they can regard as X_1 and X_2. And P(Y) is the probability of the presence of reaction time.

Decision tree is another classification model in this study. There are two types of decision tree: Regression mode decision tree and Classification mode decision tress. In this study, we use the classification decision tree to classify our outcome value. The purpose of decision tree is that we can estimate the output of the classification based on every node that we have input. The name of decision tree originates from the tree appearance, the leaves of tree represent the node in the decision path. For simply example, when we choose a worker maybe we have so many conditions about him. And at first node, we regard the strong technical skill as the most important standard. If a man can satisfy this standard, then he can participate in the next node assessment – strong motivation to work. And the other men who didn't satisfy the first standard can not participate in the second assessment. So, in this way, we can finally classify the different categories. In this study, we have two nodes" phasic EDA data" and "Tonic EDA data".

Support vector machine is a newer algorithm than two classification models mentioned above. In the classification research. In this study, we have two classification parts"0" and "1". It has the decision boundary based on the two points of each classification parts. The decision boundary is called hypersurface in SVM. The decision boundary is decided based on two closest points of each classification parts. The distance between two classification parts' closest points will be mostly same. In this way, the hypersurface is decided. There are some categories in SVM such as SVC which is short for support vector classification and SVR which is short for support vector regression. In this study, we implemented the SVC to classify. And the linear SVC model is involved. Finally, we utilize this model to classify the early stage and late stage. In this study, the researchers utilize the rbf Kernel SVM which can measure the accuracy higher. The rbf kernel SVM's advantages are that it can work in the infinite dimensions. And the similar kernel SVM named poly kernel SVM can work in the three dimensions. So, the accuracy performance of rbf kernel SVM is better and actually the result of this experiment did obeyed this theory.

The k-nearest neighbor (kNN) classification algorithm is one of the simplest methods in data mining classification techniques. By k-nearest neighbor, it means k nearest neighbors, which says that each sample can be represented by its k nearest neighbors.

kNN is used to classify by measuring the distance between different feature values. Specifically, if the majority of a sample's k nearest neighbors in the feature space belong to a certain class, the sample is classified in that class as well. The neighbors selected in the kNN algorithm are objects that have been correctly classified. The method relies only on the category of the nearest neighboring sample or samples to decide the category to which the sample to be classified belongs in the class decision. The Euclidean Distance is used as the measurement of distance.

The steps of kNN is listed as follows:

(1) Calculate the distance between the samples to be classified and the samples of the known class.

(2) Sort the samples in increasing order of Euclidean distance.

(3) Select the k samples with the smallest distance from the sample to be classified.

(4) Determine the number of occurrences of the class in which the first k samples are located.

(5) Return the class with the highest number of occurrences of the first k samples as the predicted classification of the samples to be classified.

Random forest is a forest with many decision trees, and each tree in the random forest is unrelated to each other. After obtaining the forest, when a new input sample comes, each decision tree in the forest makes a classification to decide which class the sample should belong to, and then check which class is selected the most, and then predict the sample to be in that class. A random forest consists of decision trees, which are a way of dividing the space by a hyperplane, dividing the current space in two each time. Random forest is composed by three main steps.

First step is the sample set selection. Suppose the original sample set has a total of N samples, then N samples are drawn from the original sample set by bootstrapping (with put-back sampling) in each round, and a training set of size N is obtained. In the process of drawing the original sample set, there may be samples that are repeatedly drawn, or there may be samples that are not drawn even once. The total number of rounds is k, and the training set of each round is denoted as T1, T2..., Tk.

Second step is the decision tree generation. If there are D features in the feature space, then in each round of decision tree generation, d features ($d<D$) are randomly selected from the D features

to form a new feature set, and the decision tree is generated by using the new feature set. A total of k decision trees is generated in k rounds, as these k decision trees are random in the selection of training set and the selection of features, because these k decision trees are independent of each other.

Third step is the combination of models. Since the generated k decision trees are independent of each other, the importance of each decision tree is equal, thus when combining them, there is no need to consider their weights, or they can be considered to have the same weights. For the classification problem, the final classification result is determined using all the decision trees voted to determine the final classification result.

This study measures the accuracy implemented the function:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

The author measures the correct times to measure the reaction time as the ground truth and calculates the proportion between the correct measure time and all experiment time.

3. RESULT

The authors applied the 5 classification models to classify the early stage and late stage as label "0" and "1" to measure reaction time and then divided the 90% of data to training data and 10% to test data. Figure 1 to figure 5, and table 1 shows the results of the classification accuracy of each model. The accuracy calculated function is mentioned in method part. As the result, the authors can find the highest accuracy comes from the random forest model with 72%, and the lowest accuracy comes from logistic regression with 56%. And the other accuracy is averagely in the range from 63% to 65%. So, the random forest is the most available classification model in this research.

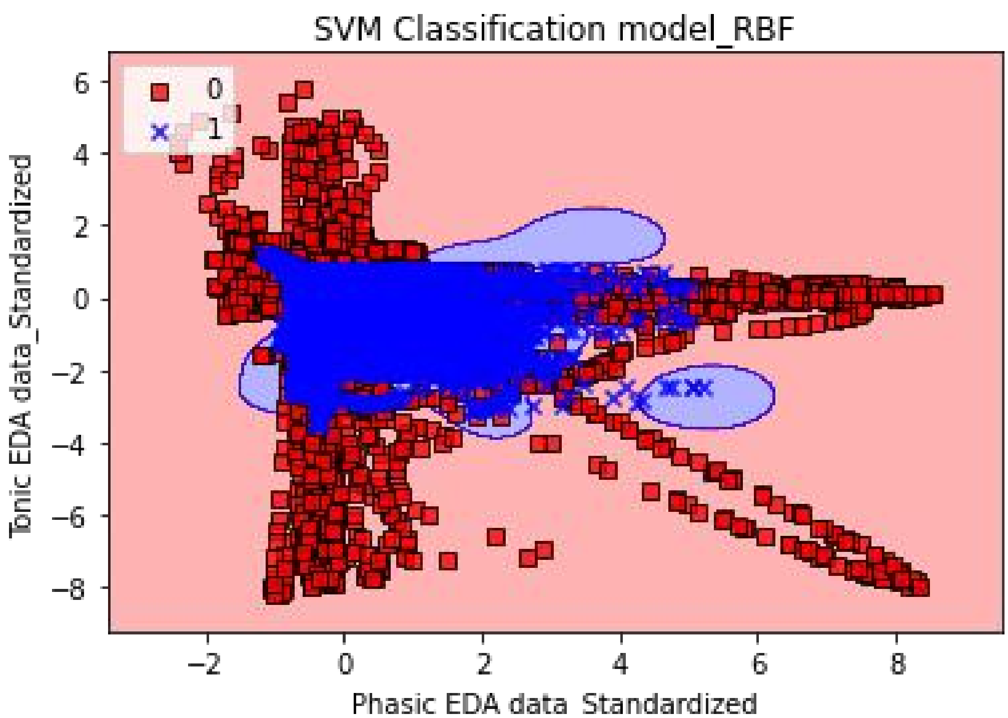

Figure 1. Rbf SVM classification result

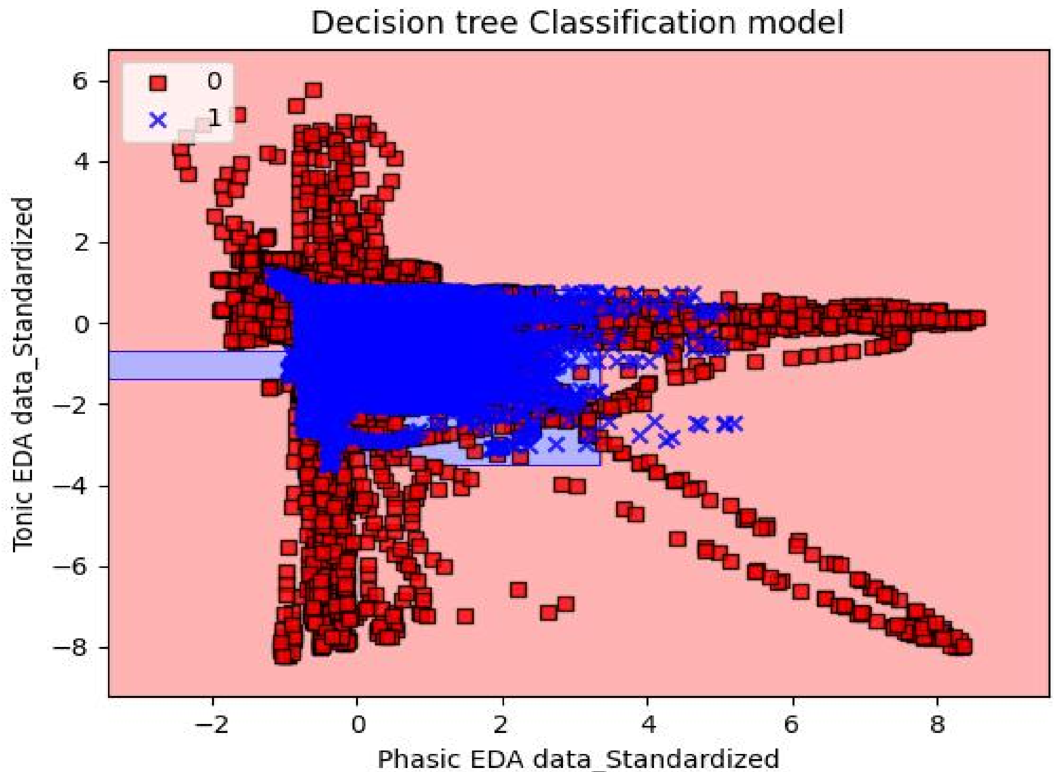

Figure 2. Decision tree classification result

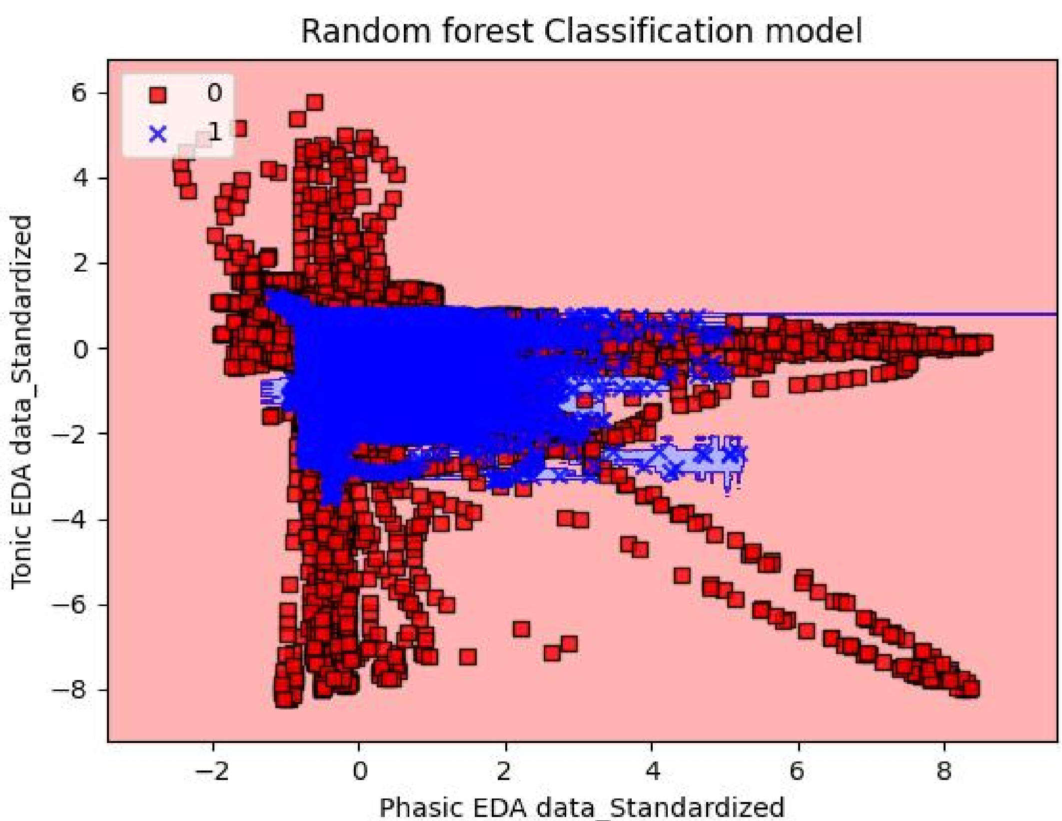

Figure 3. Random forest classification result

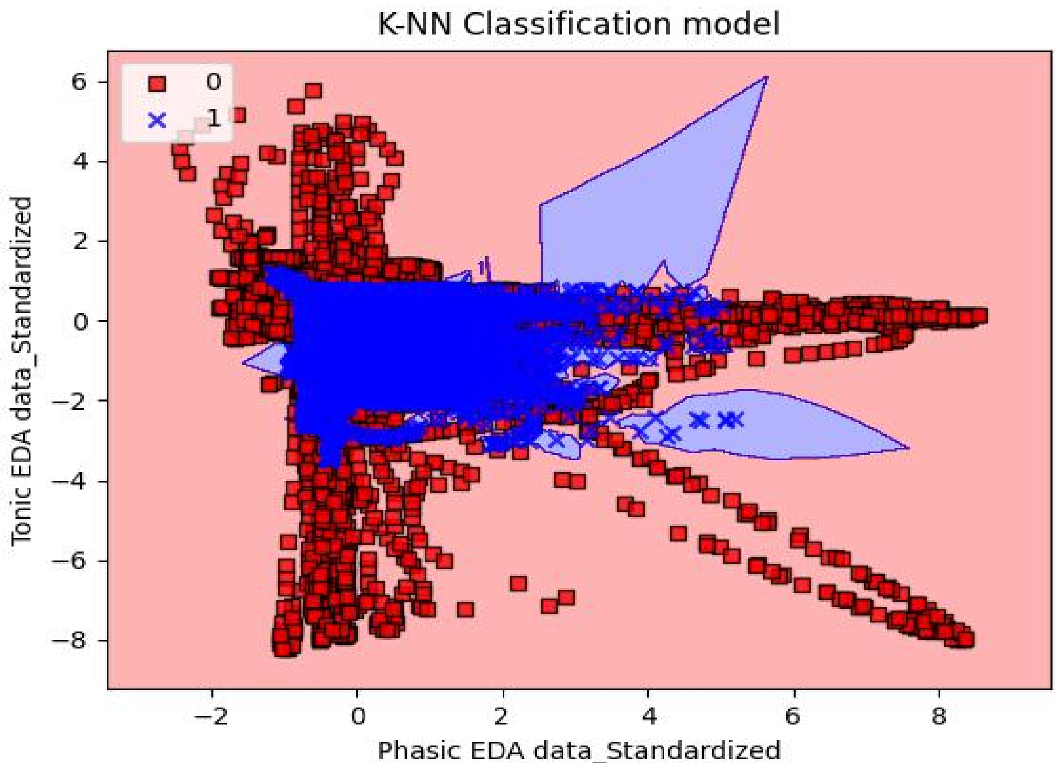

Figure 4. K-NN classification result

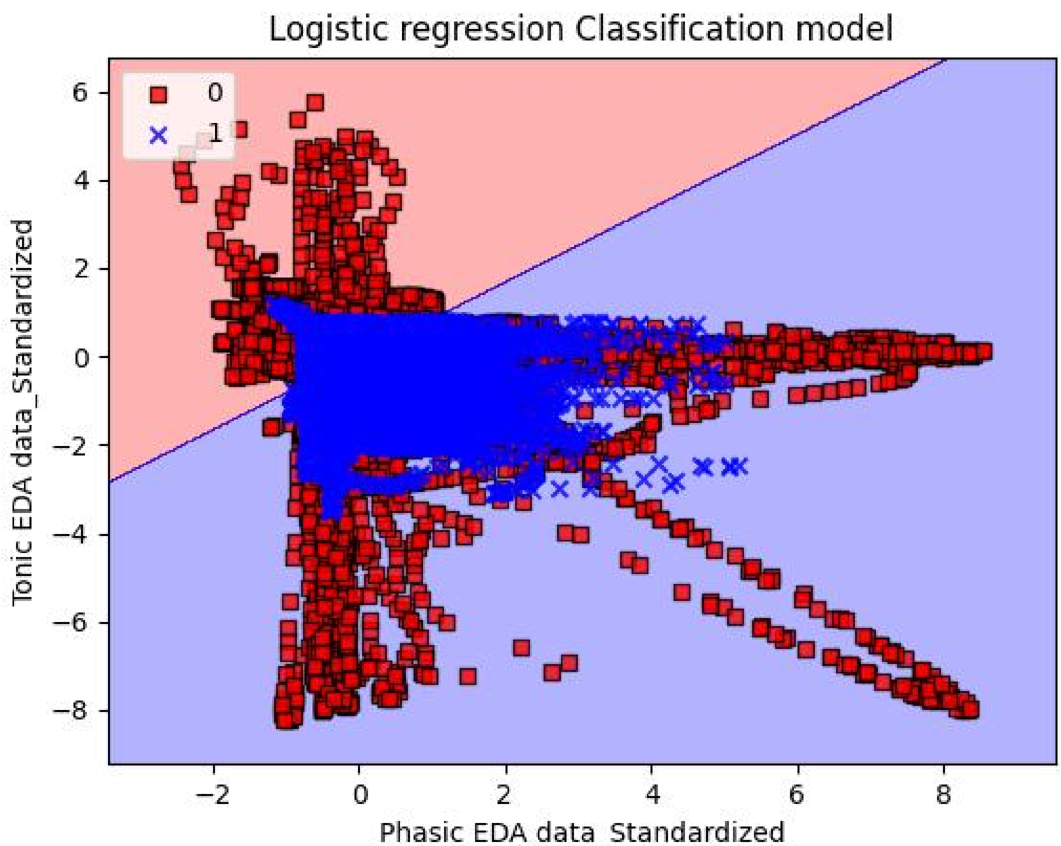

Figure 5. Logistic regression classification result

Table 1. Classification results

No.	Algorithm	Classification accuracy
1	Rbf SVM	64.5%
2	K-NN	63.9%
3	Decision Tree	63.1%
4	Random Forest	72.0%
5	Logistic Regression	56.9%

And to analysis the indicator with highest accuracy 72%, the parameter and data set distribution played an essential role in this process. The best result with random forest, the author uses the 90% training set and 10% training set. And according to many times experiments, the author found that n_estimators is the most influential parameter, when increase the n_estimator to 150, the result will be better. But when the number of estimators exceeded 150 such like modify the n_estimator as 200, the accuracy will stop increasing and have some fluctuation. So finally, the author set the n_estimator as 150. And the second influential parameter is max_depth. When max_depth decreased, the model will be simpler. In this research, the max_depth can set as 100 as the best result. And then the min_samples_leaf and min_samples_split is the third influential parameter. These two parameters are initially set as minimum indicator as 1 and 2. And in this experiment, the author modify these parameters to 30 and 8. The minimum limited indicator means that in this situation the complexity will be the highest. So, the author tried to increase these to parameter and finally earned better results. Genelization error is the indicator which can measure the accuracy of the unknown dataset. And the genelization error's change is deeply related to the model complexity.

4. DISCUSSION

5.1. Contribution

This study suggests that by using EDA sensor and supervised machine learning algorithms, it is possible to minimizing the inter-personal difference problem in physiological signal research area. Also, by minimizing the inter-personal difference, the suggested model and result can be applied to real construction site to measure and detect driver fatigue of the construction equipment operators. As detecting driver fatigue occurred in construction site with physiological signal has not been researched, this paper has contribution to the current body of knowledge. Also, by recognizing driver fatigue, the construction equipment operators can rest to recover their alertness and such rest can leads to lower construction equipment-related accidents. Therefore, eventually, this research can contribute to the safer construction industry.

4.2. Limitation and future study

This study is not free from limitations. Even though the authors propose that the result of this study shows minimized inter-personal difference problem, it is not enough to generalize the result of this study into industry implementations. Also, data size is another obstacle for generalizing the result. 20 subject and 48,000 data point is not enough to be generalized. The experiment time is short, therefore EDA sensor cannot measure the strong effect of driver fatigue, as one of the major causes of driver fatigue is prolong driving. Also, to improve the performance and reliability of the result, data labeling criteria should be more researched. Current models use only two features to detect driver fatigue. Limited number of features can lead to low detection performance.

Based on the suggested limitations, the future study should use larger sample size to generalize the result. Also, using more physiological features can lead to precise detection of driver fatigue. Finally, more reliable data labeling criteria should be used. For example, there exist some study using saliva test during experiments to measure the cortisol level of the subject to measure the stress level.

REFERENCES

[1] Korean Ministry of Employment and Labor. "Press realease_Ministry of Employment and Labor, Ahead of Enforcement of Serious Accident Punishment Act Held a meeting to prevent serious accidents with mid-sized construction companies"., 2021

[2] J. Teizer, B.S. Allread, C.E. Fullerton, J. Hinze, "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system", Automation in construction, vol. 19, no. 5, pp. 630-640, 2010.

[3] B. Ivan. "Driver fatigue", Human factors 36.2, pp.298-314, 1994

[4] J. Xi, S. Wang, T. Ding, J. Tian, H. Li, H. Shao, T. Liu. "Detection Model on Fatigue Driving Behaviors Based on the Running State of Freight Vehicles". CICTP 2020, pp. 3680-3691, 2020

[5] F. Zhou, A. Alsaid, M. Blommer, R. Curry, R. Swaminathan, D. Kochhar, W. Talamonti, L. Tijerina, B. Lei. "Driver Fatigue Transition Prediction in Highly Automated Driving Using Physiological Features", Expert Systems with Applications, vol. 147, 2020

[6] K. Wang, Y. Ma, J. Huang, C. Zhang. "Driving Performance of Heavy-Duty Truck Drivers under Different Fatigue Levels at Signalized Intersections", CICTP2019, pp. 581-592, 2019

[7] J. Chae, Y. Kang. "Designing an Experiment to Measure the Alert Fatigue of Different Alarm Sounds Using the Physiological Signals". in Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC). International Association for Automation and Robotics in Construction (IAARC), pp. 545-552, 2021.

[8] X. Xianggang, G. Yan, S. Xianghong, Z. Kan, "Influences of On-Road Driving Fatigue, Mental Workload on Driver's Performance", International Conference on Transportation Engineering. 2009

[9] J.D. Bronzino, "Biomedical Engineering Handbook 2″. Vol. 2. Springer Science & Business Media. pp., 2000.

[10] W. Boucsein, "Electrodermal activity″. Springer Science & Business Media. pp., 2012.

[11] B. Choi, H. Jebelli, S. Lee, "Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk", Safety science, vol. 115, no. pp. 110-120, 2019.