

Original Article
Occupation &
Environmental Medicine



Prediction of Return-to-original-work after an Industrial Accident Using Machine Learning and Comparison of Techniques

Jongin Lee ^{1,2} and Hyoung-Ryoul Kim ³

¹Cheongsong Health Center and County Hospital, Cheongsong, Korea

²Department of Medicine, Graduate School, The Catholic University of Korea, Seoul, Korea

³Department of Occupational and Environmental Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea



Received: Nov 9, 2017

Accepted: Mar 26, 2018

Address for Correspondence:

Hyoung-Ryoul Kim, MD, PhD

Department of Occupational and Environmental Medicine, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Korea.
E-mail: cyclor@catholic.ac.kr

© 2018 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Jongin Lee
<https://orcid.org/0000-0003-4337-8444>
Hyoung-Ryoul Kim
<https://orcid.org/0000-0001-7535-3140>

Disclosure

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Lee J. Data curation: Lee J, Kim HR. Formal analysis: Lee J. Methodology: Lee J, Kim HR. Project administration: Lee J, Kim HR. Validation: Lee J. Visualization: Lee J. Supervision: Kim HR. Writing - original draft: Lee J. Writing - review & editing: Kim HR.

ABSTRACT

Background: Many studies have tried to develop predictors for return-to-work (RTW). However, since complex factors have been demonstrated to predict RTW, it is difficult to use them practically. This study investigated whether factors used in previous studies could predict whether an individual had returned to his/her original work by four years after termination of the worker's recovery period.

Methods: An initial logistic regression analysis of 1,567 participants of the fourth Panel Study of Worker's Compensation Insurance yielded odds ratios. The participants were divided into two subsets, a training dataset and a test dataset. Using the training dataset, logistic regression, decision tree, random forest, and support vector machine models were established, and important variables of each model were identified. The predictive abilities of the different models were compared.

Results: The analysis showed that only earned income and company-related factors significantly affected return-to-original-work (RTOW). The random forest model showed the best accuracy among the tested machine learning models; however, the difference was not prominent.

Conclusion: It is possible to predict a worker's probability of RTOW using machine learning techniques with moderate accuracy.

Keywords: Return to Work; Accidents, Occupational; Machine Learning

INTRODUCTION

The first article of the Korean Industrial Accident Compensation Insurance Act states that its purpose is for injured workers to return to society, as well as to compensate workers and facilitate their rehabilitation.¹ Since return-to-original-work (RTOW) is considered the most desirable for injured workers in return to society,² the Korean government has set as a policy goal to support individuals returning to original work. Previous studies of return-to-work (RTW) have identified factors related to this outcome, including sex, age, company characteristics, working period, injury type, period for medical care,³ average wage, disability grade, job group (manual worker or not), company size,⁴ education, consultation experience,⁵ comorbidities,⁶ and socio-economic status.⁷ A systematic review⁸ concluded

that these factors affect RTW after musculoskeletal injury, and another study⁹ reported that self-efficacy also plays a role in RTW. Since complex factors have been demonstrated as predictors of RTW, it is difficult to use all of these factors practically.

Previous studies have aimed to identify meaningful factors using several statistical methods, e.g., multivariate logistic regression (MLR) has been widely used as a methodological standard in medicine.¹⁰ However, machine learning techniques are currently in practice in medical sciences especially for problems on prediction. A problem on prediction is a type of classification problem that can be solved with machine learning techniques. Additional machine learning techniques are becoming widely used, some of which have demonstrated better predictive ability on certain classification problems.^{11,12}

This study aims to establish prediction models of RTOW and comparing each model and to apply machine learning techniques to identify factors related to RTOW. By comparing the characteristics of the predictive models, a simple and/or accurate model is identified. This analysis, coupled with a discussion of the advantages and disadvantages of machine learning models, is expected to help establish policy on industrial accidents.

METHODS

Study participants

We used data from the fourth Panel Study of Worker's Compensation Insurance (PSWCI), a panel survey conducted by the Korea Workers' Compensation and Welfare Service (KCOMWEL). The nationwide survey sampled 2,000 of 89,921 patients who completed their recovery period in 2012 after an industrial accident. Patients were sampled by the Bellwether method after stratification according to specific variables. The detailed PSWCI sampling methodology is described elsewhere.¹³ All sampled participants were followed up annually, and the fourth wave of PSWCI data was collected in 2016, four years after termination of the official recovery period.

In the fourth survey, 1,660 patients were followed up successfully, yielding a follow-up rate of 83%. Patients with disability grade 1–3 ($n = 25$) or 4–7 ($n = 68$) were excluded from our analysis since high disability grade has been shown to be a strong predictive factor for failure of RTW.³ Moreover, job demand in these patients could be different from that in other patients without a high disability grade, since patients with disability grade 1–7 can receive a monthly pension. Therefore, we analyzed data from 1,567 patients.

Baseline characteristics

Among the individual characteristics examined, age, education level, earned income, and self-efficacy showed an association with RTOW (**Table 1**). Alcohol consumption showed a significant P value; however, no trend was observed. All occupational and supportive variables, with the exceptions of 'detailed explanation from medical doctors' and 'regular assessment of recovery,' showed a significant association with RTOW (**Tables 2 and 3**).

Methods

The total dataset was randomly divided into two subsets, a training dataset (75%, $n = 1,175$) to make prediction models and a test dataset (25%, $n = 392$) to report the prediction ability of each model. The random selection was done by the function 'sample' of R with setting seed

Table 1. Individual variables (demographical characteristics) related to return-to-original-work of the total dataset (n = 1,567) and χ^2 test

Variables	Return to original work		P value ^a
	Succeeded (n = 442)	Failed (n = 1,125)	
Sex			0.060
Male (n = 1,299)	379 (85.7)	920 (81.8)	
Female (n = 268)	63 (14.3)	205 (18.2)	
Age	47.54 ± 9.78	50.74 ± 11.48	< 0.001
Education level			< 0.001
Below middle school (n = 630)	115 (26.0)	515 (45.8)	
High school (n = 695)	229 (51.8)	466 (41.4)	
College or higher (n = 242)	98 (22.2)	144 (12.8)	
Earned income			< 0.001
1st quartile (n = 364)	30 (6.8)	334 (29.7)	
2nd quartile (n = 517)	137 (31.0)	380 (33.8)	
3rd quartile (n = 281)	91 (20.6)	190 (16.9)	
4th quartile (n = 405)	184 (41.6)	221 (19.6)	
Smoking			0.216
Never (n = 498)	127 (28.7)	371 (33.0)	
Past (n = 321)	99 (22.4)	222 (19.7)	
Current (n = 748)	216 (48.9)	532 (47.3)	
Alcohol consumption			0.016
Never (n = 310)	77 (17.4)	233 (20.7)	
Past (n = 115)	22 (5.0)	93 (8.3)	
Current (n = 1,142)	343 (77.6)	799 (71.0)	
Disability			0.084
None (n = 295)	90 (20.4)	205 (18.2)	
Grade 13–14 (n = 467)	143 (32.3)	324 (28.8)	
Grade 10–12 (n = 672)	182 (41.2)	490 (43.6)	
Grade 8–9 (n = 133)	27 (6.1)	106 (9.4)	
Self-efficacy	81.19 ± 9.84	78.14 ± 10.91	< 0.001

Values are presented as number (%) or mean ± standard deviation.

^aP value of χ^2 test.

numbers in order to assure reproducibility. The dependent variable in our analysis was working status at four years after termination of the official recovery period in 2012. The independent variables examined are listed in **Tables 1-3**; χ^2 tests were performed for each variable. A detailed description of the questionnaire variables from PSWCI is provided elsewhere.¹⁴

Two MLR models were established, one with the total dataset and another with the training dataset. The model with the total dataset was compared with the model from the training dataset, and the odds ratio (OR) from the two models were assessed. ORs from the total model were used to describe the results, while ORs from the training model were used to adjust and optimize the model's predictive ability. In the classical analytic models, *P* values below 0.05 were considered significant.

Models using machine learning techniques including decision tree (DT),¹⁵ random forest (RF),¹⁶ and support vector machine (SVM)¹⁷ were formed with the training dataset. To assess the predictive ability of each training model, we transformed and reconstructed the models by excluding unimportant variables and/or tuning hyper-parameters to improve the model's performance with automated process using the R package caret (R Foundation, Vienna, Austria).¹⁸ Variable importance was assessed by visualizing the varImp function of the package.

The best models with the highest predictive ability for each technique were selected as the representative models. We compared all necessary variables and the model performances by calculating the accuracy, sensitivity, specificity, kappa value, and area under the curve (AUC). The

final predictive values were reported with mean \pm standard deviation from 10 iterations of the same procedure after random allocation of training/test datasets with different seed numbers.

Table 2. Occupational variables related to return-to-original-work of the total dataset (n = 1,567) and χ^2 test

Variables	Return to original work		P value ^a
	Succeeded (n = 442)	Failed (n = 1,125)	
Occupational category			< 0.001
Non-manual (n = 151)	70 (15.8)	81 (7.2)	
Service or sales (n = 117)	31 (7.0)	86 (7.6)	
Manual (n = 1,299)	341 (77.2)	958 (85.2)	
No. of workers in company			< 0.001
< 5 (n = 240)	54 (12.2)	186 (16.5)	
5–9 (n = 286)	67 (15.2)	219 (19.5)	
10–29 (n = 393)	94 (21.3)	299 (26.6)	
\geq 30 (n = 648)	227 (51.3)	421 (37.4)	
Working status			< 0.001
Regular (n = 865)	364 (82.4)	501 (44.5)	
Temporary (n = 218)	27 (6.1)	191 (17.0)	
Daily (n = 479)	48 (10.9)	431 (38.3)	
Self-employed or other (n = 5)	3 (0.6)	2 (0.2)	
Type of employment			0.026
Full-time (n = 1,495)	430 (97.3)	1,065 (94.7)	
Part time (n = 72)	12 (2.7)	60 (5.3)	
Shift work			< 0.001
Yes (n = 174)	73 (16.5)	101 (9.0)	
No (n = 1,393)	369 (83.5)	1,024 (91.0)	
Formal contract for employment			< 0.001
Yes (n = 754)	271 (61.3)	483 (42.9)	
No (n = 813)	171 (38.7)	642 (57.1)	
Average working days per month	22.43 \pm 3.80	21.49 \pm 5.80	< 0.001
Average working hours per day	9.17 \pm 2.38	9.50 \pm 2.54	0.018

Values are presented as number (%) or mean \pm standard deviation.

^aP value of χ^2 test.

Table 3. Supportive variables related to return-to-original-work of the total dataset (n = 1,567) and χ^2 test

Variables	Return to original work		P value ^a
	Succeeded (n = 442)	Failed (n = 1,125)	
Satisfaction level with employer			< 0.001
High (n = 321)	159 (36.0)	162 (14.4)	
Moderate (n = 778)	233 (52.7)	545 (48.4)	
Low (n = 468)	50 (11.3)	418 (37.2)	
Maintenance of relationship with company			< 0.001
Yes (n = 997)	388 (87.8)	609 (54.1)	
No (n = 570)	54 (12.2)	516 (45.9)	
Detailed explanation from MD			0.723
Yes (n = 1,484)	420 (95.0)	1,064 (94.6)	
No (n = 83)	22 (5.0)	61 (5.4)	
Regular assessment of recovery			0.431
Yes (n = 1,366)	390 (88.2)	976 (86.8)	
No (n = 201)	52 (11.8)	149 (13.2)	
Opinion of duration of medical treatment			< 0.001
Appropriate (n = 867)	286 (64.7)	581 (51.6)	
Not appropriate (n = 700)	156 (35.3)	544 (48.4)	
Consultation on RTW with MD			0.007
Yes (n = 397)	133 (30.1)	264 (23.5)	
No (n = 1,170)	309 (69.9)	861 (76.5)	
Work ability assessment referral for RTW			0.001
Received (n = 142)	57 (12.9)	85 (7.6)	
Not received (n = 1,425)	385 (87.1)	1,040 (92.4)	

MD = medical doctor, RTW = return-to-work.

^aP value of χ^2 test.

All analyses were performed in R, version 3.4.2 (R Foundation). The R packages caret,¹⁸ ROCR,¹⁹ rpart,²⁰ rpart.plot,²¹ e1071,²² kernlab,²³ and randomForest¹⁶ were used.

Ethics statement

This study did not need an approval from an Institutional Ethics Review Board because all data used are from the PSWCI that is a public statistical source open to everyone with a permission of KCOMWEL.

RESULTS

MLR analysis

Although many variables showed a statistically significant association with RTOW by the χ^2 test, only six factors showed significant ORs in the MLR models. A potential explanation for this finding is that some variables might confound other variables. The trends were very similar between the total model and the training model. Moreover, all significant factors in the total model also showed significance in the training model (Table 4).

Earned income was the strongest predictor for RTOW and showed a directly proportional relationship: The greater was an individual's income, the better was the chance that he/she would return to his/her original work. Temporary or daily working status acted as a negative predictor of RTOW. Average hours worked per day showed an OR less than 1.0, meaning that the more hours per day an individual works, the more difficult it is for him/her to return to his/her original work. When a formal contract of employment was present, RTOW was easy. Moreover, when a worker was satisfied with the employer's support during the

Table 4. Logistic regression analysis for return-to-original-work using the total and/or training model

Variables	Odds ratio ^a (95% confidence interval)	
	Total model	Training model
Earned income		
1st quartile	1.00	1.00
2nd quartile	3.40 (2.17–5.48)	4.39 (2.57–7.83)
3rd quartile	3.42 (2.11–5.68)	4.50 (2.52–8.33)
4th quartile	5.37 (3.39–8.73)	7.18 (4.14–12.95)
Working status		
Regular	1.00	1.00
Temporary	0.24 (0.15–0.38)	0.29 (0.17–0.48)
Daily	0.33 (0.22–0.50)	0.31 (0.19–0.51)
Self-employed or other	2.49 (0.31–27.30)	2.50 (0.22–57.84)
Average working hours per day ^b	0.84 (0.71–0.97)	0.80 (0.65–0.96)
Formal contract for employment		
Yes	1.42 (1.07–1.89)	1.56 (1.12–2.17)
No	1.00	1.00
Satisfaction level with employer		
High	4.22 (2.77–6.51)	4.10 (2.53–6.73)
Moderate	2.36 (1.63–3.47)	2.31 (1.51–3.60)
Low	1.00	1.00
Maintenance of relationship with company		
Yes	3.11 (2.19–4.46)	3.57 (2.38–5.46)
No	1.00	1.00

^aAdjusted for sex, age, education level, smoking, alcohol consumption, disability, self-efficacy, occupational category, number of workers in company, type of employment, shift work, average working days per month, detailed explanation from a medical doctor, regular assessment of recovery, opinion of duration of medical treatment, consultation for return-to-work with a medical doctor, work ability assessment referral for return-to-work, and all variables in this table; ^bOdds ratio with normalized value.

duration of the official recovery, the worker could return easily. Maintenance of a relationship with the original company also showed a high OR, although this variable potentially has multicollinearity with 'satisfaction level with employer.' However, no significant variables were associated with any of the supportive factors from hospitals or KCOMWEL.

We next compared a predictive MLR model using all variables with another model using only the six most important variables (as assessed by ORs) and the varImp function; the former showed better predictive ability.

Inter-technique comparison of machine learning techniques

Prediction models with DT, RF, and SVM were established. For a DT model, a pathway was established. Fig. 1 displays a visual depiction of the DT model algorithm. A variable-selected model of RF showed an inferior result to the initial model. The best RF model with the model containing all variables was established via grid search, of which the hyperparameters was the following: 3 nodes, 500 trees and 3 tries.

An initial SVM was established with all the variables included, which yielded the accuracy of 0.7270, however, the very low sensitivity of 0.0273. The importance plot for the SVM showed the order of importance of the variables: the first was working status, and the following variables

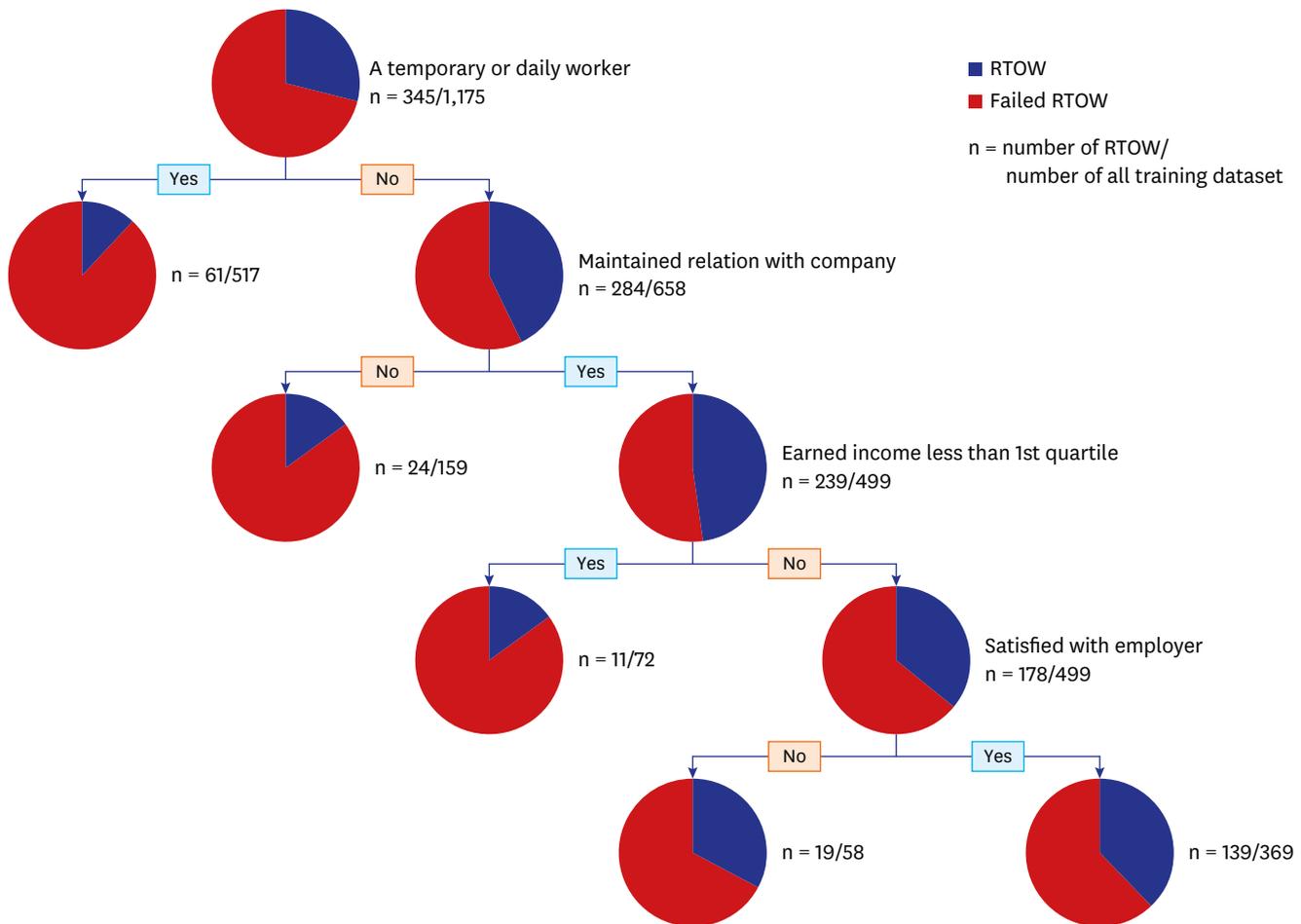


Fig. 1. Algorithm of the decision tree model. RTOW = Return-to-original-work.

Table 5. Performance of machine learning techniques

Machine learning technique	Mean \pm standard deviation of 10 iterations				
	Accuracy	Sensitivity	Specificity	Kappa	AUC
Logistic regression ^a	0.7702 \pm 0.0104	0.5514 \pm 0.0437	0.8543 \pm 0.0180	0.4137 \pm 0.0327	0.812 \pm 0.012
Logistic regression ^b	0.7796 \pm 0.0087	0.5323 \pm 0.0311	0.8747 \pm 0.0170	0.4251 \pm 0.0215	0.805 \pm 0.011
Decision tree ^a	0.7630 \pm 0.0117	0.6734 \pm 0.0395	0.7973 \pm 0.0281	0.4431 \pm 0.0191	0.761 \pm 0.019
Random forest ^b	0.7875 \pm 0.0122	0.4580 \pm 0.0484	0.9138 \pm 0.0194	0.4115 \pm 0.0359	0.811 \pm 0.010
Support vector machine ^a	0.7712 \pm 0.0095	0.4727 \pm 0.0361	0.8856 \pm 0.0167	0.3851 \pm 0.0248	0.798 \pm 0.010
Average	0.7743	0.5376	0.8651	0.4157	0.797

AUC = area under the curve.

^aModels used only their own significant variables selected by authors for prediction; ^bModels used all variables for prediction.

were earned income, maintenance of relationship with company, and satisfaction level with employer, in order (**Supplementary Table 1**). We could upgrade the function of SVM by selecting these four variables, which yielded better accuracy and sensitivity (0.7602 and 0.5909). After hyperparameter tuning by grid search algorithm, the final SVM model was established. We showed detailed process of establishing the DT, RF, and SVM model in **Supplementary Table 2**.

The MLR and RF models showed optimal prediction when they used all variables, while the others showed optimal prediction when they used only the significant variables. **Table 5** shows the final performance comparison of the models from each machine learning technique. The RF model showed the best performance in terms of accuracy and specificity (0.7875 and 0.9138, respectively). The DT model showed the highest sensitivity and kappa value (0.6734 and 0.4431, respectively). The MLR model using significant variables showed the best AUC (0.812). Ranked in order of predictive accuracy, the models were: RF, MLR, SVM, and DT. However, the accuracy of the models had only minor variation (maximally 0.025) despite statistical significance (**Supplementary Table 3**).

Supplementary Table 1 lists the factors that were selected finally in relation with RTOW in the MLR, DT, RF, and SVM models. Each model contained four to six significant variables. The DT and SVM models recognized only four variables as significant; the RF model, five; and the MLR model, six. When we analyzed the first random seed, the order of significance of the factors was different in each model. As also reflected by the ORs of the MLR analysis, the MLR model identified earned income as the most important variable, whereas the DT and SVM models identified working status as the most important variable. The RF model identified 'maintenance or relation with company' as the most significant variable. This order of significance differed by a new random allocation of training and test dataset. However, the four variables were always on the top of importance in all the models.

DISCUSSION

This study showed that machine learning techniques can predict the probability of RTOW by four years from the end of the official recovery period. The RF model showed the best performance, but the other models showed almost comparable performance. This is the first study using machine learning techniques to predict RTW. Machine learning techniques have a more complex mathematical basis than classical MLR modeling; however, since they do not assume linearity, machine learning models are more flexible and require fewer variables.

Predictive models with machine learning techniques are anticipated to be adopted more frequently in medicine and social sciences in the future. Prediction of various outcomes by

machine learning techniques has already become common in medicine, such as mortality,¹² surgical outcomes,²⁴ and psychiatric illness.²⁵ This study used similar methodologies of these previous trials: comparing variables with conventional statistical methods, establishing machine learning models with selecting all the variables or selected variables for efficiency, and hyperparameter tuning process to maximize the predictive accuracy.

When optimizing hyperparameters of RF and SVM models, we used the grid search mechanism which is a traditional method with some weaknesses. The grid search does not guarantee that the results were mathematically the most optimal. Therefore, novel methodologies such as random search, Bayesian optimization, gradient-based optimization, and evolutionary optimization are being developed and applied currently. This study did not use these current methods for hyperparameter tuning because of limited software resources of the R package caret. However, although our methodology could not guarantee a mathematical robustness, the grid search mechanism has been widely used, so that we estimate that our results are not far from the very optimized values. We also expect that this simple method could be applied practically in KCOMWEL.

If our models be applied in practice of KCOMWEL, it would be possible making an individual policy of RTW of an injured worker; for example, providing an early reemployment administration service to a worker who is expected to fail of RTOW. However, there might be considerations other than a simple prediction. In this study, we just showed the possibility of prediction and it is necessary more detailed discussion on availability of the prediction.

Comparing all the models introduced in this study, the RF model showed the best performance (**Table 5**) in terms of accuracy significantly (**Supplementary Table 3**). However, since this model used all variables in its analysis and its tuning process is automated, and RF sometimes yields unstable results by each trial, its practical applications are limited.

The SVM model showed results that were similar to those of the MLR model, even if they used only four variables in the analyses. This feature of SVM modeling — i.e., better efficiency with fewer variables¹¹ — has potential practical advantages. A modeling method using SVM is referred to as a ‘black box’ model because of its mathematical complexity²⁶; thus, the interpretation is limited.

Classical MLR modeling yields ORs of independent variables that affect dependent variables. The theoretical assumption behind this method is that ORs approximate risk ratios if the sample size is sufficient. However, regarding epidemiology, it is still unknown what is the meaning of features selected in SVM and RF. Future efforts should investigate how to interpret factors identified by other machine learning techniques in the context of social epidemiology.

Even if the accuracy is inferior, the results of the DT model seem to represent an acceptable compromise of practice and interpretation, because the mathematical methodology used to discriminate variables that maximize information gain is simpler than that used in other machine learning techniques. In particular, the DT model in this study showed a one-way path that is very easy to implement in practice. If an employee is not a temporary or daily worker, maintains a relationship with the company during the recovery period, has earnings in the upper quartile, and is satisfied with the employer's convenience, the DT model predicts that he/she will return to his/her original work with a sensitivity and specificity of 67% and

80%, respectively. Therefore, the DT model's four questions could be used when assessing how to best support patients from industrial accidents.

The differences between the techniques were not prominent; only 0.025 of difference of accuracy between the superior model (RF) and the inferior one (DT). These results are similar to a comparison study that showed statistically not significant differences of methodologies with naïve Bayes, RF, DT, SVM, and MLR even though the MLR model showed the best performance.²⁷ It means that there is not so much to expect in choosing the most elaborate method. It is necessary to select an appropriate model by advantages and disadvantages in order to predict a worker's possibility of RTOW. For example, RF model for the best accuracy; DT model for clear and easy-to-use standard; SVM model for a simple composition of necessary variables.

Regarding the characteristics of important predictors for RTOW, all models selected similar variables as important for predicting RTOW, although some differences were observed among the models. Four variables — earned income, maintenance of relation with company, satisfaction level with employer, and working status — were selected in all models. These variables can be classified as 'variables related to the company itself.' While nearly all the variables except these important variables were also significantly associated with RTOW in the χ^2 analyses, no significant effects were observed for any of the variables in the machine learning models.

This finding means that the strongest component influencing RTOW and its durability is a worker's company. This finding is consistent with a previous study³ that identified workplace characteristics as significant predictive factors for RTW. Another study from the first PSWCI¹⁴ reported that sex, age, regular recovery assessment, and opinion of duration of medical treatment were related to RTW, similar to the findings of our study. However, the data obtained from the longer follow-up in our study suggest that company-related factors were more influential than individual or supportive factors.

As well as the four variables, formal contract for employment and average working hours per day were selected in the MLR model and age in the RF model. However, the additional variables were not more significant than the four variables in each model while the four variables showed little differences among the models and internal fluctuations. Therefore, although there were some differences due to technical methods, all models indicated that variables related to the company itself would influence RTOW. In this point, staffs and medical practitioners who are working for patients aiming RTW should consider the characteristics of the patients' original companies.

Earned income showed the biggest ORs in the MLR model and had a positive gradient, as all models selected it as an important variable. This feature should be discussed in the context of social justice, e.g., whether workers' compensation insurance is adequate in the social security system. KCOMWEL provides various services and programs for RTW, as mentioned in the Introduction. Of these services and programs, consultation for RTW with medical doctors and work ability assessment referral for RTW were associated with RTW in the χ^2 analysis (Table 3). However, this effect was not observed in the MLR model or in the others. This finding suggests that these services were closely associated with company-related factors, i.e., the services necessary for RTW were provided to workers who already had a good base for RTW. If the services had similar effects on all workers, related factors should have been detected in the

machine learning analyses. This result is strongly supported by the segmented labor market in Korea: a worker's situation before injury influences the quality of employment after RTW.²⁸

This study has some limitation. It did not use variable weights, because it is unclear how to handle these weights when developing training models. Weights are used to assure representativeness, but it is unknown how these weights should be used to guide training and test datasets. Moreover, machine learning techniques are not typically used with weighted variables. The follow-up rate was 83%, and this study did not consider dropouts. Therefore, the results of our study are not necessarily applicable to all workers nationwide. However, the results of previous studies and the results of the MLR model without weighted variables in this study were similar.

We showed that, by using only four to six variables at the end of a worker's recovery period, the probability of RTOW can be predicted with an accuracy of 76%–78%. Since additional factors that were not included in our study might change the predictive performance of the models, future studies are needed to investigate and predict RTW more precisely.

ACKNOWLEDGMENTS

The authors thank to Korea Workers' Compensation and Welfare Service (KCOMWEL) for providing the raw data of Panel Study of Worker's Compensation Insurance (PSWCI) and all the participants of the survey.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

Significant variables by machine learning techniques

[Click here to view](#)

Supplementary Table 2

Process of establishing a decision tree and a random forest model

[Click here to view](#)

Supplementary Table 3

P value for ANOVA of accuracy among machine learning techniques

[Click here to view](#)

REFERENCES

1. Industrial accident compensation insurance act. http://elaw.klri.re.kr/kor_service/lawView.do?hseq=41243&lang=ENG. Updated 2016. Accessed March 12, 2018.
2. Ahn JK, Oh SM. What happens after the injury? Self-esteem and change in employment status of occupationally injured workers. *Korean Soc Secur Stud* 2015;31(1):109-35.

3. Jeong WM, Park CY, Koo JW, Roh YM. Predictors of return to work in occupational injured workers. *Korean J Occup Environ Med* 2003;15(2):119-31.
4. Kang HT, Im HJ, Kim YK, Ju YS, Lee HP, Kim JM, et al. Predictors of return to work and job retention after work-related injury or illness. *Korean J Occup Environ Med* 2006;18(3):221-31.
5. Park SK, Ahn CM. Predictors of job retension among worker's compensation clients in South Korea. *Soc Welf Policy* 2006;12:177-94.
6. Ervasti J, Joensuu M, Pentti J, Oksanen T, Ahola K, Vahtera J, et al. Prognostic factors for return to work after depression-related work disability: a systematic review and meta-analysis. *J Psychiatr Res* 2017;95:28-36.
[PUBMED](#) | [CROSSREF](#)
7. Gragnano A, Negrini A, Miglioretti M, Corbière M. Common psychosocial factors predicting return to work after common mental disorders, cardiovascular diseases, and cancers: a review of reviews supporting a cross-disease approach. *J Occup Rehabil* 2017. DOI: 10.1007/s10926-017-9714-1.
[PUBMED](#) | [CROSSREF](#)
8. Clay FJ, Newstead SV, McClure RJ. A systematic review of early prognostic factors for return to work following acute orthopaedic trauma. *Injury* 2010;41(8):787-803.
[PUBMED](#) | [CROSSREF](#)
9. Hou WH, Tsauo JY, Lin CH, Liang HW, Du CL. Worker's compensation and return-to-work following orthopaedic injury to extremities. *J Rehabil Med* 2008;40(6):440-5.
[PUBMED](#) | [CROSSREF](#)
10. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313(13):793-9.
[PUBMED](#) | [CROSSREF](#)
11. Salazar DA, Velez JI, Salazar JC. Comparison between SVM and logistic regression: which one is better to discriminate? *Rev Colomb Estad* 2012;35(2):223-37.
12. Verplancke T, Van Looy S, Benoit D, Vansteelandt S, Depuydt P, De Turck F, et al. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC Med Inform Decis Mak* 2008;8:56.
[PUBMED](#) | [CROSSREF](#)
13. Seok H, Yoon JH, Lee W, Lee JH, Jung PK, Roh J, et al. The perceived socioeconomic status is an important factor of health recovery for victims of occupational accidents in Korea. *J Korean Med Sci* 2016;31(2):164-70.
[PUBMED](#) | [CROSSREF](#)
14. Lee W, Yoon JH, Roh J, Kim YK, Seok H, Lee JH, et al. Factors related to the physician and the employer influencing successful return to work in Korea: results from the first panel study of workers' compensation insurance (PSWCI). *Ann Occup Environ Med* 2015;27:27.
[PUBMED](#) | [CROSSREF](#)
15. Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1(1):81-106.
[CROSSREF](#)
16. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.
[CROSSREF](#)
17. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273-97.
[CROSSREF](#)
18. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw* 2008;28(5):1-26.https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=27774042&dopt=Abstract
[PUBMED](#) | [CROSSREF](#)
19. Package "ROCR". <https://cran.r-project.org/web/packages/ROCR/ROCR.pdf>. Updated 2015. Accessed March 12, 2018.
20. Package "rpart". <https://cran.r-project.org/web/packages/rpart/rpart.pdf>. Updated 2018. Accessed March 12, 2018.
21. Package "rpart.plot". <https://cran.r-project.org/web/packages/rpart.plot/rpart.plot.pdf>. Updated 2017. Accessed March 12, 2018.
22. Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. *J Stat Softw* 2006;15(9):1-28.
[CROSSREF](#)
23. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - an S4 package for kernel methods in R. *J Stat Softw* 2004;11(9):1-20.
[CROSSREF](#)
24. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman ML, et al. Machine learning and neurosurgical outcome prediction: a systematic review. *World Neurosurg* 2018;109:476-486.e1
[PUBMED](#) | [CROSSREF](#)

25. Rosellini AJ, Dussaillant F, Zubizarreta JR, Kessler RC, Rose S. Predicting posttraumatic stress disorder following a natural disaster. *J Psychiatr Res* 2018;96:15-22.
[PUBMED](#) | [CROSSREF](#)
26. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49(11):1225-31.
[PUBMED](#) | [CROSSREF](#)
27. Pranckevičius T, Marcinkevičius V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Balt J Mod Comput* 2017;5(2):221-32.
28. Lee JH. “Decent” Return-To-Work (RTW): the relationship among individual, segmented labor market, and RTW program. *Health Soc Welf Rev* 2017;37(2):389-422.
[CROSSREF](#)