Heliyon 9 (2023) e20138

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Characteristics of fatal occupational injuries in migrant workers in South Korea: A machine learning study

Ju-Yeun Lee [a], Woojoo Lee [a], Sung-il Cho [a,b,*]

[a] The Department of Public Health, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea
[b] Institute of Health and Environment, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Occupational injuries
Occupational accidents
Migrant workers
Machine learning
SHapley additive exPlanations(SHAP)
Gender difference

ABSTRACT

Objective: Analysis of occupational injuries is essential for developing preventive strategies. However, few studies have evaluated severe occupational injuries in migrant workers from the perspective of gender. Therefore, using a new analytical method, this study was performed to identify gender-specific characteristics associated with fatal occupational injuries among migrant workers; the interactions between these factors, were also analyzed. In addition, we compared the utility of explainable artificial intelligence (XAI) using SHapley Additive exPlanations (SHAP) with logistic regression (LR) and discuss caveats regarding its use.
Materials and methods: We analyzed national statistics for occupational injuries among migrant workers (n = 67,576) in South Korea between January 1, 2007, and September 30, 2018. We applied an extreme gradient boosting model and developed SHAP and LR models for comparison.
Results: We found clear gender differences in fatal occupational injuries among migrant workers, with males in the same occupation having a higher risk of death than females. These gender differences suggest the need for gender-specific occupational injury prevention interventions for migrant workers to reduce the mortality rate. Occupation was a significant predictor of death among female migrant workers only, with care jobs having the highest fatality risk. The occupational fatality risk of female workers would not have been identified without the performance of detailed job-specific analyses stratified by gender. The major advantages of SHAP identified in the present study were the automatic identification and analysis of interactions, ability to determine the relative contributions of each feature, and high overall performance. The major caveat when using SHAP is that causality cannot be established.
Conclusion: Detailed job-specific analyses stratified by gender, and interventions considering the gender of migrant workers, are necessary to reduce occupational fatality rates. The XAI approach should be considered as a complementary analytical method for epidemiological studies, as it overcomes the limitations of traditional statistical analyses.

1. Introduction

Occupational injuries in migrant workers have received increasing attention, where the goal is to improve occupational safety and health. Occupational injuries have significant negative economic, physical, and emotional impacts on workers and their families.

* Corresponding author. The Department of Public Health, Graduate School of Public Health, Building 220, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea.
E-mail address: persontime@hotmail.com (S.-i. Cho).

https://doi.org/10.1016/j.heliyon.2023.e20138
Received 11 March 2023; Received in revised form 9 September 2023; Accepted 12 September 2023
Available online 14 September 2023

Furthermore, they may lead to disruptions and low morale among workers. The incidence of fatal occupational injuries has declined in many countries [1,2], except among migrant workers [1]. International migrant workers account for 4.7% of the total labor force worldwide [3]. International migrant workers are defined as "a person who is to be engaged, is engaged or has been engaged in a remunerated activity in a state of which he or she is not a national" under the United Nations Convention [4]. The number of migrant workers is steadily increasing, and they are highly important for certain regions and industries [5]. Accurate identification of individuals at high risk of occupational injuries, such as migrant workers, and determination of their risk of fatal accidents are essential for the prevention of occupational injuries.

Research on the health of migrant workers should consider interactions among factors [6–8]. One study [9] reported a U-shaped trend in the injury odds ratio according to the number of years of employment among migrant workers. The interaction between migration status and length of employment suggests unknown effects or undetected risk factors. Most previous quantitative studies analyzed migrant workers without stratification by gender [1,10–12]; gender was treated merely as a demographic variable. Moreover, while most previous studies on migrant workers have reported a higher risk of occupational injury among males compared to females [1,13,14], these studies did not specifically compare men and women within the same occupation, and they did not consider the intersectionality of gender differences in risk factors. The intersectionality approach looks for interactions among risk factors to determine how they influence health across groups and geographical regions [15].

New approaches are necessary for analyzing occupational injury risk factors among migrant workers, as traditional statistical methods have limitations in capturing complex interactions. Some studies on migrant workers have recognized the inadequate analysis of potentially interacting occupational risk factors using traditional methods [16,17]. For instance, in a study of migrant workers [18], limitations arose in the regression model due to multicollinearity between years of residing in the host country and language proficiency.

Machine learning (ML) is increasingly being employed for predicting and analyzing occupational injuries [19]. Many studies have demonstrated the superiority of ML-based methods over other statistical approaches in this context [19–23]. As of now, ML-based research on occupational injuries among non-migrant workers has primarily concentrated on predicting outcomes, assessing injury risk/severity [24–26], and extracting patterns [27,28]. However, ML-based analysis has low explanatory power [29], which has led to increasing interest in explainable artificial intelligence (XAI).

SHapley Additive exPlanations (SHAP) is a representative XAI method that provides a solid theoretical foundation based on Shapley values; it is commonly used as a basis for ML studies [30]. The Shapley value is a well-established measure of the contribution of a particular factor in analyses involving combinations of multiple factors [31]. The SHAP value is obtained by summing the Shapley values for a specific individual exposed to a set of these factors [30]. SHAP values can be used in both classification and regression problems, and the results can be interpreted at the global and local scales [32]. We used the SHAP method because its advantages were expected to improve our understanding of occupational injuries.

Empirical analysis of occupational injuries is essential because such injuries do not occur randomly [24,25,33–35]. Here, we analyzed nationwide occupational injury data for migrant workers in South Korea. Migrant workers have accounted for more than 4% of all workers in South Korea since 2007, and their numbers continue to increase. South Korea is a major destination for migrant workers from China and several other East Asian countries. Asia accounts for 65% of all occupational deaths worldwide [36]. Detailed analysis of fatal occupational injuries in Korea will improve our understanding of the factors related to such injuries.

This study used a new analytical method, SHAP, to identify risk factors for fatal occupational injuries among migrant workers and considered the interactions between these factors. This study had two major objectives: to assess gender differences in the contributions of various factors to occupational mortality risk; and to examine the interactions among those factors. In addition, we compare the utility of SHAP and traditional logistic regression (LR) analyses, and discuss a number of caveats.

2. Materials and Methods

2.1. Data source and acquisition

The data used in this cross-sectional study were obtained from the Korea Workers Compensation and Welfare Services (KCOMWEL), which deals with insurance claims by workers for occupational injuries and disease according to the "Industrial Accident Compensation Insurance Act" of South Korea [37]. The claims were filed by or on behalf of workers who died or required medical treatment for >4 days due to occupational accidents. In accordance with the "Labor Standards Act" of South Korea, workers are defined as individuals who perform paid mental and physical work for businesses or workplaces, regardless of nationality or legal status.

The study included incidents of migrant workers reported from January 1, 2007, to September 30, 2018. In 2007, the "Employment Permit System" (EPS) [38] emerged as a comprehensive management system for migrant workers, including those of Korean descent. In the present study, workers with migrant backgrounds, including EPS workers, were defined as migrant workers. Farmers, fishermen, and other specialized employment were partially included as they have alternative insurance. All other occupations are required to have current workers' compensation insurance, so all their compensation reports are included.

2.2. Study participants

In total, 71,884 migrant workers with occupational injuries or diseases were identified from 2007 to September 2018 among the KCOMWEL data, 67,576 of whom were included in the study; we excluded those with unclear nationality or age >100 years due to input errors in the raw data. We only analyzed occupational injuries because of the very low compensation claim (3%; 2265 cases) and approval (42%) rates for occupational diseases. According to the "Industrial Accident Compensation Insurance Act", occupational diseases are diseases caused by handling or being exposed to factors that can damage workers' health, such as physical factors, chemicals, dust, pathogens, and work that burdens the body. Therefore, infectious diseases caused by pathogens such as tuberculosis or typhus fever are classified as occupational diseases and excluded from the scope of this study. Since 2018, commuting accidents have been recognized as occupational injuries. Individuals with missing values were excluded (<3% of all data). The flowchart of the participant selection process is presented in Supplementary Material Fig. S1.

2.3. Data preprocessing

The study outcome was a binary variable (death or survival) and we aimed to identify risk factors for severe occupational injuries. Eight potential risk factors were identified, i.e., gender, age, nationality, industry, occupation, year of claim, injured body part, and injury type. The factors were all categorical, except year of claim. Detailed descriptions of these factors are presented in Supplementary Material Table S1.

2.4. Model evaluation and selection

We used LR, random forest (RF) [39], and eXtreme Gradient Boosting (XGB) [40] models, as they are the most common injury analysis models [19]. RF and XGB algorithms are often used because they are non-parametric [39–41], and can automatically identify interactions [42,43].

The area under receiver operating characteristic curve (AUROC) values [44] of the models were compared to identify the best model. When models had similar AUROC values, the model with higher sensitivity was selected because sensitive ML algorithms can predict injury accurately and be used to analyze interactions [22,45] among predictors of fatal occupational injuries.

LR models were developed for comparison with the other models. Using a stepwise algorithm, the LR model with the smallest Akaike information criterion was selected; 'year of claim' and 'industry' were excluded from the stepwise algorithm.

The XGB model was selected because of its high AUROC value (0.992), sensitivity (0.961), and specificity (0.967) according to an independent holdout test. The hyperparameters of the final model were as follows: eta = 0.3; gamma = 1; maximum depth = 6; number of estimators = 500; colsample_bytree = 1; nfold = 10; and evaluation metric = error and area under the curve. The results of the model performance and sensitivity tests are presented in Supplementary Material Table S2.

XGB is an ensemble learning algorithm, specifically one of the boosting algorithms, known for preventing overfitting and having high predictive power compared to individual decision trees and LR [40,43]. Other advantages of XGB include parallel computing, enabling cross-validation, and regularization. However, a disadvantage is that it has a built-in algorithm for extracting important features and measuring feature importance, but the degree of influence (positive/negative, size) of the feature is unknown and inconsistent. To address these disadvantages, SHAP was used as a compensation measure.

2.5. Statistical analysis using SHAP

TreeSHAP analysis was used to investigate the consistent associations and the degree of influence of demographic, occupational, and injury characteristics with fatal occupational injury based on the final XGB model. TreeSHAP is a variant of SHAP for tree-based ML models [46]. The shap.TreeExplainer function was used to interpret the predictions of the XGB classifier.

The Shapley value, rooted in cooperative game theory, signifies the relative contribution of a specific feature. It is computed using the Shapley formula [31], which equitably allocates the gains of a cooperative game among its participants. Several studies [47–49], have evaluated methods for determining the attributable risk of individual features for health outcomes among multiple interacting features. Shapley values are suitable for epidemiological investigations [47].

SHAP shares three important mathematical properties with Shapley values: local accuracy (completeness and additivity), missingness, and consistency (monotonicity) [30]. SHAP values represent the contributions of features to tree-based models [46]. The equations(Equation (1), (2)) for calculating SHAP and Shapley values are as follows (a more detailed explanation was provided previously [30]):

$$\text{SHAP value}: f(x) \approx g(z') = \varphi_0 + \sum_{i=1}^{M} \varphi_i z_i' (when\ x \approx z') \tag{1}$$

$$\text{Shapley value}: \varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \tag{2}$$

where g is the explanatory model, f is the original model, φ_0 is the model output with all features toggled to "off," φ_i is the Shapley value used for feature attribution (weighted average of all possible differences), z' refers to interpretable features, M is the number of features considered, N is the set of all features, S is a subset of features excluding the feature of interest, and f_x is the effect of a particular feature.

Python (version 3.8.10) and R software (version 3.6.3) were used for the data analyses. The "scikit-learn," [50] "xgboost," [40] "DALEX," [51] and "shap" [30] libraries were applied for model training and validation, and SHAP values were obtained using TreeSHAP. The "EIX" [52] and "ggplot2" R packages were used to analyze interactions and plot graphs.

3. Results

3.1. Demographic and occupational characteristics of the deceased and surviving migrant workers

We analyzed 67,576 cases of occupational injury of migrant workers reported in South Korea between January 1, 2007 and September 30, 2018. The chi-squared test revealed significant differences in characteristics between the deceased and surviving migrant workers who sustained occupational injuries. The most important characteristics of the deceased workers are summarized in Table 1.

Injuries and deaths were more common in male migrants (84.9%; 57,372 cases) than female migrants. Female migrants were older than male migrants (mean age = 45.6 and 40.9 years, respectively). Deceased migrants were older than surviving ones (mean age = 46.2 and 41.7 years, respectively, standard deviation = 12.1) regardless of gender.

The most commonly injured body parts were the hands and fingers (48.3%; 32,623 cases), followed by the lower limbs (11.4%; 7677 cases) and feet and toes (9.8%, 6596 cases). The most commonly injured body parts differed according to gender. In females, although the most commonly injured body parts were the hands and fingers, injuries to the upper limbs (12.6%; 1282 cases) and lower limbs (11.4%; 1161 cases) were more common than those to the feet and toes (9.4%; 963 cases). Fatal occupational injuries most

Table 1
Baseline characteristics of migrant workers sustaining occupational injuries by gender.

	All			Females			Males		
	N	*Fatal*	FROI	*N*	*Fatal*	FROI	*N*	*Fatal*	FROI
Total	67,576	*979*	1.5	*10,204*	*46*	0.5	*57,372*	*933*	1.6
Age (years)									
14–28	11,795	98	0.8	1063(10.4)	2	0.2	10,732 (18.7)	96	0.9
29–36	14,750	177	1.2	1351(13.2)	6	0.4	13,399 (23.4)	171	1.3
37–45	13,988	205	1.5	1980 (19.4)	12	0.6	12,008 (20.9)	193	1.6
46–53	13,330	233	1.7	3062 (30.0)	14	0.5	10,268 (17.9)	219	2.1
≥54	13,713	266	1.9	2748 (26.9)	12	0.4	10,965 (19.1)	254	2.3
Body part									
Head	2639	416	15.8	258 (2.5)	17	6.6	2381 (4.2)	399	16.8
Upper limb	5952	7	0.1	1282 (12.6)	0	0.0	4670 (8.1)	7	0.1
Hand/finger	32,623	22	0.1	4859 (47.6)	5	0.1	27,764 (48.4)	17	0.1
Lower limb	7677	13	0.2	1161 (11.4)	0	0.0	6516 (11.4)	13	0.2
Multiple body regions	841	99	11.8	134 (1.3)	4	3.0	707 (1.2)	95	13.4
Whole body	374	166	44.4	26 (0.3)	9	34.6	348 (0.6)	157	45.1
Other	17,470	256	1.5	2484 (24.3)	11	0.4	14,986 (26.1)	245	1.6
Type of injury									
Fracture	27,569	234	0.8	3700 (36.3)	12	0.3	23,869 (41.6)	222	0.9
Amputation	9409	25	0.3	1256 (12.3)	0	0.0	8153 (14.2)	25	0.3
Crush	6915	60	0.9	935 (9.2)	4	0.4	5980 (10.4)	56	0.9
Open wound	5239	22	0.4	950 (9.3)	1	0.1	4289 (7.5)	21	0.5
Burn	3341	72	2.2	1215 (11.9)	9	0.7	2126 (3.7)	63	3.0
Other	15,103	566	3.7	2148 (21.1)	20	0.9	12,955 (22.6)	546	4.2
Occupation									
Service	4567	16	0.4	3037 (29.8)	10	0.3	1530 (2.7)	6	0.4
Elementary	35,283	449	1.3	4688 (45.9)	15	0.3	30,595 (53.3)	434	1.4
Office	8993	119	1.3	1191 (11.7)	6	0.5	7802 (13.6)	113	1.4
Machine operation	7079	118	1.7	730 (7.2)	6	0.8	6349 (11.1)	112	1.8
Technical	11,654	277	2.4	558 (5.5)	9	1.6	11,096 (19.3)	268	2.4
Industry									
Construction	15,486	427	2.8	434 (4.3)	9	2.1	15,052 (26.3)	418	2.8
Service	10,111	91	0.9	5278 (51.7)	17	0.3	4833 (8.4)	74	1.5
Manufacturing	40,302	424	1.1	4341 (42.6)	20	0.5	35,961 (62.8)	404	1.1
Other	1521	34	2.2	148 (1.5)	0	0.0	1373 (2.4)	34	2.5
Nationality									
Chinese	9173	148	1.6	1801 (17.6)	12	0.7	7372 (12.8)	136	1.8
Korean-Chinese	33,326	526	1.6	6281 (61.6)	26	0.4	27,045 (47.1)	500	1.8
Vietnamese	4730	57	1.2	670 (6.6)	3	0.4	4060 (7.1)	54	1.3
Sri Lankan	2584	22	0.9	42 (0.4)	1	2.4	2542 (4.4)	21	0.8
Uzbekistani	2498	36	1.4	239 (2.3)	0	0.0	2259 (3.9)	36	1.6
Other	15,265	190	1.2	1171 (11.5)	4	0.3	14,094 (24.6)	186	1.3

Office workers encompass managers, professionals, and clerks. Technical workers comprise skilled agricultural, forestry, and fishery workers, as well as craft and related trade workers. Other industries encompass water and air transportation, agriculture, fishing, aquaculture, research and development, and broadcasting.FROI, fatality rate of occupational injury. FROI (%) = number of fatally injured workers / total number of injured workers × 100.

commonly involved the whole body (44.4%), followed by the head (15.8%) and multiple body regions (11.8%). The most common injury type was fracture (40.8%; 27,569 cases) followed by amputation (13.9%; 9409 cases) and crush (10.2%; 6915 cases). In female migrants, the third common injury type was burn (11.9%; 1215 cases), followed by crush (9.2%; 935 cases).

Elementary workers were the most commonly injured (52.2%; 35,283 cases), whereas technical workers had the highest fatality rate (2.4%; 277 out of 11,654 cases) regardless of gender. The industries with the highest numbers of injured workers were manufacturing (59.8%; 40,302 cases), construction (23.0%; 15,486 cases), and service (15.0%; 10,111 cases), although the fatality rate was highest in the construction industry (2.8%; 427 out of 15,486 cases). The injury rate for female migrant workers in the service industry (51.7%; 5278 cases) was higher than that of female migrant workers in the manufacturing industry (42.6%; 4341 cases).

Chinese (including Korean-Chinese) workers accounted for the largest proportion of all migrant workers (62.9%; 42,499 cases), followed by Vietnamese workers (7.0%; 4730 cases). The fatality rate was highest for Chinese workers (68.8%; 674 cases), followed by Vietnamese workers (5.8%; 57 cases). On average, there were 82 occupational fatalities per year, and the non-fatal claims rate has increased slightly since 2010. The fatality rate varied between 1.0% and 2.1% during the study period.

3.2. Global SHAP results by gender

Subgroup analysis confirmed clear gender differences in the characteristics of fatal and non-fatal injuries. Gender made an important contribution to the risk of fatal injuries (Fig. 1 (a)). In addition, gender was identified as an effect modifier on the directed acyclic graph. In descending order of importance, injured body part, occupation, type of injury, and industry were risk factors for fatal occupational injuries in female migrant workers. Caregiving/health/personal service occupations, including geriatric, child, and postpartum care, skincare/bathing services (hereafter "care jobs"), and pharmaceutical/medicinal chemical/botanical product manufacturing, were associated with an increased risk of fatal occupational injuries (Fig. 1 (c)). Occupation and industry did not have significant effects on the risk of fatal occupational injuries in male migrant workers (Fig. 1 (b)). Nationality was a minor risk factor, but only in male migrant workers (Fig. 1 (b)). Table 2 shows the relative importance of each factor according to gender and the gender differences in occupational injuries.

Summary plot of SHAP values for the final XGB model showed the relative effects of each factor on the likelihood of occupational injury or death (Fig. 1). The absolute mean SHAP values of the features were aligned vertically based on their overall impact. Then, horizontal lines were drawn from the points representing the SHAP values. Overlapping points on the y-axis with the same SHAP value appear thicker. Hand and finger injuries were associated with a lower risk of death and higher survival rate for all migrant workers (Fig. 1 (a)). In the bar in the bottom right of the figure, purple corresponds to "Yes" for the feature value (i.e., 1). Positive and negative

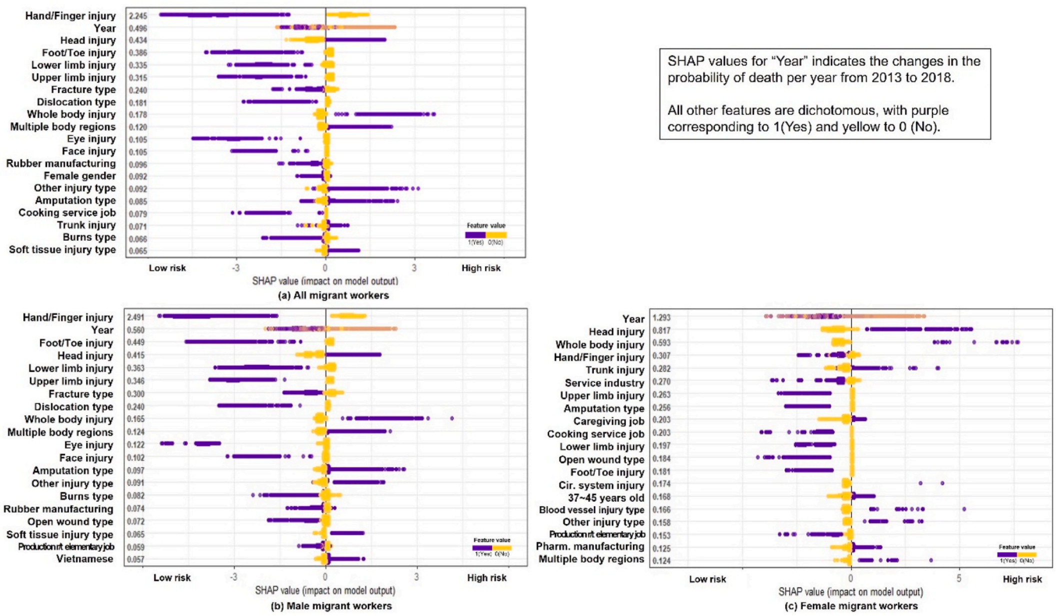

Fig. 1. SHAP summary plots for occupational injuries of all (a), male (b), and female migrant workers (c) in South Korea. The top 20 factors are shown (in descending order of mean SHAP value, where higher SHAP value indicates higher probability of death). Plots (b) and (c) indicate gender differences in risk factors for fatal occupational injuries.

Abbreviations: Dislocation type, dislocation sprain, vitreous prolapse; Rubber manufacturing, manufacture of rubber and plastic products; Other type, other and unspecified injuries; Cooking service job, cooking and food service occupations; Soft tissue injury type, soft tissue disorders; Production r/t elementary job, production-related elementary occupations; Service industry, creative, arts, and recreation-related service industries; Caregiving job, caregiving and personal service workers; Cir. system injury, circulatory system injury; Pharm. manufacturing, pharmaceutical, medicinal chemical, and botanical product manufacturing industries.

Table 2
Top 20 features associated with fatal and non-fatal occupational injuries among migrant workers in South Korea according to gender.

	Female migrant workers	Male migrant workers
↑	Injured body part: head, whole body, **trunk**, **circulatory system**, or multiple body regions Occupation: **caregiving/health/personal service** Age: **37–45 years** Injury type: **blood vessel injur**y, other/unspecified injury Industry: **manufacture of pharmaceuticals/medicinal chemicals/botanical products**	Injured body part: head, whole body, or multiple body regions Injury type: **amputation**, other/unspecified injury, and **soft tissue injury** Nationality: **Vietnamese**
↓	Injured body part: hand/finger, upper limb, lower limb, and foot/toe Industry: **creative, arts and recreation- services** Type of injury: **amputation** and open wound Occupation: **cooking and food service**, and production-related occupations	Injured body part: hand/finger, foot/toe, lower limb, upper limb, **eye, and face** Type of injury: **fracture, dislocation sprain or strain, burns**, and open wound Industry: **manufacture of rubber and plastic products** Occupation: production-related occupations

Note: Features are listed in descending order of absolute Shapley values. Features with gender-related differences are shown in bold.
The upward-pointing arrow signifies an increase in the risk of fatal occupational injuries, while a downward-pointing arrow indicates the opposite.

SHAP values indicate increased and decreased risk of death, respectively.

Injured body part was the most important predictor of fatal occupational injuries, but there were gender differences. For example, trunk injuries significantly increased the risk of fatal occupational injury in female, but not male migrant workers, while amputation increased the risk of death in male, but not female migrant workers. Local explanations further confirm that the gender difference in amputation can be explained by gender-specific interactions across occupations and industries (for more details, refer to Fig. S2 in the Supplementary Material). Age and nationality were not significant predictors of fatality. As shown in Fig. 1 (a), there were interactions among certain features (e.g., amputation and trunk injury). After reviewing the summary plot, interactions were confirmed using the "DALEX" or "EIX" package.

3.3. Comparison with LR analysis

We compared ML and LR models developed using the same data. LR models were constructed without interaction terms because it was not possible to incorporate all of the interaction terms into them, and it was not clear which interaction terms should be entered first.

The LR model has the advantage of being able to easily identify subgroups and derive odds ratios, but also has several limitations compared to SHAP. First, it is difficult to determine the importance of global features based on the results of LR models. Moreover, the LR results exhibited inconsistencies regarding feature importance. While the LR results showed that, relative to fracture type injuries, amputation and soft tissue injuries were greater risk factors for fatality (6.7- and 6.2-fold increased risk, respectively) than burns and dislocation injuries (0.4- and 0.2-fold increased risk, respectively), the LR results did not show that fracture and dislocation injuries were more important than other injury types for predicting non-fatal occupational injury risk. When not considering the injured body parts, the LR results exhibited varying feature importance across different models, with amputations and burns showing a 0.4- and 4.1-fold increased risk of death, respectively, relative to the fracture type. Second, the LR models had difficulty accounting for interaction terms when the number of features increased. This is an important limitation because multiple features contribute to the risk of occupational death in many cases. Moreover, developing LR models is time-consuming because they require >20 dummy variables and interactions with occupational groups must also be considered; such models are also prone to error. On the other hand, ML-based ensemble models rapidly and automatically detect interactions between features and can also estimate the contribution of each individual feature. Third, LR models often show inferior performance to ML models. The predictive performance of the XGB model (specificity and AUROC values of 0.967 and 0.992, respectively) was generally higher than the LR model for fatal occupational injuries (see Supplementary Material Table S2).

4. Discussion

We identified characteristics risk factors for fatal occupational injuries among migrant workers, and considered the interactions among the factors, using the SHAP framework. This study yielded two main findings. First, we identified clear gender differences in the contributions of various factors and their interactions to occupational injuries among migrant workers. In terms of occupational injuries, migrant workers have only been analyzed previously as a homogeneous group. Our results suggest that gender-specific interventions for occupational injury prevention are required to reduce the mortality rate among migrant workers. Second, we found that the occupation of female migrant workers is a significant predictor of fatal occupational injuries; care jobs were associated with a higher risk of fatal occupational injuries among females. To date, studies have only reported on the risk of musculoskeletal disorders in care workers [53,54]; the high risk of fatal occupational injuries in female migrant workers with care jobs is a new finding.

We identified clear gender differences in fatal occupational injuries among migrant workers, with males in the same occupation having a higher risk of death compared to females. This result is consistent with previous studies of non-migrant workers [55], although it has not been previously revealed in studies targeting migrant workers.

The gender differences seen in this study are assumed to be due to unmeasured factors, such as differences in tasks, injury severity, and treatment or compensation. For example, cleaners who work in the same workplace may be assigned different tasks based on their gender, so male workers who lift heavy objects can be expected to have a higher risk of fatal injuries than female workers who perform repetitive tasks [56]. It also suggests that there may be differences in the severity of bleeding or associated infections that occur in the context of injuries among male migrant workers. Many female migrant workers may also be ineligible for compensation in South Korea if they work in unlicensed workplaces, or in informal roles such as domestic work [57,58]. Female migrant workers who do not qualify for claims are highly unlikely to receive sufficient medical treatment due to high medical costs. These results suggest that further research is needed to understand the causes of gender differences in occupational fatalities within the same occupation.

The second major finding of this study was that occupation is a significant predictive factor for deaths among female migrant workers only; moreover, care jobs have a relatively high fatality risk. This may be surprising, as traditionally, men's work has been associated with a higher risk of death from occupational injuries [59]. The deceased migrant female caregivers suffered injuries such as hemorrhages and burns, and it is assumed that timely or sufficient response and treatment were not provided. Our findings were similar to those of a previous study of non-migrant workers showing that occupation was a more important determinant of occupational injury rates than gender [55].

Stratification based on job title is essential to ascertain the occupations carrying the highest risk. The relatively elevated fatality risk among female workers, as revealed in this study, would not have been evident without conducting a job title-specific stratified analysis. The categories mentioned in previous studies, such as mining and rural jobs [60], day laborers, or temporary positions in construction or fishing [61], fall short in identifying occupations with substantial injury risk. This leaves us with lingering questions about the labor circumstances of migrant workers.

Furthermore, classifying migrant workers according to industries in this study did not help predict the risk of fatality. This can be attributed to the increase in migrant employment in industries with relatively higher mortality rates (such as construction [14,62] and agriculture, forestry, and fisheries [61]), and also due to the limited number of industries in which migrant workers can work in Korea [35,61].

The male-centered analyses [11,12] were another limitation of the previous studies on occupational injuries among migrant workers. It has been repeatedly emphasized that migrant status and the female gender should be evaluated in the context of relatively low-paying jobs, institutionalized prejudice, sexism, and racism [4–6]. Moreover, job segregation by gender is common among migrant workers [63], emphasizing the need for analyses that fully consider the interaction between gender and occupation.

Age was not a significant predictive factor for fatal occupational injuries in this study. Previous studies reported that the average age of injured migrant workers was 4 years older than that of non-injured migrant workers [10], which aligns with our findings; 69% of injured migrant workers are 25–44 years old [11]. Unlike non-migrant workers, most migrant workers are young and their job options do not increase with age [10,42,64–66]. Consequently, the age of migrant workers is not a significant predictor of fatal occupational injuries, given these factors.

There were several limitations to LR models compared to XGB ones, especially in relation to interaction terms and prediction performance. Complex nonlinear interactions between features may be difficult to analyze using parametric methods but can be analyzed using ML models [43]. In addition, although there are some equivocal findings regarding model prediction performance [67,68], the performance of our LR model was lower than that of the ML model in most cases. Consistent with our results, previous studies reported several advantages of ML-based approaches compared to traditional statistical analyses [22,45].

In this study, we used XAI to predict occupational fatality risk for migrant workers and interpret the results. Several previous studies used XAI for disease detection [45] and prediction [48], and to interpret results [69]. The findings validated that the XAI method aids in the interpretation of contextual factors related to occupational injuries among migrant workers.

SHAP is model-agnostic, and the solid theoretical background of SHAP enhances the interpretability of the results, even though it is a "black box" model. SHAP can determine the relative contributions of different features to the outcome of interest and, unlike the LR model, consistently demonstrates feature importance. TreeSHAP is recommended when using a tree-based model, such as XGB. KernelSHAP, which was not used in this study, also has the advantage of being model-agnostic [30], but its processing speed is slow and there are issues related to feature dependence. In addition, as XGB is a stochastic algorithm and SHAP requires additional random sampling, randomness can be an issue for KernelSHAP. TreeSHAP is much faster than KernelSHAP [46], and has no feature dependence issues [32] and good consistency [46].

The major limitation of SHAP is the inability to make causal inferences [70]. We also recommend checking the results for a few cases before large-scale interpretation because TreeSHAP can produce non-zero Shapley values for non-contributing features. Despite its limitations, we can confirm that SHAP has significant advantages for exploratory studies based on domain knowledge.

This study had some limitations. First, the fatality rate and risks of non-fatal injuries may have been underestimated because underreporting of occupational injuries and diseases is a major problem in Korea. Based on Heinrich's Law, an estimated 76.9% of occupational injuries sustained by migrant workers in South Korea by the end of 2020 were not reported. Therefore, we used the most reliable fatal injury data [71–73] given the possibility of underreporting. Furthermore, it is important to note that farmers and fishermen may have been underrepresented in our study due to the specific nature of their insurance coverage.

The second limitation of this study was the lack of identification of the size of each migrant worker's company. It is well-established that small-size companies are at a higher risk of serious and fatal occupational injuries due to their limited resources compared to larger companies, making it more challenging for them to comply with work safety and health regulations prevalent in high-income countries [35]. A survey conducted by Statistics Korea in 2022 revealed that approximately 70% of migrant workers are employed in small companies with fewer than 30 employees. Consequently, the risk of occupational injuries among migrant workers in our study may have been underestimated.

Another limitation of this study was that it did not determine the legal status of migrant workers, or their reasons for migration, which are important factors in terms of health behaviors and status [74]. Access to healthcare and risk of fatal occupational injuries likely differ between undocumented and documented migrant workers. Migration can hinder accident reporting and needs to be evaluated in related qualitative studies.

Despite its limitations, this study constitutes a significant contribution to the literature on occupational health among migrant workers, as it leveraged the most reliable national data and introduced ML-based analysis for the first time.

5. Conclusions

We confirmed the advantages of ML and the utility of ML-based methods for analyzing occupational injuries among migrant workers, which requires the exploration of various contextual factors. The XAI approach can be considered a complementary analytical method to epidemiological analyses because it addresses the limitations of traditional statistical analyses. The gender differences detected in this study indicate that policies need to take gender into consideration; moreover, preventive interventions should be tailored to each group. Misunderstanding of the occupational health of migrant workers can lead to failure to prevent fatal injuries among female workers. Our results demonstrate the need to focus on the jobs done by female migrant workers, who have largely been overlooked by safety regulations and interventions aimed at preventing occupational injuries.

Ethics statement

This study was approved by the institutional review board of Seoul National University (approval no. E2204/003–005).

Author contribution statement

Ju-Yeun Lee: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Woojoo Lee: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Sung-il Cho: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Data availability statement

The authors do not have permission to share data.

Funding statement

This work was supported by the National Research Foundation of Korea (BK21 Center for Integrative Response to Health Disasters, Graduate School of Public Health, Seoul National University) (NO. 419,999 0514025).

The following is the Supplementary data to this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20138.

References

[1] J.Y. Lee, S.I. Cho, Prohibition on changing workplaces and fatal occupational injuries among Chinese migrant workers in South Korea, Int. J. Environ. Res. Publ. Health 16 (18) (2019) 3333, https://doi.org/10.3390/ijerph16183333.

[2] A. Mekkodathil, A. El-Menyar, H. Al-Thani, Occupational injuries in workers from different ethnicities, Int. J. Crit. Illn. Inj. Sci. 6 (1) (2016) 25, https://doi.org/10.4103/2229-5151.177365.

[3] N. Popova, M.H. Özel, ILO Global Estimates on International Migrant Workers: Results and Methodology, International Labour Office, 2018, 978-92-2-132672-4 (web pdf).

[4] United Nations, International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, OHCHR, 1990. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-protection-rights-all-migrant-workers.

[5] T. Brian, Occupational Fatalities Among International Migrant Workers: A Global Review of Data Sources, International Organization for Migration (IOM), Geneva, 2021. https://publications.iom.int/books/occupational-fatalities-among-international-migrant-workers.

[6] S. Mousaid, D. De Moortel, D. Malmusi, C. Vanroelen, New perspectives on occupational health and safety in immigrant populations: studying the intersection between immigrant background and gender, Ethn. Health 21 (3) (2016) 251–267, https://doi.org/10.1080/13557858.2015.1061103.

[7] B. Gazard, S. Frissa, L. Nellums, M. Hotopf, S.L. Hatch, Challenges in researching migration status, health and health service use: an intersectional analysis of a South London community, Ethn. Health 20 (6) (2014) 564–593, https://doi.org/10.1080/13557858.2014.961410.
[8] E.A. Viruell-Fuentes, P.Y. Miranda, S. Abdulrahim, More than culture: structural racism, intersectionality theory, and immigrant health, Soc. Sci. Med. 75 (12) (2012) 2099–2106, https://doi.org/10.1016/j.socscimed.2011.12.037.
[9] M.A. Salvatore, G. Baglio, L. Cacciani, A. Spagnolo, A. Rosano, Work-related injuries among immigrant workers in Italy, J. Immigr. Minority Health 15 (1) (2013) 182–187, https://doi.org/10.1007/s10903-012-9673-8.
[10] G. Pransky, D. Moshenberg, K. Benjamin, S. Portillo, J.L. Thackrey, C. Hill-Fotouhi, Occupational risks and injuries in non-agricultural immigrant Latino workers, Am. J. Ind. Med. 42 (2) (2002) 117–123, https://doi.org/10.1002/ajim.10092.
[11] P. Barss, K. Addley, M. Grivna, C. Stanculescu, F. Abu-Zidan, Occupational injury in the United Arab Emirates: epidemiology and prevention, Occup. Med. 59 (7) (2009) 493–498, https://doi.org/10.1093/occmed/kqp101.
[12] H. Al-Thani, A. El-Menyar, H. Abdelrahman, A. Zarour, R. Consunji, R. Peralta, R. Latifi, Workplace-related traumatic injuries: insights from a rapidly developing middle eastern country, J Environ Public Health 2014 (2014) (2014) 430832–430838, https://doi.org/10.1155/2014/430832.
[13] S.C. Moyce, M. Schenker, Migrant workers and their occupational health and safety, Annu. Rev. Publ. Health 39 (2018) 351–365, https://doi.org/10.1146/annurev-publhealth-040617-013714.
[14] A. Mekkodathil, A. El-Menyar, H. Al-Thani, A. Mekkodathil, A. El-Menyar, H. Al-Thani, Occupational injuries in workers from different ethnicities, Int. J. Crit. Illn. Inj. Sci 6 (1) (2016) 25–32, https://doi.org/10.4103/2229-5151.177365.
[15] A. Kapilashrami, O. Hankivsky, Intersectionality and why it matters to global health, Lancet 391 (10140) (2018) 2589–2591, https://doi.org/10.1016/S0140-6736(18)31431-4.
[16] J. Oh, E. Shin, (2003). Inequalities in nonfatal work injury: the significance of race, human capital, and occupations, Soc. Sci. Med. 57 (11) (1982) 2173–2182, https://doi.org/10.1016/S0277-9536(03)00073-X.
[17] V. Villanueva, A.M. Garcia, Individual and occupational factors related to fatal occupational injuries: a case-control study, Accid. Anal. Prev. 43 (1) (2011) 123–127, https://doi.org/10.1016/j.aap.2010.08.001.
[18] P.M. Orrenius, M. Zavodny, Do immigrants work in riskier jobs, Demography 46 (3) (2009) 535–551, https://doi.org/10.1353/dem.0.0064.
[19] S. Sarkar, J. Maiti, Machine learning in occupational accident analysis: a review using science mapping approach with citation network analysis, Saf. Sci. 131 (2020), 104900, https://doi.org/10.1016/j.ssci.2020.104900.
[20] S. Kulshrestha, D. Dligach, C. Joyce, R. Gonzalez, A.P. O'Rourke, J.M. Glazer, M. Afshar, Comparison and interpretability of machine learning models to predict severity of chest injury, JAMIA open 4 (1) (2021), https://doi.org/10.1093/jamiaopen/ooab015 ooab015.
[21] J.M. Karnuta, B.C. Luu, H.S. Haeberle, P.M. Saluan, S.J. Frangiamore, K.L. Stearns, P.N. Ramkumar, Machine learning outperforms regression analysis to predict next-season major league baseball player injuries: epidemiology and validation of 13,982 player-years from performance and injury profile trends, 2000-2017, Orthop. J. Sports Med. 8 (11) (2020), https://doi.org/10.1177/2325967120963046.
[22] J.L. Oliver, F. Ayala, M.B.D.S. Croix, R.S. Lloyd, G.D. Myer, P.J. Read, Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players, J. Sci. Med. Sport 23 (11) (2020) 1044–1048, https://doi.org/10.1016/j.jsams.2020.04.021.
[23] Q. Xu, H. Lei, X. Li, F. Li, H. Shi, G. Wang, B. Peng, Machine learning predicts cancer-associated venous thromboembolism using clinically available variables in gastric cancer patients, Heliyon (2023), e12681, https://doi.org/10.1016/j.heliyon.2022.e12681.
[24] F.D. Kakhki, S.A. Freeman, G.A. Mosher, Evaluating machine learning performance in predicting injury severity in agribusiness industries, Saf. Sci. 117 (2019) 257–262, https://doi.org/10.1016/j.ssci.2019.04.026.
[25] A.J.P. Tixier, M.R. Hallowell, B. Rajagopalan, D. Bowman, Application of machine learning to construction injury prediction, Autom. Constr. 69 (2016) 102–114, https://doi.org/10.1016/j.autcon.2016.05.016.
[26] A.S. Sánchez, P.R. Fernández, F.S. Lasheras, F.J. de Cos Juez, P.G. Nieto, Prediction of work-related accidents according to working conditions using support vector machines, Appl. Math. Comput. 218 (7) (2011) 3539–3552, https://doi.org/10.1016/j.amc.2011.08.100.
[27] M. Bevilacqua, F.E. Ciarapica, G. Giacchetta, Industrial and occupational ergonomics in the petrochemical process industry: a regression trees approach, Accid. Anal. Prev. 40 (4) (2008) 1468–1479, https://doi.org/10.1016/j.aap.2008.03.012.
[28] M. Amiri, A. Ardeshir, M. Fazel Zarandi, E. Soltanaghaei, Pattern extraction for high-risk accidents in the construction industry: a data-mining approach, Int. J. Inj. Control Saf. Promot. 23 (3) (2016) 264–276, https://doi.org/10.1080/17457300.2015.1032979.
[29] J. Burrell, How the machine 'thinks': understanding opacity in machine learning algorithms, Big Data Soc 3 (1) (2016), 2053951715622512, https://doi.org/10.1177/2053951715622512.
[30] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017), https://doi.org/10.48550/arxiv.1705.07874.
[31] L. Shapley, 7. A value for n-person games. Contributions to the theory of games II (1953) 307-317, in: Classics in Game Theory, Princeton University Press, 2020, pp. 69–79, https://doi.org/10.1515/9781400829156-012.
[32] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006, 0-08-051363-8.
[33] M.B. Schenker, A global perspective of migration and occupational health, Am. J. Ind. Med. 53 (4) (2010) 329–337, https://doi.org/10.1002/ajim.20834.
[34] M. Schenker, Work-related injuries among immigrants: a growing global health disparity, Occup. Environ. Med. 65 (11) (2008) 717–718, https://doi.org/10.1136/oem.2008.040907.
[35] G. Campo, L. Cegolon, D. De Merich, U. Fedeli, M. Pellicci, W.C. Heymann, G. Mastrangelo, The Italian national surveillance system for occupational injuries: conceptual framework and fatal outcomes, 2002–2016, IJERPH 17 (20) (2020) 7631, https://doi.org/10.3390/ijerph17207631.
[36] P. Hämäläinen, J. Takala, T.B. Kiat, Global Estimates of Occupational Accidents and Work-Related Illnesses 2017, Workplace Safety and Health Institute. Finland. Finland, 2017. https://www.icohweb.org/site/images/news/pdf/Report%20Global%20Estimates%20of%20Occupational%20Accidents%20and%20Work-related%20Illnesses%202017%20rev1.pdf.
[37] Korean Worker's Compensation & Welfare Service. (n.d.). www.comwel.or.kr. Retrieved March 6, 2023, from https://www.comwel.or.kr/eng/index.jsp.
[38] Employment Permit System. (n.d.). www.eps.go.kr. Retrieved October 19, 2022, from https://www.eps.go.kr/eo/langMain.eo?langCD=ph.
[39] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/a:1010933404324.
[40] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.
[41] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 56–67, https://doi.org/10.1038/s42256-019-0138-9.
[42] M. Carangan, K.Y. Tham, E. Seow, Work-related injury sustained by foreign workers in Singapore, Ann. Acad. Med. Singapore 33 (2) (2004) 209–213.
[43] W. Qiu, H. Chen, A.B. Dincer, S. Lundberg, M. Kaeberlein, S.I. Lee, Interpretable machine learning prediction of all-cause mortality, Commun. Med. 2 (1) (2022) 125, https://doi.org/10.1038/s43856-022-00180-x.
[44] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (7) (1997) 1145–1159, https://doi.org/10.1016/S0031-3203(96)00142-2.
[45] R. Li, A. Shinde, A. Liu, S. Glaser, Y. Lyou, B. Yuh, A. Amini, Machine learning–based interpretation and visualization of nonlinear interactions in prostate cancer survival, JCO Clin. Cancer Inform. 4 (2020) 637–646, https://doi.org/10.1200/CCI.20.00002.
[46] S.M. Lundberg, G.G. Erion, S.I. Lee, Consistent Individualized Feature Attribution for Tree Ensembles, 2018, https://doi.org/10.48550/arxiv.1802.03888 arXiv preprint arXiv:1802.03888.
[47] C. Rabe, O. Gefeller, The attributable risk in a multifactorial situation, Methods Inf. Med. 45 (4) (2006) 404–408, https://doi.org/10.1055/s-0038-1634095.
[48] L. Lama, O. Wilhelmsson, E. Norlander, L. Gustafsson, A. Lager, P. Tynelius, C.G. Östenson, Machine learning for prediction of diabetes risk in middle-aged Swedish people, Heliyon 7 (7) (2021), e07419, https://doi.org/10.1016/j.heliyon.2021.e07419.
[49] Louis A. Cox Jr., A new measure of attributable risk for public health applications, Manag. Sci. 31 (7) (1985) 800–813, https://doi.org/10.1287/mnsc.31.7.800.

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830, https://doi.org/10.5555/1953048.2078195.

[51] H. Baniecki, W. Kretowicz, P. Piatyszek, J. Wisniewski, P. Biecek, Dalex: Responsible Machine Learning with Interactive Explainability and Fairness in python, 2020 arXiv preprint arXiv:2012.14406.

[52] Szymon Maksymiuk, Package 'EIX', 2021. https://github.com/ModelOriented/EIX.

[53] W.S. Marras, K.G. Davis, B.C. Kirking, P.K. Bertsche, A comprehensive analysis of low-back disorder risk and spinal loading during the transferring and repositioning of patients using different techniques, Ergonomics 42 (7) (1999) 904–926, https://doi.org/10.1080/001401399185207.

[54] A.R. Darragh, C.M. Sommerich, S.A. Lavender, K.J. Tanner, K. Vogel, M. Campo, Musculoskeletal discomfort, physical demand, and caregiving activities in informal caregivers, J. Appl. Gerontol. 34 (6) (2015) 734–760, https://doi.org/10.1177/0733464813496464.

[55] Y.H. Lin, C.Y. Chen, J.L. Luo, Gender and age distribution of occupational fatalities in Taiwan, Accid. Anal. Prev. 40 (4) (2008) 1604–1610, https://doi.org/10.1016/j.aap.2008.04.008.

[56] K. Messing, C. Chatigny, J. Courville, 'Light'and 'heavy'work in the housekeeping service of a hospital, Appl. Ergon. 29 (6) (1998) 451–459, https://doi.org/10.1016/S0003-6870(98)00013-1.

[57] J. Holliday, J. Hennebry, S. Gammage, Achieving the sustainable development goals: surfacing the role for a gender analytic of migration, J. Ethnic Migrat. Stud. 45 (14) (2019) 2551–2565, https://doi.org/10.1080/1369183X.2018.1456720.

[58] D. Paiewonsky, The Feminization of International Labour Migration, INSTRAW, Santo Domingo, 2009. https://trainingcentre.unwomen.org/instraw-library/2009-R-MIG-GLO-FEM-EN.pdf (Working paper 1).

[59] World Health Organization, Gender Equality, Work and Health: a Review of the Evidence, WHO, 2006, 9241593539.

[60] C.F. Corvalan, T.R. Driscoll, J.E. Harrison, Role of migrant factors in work-related fatalities in Australia, Scand. J. Work. Environ. Health (1994) 364–370, https://doi.org/10.5271/sjweh.1385.

[61] L.S. Azaroff, C. Levenstein, D.H. Wegman, The occupational health of Southeast Asians in Lowell: a descriptive study, Int. J. Occup. Environ. Health 10 (1) (2004) 47–54.

[62] International Organization for Migration, World Health Organization, United Nations. Office of the High Commissioner for Human Rights, International Migration, Health and Human Rights, IOM, 2013, p. 38. https://publications.iom.int/books/international-migration-health-and-human-rights.

[63] T.A. McAuliffeM, World Migration Report 2022, International Organization for Migration (IOM), Geneva, 2021.

[64] E.Q. Ahonen, F.G. Benavides, Risk of fatal and non-fatal occupational injury in foreign workers in Spain, J. Epidemiol. Community Health 60 (5) (2006) 424–426, https://doi.org/10.1136/jech.2005.044099.

[65] X. Zhang, S. Yu, K. Wheeler, K. Kelleher, L. Stallones, H. Xiang, Work-related non-fatal injuries among foreign-born and US-born workers: findings from the U.S. National Health Interview Survey, 1997-2005, Am. J. Ind. Med. 52 (1) (2009) 25–36, https://doi.org/10.1002/ajim.20642.

[66] Y. Yi, Y. Liao, L. Zheng, M. Li, J. Gu, C. Hao, Y. Hao, Health selectivity and rural-urban migration in China: a nationwide multiple cross-sectional study in 2012, 2014, 2016, Int. J. Environ. Res. Publ. Health 16 (9) (2019) 1596, https://doi.org/10.3390/ijerph16091596.

[67] E. Christodoulou, J. Ma, G. Collins, E. Steyerberg, J. Verbakel, B. Van Calster, A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models, J. Clin. Epidemiol. 110 (2019) 12–22, https://doi.org/10.1016/j.jclinepi.2019.02.004.

[68] D.J. Hand, Classifier technology and the illusion of progress, Statist. Sci. 21 (1) (2006) 1–14, https://doi.org/10.1214/088342306000000060.

[69] S. Han, An analysis of Koreans' attitudes towards migrants by application of algorithmic approaches, Heliyon 8 (8) (2022), e10087, https://doi.org/10.1016/j.heliyon.2022.e10087.

[70] Y. Chou, C. Moreira, P. Bruza, C. Ouyang, J. Jorge, Counterfactuals and causability in explainable artificial intelligence: theory, algorithms, and applications, Inf. Fusion 81 (2022) 59–83, https://doi.org/10.1016/j.inffus.2021.11.003.

[71] N. Stout, C. Bell, Effectiveness of source documents for identifying fatal occupational injuries: a synthesis of studies, Am. J. Public Health 81 (6) (1991) 725–728, https://doi.org/10.2105/AJPH.81.6.725.

[72] M. Rossignol, M. Pineault, Fatal occupational injury rates: quebec, 1981 through 1988, Am. J. Public Health 83 (11) (1993) 1563–1566, https://doi.org/10.2105/AJPH.83.11.1563.

[73] M. Kelsh, J. Sahl, Sex differences in work-related injury rates among electric utility workers, Am. J. Epidemiol. 143 (10) (1996) 1050–1058, https://doi.org/10.1093/oxfordjournals.aje.a008669.

[74] H. Shen, S.B. Sorenson, Violence and Injury Among Immigrants. Handbook of Immigrant Health, 1998, pp. 545–565. The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: http://www.textcheck.com/certificate/wSDA1f.