Safety and Health at Work 15 (2024) 200–207

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

Severity Analysis for Occupational Heat-related Injury Using the Multinomial Logit Model

Peiyi Lyu, Siyuan Song*

Safety Automation and Visualization Environment (SAVE) Laboratory, Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, USA

ARTICLE INFO

Article history:
Received 20 November 2023
Received in revised form
27 February 2024
Accepted 31 March 2024
Available online 12 April 2024

Keywords:
Heat-related injuries
Impact factors
Multinomial logit model
Worker safety

ABSTRACT

Background: Workers are often exposed to hazardous heat due to their work environment, leading to various injuries. As a result of climate change, heat-related injuries (HRIs) are becoming more problematic. This study aims to identify critical contributing factors to the severity of occupational HRIs.
Methods: This study analyzed historical injury reports from the Occupational Safety and Health Administration (OSHA). Contributing factors to the severity of HRIs were identified using text mining and model-free machine learning methods. The Multinomial Logit Model (MNL) was applied to explore the relationship between impact factors and the severity of HRIs.
Results: The results indicated a higher risk of fatal HRIs among middle-aged, older, and male workers, particularly in the construction, service, manufacturing, and agriculture industries. In addition, a higher heat index, collapses, heart attacks, and fall accidents increased the severity of HRIs, while symptoms such as dehydration, dizziness, cramps, faintness, and vomiting reduced the likelihood of fatal HRIs.
Conclusions: The severity of HRIs was significantly influenced by factors like workers' age, gender, industry type, heat index , symptoms, and secondary injuries. The findings underscore the need for tailored preventive strategies and training across different worker groups to mitigate HRIs risks.

1. Introduction

Occupational heat-related injuries (HRIs) are a crucial societal concern since they adversely affect the health and productivity of workers. Outdoor and indoor workers in various industries, such as construction, agriculture, manufacturing, public safety service, and transportation, are at risk of dangerous heat exposure and are more likely to suffer from HRIs [1]. According to the U.S. Bureau of Labor Statistics, heat stress caused 943 fatalities among U.S. employees between 1992 and 2021 [2]. Also, HRIs have shown an increasing trend over the last three decades, with a significant increase in recent years [3]. In addition, HRIs have adverse effects on productivity and socioeconomic factors by lowering the size of the labor force, reducing work hours, and raising healthcare costs [4]. Furthermore, the morbidity and mortality of HRIs are expected to rise as average daily temperatures and extreme heat events rise due to climate change [5].

Workers experience heat stress due to frequent exposure to hot environments or heavy job-related activities, which can result in various injuries or fatalities without appropriate actions. Occupational Safety and Health Administration (OSHA) categorizes the severity of occupational injuries into three levels: non-hospitalized, hospitalized, and fatality. Attempting to reduce the occurrence and severity of HRIs, studies have focused on identifying contributing factors and understanding their complex relationships. General risk factors for HRIs have been identified and categorized as demographic, environmental, behavioral, and physical factors. Demographic attributes of worker age, gender, and occupation correlate closely with HRIs [6,7]. Environmental factors such as hot and humid conditions can increase the occurrence of HRIs [8].

Peiyi Lyu: https://orcid.org/0000-0003-2560-8664; Siyuan Song: https://orcid.org/0000-0002-6444-3914
* Corresponding author. Safety Automation and Visualization Environment (SAVE) Laboratory, Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA
E-mail addresses: plyu@crimson.ua.edu (P. Lyu), siyuan.song@ua.edu (S. Song).

2093-7911/$ — see front matter
https://doi.org/10.1016/j.shaw.2024.03.005

Behavioral factors are also critical contributing factors, where wearing PPE, insufficient rest, inadequate hydration, and poor nutrition can promote the probability of HRIs [9]. Other contributing factors include physical factors, medical conditions, heat acclimatization, and so on [10]. Furthermore, researchers believe that identifying key risk factors and taking appropriate improvement actions can be effective countermeasures for safety improvements in HRIs. For instance, being aware of the level of physical exhaustion that workers experience in the morning can help prevent HRIs [11]. Workers who obtain adequate rest are more productive and have a lower risk of experiencing HRIs [12]. However, the influence of more process impact factors on the severity of HRIs has not been further investigated, such as heat illness symptoms and workers' actions. Therefore, further research is needed to explore the correlates of the severity of occupational HRIs.

The objective of this study is to analyze the relationships between the severity of HRIs and associated impact factors. This study considers impact factors on the severity of injury, including worker, industries, environmental, symptomatic, and behavioral factors. The methods of text mining, model-free machine learning, and Multinomial Logit Model (MNL) are integrated to explore more specific and essential HRIs impact factors from OSHA historical accident report data. The findings of this study can promote administrative control methods for minimizing and mitigating heat hazards in the workplace. The results can also provide additional information to improve heat stress training materials.

2. Materials and methods

2.1. Research framework

This study combined OSHA accident data with weather information to provide comprehensive information and a detailed description of each accident. Initial HRIs data were extracted based on keywords. Subsequently, data cleaning, text mining, feature engineering, and MNL were used to uncover more impact factors. Fig. 1 summarizes the overall research framework. Fundamental accident factors were collated from the combined dataset of OSHA enforcement data, including demographic features, industries, and environmental attributes. In order to discover high-frequency keywords from complex descriptive data, text mining was first performed on the descriptive data. Furthermore, based on these high-frequency keywords, multiple machine learning methods were used simultaneously to explore the influence factors with high correlation with the severity of HRIs. Finally, the MNL was utilized to quantify the effect of each relevant risk factor on HRIs severity.

2.2. Data acquisition and cleaning

The U.S. Department of Labor provides a range of OSHA enforcement datasets, including accident abstract, injury data, and inspection details [13]. Injury data, serving as the foundational dataset for this study, contains summary number, inspection number, age, gender, and the source of injury. To get more comprehensive information such as date, time, state, city, industry code, and description report of injuries, this study merged accident abstract and inspection data into injury data based on the summary number and inspection number. Subsequently, depending on the source of injury and descriptive keywords, a total of 1,698 HRIs records were identified between 1984 and 2022. After that, the historical timestamped weather data was obtained utilizing weather Application Programming Interface (API) calls based on injuries' time and location [14]. The weather data included variables such as air temperature, humidity, and wind speed.

The final data include injury degrees, demographic variables, industry types, environmental variables, and a complete textual accident description. These descriptions typically detail the events' process and their causative factors. The severity of HRIs has three levels: non-hospitalized, hospitalized, and fatal. Demographic variables cover the age and gender of workers and environmental variables include date, time, and weather conditions. Specifically, for assessing weather factors, the study utilized heat index (HI) to measure ambient heat in workplaces. To improve estimation efficiency, avoid mistakes in interpretation, and preserve the population size of available HRIs, error checking and data cleaning were applied to improve the data.

2.3. Variables extraction and selection

Textual reports involve ambiguous and vague natural language descriptions. Text mining and model-free machine learning methods were applied to process text and extract critical factors from reports. Before beginning to mine text data, it was required to correct or remove spelling errors and superfluous characters. In addition, the number of potential keywords was reduced by merging words of varying tenses with similar meanings. In this analysis, features and essential variables of HRIs were extracted from investigation summaries using text mining.

It is impractical to include all available variables in modeling. Therefore, various model-free machine learning methods were utilized to find more essential variables after text mining, including Classification and Regression Trees, Bagging, Random Forest, and Bayesian Additive Regression Trees. Model-free solutions generate nonparametric representations utilizing machine learning algorithms or ensembles of many base learners without first simplifying the problem [15]. Subsequently, possible variables with the top 200 frequencies in text mining results were examined by four selected model-free machine learning methods. According to the results, these four approaches maintained the most important variables. In addition, to prevent errors brought on by one model, variables with high frequencies that occur in all machine learning models were chosen here as their focus.

2.4. Multinomial logit model

The severity of occupational HRIs in this paper consists of three outcome categories and is assumed to be the dependent variable. Independent variables, assumed to be not reliant on other variables, encompass all impact factors that impact the severity of HRIs. The MNL explains the relationship between the potential outcomes of the dependent variable and each independent variable. The MNL works by choosing one outcome category as the reference category for other categories. This study decided on non-hospitalized as the reference group because it was the mildest outcome of the severity of HRIs.

In the context of HRIs severity, assume a worker n experiencing HRIs severity level of j. The severity propensity function for the outcome is:

$$Y_{nj} = \beta X_{nj} + \varepsilon_{nj} \tag{1}$$

Where Y_{nj} is a function of covariates that determines the severity, β is a vector of estimable coefficients for HRIs accident severity level of j, X_{nj} is a vector of coefficients to be estimated for severity level j, and ε_{nj} is a random error term that accounts for unobserved factors influencing HRIs' severity level. The errors are assumed to be independently and identically distributed with identical type 1 extreme value distribution.

Based on the above specification, let P_i as the probability of worker n experiencing an HRIs severity level of j, and then the MNL probability is expressed as:

$$P_i = e^{\beta X_{nj}} \Big/ 1 + \sum_{n=1}^{j} e^{\beta X_{nj}} \quad (2)$$

Marginal effects provide the change in response values associated with one unit change in explanatory variables. For ordinary least squares models, parameters are marginal effects. However, for the MNL, estimated coefficients are the log of odds ratio [16]. In order to interpret modeling results and obtain the effects of variables, marginal effects for the MNL model can be calculated as follows:

$$\frac{\partial p_i}{\partial X_j} = e^{\beta X_{nj}\beta'} \Big/ \left(1 + \sum_{n=1}^{N-1} e^{\beta X_{nj}}\right)^2 \quad (3)$$

In the MNL model, the signs of marginal effects are not always consistent with the sign of coefficients (β), while in some logit regressions (e.g., binary logit), they are consistent. This inconsistency is because the marginal effect depends on the values and levels of other variables. As the values of other variables and variables in the equation change, the signs of marginal effect can also change. The marginal effects show the probability of change in one outcome compared with the base level of HRIs [17].

3. Results

3.1. Descriptive analysis

Table 1 shows a summary of HRIs characteristics after data imputation. The import data were error-checked, correcting problems such as misclassifying accident classes or imprecise occupations in the original data. Workers' age, time of year, time of day, and heat index were classified to examine whether a non-linear relationship exists between them and HRIs' degree. Further, the classification of factors helped to highlight sample groups of interest, such as firefighters. Among the 1698 HRIs, the degree distribution showed that only 18.37% of HRIs resulted in non-hospitalized, and 38.75% caused workers to suffer from hospitalized injuries. Of sampled accidents, 42.87% resulted in worker fatalities.

Basic information on the data includes demographic characteristics, industries, and environmental factors. Regarding the age of workers, the Bureau of Labor Statistics categorized ages into three groups: "youth" (age 16 to 24), "middle-aged" (age 25 to 54), and "older" (age 55 and older) [18]. Most workers with HRIs were middle-aged, and the highest fatality rate was among older workers. Male workers sustained injuries more than female workers, and male workers had significantly higher fatality rates than female workers. In addition, this study extracted several hazardous industries. Industries with a higher frequency of HRIs reports included agriculture, construction, manufacturing, services,

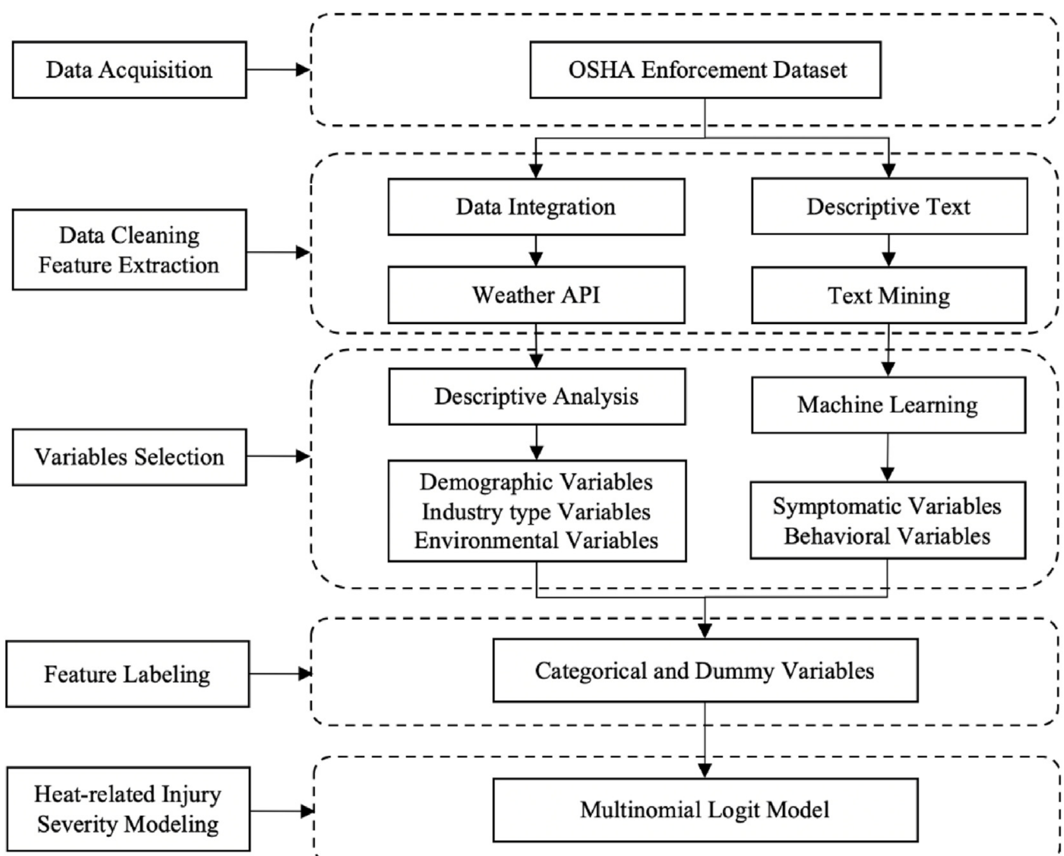

Fig. 1. Research framework.

Table 1
Descriptive statistics for HRIs

Variable (Total $n = 1{,}698$)			Frequency	Percent	Non-hospitalized	Hospitalized	Fatal
Degree		Non-hospitalized	312	18.37%	—	—	—
		Hospitalized	658	38.75%	—	—	—
		Fatal	728	42.87%	—	—	—
Demographics	Worker age	Youth	225	13.25%	54 24.00%	96 42.67%	75 33.33%
		Middle-aged	1176	69.26%	221 18.79%	450 38.27%	505 42.94%
		Older	271	15.96%	32 11.81%	106 39.11%	133 49.08%
		Unknown	26	1.53%	—	—	—
	Worker gender	Male	1533	90.28%	240 15.66%	595 38.81%	698 45.53%
		Female	141	8.30%	67 47.52%	58 41.13%	16 11.34%
		Unknown	24	1.41%	—	—	—
Industry		Construction	434	25.56%	48 11.06%	132 30.41%	254 58.53%
		Agriculture	328	19.32%	76 23.17%	128 39.02%	124 37.81
		Manufacturing	189	11.13%	18 9.52%	91 48.15%	80 42.33%
		Service	189	11.13%	30 15.87%	62 32.81%	97 51.32%
		Fire protection	186	10.95%	72 38.71%	96 51.61%	18 9.68%
		Transportation	107	6.30%	12 11.21%	61 57.01%	34 31.78%
		Others	265	15.61%	56 21.13%	88 33.21%	121 45.66%
Environment factors	Time of year	Spring (Mar, Apr, May)	204	12.01%	50 24.51%	76 37.25%	78 38.24%
		Summer (Jun, Jul, Aug)	1290	75.97%	228 17.67%	487 37.75%	575 44.57%
		Fall (Sep, Oct, Nov)	188	11.07%	32 17.02%	86 45.74%	70 37.23%
		Winter (Dec, Jan, Feb)	16	0.94%	2 12.50%	9 56.25%	5 31.25%
	Time of day	Morning	292	17.20%	65 22.26%	121 40.88%	106 36.30%
		Afternoon	1034	60.90%	168 16.25%	406 39.26%	460 44.49%
		Night	278	16.37%	62 22.30%	106 38.13%	110 39.57%
		Unknown	94	5.53%	—	—	—
	Heat index	<80	296	17.43%	73 24.66%	124 41.89%	99 33.45%
		80–90	473	27.86%	118 24.95%	194 41.01%	161 34.04%
		90–100	508	29.92%	75 14.76%	189 37.20%	244 48.03%
		≥ 100	327	19.26%	29 8.87%	126 38.53%	172 52.60%
		Unknown	94	5.54%	—	—	—

fire protection, and transportation. The construction industry had the highest number of HRIs and fatality rates, whereas fire protection had a notably lower fatality rate than other industries. Among the environment-related factors, there were more HRIs during the summer months of the year, and the peak of HRIs occurred in the afternoon of the day. Regarding the HI, an increase in its value corresponded to a heightened risk of severe HRIs.

Following text mining, variables with a high correlation with HRIs severity degree were obtained from the top 200 keywords using multiple machine learning methods. The correlates were obtained from four machine learning models, and these variables were arranged from highest to lowest in terms of feature importance. Among the main symptoms shown in Table 2, "dehydration", "dizziness", "cramps", "vomiting", and "faintness" were mainly associated with hospitalized accidents and fatal accidents mainly caused by "collapse" and "heart attack". Also, "fall accident" was a frequent secondary injury associated with HRIs, occurring in about 7.54% of all HRIs. And about 52% of the fall accidents related to heat resulted in fatalities. Moreover, the results of machine learning models showed that "back to work" was an essential factor in the severity of HRIs. "Back to work" refers to workers who had heat illness symptoms but continued to work after a short break because they failed to recognize heat illness or take heat illness seriously enough. The statistical results showed that around 43% of "back to work" resulted in fatalities.

3.2. MNL modeling results and discussion

Table 3 presents correlates of HRIs degree tested by the MNL model, including coefficients, standard errors, P values, and marginal effects developed at different severity levels. Many variables from the row dataset and text mining and machine learning results were selected. Three HRIs severity levels were considered as dependent variables, and non-hospitalized was the base category.

Table 2
Variables selection results

Variable (Total n = 1,698)	Frequency	Percent	Non-hospitalized	Hospitalized	Fatal	
Collapse	1 = Yes, 0 = Otherwise	255	15.02%	12 4.71%	43 16.86%	200 78.43%
Dehydration	1 = Yes, 0 = Otherwise	226	13.31%	69 30.53%	142 62.83%	15 6.64%
Dizziness	1 = Yes, 0 = Otherwise	162	9.54%	57 35.18%	71 43.83%	34 20.99%
Cramps	1 = Yes, 0 = Otherwise	153	9.01%	33 21.57%	106 69.28%	14 9.15%
Vomiting	1 = Yes, 0 = Otherwise	111	6.54%	24 21.62%	58 52.25%	29 26.13%
Faintness	1 = Yes, 0 = Otherwise	80	4.71%	22 27.50%	43 53.75%	15 18.75%
Heart attack	1 = Yes, 0 = Otherwise	96	5.65%	3 3.13%	13 13.54%	80 83.33%
Fall accident	1 = Yes, 0 = Otherwise	128	7.54%	13 10.16%	49 38.28%	66 51.56%
Back to work	1 = Yes, 0 = Otherwise	68	4.00%	12 17.65%	27 39.70%	29 42.65%

Therefore, all estimated coefficients and marginal effects for selected variables indicated the influence of variables on specific injury levels compared with the non-hospitalized group. Since the dependent variable has a reference group in the MNL model, the use of coefficient interpretation to analyze MNL results has apparent shortcomings. Therefore, this study focused on explaining the effects of the variables in terms of marginal effect results.

The final MNL model specification was selected based on theoretical considerations and empirical properties of the model. All variables from descriptive statistics were considered and used in the model development. However, some of the variables had p values higher than 0.05 were taken off the list of essential variables. In addition, the model only kept variables with a number greater than 20. To enhance the robustness and comprehensiveness of the model, data augmentation techniques were employed to enrich the dataset, mitigating the potential effects of overfitting and improving the model's ability to generalize new, unseen data. Following tests for multicollinearity, variables with a variance inflation factor less than five were retained. Moreover, categorical variables were changed into dummy variables before performing the MNL model. The base levels of dummy variables were selected based on the principles of the highest value, lowest value, or precise definition. In this study, the base categories of independent variables were "youth worker", "male", "construction", "spring", and "HI (<80)". With application of dummy variables, 23 variables were found to be significant in predicting the severity of HRIs in the MNL model. The variables' goodness of fit for the final model was reasonable, and parameter signs were as expected. The selected model contained some statistically insignificant variables because they were part of a group of variables. These data were imported into statistical analysis software R [19].

3.2.1. *Demographic attributes*

In view of demographic attributes, middle-aged, older, and female workers showed significant impacts on the severity of HRIs. As shown in Table 3, HRIs that happened to middle-aged workers faced about a 3.3% decrease in the chance of hospitalized HRIs and a 7.5% increase in the chance of fatal HRIs compared to youth workers. Older workers were 2.1% and 9.8% more likely to be involved in hospitalized HRIs and fatal HRIs compared to youth workers, respectively. Similar findings were found in the literature, which indicated that older people are at higher risk of HRIs [20]. In addition, females were associated with a 1.6% higher chance of hospitalized HRIs, but a 22.4% lower chance of fatal HRIs compared to male workers. Similar results were shown in an existing literature, Gifford et al [21] found that the rate of HRIs was significantly increased in men compared with women, and the mortality rate of HRIs was higher in men than in women.

3.2.2. *Industry attributes*

Compared to the construction industry, the likelihood of hospitalized HRIs in the agriculture, manufacturing, fire protection, and transportation industries increased by 4.9%, 17.4%, 22.8%, and 19.4%, respectively. Also, the likelihood of hospitalized injury in service and other industries decreased by 3.1% and 2.94%, respectively. In addition, the construction industry had the highest fatal HRIs likelihood. Agriculture, manufacturing, service, fire protection, and transportation were less likely to suffer fatal injuries. Studies also found that the construction industry had a higher rate of fatality due to HRIs than other industries [22,23]. Interestingly, HRIs severity decreased significantly in the fire protection industry. Firefighters were less likely to suffer fatal injuries, with a 38.3% reduction compared to construction workers. This might be because firefighters are often exposed to heat environments and have better heat acclimatization. Moreover, during the statistical data analysis, this study also found that firefighters were severely exposed to extreme heat hazards, not only in their firefighting actions but also during their daily physical training. The injury reports showed that around 30% of firefighters' HRIs occurred during regular exercise, and there were cases of fatality. Therefore, more scientific training methods and effective preventive measures are crucial for firefighters.

3.2.3. *Environmental attributes*

In terms of time of year, there were more hospitalized HRIs during fall and winter compared to spring. Fatal injuries were more likely to increase during summer and fall. Seasonal variations in HRIs may arise from differing levels of preparedness and awareness. Inadequate acclimatization during fall and winter contrasted with heightened precaution in summer. However, prolonged exposure to high temperatures, especially during sporadic fall heatwaves, may result in severe or fatal outcomes. This underscores the necessity of adaptive risk mitigation across seasons.

The injury severity increased when HI increased. Marginal effect results showed that as HI increased, there were declines in hospitalized HRIs. Specifically, reductions of 0.4% and 0.5% observed for

Table 3
Multinomial logit regression model results

Dependent variable:		β		Std. error		Marginal effects		
		Hospitalized	Fatal	Hospitalized	Fatal	Non-hospitalized	Hospitalized	Fatal
Demographics	Middle-aged	0.169***	0.606***	0.076	0.090	−0.042	−0.033	0.075
	Older	0.828***	1.215***	0.111	0.124	−0.119	0.021	0.098
	Female	−1.323***	−2.358***	0.088	0.123	0.208	0.016	−0.224
Industries	Agriculture	−0.211**	−0.766***	0.096	0.100	0.050	0.049	−0.099
	Manufacturing	0.861***	−0.004	0.118	0.127	−0.070	0.174	−0.104
	Service	−0.339***	−0.314***	0.114	0.115	0.041	−0.031	−0.010
	Fire protection	−0.484***	−2.721***	0.096	0.138	0.155	0.228	−0.383
	Transportation	0.638***	−0.550**	0.147	0.166	−0.029	0.194	−0.166
	Other industries	−0.537***	−0.657***	0.100	0.102	0.072	−0.029	−0.042
Environment	Summer	0.090	0.281***	0.087	0.097	−0.019	−0.016	0.035
	Fall	0.302***	0.205	0.112	0.128	−0.033	0.036	−0.003
	Winter	2.139***	1.948***	0.373	0.442	−0.257	0.197	0.060
	HI (80−90)	0.140*	0.201**	0.077	0.092	0.020	−0.004	−0.016
	HI (90−100)	0.306***	0.554***	0.086	0.097	−0.048	−0.005	0.053
	HI (≥100)	0.782***	0.924***	0.105	0.115	−0.103	0.047	0.056
Symptoms	Collapse	0.674***	2.198***	0.137	0.134	−0.148	−0.128	0.276
	Dehydration	−0.032	−2.384***	0.070	0.123	0.104	0.279	−0.383
	Dizziness	−0.656***	−1.462***	0.083	0.105	0.116	0.043	−0.159
	Cramps	0.242***	−1.593***	0.088	0.133	0.048	0.239	−0.287
	Faintness	0.615***	−0.672***	0.122	0.166	−0.022	0.204	−0.183
	Vomiting	0.328***	−0.001	0.110	0.132	−0.027	0.066	−0.039
	Heart attack	0.742**	2.168***	0.254	0.247	−0.153	−0.110	0.263
	Fall accident	0.532***	0.819***	0.137	0.142	−0.078	0.009	0.069
	Constant	0.498***	0.616***	0.131	0.147			
	Observations number	1579						
	Pseudo−R^2	0.427						

Note: * >90% level of significance. ** >95% level of significance. *** >99% level of significance.

HI ranges of 80–90 and 90–100, respectively, compared to cases with HI below 80. In contrast, the prevalence of fatal injuries showed proportional increases of 5.3% and 5.6% for the HI ranges of 90–100 and more than 100, respectively. The previous study has also found that a 1°C augmentation in temperature corresponded to an average increment of 18% in heat illness morbidity, alongside a 35% escalation in mortality [24].

3.2.4. Symptom attributes

According to model results, the possibility of hospitalized HRIs decreased by 12.8% and the possibility of fatal HRIs increased by 27.6% when the worker had collapse symptoms. Furthermore, when the workers had the symptoms of dehydration, dizziness, cramps, faintness, or vomiting, hospitalized HRIs probability was increased. Meanwhile, when the workers have the symptoms of dehydration, dizziness, cramps, faintness, or vomiting, the probability of resulting in fatal HRIs decreased by 38.3 %, 15.9%, 28.7%, 18.3%, and 3.9%, respectively. Moreover, with heart attack, workers experienced a significant increase in HRIs fatality. Heart attacks caused by heat resulted in a 26.3% increase in the probability of fatality among workers. In addition, the results showed that there was a 0.9% higher possibility of hospitalized and a 6.9% higher possibility of fatal injuries with a fall accident caused by heat. Therefore, appropriate actions shall be taken to prevent more severe HRIs, such as stopping working immediately, taking sufficient rest, or receiving appropriate medical attention.

4. Conclusion

This research effort explored the impact factors of the severity of HRIs based on the injury data collected by the OSHA. More impact factors were explored from descriptive summary reports using text mining and machine learning. The MNL model was applied to explore the impact of a variety of demographics, industry, environment, and symptom factors on HRIs severity. The results indicated that middle-aged and older workers were more likely to die in HRIs than youth workers. Males were more likely to get fatal HRIs than females. Among the different industries, construction workers faced the most dangerous HRIs, followed by service workers, manufacturing workers, and farmers. Moreover, firefighters had the lowest HRIs fatality rate. In addition, the likelihood of fatal HRIs increased as the heat index rose. Among the symptoms and secondary injury associated with HRIs severity, the presence of collapse, heart attack, and fall accidents resulted in more severe HRIs. In contrast, the presence of dehydration, dizziness, cramps, faintness, and vomiting symptoms resulted in a lower probability of fatal HRIs.

This study provides new insights into the contributing factors to the severity of occupational HRIs. By integrating text mining, model-free machine learning, and MNL methods, it innovatively explores the multifaceted impact factors contributing to the severity of HRIs. This methodology leverages OSHA's historical accident reports to uncover nuanced insights into HRIs' contributing factors. In particular, considering the variation in the impacts across observations will lead to better estimation and a more efficient design of safety countermeasures. Recommendations may include specific educational programs like applying different prevention measures for different industries, taking different measures after the appearance of different symptoms, and so on. Another novel aspect of our study is that most of the existing studies on occupational HRIs are on the influencing factors of HRIs themselves. This study can offer plausible suggestions on prevention of HRIs to legislatures and employers for different workers and time of year. Moreover, the findings can be used to provide practical, specific suggestions on how heat stress training regimens could be improved. For instance, training programs can include guidance on identifying early symptoms of heat stress and implementing immediate, symptom-specific interventions to prevent escalation.

This study has several limitations and could be further improved through future work. As the data covers a relatively long period (1984–2021), temporal instability could be associated with the data, which may affect the model's reliability. The finding of random parameters may capture temporal variations and not heterogeneity. Future work could benefit from addressing temporal instability. Besides, due to the imbalanced nature of HRIs data, future work may also benefit from applying the data-driven method to address this issue and thereby improve model performance.

Conflicts of interest

We confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

[1] Occupational Safety and Health Administration (OSHA) [Internet]. Safety and Health Topics-heat [cited 2023 Oct 15]; 2023., https://www.osha.gov/heat-exposure.
[2] Bureau of Labor Statistics (BLS) [Internet]. Fatal Occupational Injuries by Selected Worker Characteristics and Selected Event or Exposure, All U.S., All Ownerships, 1992–2020 [cited 2023 Oct 10]; 2023., https://data.bls.gov/gqt/InitialPage.
[3] Lyu P, Song S, Watts P. Risk analysis of occupational heat-related illness. In: 58th Annual Associated Schools of Construction International Conference vol. 3; 2022. p. 47–55. https://doi.org/10.29007/zxhz.
[4] Borg MA, Xiang J, Anikeeva O, Pisaniello D, Hansen A, Zander K, Dear K, Sim MR. Bi P. Occupational heat stress and economic burden: a review of global evidence. Environ Res 2021;195:110781. https://doi.org/10.1016/j.envres.2021.110781.
[5] Kenny GP, Flouris AD, Yagouti A, Notley SR. Towards establishing evidence-based guidelines on maximum indoor temperatures during hot weather in temperate continental climates. Temperature 2019;6:11–36. https://doi.org/10.1080/23328940.2018.1456257.
[6] Varghese BM, Hansen A, Bi P, Pisaniello D. Are workers at risk of occupational injuries due to heat exposure? A comprehensive literature review. Saf Sci 2018;110:380–92. https://doi.org/10.1016/j.ssci.2018.04.027.
[7] Binazzi A, Levi M, Bonafede M, Bugani M, Messeri A, Morabito M, Marinaccio A, Baldasseroni A. Evaluation of the impact of heat stress on the occurrence of occupational injuries: meta-analysis of observational studies. Am J Ind Med 2019;62:233–43. https://doi.org/10.1002/ajim.22946.
[8] Nelson NG, Collins CL, Comstock RD, McKenzie LB. Exertional heat-related injuries treated in emergency departments in the U.S., 1997–2006. Am J Prev Med 2011;40:54–60. https://doi.org/10.1016/j.amepre.2010.09.031.
[9] Varghese BM, Hansen AL, Williams S, Bi P, Hanson-Easey S, Barnett AG, Heyworth JS, Sim MR, Rowett S, Nitschke M, Di Corleto R, Pisaniello DL. Heat-related injuries in Australian workplaces: perspectives from health and safety representatives. Saf Sci 2020;126:104651. https://doi.org/10.1016/j.ssci.2020.104651.
[10] Nelson DA, Deuster PA, O'Connor FG, Kurina LM. Timing and predictors of mild and severe heat illness among new military enlistees. Med Sci Sports Exerc 2018;50:1603–12. https://doi.org/10.1249/MSS.0000000000001623.
[11] Kakamu T, Endo S, Hidaka T, Masuishi Y, Kasuga H, Fukushima T. Heat-related illness risk and associated personal and environmental factors of construction workers during work in summer. Sci Rep 2021;11:1119. https://doi.org/10.1038/s41598-020-79876-w.
[12] Ebi KL, Capon A, Berry P, Broderick C, De Dear R, Havenith G, Honda Y, Kovats RS, Ma W, Malik A, Morris NB, Nybo L, Seneviratne SI, Vanos J, Jay O. Hot weather and heat extremes: health risks. The Lancet 2021;398:698–708. https://doi.org/10.1016/S0140-6736(21)01208-3.
[13] Department of Labor (DOL) [Internet]. OSHA Enforcement Data 2023 [cited 2023 Sep 4]. https://enforcedata.dol.gov/views/data_summary.php.
[14] OpenWeather [Internet]. History API for Timestamp 2023 [cited 2023 Sep 18]. https://openweathermap.org/api/history-api-timestamp.
[15] Gao C, Sun H, Wang T, Tang M, Bohnen NI, Müller MLTM, Herman T, Giladi N, Kalinin A, Spino C, Dauer W, Hausdorff JM, Dinov ID. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. Sci Rep 2018;8:7129. https://doi.org/10.1038/s41598-018-24783-4.

[16] Fiebig DG, Keane MP, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Mark Sci 2010;29: 393–421. https://doi.org/10.1287/mksc.1090.0508.

[17] Shankar V, Manner F. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. J Safety Res 1996;27:183–94. https://doi.org/10.1016/0022-4375(96)00010-2.

[18] Bureau of Labor Statistics (BLS) [Internet]. Labor Force Statistics From the Current Population Survey, Demographic; 2023 [cited 2023 Oct 9], https://www.bls.gov/cps/demographics.htm#age.

[19] R Core Team. R: A Language and Environment for Statistical Computing 2021.

[20] Langer CE, Mitchell DC, Armitage TL, Moyce SC, Tancredi DJ, Castro J, Vega-Arroyo AJ, Bennett DH, Schenker MB. Are Cal/OSHA regulations protecting farmworkers in California from heat-related illness? J Occup Environ Med 2021;63:532–9. https://doi.org/10.1097/JOM.0000000000002189.

[21] Gifford RM, Todisco T, Stacey M, Fujisawa T, Allerhand M, Woods DR, Reynolds RM. Risk of heat illness in men and women: a systematic review and meta-analysis. Environ Res 2019;171:24–35. https://doi.org/10.1016/j.envres.2018.10.020.

[22] Acharya P, Boggess B, Zhang K. Assessing heat stress and health among construction workers in a changing climate: a review. Int J Environ Res Public Health 2018;15:247. https://doi.org/10.3390/ijerph15020247.

[23] Gubernot DM, Anderson GB, Hunting KL. Characterizing occupational heat-related mortality in the United States, 2000–2010: an analysis using the census of fatal occupational injuries database. Am J Ind Med 2015;58:203–11. https://doi.org/10.1002/ajim.22381.

[24] Faurie C, Varghese BM, Liu J, Bi P. Association between high temperature and heatwaves with heat-related illnesses: a systematic review and meta-analysis. Sci Total Environ 2022;852:158332. https://doi.org/10.1016/j.scitotenv.2022.158332.