Received: 27 September 2022 | Revised: 29 December 2022 | Accepted: 25 January 2023

DOI: 10.1002/1348-9585.12392

Journal of Occupational Health Open Access WILEY

ORIGINAL ARTICLE

Status and prediction of disapproval of the Korean workers' compensation insurance for diseases and injuries

Soo Beom Choi[1] | **Seunghyun Lee**[2] | **Wanhyung Lee**[2]

[1]Department of Smart City Research, Seoul Institute of Technology, Seoul, Republic of Korea

[2]Department of Occupational and Environmental Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

Correspondence
Wanhyung Lee, Department of Occupational and Environmental Medicine, Gil Medical Center, Gachon University College of Medicine, 21, Namdong-daero 774, Namdong-gu, Incheon, 21565 Republic of Korea.
Email: wanhyung@gmail.com

Abstract

Objective: Workers' diseases and injuries are often highly related to work. However, due to limited resources and unclear work relatedness, workers' compensation insurance cannot cover all diseases or injuries among workers. This study aimed to estimate the status and probability of disapproval from national workers' compensation insurance using basic information from Korean workers' compensation system.

Methods: The compensation insurance data for Korean workers consists of personal, occupational, and claims data. We describe the status of disapproval by workers' compensation insurance according to the type of disease or injury. A prediction model for disapproval by workers' compensation insurance was established by applying two machine-learning methods with a logistic regression model.

Results: Among 42 219 cases, there were significantly higher risks of disapproval by workers' compensation insurance for women, younger workers, technicians, and associate professionals. We established a disapproval model for workers' compensation insurance after the feature selection. The prediction model for workers' disease disapproval by the workers' compensation insurance showed a good performance, and the prediction model for workers' injury disapproval showed a moderate performance.

Conclusion: This study is the first attempt to demonstrate the status and prediction of disapproval by workers' compensation insurance using basic information from the Korean workers' compensation data. These findings suggest that diseases or injuries have a low level of evidence of work relatedness or there is a lack of research on occupational health. It is also expected to contribute to the efficiency of the management of workers' diseases or injuries.

KEYWORDS

Korea and Welfare Service, occupational disease, occupational injury, prediction, Workers' compensation, workers' compensation

Soo Beom Choi & Seunghyun Lee have contributed equally to this work.

1 | INTRODUCTION

Worker's health is closely related to their occupational environment. Occupational accidents occur at work, leading to nonfatal or fatal injuries. Occupational disease is a health condition or disorder caused by exposure to risk factors from work activity or the workplace environment. Occupational injury that is distinct from an occupational disease includes worker's injury and disease as well as death due to occupational accidents. In terms of loss of workforce and financial burden, these are public concerns. These events have a major impact on workers' livelihoods and health-related quality of life.[1]

According to the World Health Organization and International Labor Organization, occupational injuries account for almost 20% of deaths annually. More than 80 percent of deaths are due to occupational risk factors for noncommunicable diseases, including stroke, heart disease, and chronic obstructive pulmonary disease. The burden of occupational injuries accounts for almost 20% of deaths annually. This factor accounted for the highest number of DALYs in thousands and the third-highest number of occupational deaths.[2] In Korea, 20 435 cases of occupational diseases, and 2080 cases of occupational injuries were reported in 2021.

As industrialization progresses, most nations have compensation insurance systems for industrial accidents. Although the laws and regulations concerning occupational diseases and injuries differ in each country, the ultimate principle is to compensate workers for occupational injury or disease and facilitate rehabilitation and return to society.[3,4] In Korea, industrial accident compensation insurance is applied to all workplaces. The act is based on the no-fault liability principle, whether the accident occurs intentionally or negligently. Occupational injury covers any injury that occurs when a worker performs a duty under their employment contract when it occurs while a worker uses a facility provided by the employer due to any defect or reckless management of such a facility etc. The scope of occupational disease includes any disease caused by handling or exposure to any physical or chemical substance, dust, or pathogen or work imposing a burden on workers' bodies, and causing occupational injury. There are eight categories of insurance benefits: medical care, temporary disability compensation, disability, nursing, survivors', and vocational rehabilitation benefits; injury/disease compensation annuities; and funeral expenses. Under the Industrial Accident Compensation Insurance Act, only paid employees in Korea are subject to compensation. The compensation range depends on the workers' compensation insurance.[5]

Covering all workers' diseases and injuries by the compensation system is the best, but it is sometimes limited because of limited resources. Work relatedness is a key factor in deciding the approval or disapproval of workers' compensation insurance in Korea. Although the Act has detailed criteria for approval or disapproval, its results are not consistent and depend on the situation of the case. To be recognized as an occupational disease or injury, every single case needs to prove a causal relationship between the accident and the occupational factors, and how or what can affect the course of the case in their working environment.[6]

For cases to be recognized as occupational diseases or injuries, complicated administrative processes.[7] For some occupational disease approval processes, it takes approximately a year from the occurrence of the disease to the receipt of an application for insurance benefits. In some cases, a series of compensation systems may intensify ill health among injured workers and have negative effects on them.[8,9] Therefore, patients affected by a possible occupational disease often find themselves in a difficult situation that is characterized by economic problems and uncertainty related to the future. This is particularly true in the period when the patient is examined for a final diagnosis and function level while waiting for a final decision on whether the disease will be recognized as an occupational disease with economic compensation. In this period, people may experience a potential conflict between treatment and recovery, on the one hand, and the possible economic benefits of being sick, on the other. Thus, the prediction of disapproval of claims of occupational diseases or injuries could have a proactive impact at both the individual and national levels. As swift decisions on compensation claims could aid a fast return to work, the prediction could curtail the monetary and productivity loss to workers or the bereaved family.[6] At the same time, the prediction of occupational disease or injury reduces waste of both human and federal resources from application to disbursement of benefits and helps in making faster decisions at the national level. Furthermore, this study encourages scholarly attention in the field of occupational health and provides a direction for occupational health research.

This study aims to identify the status and probability of disapproval of claims by national workers' compensation insurance using basic information from the Korean workers' compensation system.

2 | METHODS

2.1 | Data and study participants

The Korea Workers' Compensation and Welfare Service (KCOMWEL, https://www.comwel.or.kr/) has provided various social security and labor welfare services and

programs, as specified in relevant laws and regulations, including the integrated assessment and billing of workers' compensation insurance premiums, medical care and rehabilitation services, and credit guarantee and wage claim guarantee programs among other programs. Workers' compensation insurance in the Korean scheme is a social security scheme to compensate workers exposed to a work-related accident for their consequent injury or disease on a timely and fair basis, thereby facilitating their rehabilitation and return to society. KCOMWEL has opened the database of workers' compensation insurance to the public. Information includes the date of claim, injury, diseases, age at event, sex, type of work, migrant workers, International Standard Classification of Occupations (ISCO), job duration, International Classification of Diseases (ICD), and results of insurance judgment. This study used data from KCOMWEL to access the current three-year dataset (2017, 2018, and 2019).

2.2 | Measurement and variables

Insurance claim-related data contained three variables: year of claim, results of judgment (approval, partial approval, or disapproval), and type of claim (injury or disease). Personal information included age at the time of the event and sex. Occupation-related information consisted of the type of work (full time or part time), migrant workers, job duration, and ISCO. Injury or disease was categorized into 22 major domains based on the ICD-10.

The list of 22 major domains with codes is as follows: I. Certain infectious and parasitic diseases (A00–B99); II. Neoplasms (C00–D48); III. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50–D89); IV. Endocrine, nutritional, and metabolic diseases (E00–E90); V. Mental and behavioral disorders (F00–F99); VI. Diseases of the nervous system (G00–G99), VII. Diseases of the eye and adnexa (H00–H59); VIII. Diseases of the ear and mastoid processes (H60–H95); IX. Diseases of the circulatory system (I00–I99); X. Diseases of the respiratory system (J00–J99); XI. Diseases of the digestive system (K00–K93); XII. Diseases of the skin and subcutaneous tissue (L00–L99); XIII. Diseases of the musculoskeletal system and connective tissue (M00–M99); XIV. Diseases of the genitourinary system (N00–N99); XV. Pregnancy, childbirth, puerperium (O00–O99); XVI. Certain conditions originating in the perinatal period (P00–P96); XVII. Congenital malformations, deformations, chromosomal abnormalities (Q00–Q99); XVIII. Symptoms, signs, and abnormal clinical and laboratory findings were not classified elsewhere (R00–R99); XIX. Injury, poisoning, and other consequences of external causes (S00–T98); XX. External causes of morbidity and mortality (U00–U99); XXI. Factors influencing health status and contact with health services (V01–Y98); and XXII. Codes for special purposes (Z00–Z99).

2.3 | Statistical analysis

The prevalence of disapproval of occupational disease or injury according to baseline characteristics was assessed using the chi-square test. To improve the performance in predicting disapprovals of occupational diseases and injuries, we applied two novel machine-learning methods along with a logistic regression model. Input variables were selected using multivariate logistic regression analysis with backward stepwise elimination. Extreme gradient boosting (XGBoost) and neural networks (NNs) have shown high performance in prediction models in recent clinical domains.[10]

XGBoost is a novel machine-learning algorithm that was popularized by Chen and Guestrin in February 2014. The XGBoost algorithm is an advanced version of the gradient boosting decision tree and ensemble tree method that combines multiple decision trees in a boosting manner.[10] XGBoost continuously adds each new tree to reduce the residual, which is the difference between actual and predicted values.[11] XGBoost uses hyperparameters to control overfitting and improve performance, such as the number of estimators, learning rate, and maximum tree depth. A more detailed description and hyperparameters of XGBoost are available on the website.[12] The advantage of XGBOOST is that each tree learns from its predecessor and updates residual errors. The tree that grows next in the sequence learns from the updated versions of the residuals. Distributed computing ensures faster learning.

An NN is a machine-learning method inspired by the architecture of biological neurons, such as those in the human brain.[13] The input layer receives an input signal that moves to the next layer as a modified version of the input signal. It passes through several layers composed of multiple transformations and finally through the output layer as an output signal.[14] The advantage of NN is that, compared to traditional linear and simple nonlinear analyses, self-adapting models perform best when the relationship between the input and target responses is highly nonlinear.

In this study, NN models were optimized using the Adam optimizer with a binary cross-entropy loss function. The neuron activation functions were rectified linear units for the first layer with seven nodes and a sigmoid function for the second layer with three nodes. Dropouts were not chosen because of the small number of nodes. The NN was implemented using the Python library Keras in TensorFlow (version 2.5.0P).

2.4 | Prediction model

The data sets for disapproval of claims of occupational diseases and injuries were randomly divided into two independent training and validation sets to test internal validation, which was stratified by the dependent variable. The training set, comprising 75% of the data set, was used to construct the prediction models. The validation set, comprising 25% of the data set, was used to assess the performance of the model in predicting disapproval of occupational diseases and injuries.

The data sets were imbalanced because of the low incidence of disapproval of claims of occupational diseases and injuries. Data sets with imbalanced classes tend to be difficult to handle for machine-learning algorithms.[15] We selected an oversampling technique to solve this problem by matching the ratio between the major and minor groups using Synthetic Minority Oversampling TEchnique (SMOTE).

To obtain the optimal model, we adopted a grid search in which a range of parameter values was tested using fivefold cross validation.[11] The parameter values for XGBoost were the number of estimators (100, 200, 400, and 600), the learning rate (0.01, 0.05, 0.1, 0.15, and 0.2), and the maximum depth (4,6,8,10,and 12). The parameter values for the NN were batch size (10, 50, and 100) and epochs (10, 50, and 100). To validate the performance of the prediction models, we performed accuracy and receiver operating characteristic curve-area under the curve (ROC-AUC) analyses. Accuracy is the number of correctly predicted data points from all data points.[12] The ROC-AUC was executed to verify the performance of each prediction model for the disapproval of claims of occupational diseases and injuries using validation sets. Perfect discrimination showed an AUC of 1.0, and the worst AUC was 0.5.[16] The categories of ROC analysis were as follows: 0.9–1, excellent; 0.8–0.9 and 0.7–0.8, fair; 0.6–0.7 and 0.5–0.6, fail.[17]

An overall flowchart of the prediction model is shown in Supplementary Figure S1. The scikit-learn library (https://scikit-learn.org/stable/) was used for data management and preprocessing. Preprocessing and machine-learning analyses were performed using Python 3.6.2 (Python Software Foundation).

3 | RESULTS

In this study, there were 42 219 cases of occupational diseases and injuries from 2017 to 2019, comprising 24 484 approved cases and 17 735 disapproved cases. Table 1 shows the approval status of workers diagnosed with an occupational disease. The majority of claimants have significantly approved of their disease as work-related. Among the economically active population, those who were of <20 years and were starting their working lives and careers and those aged >70 years showed a low approval rate for occupational disease. In the occupational categories, professionals, individuals in the service industry, and career soldiers were recognized as having a low risk of work-related diseases. According to ICD-10 categories, some diseases showed low approval rates, including E00–E90, H00–H59, J00–J99, K00–K93, O00–O99, Q00–Q99, R00–R99, and Z00–Z99. Table 2 shows the workers' approval status for occupational injuries resulting from accidents. Almost all workers have been approved for occupational injuries, regardless of their age, type of work, nationality, work duration, and occupational categories. In accordance with ICD code classifications, most injuries were approved as work-related accidents, except for C00–D48 and E00–E90.

Table 3 shows the disapproval of claims of occupational diseases in an assortative manner, using logistic regression. In the women worker group, the odds of disapproval for occupational diseases were 1.428 (95% confidence interval [CI], 1.337–1.522). Workers aged <20 and >70 years had a high risk of disapproval of claims for occupational diseases (odds ratio [OR], 3.477; 95% CI, 1.798–1.522 in those aged <20 years; OR, 1.593; 95% CI, 1.454–1.745 in those aged >70 years). When the results were stratified by occupational classification, the technician and associate professional groups had the highest OR (95% CI) for occupational diseases compared with other groups (OR, 4.109; 95% CI, 3.434–4.918). In accordance with the ICD code classifications, K00–K93 had the highest OR (95% CI) for occupational diseases (OR, 10.144; 95% CI, 5.819–17.683). Table 4 shows the disapproval of claims for occupational injuries using logistic regression. In contrast to occupational diseases, the odds of occupational injury were significantly higher in men, migrant workers, and workers working for 1 year. In accordance with ICD code classifications, M00–M99 had the highest OR (95% CI) for disapproval of occupational injury (OR, 19.530; 95% CI, 18.029–21.156).

After a grid search to select the optimal parameters, we selected 0.15 learning rate, 4 max depth, and 200 estimators of the XGBoost model, 10 batch sizes, and 100 epochs of NN for predicting the disapproval of occupational diseases. To predict the disapproval of occupational injury, the final XGBoost model with 0.2, 12, and 600 estimators, and the final NN model with 100 batch sizes and 100 epochs were selected.

Table 5 shows the performance of the prediction models for disapproval of occupational diseases and injuries using logistic regression, XGBoost, and NN. The XGBoost models for disapproval of occupational diseases and

TABLE 1 Characteristics of occupational disease according to the approval status.

	Occupational diseases		
	Approval (including partial approval) ($n = 24\,484$)	**Disapproval ($n = 17\,735$)**	***P*-value**
Year			<.001
2017	5901 (51.3)	5590 (48.7)	
2018	7655 (60.0)	5112 (40.0)	
2019	10 928 (60.8)	7033 (39.2)	
Sex			.0052
Male	20 688 (58.3)	14 809 (41.7)	
Female	3796 (56.4)	2926 (43.6)	
Age (years)			<.001
<20	15 (38.5)	24 (61.5)	
20–29	641 (54.7)	530 (45.3)	
30–39	2395 (63.5)	1372 (36.5)	
40–49	4630 (63.7)	2636 (36.3)	
50–59	8126 (63.2)	4719 (36.8)	
60–69	6774 (56.1)	5297 (43.9)	
≥70	1903 (37.6)	3157 (62.4)	
Type of work			<.001
Full time	20 247 (58.6)	14 313 (41.4)	
Part time	4237 (55.3)	3422 (44.7)	
Migrant workers			.0122
No	24 175 (58.0)	17 460 (42.0)	
Yes	309 (52.8)	275 (47.2)	
Job duration			.0480
<1 years	5485 (57.1)	4118 (42.9)	
≥1 years	18 999 (58.2)	13 617 (41.8)	
International standard classification of occupations			<.001
Managers	1515 (52.1)	1392 (47.9)	
Professionals	642 (48.4)	684 (51.6)	
Technicians and associate professionals	234 (32.6)	481 (67.4)	
Clerical support workers	1848 (65.9)	954 (34.1)	
Service and sales workers	264 (47.4)	293 (52.6)	
Skilled agricultural, forestry, and fishery worker	217 (68.5)	100 (31.5)	
Craft and related trade workers	5132 (70.3)	2167 (29.7)	
Plant and machine operators and assemblers	3495 (64.5)	1921 (35.5)	
Elementary occupations	11 098 (53.4)	9684 (46.6)	
Armed forces occupations	39 (39.8)	59 (60.2)	
International classification of diseases			<.001
A00–B99:Certain infectious and parasitic diseases	180 (65.9)	93 (34.1)	
C00–D48: Neoplasms	661 (69.5)	290 (30.5)	
D50–D89: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	6 (40.0)	9 (60.0)	

(Continues)

TABLE 1 (Continued)

	Occupational diseases		
	Approval (including partial approval) ($n = 24484$)	**Disapproval ($n = 17735$)**	***P*-value**
E00–E90: Endocrine, nutritional, and metabolic diseases	3 (11.5)	23 (88.5)	
F00–F99: Mental and behavioral disorders	356 (68.2)	166 (31.8)	
G00–G99: Diseases of the nervous system	556 (58.4)	396 (41.6)	
H00–H59: Diseases of the eye and adnexa	6 (15.8)	32 (84.2)	
H60–H95: Diseases of the ear and mastoid process	4028 (55.9)	3171 (44.1)	
I00–I99: Diseases of the circulatory system	2369 (40.7)	3456 (59.3)	
J00–J99: Diseases of the respiratory system	1666 (30.7)	3766 (69.3)	
K00–K93: Diseases of the digestive system	16 (19.3)	67 (80.7)	
L00–L99: Diseases of the skin and subcutaneous tissue	67 (50.4)	66 (49.6)	
M00–M99: Diseases of the musculoskeletal system and connective tissue	10 599 (70.4)	4455 (29.6)	
N00–N99: Diseases of the genitourinary system	20 (31.7)	43 (68.3)	
O00–O99: Pregnancy, childbirth, and the puerperium	2 (33.3)	4 (66.7)	
P00–P96: Certain conditions originating in the perinatal period	—	—	
Q00–Q99: Congenital malformations, deformations, and chromosomal abnormalities	1 (14.3)	6 (85.7)	
R00–R99: Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	176 (37.5)	293 (62.5)	
S00–T98: Injury, poisoning, and certain other consequences of external causes	3768 (72.9)	1398 (27.1)	
U00–U99: Codes for special purposes	4 (100.0)	0 (0.0)	
V01–Y98: External causes of morbidity and mortality	—	—	
Z00–Z99: Factors influencing health status and contact with health services	0 (0.0)	1 (100.0)	

Note: The *P*-values were calculated by the chi-square test.

injuries showed the best performance, with ROC-AUC values of 0.724 and 0.660, respectively. The differences between the accuracies of the training and validation sets were small, indicating that the prediction models fit the data well. Figure 1 shows the performance of the prediction models for the disapproval of occupational diseases and injuries in the validation sets.

4 | DISCUSSION

This study aimed to assess the prediction of the risk of disapproval of claims for occupational diseases and injuries among Korean workers. Occupational injuries and diseases have different pathophysiologies and etiologies among workers, and we conducted an analysis after stratification. We used a basic data set from KCOMWEL, and it was worthwhile to obtain meaningful results with minimal information. We demonstrated good performance of the prediction model for the disapproval of claims for occupational diseases (ROC-AUC of 0.724 by XGBoost) and moderate performance of the prediction model for the disapproval of claims for occupational injury (ROC-AUC of 0.660 by XGBoost).

This study showed that there is a higher probability of disapproval of claims for occupational diseases than

TABLE 2 Characteristics of occupational injury according to the approval status.

	Occupational injury		
	Approval (including partial approval) ($n = 253\,199$)	Disapproval ($n = 10\,418$)	*P*-value
Year			<.001
2017	72 792 (95.3)	3578 (4.7)	
2018	87 537 (96.1)	3574 (3.9)	
2019	92 870 (96.6)	3266 (3.4)	
Sex			.0010
Male	187 372 (96.0)	7860 (4.0)	
Female	65 827 (96.3)	2558 (3.7)	
Age (years)			<.001
<20	2806 (96.0)	118 (4.0)	
20–29	29 761 (96.8)	992 (3.2)	
30–39	35 660 (96.4)	1346 (3.6)	
40–49	49 731 (96.0)	2068 (4.0)	
50–59	78 060 (96.0)	3219 (4.0)	
60–69	48 290 (95.7)	2176 (4.3)	
≥70	8891 (94.7)	499 (5.3)	
Type of work			.3342
Full time	150 968 (96.0)	6261 (4.0)	
Part time	102 231 (96.1)	4157 (3.9)	
Migrant workers			<.001
No	235 734 (95.9)	10 017 (4.1)	
Yes	17 465 (97.8)	401 (2.2)	
Job duration			<.001
<1 years	160 046 (96.2)	6237 (3.8)	
≥1 years	93 153 (95.7)	4181 (4.3)	
International standard classification of occupations			<.001
Managers	19 907 (94.5)	1160 (5.5)	
Professionals	10 728 (94.6)	608 (5.4)	
Technicians and associate professionals	5566 (92.8)	434 (7.2)	
Clerical support workers	29 549 (96.7)	999 (3.3)	
Service and sales workers	3966 (95.3)	194 (4.7)	
Skilled agricultural, forestry, and fishery worker	4614 (97.1)	138 (2.9)	
Craft and related trade workers	49 097 (96.6)	1717 (3.4)	
Plant and machine operators and assemblers	25 978 (96.6)	926 (3.4)	
Elementary occupations	101 134 (96.1)	4069 (3.9)	
Armed forces occupations	2660 (93.9)	173 (6.1)	
International classification of diseases			<.001
A00–B99: Certain infectious and parasitic diseases	33 (76.7)	10 (23.3)	
C00–D48: Neoplasms	3 (42.9)	4 (57.1)	
D50–D89: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	3 (100.0)	0 (0.0)	

(Continues)

TABLE 2 (Continued)

	Occupational injury		
	Approval (including partial approval) ($n = 253199$)	Disapproval ($n = 10\,418$)	*P*-value
E00–E90: Endocrine, nutritional, and metabolic diseases	5 (20.0)	20 (80.0)	
F00–F99: Mental and behavioral disorders	8 (66.7)	4 (33.3)	
G00–G99: Diseases of the nervous system	128 (67.0)	63 (33.0)	
H00–H59: Diseases of the eye and adnexa	520 (92.7)	41 (7.3)	
H60–H95: Diseases of the ear and mastoid process	82 (75.9)	26 (24.1)	
I00–I99: Diseases of the circulatory system	232 (80.0)	58 (20.0)	
J00–J99: Diseases of the respiratory system	132 (93.0)	10 (7.0)	
K00–K93: Diseases of the digestive system	202 (85.2)	35 (14.8)	
L00–L99: Diseases of the skin and subcutaneous tissue	559 (88.6)	72 (11.4)	
M00–M99: Diseases of the musculoskeletal system and connective tissue	1587 (58.3)	1134 (41.7)	
N00–N99: Diseases of the genitourinary system	21 (87.5)	3 (12.5)	
O00–O99: Pregnancy, childbirth, and the puerperium	4 (80.0)	1 (20.0)	
P00–P96: Certain conditions originating in the perinatal period	-	-	
Q00–Q99: Congenital malformations, deformations, and chromosomal abnormalities	2 (100.0)	0 (0.0)	
R00–R99: Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	271 (83.1)	55 (16.9)	
S00–T98: Injury, poisoning, and certain other consequences of external causes	249 403 (95.6)	8882 (3.4)	
U00–U99: Codes for special purposes	-	-	
V01–Y98: External causes of morbidity and mortality	2 (100.0)	0 (0.0)	
Z00–Z99: Factors influencing health status and contact with health services	2 (100.0)	1 (0.0)	

Note: The *P*-values were calculated by the chi-square test.

that for occupational injuries. Following the Industrial Accident-Related Act, recognition of occupational diseases includes objective evidence that workers are exposed to harmful physical and chemical risk factors that are likely to cause diseases in the workplace. However, it is difficult for workers to establish an association between occupational risk factors and disease.

In the case of digestive system diseases, digestive disease-causing substances, including asbestos, crystalline silica, and heavy metals, are limited to hazardous chemicals recognized by the International Agency for Research on Cancer. Simultaneously, it is necessary to determine how exposure to the working environment affects the disease status to be recognized as an occupation-related digestive system disease.[18–20] Moreover, it is difficult to perceive digestive problems caused due to work. This makes workers seldom register in the compensation insurance systems. We found a small number of cases related to digestive system disease (n = 83) in the current analysis. Given the increased incidence and mortality rates of digestive diseases, including cancers, unawareness of the causal relationship between work and disease may be a barrier for victims.[21]

In another case, respiratory system diseases (J00-J99) had the second highest risk of disapproval for occupational

TABLE 3 Feature selection for disapproval of claims for occupational diseases using logistic regression.

	Odds ratio (95% confidence interval) for disapproval of occupational diseases	
Sex		
Male	Reference	
Female	1.428	(1.339–1.522)
Age (years)		
<20	3.477	(1.798–6.724)
20–29	1.459	(1.276–1.667)
30–39	1.120	(1.027–1.222)
40–49	Reference	
50–59	0.969	(0.908–1.033)
60–69	1.047	(0.975–1.123)
≥70	1.593	(1.454–1.745)
Type of work		
Full time	Reference	
Part time	1.099	(1.036–1.166)
Migrant workers		
No	Reference	
Yes	1.234	(1.036–1.469)
Job duration		
<1 years	1.317	(1.245–1.394)
≥1 years	Reference	
International standard classification of occupations		
Managers	2.041	(1.854–2.247)
Professionals	2.256	(1.979–2.572)
Technicians and associate professionals	4.109	(3.434–4.918)
Clerical support workers	1.193	(1.075–1.325)
Service and sales workers	2.422	(2.015–2.913)
Skilled agricultural, forestry, and fishery worker	1.218	(0.945–1.572)
Craft and related trade workers	Reference	
Plant and machine operators and assemblers	1.461	(1.348–1.582)
Elementary occupations	1.518	(1.427–1.615)
Armed forces occupations, etc.	2.949	(1.921–4.527)
International classification of diseases		
A00–B99:Certain infectious and parasitic diseases	1.007	(0.771–1.314)
C00–D48: Neoplasms	1.217	(1.042–1.422)
F00–F99: Mental and behavioral disorders	0.817	(0.666–1.003)
G00–G99: Diseases of the nervous system	1.776	(1.535–2.055)
H60–H95: Diseases of the ear and mastoid process	2.466	(2.247–2.694)
I00–I99: Diseases of the circulatory system	3.752	(3.451–4.079)
J00–J99: Diseases of the respiratory system	6.200	(5.627–6.832)
K00–K93: Diseases of the digestive system	10.144	(5.819–17.683)
L00–L99: Diseases of the skin and subcutaneous tissue	2.294	(1.613–3.262)
M00–M99: Diseases of the musculoskeletal system and connective tissue	1.231	(1.145–1.323)
N00–N99: Diseases of the genitourinary system	4.731	(2.745–8.153)
R00–R99: Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	3.860	(3.156–4.721)
S00–T98: Injury, poisoning, and certain other consequences of external causes	Reference	

diseases, even though respiratory hazards are common exposure factors in most workplaces, including construction and mineral industries.[22] According to the World Health Organization, occupational respiratory diseases, including pneumoconiosis, asbestosis, and silicosis, have been considered a global issue, as estimated by the Global Burden of Disease project in 2000.[23] Regardless of this global concern, workers with respiratory diseases must be accompanied by supporting documents regarding their working history and accurate diagnoses to approve claims of work-related respiratory diseases. Respiratory diseases have been largely observed in miners and construction workers in Korea.[24] However, accurate and complete data regarding past exposures are limited in Korea.[25] Therefore, many workers may have difficulty providing specific evidence of their injuries.

Conversely, the causality of occupational injuries, such as falls and collisions in the workplace is relatively definite; thus, the claims for almost all injuries can be approved compared with occupational diseases.

TABLE 4 Feature selection for disapproval of claims for occupational injury using logistic regression.

	Odds ratio (95% confidence interval) for disapproval of occupational injury	
Sex		
Male	1.192	(1.131–1.256)
Female	Reference	
Age (years)		
<20	1.320	(1.083–1.610)
20–29	Reference	
30–39	1.046	(0.959–1.140)
40–49	1.212	(1.119–1.313)
50–59	1.299	(1.205–1.401)
60–69	1.431	(1.321–1.550)
≥70	1.679	(1.496–1.883)
Migrant workers		
No	1.589	(1.433–1.761)
Yes	Reference	
Job duration		
<1 years	Reference	
≥1 years	1.049	(1.005–1.095)
International standard classification of occupations		
Managers	2.015	(1.676–2.423)
Professionals	2.217	(1.823–2.697)
Technicians and associate professionals	3.231	(2.634–3.963)
Clerical support workers	1.370	(1.136–1.653)
Service and sales workers	1.880	(1.493–2.367)
Skilled agricultural, forestry, and fishery worker	Reference	
Craft and related trade workers	1.234	(1.031–1.478)
Plant and machine operators and assemblers	1.304	(1.082–1.572)
Elementary occupations	1.497	(1.255–1.787)
Armed forces occupations, etc.	2.573	(2.033–3.255)
International classification of diseases		
A00–B99: Certain infectious and parasitic diseases	8.098	(3.974–16.503)
G00–G99: Diseases of the nervous system	13.495	(9.950–18.303)
H00–H59: Diseases of the eye and adnexa	2.221	(1.613–3.058)
H60–H95: Diseases of the ear and mastoid process	8.780	(5.628–13.696)
I00–I99: Diseases of the circulatory system	6.592	(4.928–8.819)
J00–J99: Diseases of the respiratory system	2.033	(1.067–3.875)
K00–K93: Diseases of the digestive system	4.783	(3.332–6.865)
L00–L99: Diseases of the skin and subcutaneous tissue	3.651	(2.851–4.646)
M00–M99: Diseases of the musculoskeletal system and connective tissue	19.530	(18.029–21.156)
N00–N99: Diseases of the genitourinary system	4.363	(1.295–14.700)
R00–R99: Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	5.620	(4.194–7.531)
S00–T98: Injury, poisoning, and certain other consequences of external causes	Reference	

However, some injuries had a high OR for the disapproval of claims for occupational injuries. In particular, our results show that diseases of the musculoskeletal system and connective tissue, which are ICD-10 codes for M00-M99, have the highest risk for disapproval.

Musculoskeletal disorder (MSD) is a disorder of the cartilage, muscles, joints, tendons, and supporting structures such as intervertebral discs. Work-related MSDs occur while lifting heavy objects, performing repetitive forceful motions, or working with the neck in a chronically flexed position. To recognize occupational injury with MSD, it is important to satisfy certain conditions, such as diagnosis, occupational category, and detailed tenure of the work. For example, a plumber with a herniated intervertebral disk must prove over a tenure of a decade and diagnosis of lumbar and other intervertebral disc disorders with myelopathy or radiculopathy (M51.0 or M51.1). For the recognition of occupational

TABLE 5 Performance of the prediction models for disapproval of claims for occupational diseases and injury.

	Disapproval of occupational disease			Disapproval of occupational Injury		
	Accuracy (Train)	Accuracy (Validation)	ROC-AUC (Validation)	Accuracy (Train)	Accuracy (Validation)	ROC-AUC (Validation)
LR	0.594	0.590	0.603	0.570	0.572	0.599
XGBoost	0.673	0.668	0.724	0.605	0.604	0.660
NN	0.616	0.627	0.655	0.583	0.588	0.620

Abbreviations: LR, logistic regression; NN, neural network; ROC-AUC, receiver operating characteristic curves and area under the curve; XGBoost, extreme gradient boosting.

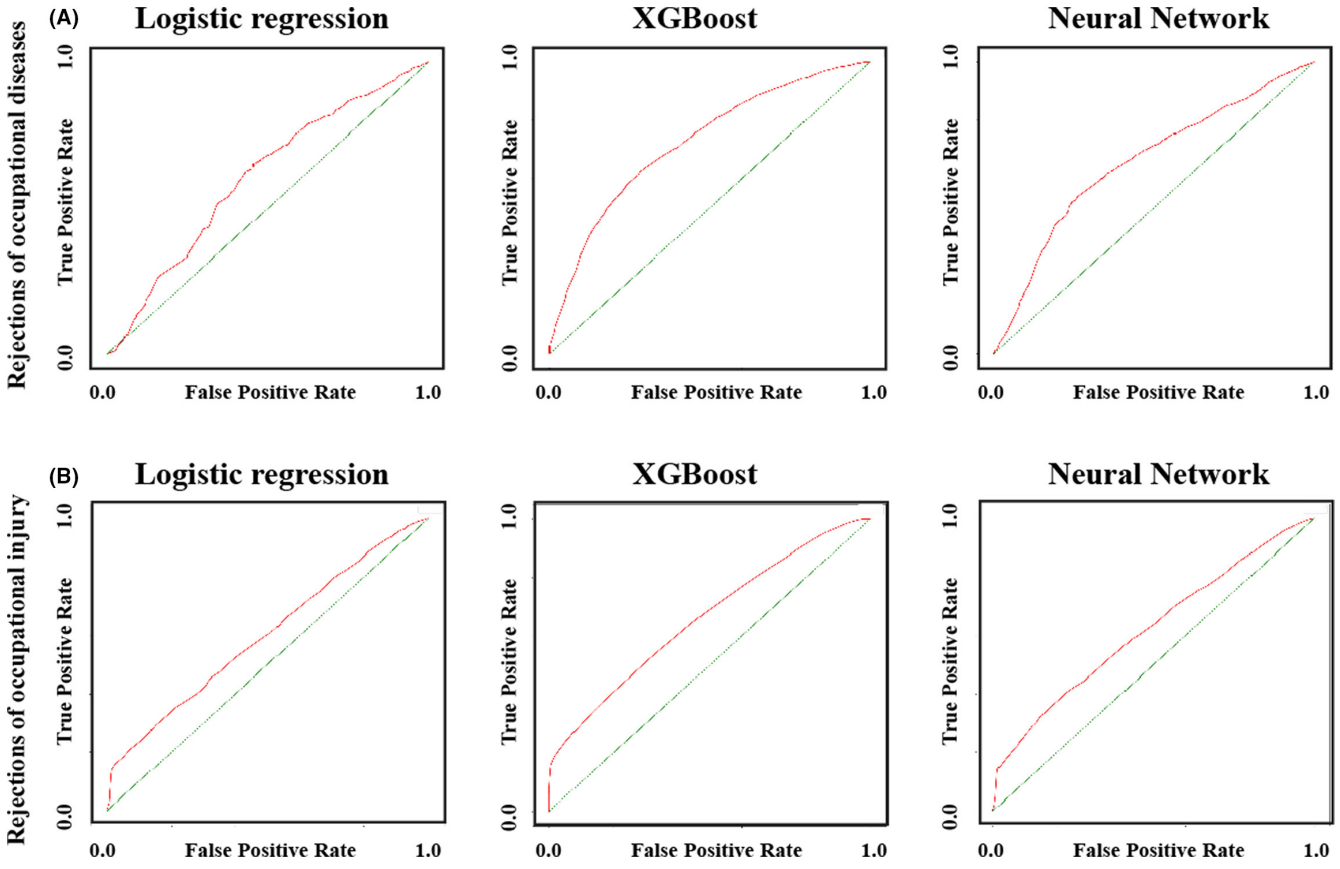

FIGURE 1 Comparison of receiver operating characteristic curves for the prediction models for disapprovals of claims of occupational diseases (A) and injury (B).

meniscal injury, a courier disk must prove more than a 5-year work duration and diagnosis of derangement of the meniscus and anterior or posterior horn of the medial meniscus due to old tears or injuries (M23.2, M23.21, and M23.22).

However, clinical features and objective physical examinations are often inconsistent. Moreover, MSD could develop not only because of occupational risk factors but also degenerative changes due to the aging process. Thus, workers may have difficulty providing accurate evidence of occupational causality with MSD.

This study has important strengths that must be considered. To date, our study is the first attempt to reflect on the features of disapproval of claims of occupational diseases or injuries according to specific disease classifications. New insights into this from scholarly attention would help applicants as well as judges understand and apply workers' compensation systems. Second, we rendered a high-level prediction model using this data set. Accurate predictive analysis can lead to shorter approval time and lower capital costs. This could be helpful for making precise decisions at the national level. Contrary to our expectations, even though machine learning was applied, it was difficult to find a critical improvement in performance compared with the existing model. This suggests that more detailed information on the claims of

occupational diseases and injuries is needed. Beyond data, such as social environment, can be reflected in decisions regarding occupational diseases and injuries. Therefore, it is necessary to make predictions by constructing more elaborate data for future research, and the influence of social factors should be considered.

Despite its strengths, our study has several limitations. Our results may have limitations owing to the nature of the data from the Korean workers' compensation. Much information about the working environment and workers, and these findings cannot be extrapolated to all occupational diseases and injuries in Korea. There is a possibility of an information error. For example, occupational injury data do not consist of injury-specific codes (S, T, or M). The misclassification error of the disease category was <1% (for injury, 261 006 cases of the S, T, and M codes from 263 617 total cases). Following KCOMWEL's confirmation, there are some errors in the current publicly available open data, and there is the possibility of some omissions or duplications in the process of data integration, especially if there are multiple disease codes. In addition, a case's causal relationship with occupation and not only the objective information of the applicant may affect the judgment error of the expert. There is a possibility of improvement of occupational disease and injury disapproval situations with long-term efforts of the KCOMWEL and occupational medicine researchers with the establishment of detailed data on occupational disease and injury claims and further analysis.

5 | CONCLUSIONS

In conclusion, this study aimed to demonstrate the status and prediction of the disapproval of claims of workers' diseases or injuries by Korean compensation insurance. The results of this investigation showed good performance for the occupational disease disapproval prediction model and moderate performance for the occupational injury disapproval prediction model. These findings suggest disapproval of diseases or injuries among workers from the current Korean compensation insurance system with a low level of evidence of work relatedness or a lack of research in occupational health. It is also expected to contribute to the efficiency of management of workers' diseases or injuries focused on the workers' compensation system.

AUTHOR CONTRIBUTIONS

W.L. and S.B.C.: conceptualization, data curation, formal analysis, investigation, software, and validation; S.L.: methodology, visualization, writing—original draft; W.L. and S.L.: project administration; W.L.: resources; S.B.C. and S.L.: supervision; all authors: writing—review and editing.

ACKNOWLEDGMENTS

This research was funded by the Seoul Institute of Technology (SIT) (2022-BA-019, A Study on Improvement of Public Health Care Operation System in Seoul).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Soo Beom Choi https://orcid.org/0000-0003-3892-0630
Seunghyun Lee https://orcid.org/0000-0001-7497-4930
Wanhyung Lee https://orcid.org/0000-0001-6408-7668

REFERENCES

1. Chin W-S, Guo YL, Liao S-C, et al. Quality of life at 6 years after occupational injury. *Qual Life Res*. 2018;27(3):609-618.
2. World Health Organization. *WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury, 2000-2016: Global Monitoring Report*. World Health Organization and the International Labour Organization; 2021. https://www.who.int/publications/i/item/9789240034945. Accessed February 1, 2023
3. Lippel K, Lötters F. *Public Insurance Systems: a Comparison of Cause-Based and Disability-Based Income Support Systems*. Handbook of Work Disability: Springer; 2013:183-202.
4. Mittag O, Kotkas T, Reese C, et al. Intervention policies and social security in case of reduced working capacity in The Netherlands, Finland and Germany: a comparative analysis. *Int J Public Health*. 2018;63(9):1081-1088.
5. Kang S-K, Kim EA. Occupational diseases in Korea. *J Korean Med Sci*. 2010;25(Suppl):S4-S12.
6. Lee K, Lee S, Min J, Kim I. Occupational cancer claims in Korea from 2010 to 2016. *Ann Occup Environ Med*. 2018;30(1):1-7.
7. O'Leary P, Boden LI, Seabury SA, Ozonoff A. Workplace injuries and the take-up of social security disability benefits. *Soc Secur Bull*. 2012;72(3):1-7.
8. Grant GM, O'Donnell ML, Spittal MJ, Creamer M, Studdert DM. Relationship between stressfulness of claiming for injury compensation and long-term recovery: a prospective cohort study. *JAMA Psychiat*. 2014;71(4):446-453.
9. Harris I, Mulford J, Solomon M, van Gelder JM, Young J. Association between compensation status and outcome after surgery: a meta-analysis. *JAMA*. 2005;293(13):1644-1652.
10. Davagdorj K, Pham VH, Theera-Umpon N, Ryu KH. XGBoost-based framework for smoking-induced noncommunicable disease prediction. *Int J Environ Res Public Health*. 2020;17(18):6513.

11. Li W, Yin Y, Quan X, Zhang H. Gene expression value prediction based on XGBoost algorithm. *Front Genet*. 2019;10:1077.
12. Pesantez-Narvaez J, Guillen M, Alcañiz M. Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks*. 2019;7(2):70.
13. Zou J, Han Y, So S-S. Overview of artificial neural networks. In: Livingstone DJ, ed. *Artificial Neural Networks: Methods and Applications*. Humana Press; 2009. doi:10.1007/978-1-60327-101-1_2 Accessed February 1, 2023.
14. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
15. Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: mining adverse drug reactions. *J Biomed Inform*. 2015;56:318-332.
16. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4(2):627-635.
17. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol*. 2022;75(1):25-36.
18. Lee W, Ahn Y-S, Lee S, Song BM, Hong S, Yoon J-H. Occupational exposure to crystalline silica and gastric cancer: a systematic review and meta-analysis. *Occup Environ Med*. 2016;73(11):794-801.
19. Steenland K, Barry V, Anttila A, et al. Cancer incidence among workers with blood lead measurements in two countries. *Occup Environ Med*. 2019;76(9):603-610.
20. Fang Y-J, Chuang H-Y, Pan C-H, et al. Increased risk of gastric cancer in Asbestos-exposed workers: a retrospective cohort study based on Taiwan cancer registry 1980–2015. *Int J Environ Res Public Health*. 2021;18(14):7521.
21. Ricciardiello L. Digestive diseases: big burden, low funding? Results of the new united European gastroenterology white book on digestive diseases. *United European Gastroenterol J*. 2022;10(7):627-628.
22. Cummings KJ, Johns DO, Mazurek J, Hearl FJ, Weissman DN. Respiratory health research at NIOSH. 2021. https://stacks.cdc.gov/view/cdc/112512. Accessed February 1, 2023.
23. Murray CJ, Lopez AD, Mathers CD, Stein C. *The Global Burden of Disease 2000 Project: Aims, Methods and Data Sources*. Vol 36. World Health Organization; 2001:1-57.
24. Baik J. Analysis of industrial accidents data with survival model. *Ind Promot Res*. 2020;5(1):1-11.
25. Kang D-M, Kim J-E, Kim J-Y, et al. Environmental asbestos exposure sources in Korea. *Int J Occup Environ Health*. 2016;22(4):307-314.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Choi SB, Lee S, Lee W. Status and prediction of disapproval of the Korean workers' compensation insurance for diseases and injuries. *J Occup Health*. 2023;65:e12392. doi:10.1002/1348-9585.12392