Novel Surveillance of Occupational Injury Powered by Machine Learning Using Chief Complaint at Emergency Triage

Ting-Chia Weng

National Cheng Kung University Hospital

Yu-Ting Wei

Taichung Tzu Chi Hospital

Tsung-Yu Chan

National Cheng Kung University Hospital

Wen-Chau Chen

National Cheng Kung University Hospital

Ming-Hong Chen

National Cheng Kung University Hospital

Jung-Der Wang

National Cheng Kung University

Chung-I Li

National Cheng Kung University

Yau-Chang Kuo (✉ kuo550221@gmail.com)

National Cheng Kung University Hospital

Research

Keywords: Occupational Injury, Surveillance, Chief Complaint, Supervised Machine Learning, Natural Language Processing, Emergency Department

Posted Date: June 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29316/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Background: Underreporting of occupational injuries remains a global health issue, and warrants social awareness. To assist surveillance of occupational injuries, we developed an automatic screening model using chief complaint at emergency department to improve detection and reporting of occupational injuries.

Methods: A total of 181,843 emergency visits aged 15 to 65 year-old were included at a medical center in Taiwan from April, 2015 to March, 2018. The retrospective cohort composed of 9,307 cases of occupational injuries and 172,536 controls. We applied the first 30 months as training dataset, and the following 6 months for prospective testing. Natural Language Processing (NLP) was applied to analyze patients' chief complaints. The sentences were processed by JiebaR, a Chinese text segmentation technique, reviewed by two occupational physicians, and transformed by a word-embedding model using Glove. Logistic regression was conducted to predict suspected cases of occupational injuries.

Results: The prediction model using the chief complaint alone can achieve an overall AUC of 0.936 [95% confidence interval (CI): 0.931 to 0.941], with high sensitivity (93.1%, 95% CI: 91.8 to 94.4) and adequate specificity (84.8%, 95% CI: 84.4 to 85.2). Patients in the most urgent or severe conditions showed the highest accuracy around 90%. We also observed increasing trends of referral to specialists of occupational medicine, and reimbursement rate.

Conclusions: Interdepartmental coordination and integration of machine learning may augment the detection of occupational injuries at the emergency triage and improve the reporting, compensation, and prevention of occupational injuries.

Contributions to the literature

- I This novel surveillance powered by supervised machine learning can help to identify occupational injuries at the emergency triage, and it achieves 93.1% sensitivity and 84.8% specificity.
- I Along with the system implementation, there are trends of increasing reimbursement rate in the National Labor Insurance and increasing case referrals to specialists of occupational medicine.
- I Combining the use of Artificial Intelligence (AI) with conventional medical information, such as chief complaints, may provide an effective surveillance system for occupational injuries at hospitals.
- I Effective screening of labors who suffer from occupational injuries will promote Total Worker Health.

Background

Occupational injuries have posed substantial health and economic burdens on individuals, employers, and the society(1, 2). It remains under-reported and further impedes the effectiveness of intervention to reduce potential hazards in the workplace(3, 4). Although the National Labor Insurance (NLI) has been in

practice for more than 20 years in Taiwan, there is still a high proportion of occupational diseases that have not been reported in the official system(5–8). Meanwhile, the emergency department (ED) is the main unit to manage urgent injuries. It serves as an excellent venue for disease surveillance(6, 9–11). However, healthcare professionals typically have to handle an overwhelming number of patients at the ED. This situation makes it difficult to efficiently assess occupational injuries; thus an automatic screening system to facilitate clinical evaluation is needed(12, 13). To solve the under-reporting, joint support by healthcare professionals may serve to detect occupational injuries at the frontline. Finding ways to improve the reporting rate of occupational injuries without burdening the clinical staff in the overcrowded ED is a challenging but important task.

Machine learning has been applied in biomedical fields to assist clinical decision-making(14–16). Natural Language Processing (NLP) is a subset algorithm which has specific strengths in non-structural text mining, and has been applied in the surveillance of infectious diseases(12, 17, 18). We hypothesize that the analyses of the chief complaints using a NLP algorithm will not only augment the detection of occupational injuries at the ED, but also save time for healthcare professionals.

Chief complaints at the ED are patient-reported outcomes written in Chinese by experienced triage nurses, and comprise key information in the form of brief sentences, including causes of illness, type of injuries or mechanisms, severity, duration of symptoms, time and place of onset(19, 20). Starting from February, 2015, a novel occupational surveillance system was established in a tertiary medical center in southern Taiwan that includes universal screening and documentation of work-related injuries and illnesses. Using the system with both unique and universal labels of occupational injuries, we were able to perform chart review and analyze the association between chief complaint and occupational injury.

The specific aim of the study was to build a novel surveillance system that automatically detects occupational injuries by using NLP and statistics machine learning algorithms to facilitate the identification, reporting, compensation and prevention of occupation injuries. Further evaluation of the database may help us analyze risk factors at the work place and promote workers' overall health.

Methods

This study was approved by the Institutional Review Board of National Cheng Kung University Hospital (NCKUH), approval number: B-ER-106-158, before commencement. We analyzed chief complaints at the triage station of the emergency department retrospectively by abstracting electronic records from NCKUH, a tertiary medical center in southern Taiwan, from April, 2015 to March, 2018. Prior to this, an Occupational Diseases Surveillance System (ODSS) has been in place at NCKUH since February, 2015. Three groups of healthcare professionals used the ODSS to detect occupational injuries, involving (1) triage nurses, (2) staff at the cashier/ registration counter, and (3) ED doctors during consultations to detect if patients visited ED due to injuries at the work place. Although the conventional definition of occupational injury must fulfill both criteria of “arising out of work” and/or “in the course of work”, the ODSS can only be applied for the second definition in daily practice. Briefly, the triage nurses identified

injuries that occurred in the course of work and recorded the chief complaints, which were supplemented by complaints identified by doctors and staff at the cashier/ registration counter. Injuries which resulted from domestic work without payment or without commuting to or from work was not included. Data from the initial two months (February and March, 2015) were excluded to assure the validity.

As outlined in Fig. 1(A), there were 309,644 emergency visits in total at the hospital, from April, 2015 to March, 2018. The ED visits of populations aged between 15- to 65-year-old were included, resulting in a total of 181,843 patients. Among them, there were 9,307 suspected cases of occupational injuries identified by our healthcare professionals. Among the 9,307 suspected cases of occupational injuries, 6,495 (69.79%) cases were identified by nurses at the triage station, 2,698 (28.99%) by doctors, and 2,055 (22.08%) by staff at the cashier/ registration counter. We excluded duplicated data for cases re-admitted to the ED within 36 hours, and only the chief complaint at the first visit for patients with multiple visits were included. The study period was from April, 2015 to March, 2018. We applied events in the first 30 months, from April 2015 to September 2017, as the training dataset (7,747 cases and 141,081 controls). The events of the following 6 months, from October, 2017 to March, 2018, were classified as the testing dataset (1,560 cases and 31,455 controls) for prospective evaluation.

The main outcome of interest was to evaluate if the algorithm could automatically differentiate emergency visits due to occupational injuries from visits due to other causes according to chief complaints recorded at the triage station. The algorithm of the screening model is summarized in Fig. 1B: The Chinese text segmentation technique JiebaR(21) was first applied to identify all stop words to be omitted from the chief complaints. To assure the quality of the results of segmentation and to save the resources of computer processing, two occupational physicians in our team carefully selected meaningless words out and added them into the pool of stop words. Then, each chief complaint was represented as a vector that considered the summed effect of leftover words and its relative location in a sentence, which was processed with the word-embedding model with Glove in a dimensionality of 100(22).

A third step of constructing a classification model was developed with logistic regression to predict occupational injuries. Three models with different input variables were developed. In *Model 1*, only the sentence of the chief complaint was applied to predict occupational injuries. In *Model 2*, additional age and sex were incorporated into the chief complaint. In *Model 3*, the model included chief complaint, age, sex and additional triage level, time of ED visit (within working hours versus non-working hours), and weekday versus weekend. The seven metrics of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio [LR (+)], negative likelihood ratio [LR (-)], and area under curve (AUC) were calculated to evaluate the overall model performance, and subgroups stratified by age, gender, triage level and time of visit.

Results

In total, there were 9,307 cases of emergency visits of occupational injuries and 172,536 controls of non-occupational visits for working populations at NCKUH from April 1, 2015 to March 31, 2018 (Table 1). Men accounted for 59.62% of visits related to occupational injuries, while women accounted for 52.71% of other emergency visits during the study period. In the subset of triage level 1 (most urgent in severity), there were more ED visits related to other causes (the control group) than those with occupational injuries (the study group). On average, cases in the study group were generally younger than controls. The time of visit for the study group was mainly during working hours 8:00–17:00 and on weekdays (Monday to Friday), while those of the control group were during non-working hours and on weekends. There was no major difference in the distributions of age, sex, triage severity, and time of visit for both the training and testing datasets. The average length of chief complaints in Chinese character word count was 33 (SD = 12) in the training period and 35 (SD = 13) in the testing period.

Table 1

Frequency distributions of visits to emergency department (ED) stratified by relatedness to occupational injuries and datasets of training versus testing.

	Cases	Controls	Training	Testing
Number of visits				
Suspected cases of occupational injury	9,307	0	7,747 (5.21)	1,560 (4.73)
Other visits	0	172, 536	141,081 (94.79)	31,455 (95.27)
Gender, N (%)				
Female	3,758 (40.38)	90,950 (52.71)	77,495 (52.07)	17,213 (52.14)
Male	5,549 (59.62)	81,586 (47.29)	71,333 (47.93)	15,802 (47.86)
Age (mean, SD)	35.30 (13.07)	40.64 (15.14)	40.31 (15.09)	40.61 (15.07)
Age-group, N (%)				
15–25	2,791 (29.99)	39,197 (22.72)	34,549 (23.21)	7,439 (22.53)
26–35	2,544 (27.33)	32,634 (18.91)	28,947 (19.45)	6,231 (18.87)
36–45	1,640 (17.62)	28,303 (16.40)	24,326 (16.35)	5,617 (17.01)
46–55	1,377 (14.80)	32,411 (18.79)	27,607 (18.55)	6,181 (18.72)
56–65	955 (10.26)	39,991 (23.18)	33,399 (22.44)	7,547 (22.86)
Triage Level, N (%)				
Triage Level 1	72 (0.77)	4,466 (2.59)	3,665 (2.46)	873 (2.64)
Triage Level 2	1,218 (13.09)	19,382 (11.23)	16,218 (10.90)	4,382 (13.27)
Triage Level 3	6,466 (69.47)	130,867 (75.85)	111,471 (74.90)	25,862 (78.34)
Triage Level 4	1,110 (11.93)	17,247 (10.00)	16,646 (11.18)	1,711 (5.18)
Triage Level 5	441 (4.74)	574 (0.33)	828 (0.56)	187 (0.57)
Day of week, N (%)				

	Cases	Controls	Training	Testing
Monday	1,340 (14.40)	24,323 (14.10)	20,982 (14.11)	4,681 (14.18)
Tuesday	1,434 (15.41)	22,289 (12.92)	19,468 (13.08)	4,255 (12.89)
Wednesday	1,441 (15.48)	22,457 (13.02)	19,616 (13.18)	4,282 (12.97)
Thursday	1,473 (15.83)	21,953 (12.72)	19,230 (12.92)	4,196 (12.71)
Friday	1,537 (16.51)	22,731 (13.17)	19,783 (13.29)	4,485 (13.58)
Saturday	1,129 (12.13)	26,022 (15.08)	22,196 (14.91)	4,955 (15.01)
Sunday	953 (10.24)	32,761 (18.99)	27,553 (18.51)	6,161 (18.66)
Visiting hour, N(%)				
08:00–17:00	5,979 (64.24)	76,529 (44.36)	67,370 (45.27)	15,138 (45.85)
17:00–08:00	3,328 (35.76)	96,007 (55.64)	81,458 (54.73)	17,877 (54.18)

Table 2 summarizes the overall screening performance of different models and comparisons between the subsets of training and testing. In general, the performance of the three models with different input variables were similar. Namely, *Model 1* using chief complaints alone showed excellent performance of sensitivity (93.1%), specificity (84.8%), PPV (23.3%), NPV (99.6%), and the AUC (0.936). Slight improvements were achieved in *Model 2* and *Model 3* with additional input variables related to demographic characteristics.

Table 2

Comparison of overall screening performance in identifying occupational injuries between subsets of training and testing, and model selection from 3 different combinations of variables.

	Model 1		Model 2		Model 3	
	Training	Testing	Training	Testing	Training	Testing
AUC	0.939 (0.937, 0.941)	0.936 (0.931, 0.941)	0.939 (0.937, 0.941)	0.936 (0.931, 0.941)	0.944 (0.943, 0.946)	0.941 (0.937, 0.946)
Accuracy (%)	85.2 (84.1, 85.3)	85.2 (84.8, 85.6)	85.3 (85.1, 85.4)	85.2 (84.8, 85.6)	85.9 (85.7, 86.0)	85.7 (85.3, 86.1)
Sensitivity (%)	92.3 (91.7, 92.9)	93.1 (91.8, 94.4)	92.5 (91.9, 93.0)	93.1 (91.8, 94.4)	92.4 (91.8, 93.0)	92.9 (91.5, 94.1)
Specificity (%)	84.8 (84.6, 85.0)	84.8 (84.4, 85.2)	84.9 (84.7, 85.1)	84.8 (84.4, 85.2)	85.5 (85.3, 85.7)	85.3 (84.9, 85.7)
PPV (%)	25.0 (24.5, 25.5)	23.3 (22.3, 24.4)	25.1 (24.6, 25.6)	23.3 (22.3, 24.4)	25.9 (25.4, 26.4)	23.9 (22.8, 25.0)
NPV (%)	99.5 (99.5, 99.5)	99.6 (99.5, 99.7)	99.5 (99.5, 99.5)	99.6 (99.5, 99.7)	99.5 (99.5, 99.6)	99.6 (99.5, 99.7)
LR (+)	6.06 (5.98, 6.15)	6.13 (5.95, 6.31)	6.11 (6.03, 6.19)	6.14 (5.96, 6.32)	6.37 (6.28, 6.46)	6.33 (6.15, 6.53)
LR (-)	0.09 (0.08, 0.10)	0.08 (0.07, 0.10)	0.09 (0.08, 0.10)	0.08 (0.07, 0.10)	0.09 (0.08, 0.10)	0.08 (0.07, 0.10)
Model 1: input variable with chief complaint at triage only;						
Model 2: chief complaint plus age (in continuous variable) and gender;						
Model 3: a robust model with chief complaint, age, gender plus triage level, day of patient visit (Monday to Friday versus Saturday and Sunday), and visiting hours (8 am to 5 pm, versus 5 pm to 8 am on the next day).						
*AUC: area under curve of receiver operating characteristic; PPV: positive predictive value; NPV: negative predictive value; LR (+): positive likelihood ratio; LR (-): negative likelihood ratio						

We performed subgroup evaluations to test the robustness and utility of prediction performance across different subgroups by age, sex, severity and time of visit. The results of *Model 1* are summarized in Table 3, which showed similar performance results after stratification by patient age, sex, disease severity, and time-frames in all three models. Similar performance was achieved in *Model 3* (summarized in the Supplementary Table 1). As male and young adults (26–35 years old) had the highest proportions of occupational injuries, they were accompanied with the highest PPV, in contrast to females and other age groups. Most importantly, *Model 1* yielded better accuracy (90.6%) and PPV (30.4%) in patients with the most urgent or severe conditions (triage levels 1 and 2) than patients with milder clinical problems.

Table 3

Comparison of screening performance of Model 1 in different subgroups stratified by age, gender, triage level and time of visit.

Group	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR (+)	LR (-)
Women	Training	86.4	92.4	86.2	21.9	99.6	6.68	0.09
		(86.2, 86.7)	(91.4, 93.3)	(85.9, 86.4)	(21.2, 22.6)	(99.6, 99.7)	(6.55, 6.82)	(0.08, 0.10)
	Testing	86.4	92.2	86.2	20.4	99.7	6.67	0.09
		(85.9, 86.9)	(89.8, 94.1)	(85.7, 86.7)	(19.0, 22.0)	(99.5, 99.7)	(6.39, 6.98)	(0.07, 0.12)
Men	Training	83.8	92.3	83.2	27.6	99.4	5.50	0.09
		(83.5, 84.1)	(91.5, 93.0)	(82.9, 83.5)	(26.9, 8.3)	(99.3, 99.4)	(5.39, 5.60)	(0.08, 0.10)
	Testing	83.9	93.8	83.3	25.8	99.5	5.60	0.07
		(83.3, 84.5)	(92.1, 95.3)	(82.7, 83.9)	(24.3, 27.3)	(99.4, 99.7)	(5.39, 5.83)	(0.06, 0.10)
15–25	Training	75.2	94.9	73.7	20.8	99.5	3.61	0.07
		(74.7, 75.6)	(93.9, 95.8)	(73.3, 74.2)	(20.0, 21.6)	(99.4, 99.6)	(3.54, 3.69)	(0.06, 0.08)
	Testing	74.5	92.0	73.3	18.3	99.3	3.45	0.11
		(73.5, 75.5)	(89.1, 94.4)	(72.3, 74.4)	(16.7, 19.9)	(99.0, 99.5)	(3.29, 3.62)	(0.08, 0.15)
26–35	Training	86.7	91.4	86.3	34.4	99.2	6.67	0.10
		(86.3, 87.1)	(90.1, 92.5)	(85.9, 86.7)	(33.2, 35.7)	(99.1, 99.3)	(6.46, 6.90)	(0.09, 0.12)
	Testing	87.0	94.9	86.4	34.3	99.6	6.99	0.06
		(86.2, 87.8)	(92.4, 96.8)	(85.5, 87.3)	(31.6, 37.0)	(99.3, 99.7)	(6.53, 7.49)	(0.04, 0.09)
36–45	Training	88.0	92.7	87.8	30.8	99.5	7.57	0.08
		(87.6, 88.4)	(91.2, 94.0)	(87.3, 88.2)	(29.4, 32.3)	(99.4, 99.6)	(7.29, 7.86)	(0.07, 0.10)
	Testing	87.3	91.3	87.1	27.6	99.5	7.07	0.10
		(86.4, 88.2)	(87.4, 94.3)	(86.2, 88.0)	(24.8, 30.5)	(99.2, 99.7)	(6.54, 7.65)	(0.07, 0.15)

*PPV: Positive predictive value; NPV: Negative predictive value; LR (+): Positive likelihood ratio; LR (-): Negative likelihood ratio

Group	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR (+)	LR (-)
46–55	Training	88.6	90.4	88.5	25.6	99.5	7.86	0.11
		(88.2, 89.0)	(88.6, 92.0)	(88.1, 88.9)	(24.3, 27.0)	(99.4, 99.6)	(7.57, 8.17)	(0.09, 0.13)
	Testing	89.1	93.6	88.9	23.7	99.7	8.42	0.07
		(88.2, 89.8)	(89.6, 96.5)	(88.1, 89.7)	(20.9, 26.7)	(99.6, 99.9)	(7.78, 9.12)	(0.04, 0.12)
56–65	Training	89.3	89.3	89.3	16.7	99.7	8.33	0.12
		(89.0, 89.6)	(86.9, 91.4)	(88.9, 89.6)	(15.6, 17.9)	(99.6, 99.8)	(8.01, 8.67)	(0.10, 0.15)
	Testing	89.6	94.1	89.4	17.0	99.9	8.91	0.07
		(88.8, 90.2)	(89.4, 97.1)	(88.7, 90.1)	(14.6, 19.5)	(99.7, 99.9)	(8.26, 9.62)	(0.04, 0.12)
Triage I & II	Training	89.8	90.6	89.8	33.5	99.4	8.88	0.10
		(89.4, 90.3)	(88.7, 92.3)	(89.4, 90.2)	(31.8, 35.2)	(99.3, 99.5)	(8.48, 9.31)	(0.09, 0.13)
	Testing	90.6	92.9	90.5	30.4	99.7	9.79	0.08
		(89.8, 91.4)	(88.7, 95.9)	(89.7, 91.3)	(26.9, 34.0)	(99.4, 99.8)	(8.93, 10.75)	(0.05, 0.13)
Triage III	Training	85.4	92.5	85.1	23.7	99.6	6.20	0.09
		(85.2, 85.6)	(91.8, 93.2)	(84.9, 85.3)	(23.1, 24.3)	(99.5, 99.6)	(6.10, 6.30)	(0.08, 0.10)
	Testing	85.1	93.0	84.7	22.1	99.6	6.07	0.08
		(84.6, 85.5)	(91.4, 94.4)	(84.2, 85.1)	(21.0, 23.3)	(99.5, 99.7)	(5.87, 6.28)	(0.07, 0.10)
Triage IV & V	Training	78.1	92.9	76.8	25.5	99.2	4.01	0.09
		(77.5, 78.7)	(91.4, 94.2)	(76.2, 77.5)	(24.3, 26.7)	(99.0, 99.4)	(3.88, 4.14)	(0.08, 0.11)
	Testing	72.1	94.4	69.8	24.7	99.2	3.13	0.08
		(70.1, 74.1)	(90.0, 97.3)	(67.6, 72.0)	(21.5, 28.1)	(98.5, 99.6)	(2.89, 3.39)	(0.04, 0.15)

*PPV: Positive predictive value; NPV: Negative predictive value; LR (+): Positive likelihood ratio; LR (-): Negative likelihood ratio

Group	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR (+)	LR (-)
Weekday	Training	85.6 (85.3, 85.8)	92.9 (92.2, 93.6)	85.1 (84.9, 85.3)	28.6 (28.0, 29.3)	99.5 (99.4, 99.5)	6.23 (6.13, 6.34)	0.08 (0.08, 0.09)
	Testing	85.5 (85.1, 86.0)	94.1 (92.6, 95.3)	85.0 (84.5, 85.5)	27.2 (25.9, 28.6)	99.6 (99.5, 99.7)	6.28 (6.06, 6.51)	0.07 (0.06, 0.09)
Weekend	Training	84.4 (84.1, 84.7)	90.3 (88.8, 91.7)	84.2 (83.8, 84.5)	17.2 (16.5, 18.0)	99.6 (99.5, 99.6)	5.70 (5.55, 5.85)	0.12 (0.10, 0.13)
	Testing	84.5 (83.9, 85.2)	89.6 (85.8, 92.7)	84.4 (83.7, 85.1)	14.9 (13.3, 16.5)	99.6 (99.5, 99.7)	5.74 (5.42, 6.08)	0.12 (0.09, 0.17)
08:00–17:00	Training	84.1 (83.8, 84.4)	93.4 (92.7, 94.1)	83.3 (83.0, 83.6)	31.0 (30.2, 31.7)	99.4 (99.3, 99.4)	5.61 (5.50, 5.71)	0.08 (0.07, 0.09)
	Testing	84.0 (83.4, 84.5)	93.9 (92.2, 95.3)	83.3 (82.6, 83.9)	28.1 (26.6, 29.7)	99.5 (99.4, 99.6)	5.61 (5.39, 5.84)	0.07 (0.06, 0.09)
17:00–08:00	Training	86.1 (85.8, 86.3)	90.3 (89.1, 91.4)	85.9 (85.7, 86.2)	18.3 (17.7, 19.0)	99.6 (99.6, 99.7)	6.41 (6.28, 6.55)	0.11 (0.10, 0.13)
	Testing	86.3 (85.7, 86.8)	91.8 (89.2, 93.9)	86.1 (85.5, 86.6)	17.9 (16.5, 19.3)	99.7 (99.6, 99.8)	6.59 (6.30, 6.89)	0.10 (0.07, 0.13)
*PPV: Positive predictive value; NPV: Negative predictive value; LR (+): Positive likelihood ratio; LR (-): Negative likelihood ratio								

Discussion

In this study, we have successfully developed a surveillance system that integrates natural language processing and supervised machine learning to automatically identify occupational injuries through analysis of chief complaints at the emergency department. By using chief complaints only, this system achieved a favorable AUC (0.936) with a sensitivity of 93.1% and specificity of 84.8% (Table 2). It also showed stable performance with accuracies above 84% in both genders, people aged > 25, urgent or severe cases (triage levels I, II, and III), during both working hours and night-time, as well as weekdays and

weekends (Table 3). In fact, following the implementation of the occupational surveillance system at the ED, we observed an increasing trend of referral to specialists of occupational medicine at outpatient clinics, and an increasing trend in the reimbursement rate from the National Labor Insurance (Fig. 2), indicating an initial effect in the improvement of identification and surveillance.

Chief complaints provided valuable information, and served as an essential element in occupational injury surveillance. Table 2 shows that adding more variables (*Models 2 & 3*) to chief complaints (*Model 1*) does not seem to improve the performance of accuracy, sensitivity, and specificity. Namely, the results of adding age and gender in *Model 2* and those of adding level of triage and time of visit in *Model 3* do not improve the AUC, sensitivity and specificity. Moreover, since this surveillance tool showed consistent performance in both training and predicting datasets even after stratification into smaller subgroups(23) (Table 3), its validity seems adequate for application in prospective clinical practice(16) at least for the time being studied.

In addition to a standardized, structured electronic medical record (EMR) system, we consider the following major reasons to also contribute to the successful application of supervised machine learning algorithms for the surveillance of occupational injuries: First, we had experienced triage nurses who provided short but structured sentences of chief complaints summarizing information of patient-reported outcomes, signs, time, place, as well as mechanism of trauma(13, 19, 24). Second, as much attention has been paid to occupational injuries, all emergency visits are universally checked by three different personnel, namely, triage nurses, doctors, and staff at the cashier/ registration counter; hence, chief complaints related to such injuries would be picked up from surveillance. Third, two independent domain scientists or occupational physicians carefully reviewed the results of word segmentation, and filtered out 28, 030 meaningless stop words and phrases. Such efforts provide excellent feedback to machine learning algorithms to improve accuracy during training and testing periods(25, 26). Moreover, this system applied conventional logistic regression which requires relatively low programming efforts, and can be carried out on commercial laptops with free R packages available online in multi-language text mining algorithms. The automated processing can reduce clinical documentation, increase workflow efficiency, and hence reduce practice time and lower costs(12, 20). Supported by the ergonomic design and user friendly EMR system, primary care doctors can devote more time to patients and facilitate the identification, referral, reporting, compensation and prevention of occupational injuries(27).

There were at least the following limitations that must be acknowledged: First, this study applied a logistic regression model on chief complaints only. As the data could be enriched through using detailed electronic medical records, the performance could possibly be improved with non-linear transformation algorithms(28). However, since the current system achieved AUC above 0.931 across the three models; a more sophisticated system might not necessarily improve accuracy significantly while increasing the cost and complexity of computation and facility requirements. Second, the sampled patients in this study were collected at an emergency department in a tertiary teaching hospital in Tainan metropolitan city. The patients' demographics, disease severity, and prevalence of injury types might be quite different to townships of agricultural communities or rural areas. Thus, our results may not be generalized to other

communities, though the method could be adopted and tested in other settings. Third, while the usual gold standard for occupational injury may be validated by worker's compensation, we were unable to apply the same criteria in this study because of the low compensation rate, namely, 21.69% in this study. The major reason may be related to the experience-rated premium charged by the National Labor Insurance in Taiwan. While medical claims of occupational injuries should be covered by the National Labor Insurance (worker's compensation), the National Health Insurance, too, can be utilized in Taiwan since 1995(29, 30), causing under-reporting of occupational injuries for the experience-rating system of the National Labor Insurance(31). There are only minor differences in copayments, 18 USD and 5 USD for general diseases and occupational diseases, respectively, which further aggravates such a discrepancy(32). To improve the performance of surveillance for occupational injury, we applied screening results by health professionals to avoid under-reporting and/or detection. As this approach has improved the referral rate, we would anticipate a consistent increase in the compensation rate in the future, which would in turn provide feedback to healthcare professionals and improve this surveillance system.

Conclusions

In our study, we successfully developed a useful surveillance system for occupational injuries using natural language processing and supervised machine learning with high sensitivity and specificity. The system utilized chief complaints collected at the emergency triage in order to strengthen the conventional surveillance system and reduce underreporting and improve occupational safety and health(29). We recommend that such a system be tested in more diverse settings to improve the efficiency and generalizability(20). Such research might also trigger studies for the detection of injuries of workers co-morbid with chronic disease and/or treated with polypharmacy, which is crucial in promoting Total Worker Health(33, 34) in the aging workforce(35, 36).

Abbreviations

NLP: Natural Language Processing; AI: Artificial Intelligence; NLI: National Labor Insurance; ED: emergency department; NCKUH: National Cheng Kung University Hospital; ODSS: Occupational Diseases Surveillance System; PPV: positive predictive value (PPV); NPV: negative predictive value; LR (+): positive likelihood ratio; LR (-): negative likelihood ratio; AUC: area under curve; EMR: electronic medical record.

Declarations

Ethical approval and consent to participate:

The study protocol was approved by the Institutional Review Board of NCKUH (Approval Number: B-ER-106-158) before commencement. Inform consent was waived since this is a retrospective, de-identified analysis.

Consent for publication:

not applicable.

Availability of supporting data:

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was partially supported by the Ministry of Science and Technology (MOST 107-2627-M-006-007) of Taiwan. The first author was also supported to attend occupational health conferences by three grants: Ministry of Science and Technology (Grant #106CIFA0900102) in Taiwan, National Cheng Kung University Hospital, and Cheng-Hsing Medical Foundation in Tainan, Taiwan.

Author contributions:

Conceived and designed this study: TCW, YTW, CIL and YCK. Constructed the surveillance system and cleaned the data: TYC, WCC, and MHC. Analyzed the data: CIL, TCW and YTW. Wrote this manuscript: TCW and YTW. Initiated this study: TCW, CIL and YCK. Revised this manuscript: TYC, WCC, JDW, CIL, and YCK. All authors critically revised the manuscript and agreed on the final version.

Acknowledgements:

Special thanks to Dr. Winda Liviya Ng, Dr. Yen-Hsiang Wang and Dr. Guo-Dong Chung for their contributions in proofreading as well as brainstorming. The reference format assistance of Ms. Flora Fang, and English editing by Ms. Lynn Lai are highly appreciated.

References

1. Concha-Barrientos M, Nelson DI, Fingerhut M, Driscoll T, Leigh J. The global burden due to occupational injury. *Am J Ind Med*. 2005;48(6):470-81.
2. Takala J, Hamalainen P, Saarela KL, Yun LY, Manickam K, Jin TW, et al. Global estimates of the burden of injury and illness at work in 2012. *J Occup Environ Hyg*. 2014;11(5):326-37.

3. Azaroff LS, Levenstein C, Wegman DH. Occupational injury and illness surveillance: conceptual filters explain underreporting. *Am J Public Health*. 2002;92(9):1421-9.
4. Wolfe D, Fairchild AL. The need for improved surveillance of occupational disease and injury. *Jama*. 2010;303(10):981-2.
5. Wu TN, Liou SH, Wang JD, Shen CY, Ko KN, Yang GY, et al. Establishment of a work-related diseases surveillance system in Taiwan, Republic of China. *Prev Med*. 1996;25(6):725-9.
6. Wu TS, Shih FY, Yen MY, Wu JS, Lu SW, Chang KC, et al. Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan. *BMC Public Health*. 2008;8:18.
7. Chu PC, Fuh HR, Luo JC, Du CL, Chuang HY, Guo HR, et al. The impact of occupational health service network and reporting system in Taiwan. *Int J Occup Environ Health*. 2013;19(4):352-62.
8. Wu TN, Liou SH, Shen CY, Hsu CC, Chao SL, Chang PY. Occupational disease surveillance in Taiwan. *Lancet*. 1996;348(9030):827.
9. Sayhan MB, Sayhan ES, Yemenici S, Oguz S. Occupational injuries admitted to the emergency department. *J Pak Med Assoc*. 2013;63(2):179-84.
10. Mascarenhas MD, de Freitas MG, Monteiro RA, da Silva MM, Malta DC, Gomez CM. Emergency room visits for work-related injuries: characteristics and associated factors - capitals and the Federal District, Brazil, 2011. *Cien Saude Colet*. 2015;20(3):667-78.
11. Marsh SM, Reichard AA, Bhandari R, Tonozzi TR. Using emergency department surveillance data to assess occupational injury and illness reporting by workers. *Am J Ind Med*. 2016;59(8):600-9.
12. Ong MS, Magrabi F, Coiera E. Automated categorisation of clinical incident reports using statistical text classification. *Qual Saf Health Care*. 2010;19(6):e55.
13. Conway M, Dowling JN, Chapman WW. Using chief complaints for syndromic surveillance: a review of chief complaint based classifiers in North America. *J Biomed Inform*. 2013;46(4):734-43.
14. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *Jama*. 2017;318(22):2211-23.
15. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8.
16. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *Jama*. 2017;318(14):1377-84.
17. Lopez Pineda A, Ye Y, Visweswaran S, Cooper GF, Wagner MM, Tsui FR. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *J Biomed Inform*. 2015;58:60-9.
18. Luther SL, McCart JA, Berndt DJ, Hahm B, Finch D, Jarman J, et al. Improving identification of fall-related injuries in ambulatory care using statistical text mining. *Am J Public Health*. 2015;105(6):1168-73.

19. Griffey RT, Pines JM, Farley HL, Phelan MP, Beach C, Schuur JD, et al. Chief complaint-based performance measures: a new focus for acute care quality measurement. *Ann Emerg Med*. 2015;65(4):387-95.
20. Mowafi H, Dworkis D, Bisanzo M, Hansoti B, Seidenberg P, Obermeyer Z, et al. Making recording and analysis of chief complaint a priority for global emergency care research in low-income countries. *Acad Emerg Med*. 2013;20(12):1241-5.
21. jiebaR [cited 2019. Available from: <https://www.r-project.org/nosvn/pandoc/jiebaR.html>.
22. C.D. PJSRM, editor GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*; 2014; Doha, Qatar.: Association for Computational Linguistics.
23. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231-7.
24. Colubri A, Silver T, Fradet T, Retzepi K, Fry B, Sabeti P. Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients. *PLoS Negl Trop Dis*. 2016;10(3):e0004549.
25. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med*. 2010;363(6):501-4.
26. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019;380(14):1347-58.
27. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. *Jama*. 2016;316(22):2368-9.
28. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44.
29. Kiselica D, Sibson B, Green-McKenzie J. Workers' compensation: a historical review and description of a legal and social insurance system. *Clin Occup Environ Med*. 2004;4(2):v, 237-47.
30. Hsiao WC. Taiwan's path to universal health coverage-an essay by William C Hsiao. *Bmj*. 2019;367:l5979.
31. Chung PH, Cheng Y. Reporting systems for occupational injuries and diseases in Taiwan: Conditions and problems. 2010;29:561-74.
32. Bureau of Labor Insurance MoL. Labor Insurance Act 2019 [Available from: <https://www.bli.gov.tw/en/0011297.html>.
33. Herbert R, Plattus B, Kellogg L, Luo J, Marcus M, Mascolo A, et al. The Union Health Center: a working model of clinical care linked to preventive occupational health services. *Am J Ind Med*. 1997;31(3):263-73.
34. Carr LJ, Leonhard C, Tucker S, Fethke N, Benzo R, Gerr F. Total Worker Health Intervention Increases Activity of Sedentary Workers. *Am J Prev Med*. 2016;50(1):9-17.
35. Jinnett K, Schwatka N, Tenney L, Brockbank CV, Newman LS. Chronic Conditions, Workplace Safety, And Job Demands Contribute To Absenteeism And Job Performance. *Health Aff (Millwood)*. 2017;36(2):237-44.
36. Farrow A, Reynolds F. Health and safety of the older worker. *Occup Med (Lond)*. 2012;62(1):4-11.

Figures

Figure 1

(A). Flow diagram of the inclusion of the study population divided into training and testing datasets. Emergency department (ED) visits aged between 15- to 65-year-old were included at a medical center in southern Taiwan from April 2015 to March 2018. We applied events that occurred in the first 30 months as the training dataset, and those of the last 6 months as the testing dataset for evaluation of performance of the screening model based on machine learning. (B). The occupational injury screening system with natural language processing (NLP) was implemented in three stages: Word segmentation was applied to Chinese texts with an R package JiebaR, which resulted in 28 030 stop words to be omitted from the training data. Then, the word-embedding model called GloVe was used to assign each chief complaint sentence a vector. In the third step, a classification model was developed using logistic regression.

Figure 2

Trend of outpatient referral to occupational medicine and compensation rate of National Labor Insurance among those suspected cases of occupational injuries.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [STROBStatement.docx](#)
- [STROBStatement.docx](#)
- [SupplementaryTable1.docx](#)
- [SupplementaryTable1.docx](#)