The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-025-16835-7

ORIGINAL ARTICLE

Analysis of occupational accidents in the construction metal products manufacturing subsector in Spain: Trends and risk factors

José Luis Fuentes-Bargues[1] · Alberto Sánchez-Lite[2] · José Manuel Geijo-Barrientos[2] · Alvaro Romero-Barriuso[3] · Blasa Maria Villena-Escribano[3] · Cristina González-Gaya[3]

Received: 28 August 2025 / Accepted: 15 October 2025

Abstract

Research on work accidents is important to determine the causes of occupational accidents to effectively prevent them in the future and improve workplace safety. This study aims to analyse the evolution of accidents in the metal products manufacturing subsector for construction (CNAE subsector 251) in Spain for the period 2009–2022 to classify accidents into different operational profiles. This will facilitate the proposal of specific preventive measures based on the severity and characteristics of each accident. Data for this study are collected from occupational accident reports via the Delt@ (Electronic declaration of injured workers) IT system. The study variables were classified into five groups: temporal, personal, business, circumstances, and consequences. Accidents at work are more common in males and in middle-aged workers (30–59 years). Companies with less than five workers, works outside the usual workplace and workers with less three months of length of the service in the company present high accident rate, both in light as fatal accidents. A semi-supervised model has been developed using the Fuzzy Cluster algorithm that can detect serious accidents with a recall rate of approximately 64% and group them into three distinct categories. This makes it possible to propose specific preventive measures for each category, of which there are 12 in total.

Keywords Accidents rate · Health & safety · Spain · Metal products · Construction

1 Introduction

An occupational accident is defined as an event occurring during the working time, resulting in a non-fatal injury with loss of working time or a fatal injury. According to reports from the International Labour Organization (ILO), every 15 s, a worker dies due to a work-related accident or disease [1]. These incidents carry a significant cost for families, employers, and society as a whole [2, 3].

Research on work accidents is important to determine the causes of occupational accidents to effectively prevent them in the future and improve workplace safety [4]. This kind of research must include the analysis of factors such as the technical and organizational conditions of companies, the adaptation of jobs to employees, and workers' attitudes toward occupational safety and health [5]. It is an essential first step in designing and implementing appropriate preventive measures to avoid similar accidents [6, 7].

The metal sector includes economic activities such as metallurgy and the production of metallic components. It covers a wide range of processes, including the manufacturing of basic ferrous and nonferrous metal products, metal casting, material coating, and the production of tools and machinery. The metal sector is one of the productive sectors with the highest accident rate, both at the world, European and national level, and in which workers are exposed to the worst and most dangerous working conditions [8, 9].

✉ José Luis Fuentes-Bargues
jofuebar@dpi.upv.es

1 Project Management, Innovation and Sustainability Research Center (PRINS), Universitat Politècnica de València, Valencia 46022, Spain

2 Department of Materials Science and Metallurgical Engineering, Graphic Expression in Engineering, Cartographic Engineering, Geodesy and Photogrammetry, Mechanical Engineering and Manufacturing Engineering, School of Industrial Engineering, Universidad de Valladolid, P° del Cauce 59, Valladolid 47011, Spain

3 Construction and Manufacturing Engineering Department, National Distance Education University (UNED), C/Juan del Rosal 12, Madrid 28040, Spain

Published online: 03 November 2025

As shown in the latest studies and research carried out in Spain [10, 11], among the subsectors that make up the metal sector, the manufacturing of fabricated metal products for construction accounts for 30% of total and light accidents, 35% of serious and very serious accidents and 23% of fatal accidents. These important values require a more detailed analysis, which is the objective of this article. The aim is to explore the evolution and the severity of occupational accidents in the subsector of manufacturing fabricated metal products for construction (subsector 251 of the National Classification of all Economic Activities CNAE [12]) and analyse the relationship between the main key factors. The results will make it possible to understand the characteristics of these accidents and to draw up preventive guidelines to improve safety conditions at manufacturing plants of fabricated metal products for construction. Based on the descriptive analysis of the data and through the development of a classification model, the objective of the present study is to categorise accidents into distinct operational profiles. The classification system is intended to facilitate the proposal of specific preventive measures under the severity and characteristics of each accident.

2 Materials and methods

The data for this analysis are based on accidents in the subsector of manufacturing metal products for construction in Spain during the 2009–2022 period. This data comes from occupational accident reports, which are required to be sent to the relevant administrative body via the Delt@ (Electronic declaration of injured workers) IT system. The information provided in occupational accident reports is structured following Act TAS/2926, 21 November 2002 [13]. The data used in this research refer to the number of accidents that occurred, not to the incidence rates, as data on the number of workers according to each of the variables studied are not available. In the data used, the accidents" in itinere" (accidents to and from home to the workplace) were not considered. The study variables were selected from the official occupational accident forms according to similar studies [14–16]. The variables were grouped into five groups: temporal, personal, business, circumstances, and consequences.

The methodology applied is shown in Fig. 1. The initial phase was the establishment of a foundation of data. The study commenced with a cleaned statistical dataset, which was obtained by removing incomplete records and ensuring variable consistency. A descriptive statistical analysis was then conducted for the purpose of characterising the dataset and identifying preliminary patterns (phase 1 – Data Groundwork). The selection of relevant variables was conducted based on statistical and domain criteria. Subsequently, an unsupervised fuzzy clustering model (Fuzzy C-Means) was trained using 70% of the randomly selected samples. After the identification of a pivotal variable during this phase, a semi-supervised model was developed to refine pattern recognition (phase 2 – Knowledge Modelling). The remaining 30% of the samples were used to validate the semi-supervised model, assessing its performance against the initial unsupervised model and established reference metrics (phase 3 – Insight Validation). All statistical calculations were performed using IBM SPSS Statistics software (Version 30) [17]. KNIME software (Version 5.4.3) [18] was used to develop the models.

In the descriptive analysis, a classification has been made according to severity, distinguishing between total, light, serious, very serious, and fatal accidents. The different accident rates are determined and expressed in percentages. The total accident rate (TAR) was obtained by dividing the number of total accidents of item i by the number of total

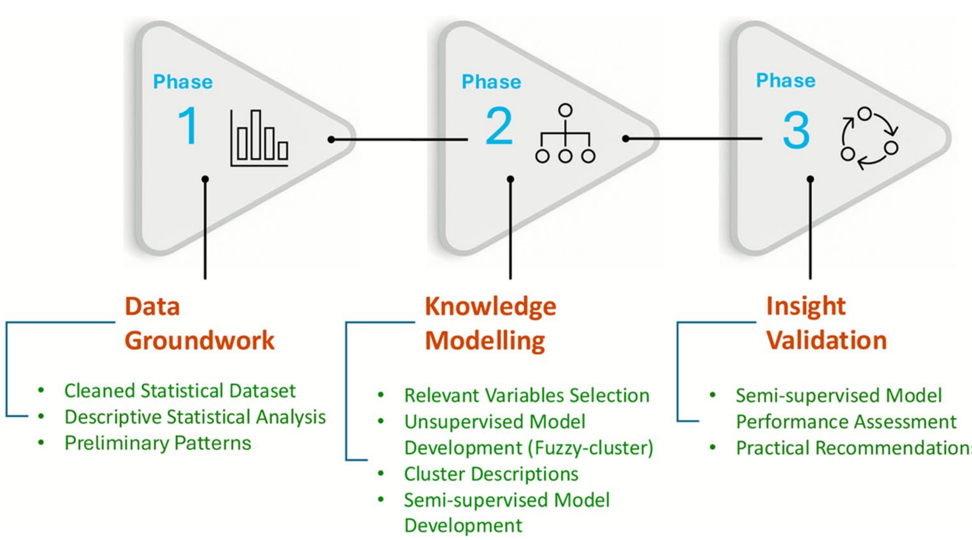

Fig. 1 Methodological Framework

accidents in the subsector. The light accident rate (LAR) was obtained by dividing the number of light accidents of item i by the number of total light accidents in the subsector. The serious accident rate (SAR) was obtained by dividing the number of serious accidents of item i by the number of total serious accidents of the subsector. The very serious accident rate (VSAR) was obtained by dividing the number of very serious accidents of item i by the total number of very serious accidents of the subsector. Lastly, the fatal accident rate (FAR) was obtained by dividing the number of fatal accidents of item i by the total number of fatal accidents of the subsector. In this phase, accidents that have occurred in the study sector in countries with different income levels have been contextualised.

3 Results

3.1 Data groundwork

To contextualise the situation of the metal sector in Spain, it is useful to compare it with recent international data from countries with different levels of economic income. As illustrated in Table 1, the data set encompasses a range of income levels, including low, medium- and high-income countries. It provides a comprehensive overview of accident incidence rates, annual prevalence, and the primary causative factors for accidents in specific subsectors such as steelmaking, primary metals, and the manufacture of metal products.

In the period of study, a total de 100,805 accidents occurred in Spain, 99,856 lights, 852 serious, 28 very serious, and 69 fatal. Figure 2 shows the evolution of total accidents in the subsector 251 between 2009 and 2022. The initial reduction due to the improvement of prevention policies from the beginning of the series until the abrupt reduction of accidents because of the economic crisis in the Spanish economy (2012–2014) and the gradual increase of accidents with the later recovery. The decrease in 2020 is explained by the confinement and reduction of activity due to COVID-19.

Table 2 shows the importance from the point of view of the accidents of the subsector of manufacturing of fabricated products for construction within the metal sector. Approximately 30% of the total accidents occur in this subsector, 33% of the serious and very serious accidents, and near of 21% of the fatal accidents.

Table 1 Accidents in the metal sector according to income level in different countries

Economic level (Income)	Country	Metal subsector	Indicators//Period//Relative vs. other sectors	Ref.
Lower-middle	Vietnam	Large-scale metal manufacturing	6,2% of all occupational accidents and 6% of all workplace deaths (749 fatal accidents and 1,485 serious injuries)//2021 (pub. 2023)//High	[19]
		National OSH policy – metal sector	Target: reduce fatal accidents in metallurgy by 5% annually. Policy target; fatal accident reduction in metal sector//Policy to 2020//High	[20]
	Tanzania	Iron & steel (factories)	12-month injury prevalence = 65.1% workers with ≥ 1 injury causing ≥ 1 day off. High injury prevalence (iron & steel)//2022 (pub. 2025)//High	[21]
Upper-middle	China	Metallurgical industry	152 fatal accidents (2001–2018), 731 deaths; ~40% poisoning/asphyxiation//2001–2018 (pub. 2020)//High	[22]
		Heavy industry (metals included)	> 10,000 new occupational disease cases; metals among highest-risk sectors. poisoning/asphyxiation dominance//1997–2007 (pub. 2010)//High	[23]
High	United States	Primary metal manufacturing	TRC = 3.4 per 100 FTE; DART = 1.9; DAFW = 1.0; DJTR = 1.0//2023 (Released 2024)//Medium-high	[24]
		Fabricated metal product manufacturing	TRC = 3.2 per 100 FTE; DART = 1.8; DAFW = 1.1; DJTR = 0.8//2023 (Released 2024)//Medium-high	[24]
	EU-27	Manufacturing (incl. basic metals & fabricated metals, NACE C24–C25)	Main causes: losing control of machines/tools/handling equip. = 20.3% (non-fatal) & 27.4% (fatal) of all accidents. Machine/handling-related accidents dominant//2022 (page updated 2024)//Medium-high	[25]

TRC (Total recordable cases); FTE (Full-Time Equivalent); DART (Days away from work, job restriction or transfer); DAFW (Days away from work); DJTR (Days of job transfer of restriction)

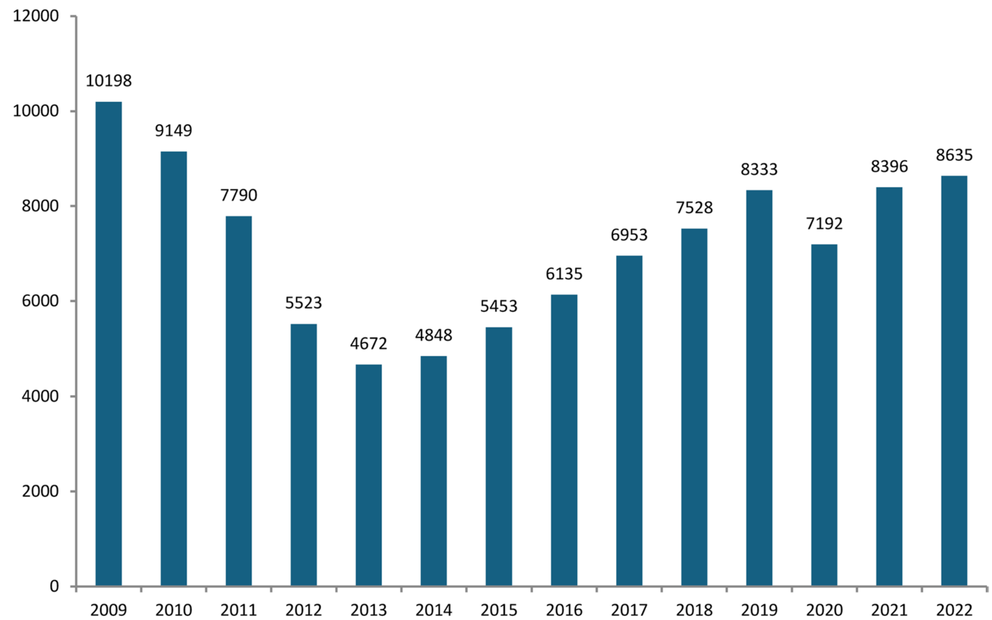

Fig. 2 Evolution of the total accidents of the subsector of manufacturing of fabricated metal products for construction between 2009 and 2022

Table 2 Comparison of the number of accidents between the subsector of manufacturing fabricated metal products for construction and the metal sector in the period 2009–2022

Year	Total			Serious + Very Serious			Fatal		
	Subsector 251	Metal Sector	%	Subsector 251	Metal Sector	%	Subsector 251	Metal Sector	%
2009	10,198	32,319	31.6	79	238	33.2	4	18	22.2
2010	9,149	29,534	31.0	79	229	34.5	8	29	27.6
2011	7,790	19,651	39.6	78	213	36.6	6	33	18.2
2102	5,523	19,585	28.2	46	140	32.9	2	10	20.0
2013	4,672	17,716	26.4	45	156	28.9	1	18	5.6
2014	4,848	18,204	26.6	51	149	34.2	4	22	18.2
2015	5,453	19,950	27.3	61	173	35.3	3	18	16.7
2016	6,135	21,461	28.6	55	150	36.7	6	15	40.0
2017	6,953	22,741	30.6	54	163	33.1	1	12	8.3
2018	7,528	27,854	27.0	58	218	26.6	8	26	30.8
2019	8,333	29,929	27.8	80	222	36.0	5	25	20.0
2020	7,192	23,958	30.0	71	201	35.3	10	38	26.3
2021	8,396	27,327	30.7	62	190	32.6	6	32	18.8
2022	8,635	26,872	32.1	51	199	25.6	5	33	15.2
Mean			29.8			33.0			20.6

3.1.1 Temporal variables

Table 3 shows a distribution of fatal accidents according to the time variables: day of the week, daytime, and time of the working day. There are a higher number of accidents on Mondays (22.5%), with a slight decrease during the week. At weekends there are almost no accidents because the activity of this industry is practically reduced. The time slots with the highest number of accidents are between 10:00 and 12:59 (40.9% of the total accidents), coinciding with the time after meal break. Many accidents occur during the second (17,513; 17.4%), third (16,001; 15.9%), and fourth hour (15,167; 15.0%) of the working day respectively, although there are a considerable number of accidents at the beginning of the working day (11,543; 11.5%). Also, it can be concluded that this is not a sector with a lot of overtime because the accident rate is drastically reduced from the eighth hour of the working day.

In the case of fatal accidents, the highest number of accidents occurred on Thursdays (18; 26.1%), and Wednesdays (17; 24.6%), while Mondays, in contrast to the total number of accidents, had the lowest accident rate of the week with only 7 fatal accidents. On weekends there was only one fatal accident. About the daytime, the highest number of fatal accidents

Table 3 Distribution of total accidents of the subsector of manufacturing fabricated metal products for construction in the period 2009–2022

Day of the week	Total Accidents		Day time	Total Accidents		Time of the working day	Total Accidents	
	N	%		*N*	%		*N*	%
	100,805	100.0		100,805	100.0		100,805	100.0
Monday	22,650	22.5	8:00 to 8:59	5,934	5.9	1	11,543	11.5
Tuesday	21,027	20.9	9:00 to 9:59	9,088	9.0	2	17,513	17.4
Wednesday	20,199	20.0	10:00 to 10:59	14,231	14.1	3	16,001	15.9
Thursday	18,068	17.9	11:00 to 11:59	12,815	12.7	4	15,167	15.0
Friday	15,823	15.7	12:00 to 12:59	14,209	14.1	5	10,393	10.3
Saturday	2,586	2.6	13:00 to 13:59	7,206	7.1	6	10,860	10.8
Sunday	452	0.4	14:00 to 14:59	3,606	3.6	7	10,122	10.0
			15:00 to 15:59	3,759	2.7	8	5,698	0.9
			16:00 to 16:59	6,878	6.8	9	883	0.8
			17:00 to 17:59	7,673	7.6	10	757	0.6
			18:00 to 18:59	4,452	4.4	>10	1,868	6.8
			19:00 to 19:59	1,602	1.6			
			20:00 to 20:59	785	0.8			
			Rest of hours	8,567	8.5			

N number of accidents; % percentage of accidents

occurred in the 12:00–12:59 time slot (8; 11.6%) followed by the first hour of the day (9; 13.0%). Regarding the time of the working day, the values are similar between the first six hours of day, with the highest value in the fourth hour (15; 21.7%).

3.1.2 Personal variables

According to the variable gender, 99.1% of light accidents occurred in men (97,913), and only 1,043 accidents in women (0.9%). It is noteworthy that there were no very serious accidents or fatalities among women in the entire study period.

Table 4 shows the distribution of accidents by age and severity. The age ranges with the highest accident rates are between 30 and 39 (32.6%) and between 40 and 49 (27.8%). On the other hand, for fatal accidents, the age groups are reversed, with 22 accidents in the 40–49 age group followed by 18 accidents in the 30–39 age group.

3.1.3 Business variables

Table 5 shows the distribution of fatal accidents according to the company staff. Micro enterprises (up to 5 workers) have the highest accident rates in all fields (TAR 25.4%; LAR 25.3%; SAR 38.1% and FAR 31.9%). Approximately 79.4% of total accidents and 84% of fatal accidents occurred in companies with less than 50 workers. These results reflect the profile of companies in the sector.

The distribution of accidents by severity and length of service can be observed in Table 6. The high number of accidents that occurred when workers are in the first two steps (when they have been with the company for less than 3 months) is remarkable. They account for 20.6% of total accidents, 22.4% of the serious, 25% of the very serious, and 20.3% of fatal accidents.

Table 4 Accidents in the subsector of manufacturing fabricated metal products for construction by age between 2009 and 2022

Age (years)	Total		Light		Serious		Very Serious		Fatal	
	N	TAR(%)	*N*	LAR(%)	*N*	SAR(%)	*N*	VSAR(%)	*N*	FAR(%)
	100,805	100	99,856	100	852	100	28	100	69	100
16–19	1,517	1.5	1,501	1.5	16	1.9	0	0.0	0	0.0
20–29	19,259	19.1	19,136	19.2	114	13.4	2	7.1	7	10.1
30–39	32,902	32.6	32,662	32.7	215	25.2	7	25.0	18	26.1
40–49	27,976	27.8	27,698	27.7	245	28.8	11	39.3	22	31.9
50–59	15,956	15.8	15,733	15.8	202	23.7	6	21.4	15	21.7
60–69	3,174	3.1	3,106	3.1	59	6.9	2	7.1	7	10.1
> 70	21	0.0	20	0.0	1	0.1	0	0.0	0	0.0

N number of accidents; % percentage of accidents

Table 5 Accidents in the subsector of manufacturing fabricated metal products for construction by company staff between 2009 and 2022

Company staff (number of workers)	Total		Light		Serious		Very Serious		Fatal	
	N	TAR(%)	*N*	LAR(%)	*N*	SAR(%)	*N*	VSAR(%)	*N*	FAR(%)
	100,805	100	99,856	100	852	100	28	100	69	100
≤5	25,616	25.4	25,255	25.3	325	38.1	14	50.0	22	31.9
6–10	15,804	15.7	15,635	15.7	153	18.0	3	10.7	13	18.8
11–25	23,536	23.3	23,345	23.4	172	20.2	6	21.4	12	17.4
26–50	15,153	15.0	15,038	15.1	102	12.0	2	7.1	11	15.9
51–100	9,838	9.8	9,787	9.8	47	5.5	0	0.0	4	5.8
100–250	7,105	7.0	7,057	7.1	42	4.9	1	3.6	5	7.2
>250	3,753	3.7	3,738	3.7	11	1.3	2	7.1	2	2.9

N number of accidents; % percentage of accidents

Table 6 Accidents in the subsector of manufacturing fabricated metal products for construction by length of service and severity between 2009 and 2022

Length service	Total		Light		Serious		Very Serious		Fatal	
	N	TAR(%)	*N*	LAR(%)	*N*	SAR(%)	*N*	VSAR(%)	*N*	FAR(%)
	100,805	100	99,856	100	852	100	28	100	69	100
<1 month	7,668	7.6	7,569	7.6	85	10.0	6	21.4	8	11.6
1–3 months	13,100	13.0	12,987	13.0	106	12.4	1	3.6	6	8.7
4–12 months	17,975	17.8	17,819	17.8	142	16.7	6	10.7	8	11.6
1–2 years	11,133	11.0	11,046	11.1	74	8.7	3	10.7	10	14.5
3–4 years	13,495	13.4	13,379	13.4	103	12.1	4	14.3	9	13.0
5–10 years	19,392	19.2	1,230	19.3	149	17.5	3	10.7	10	14.5
11–30 years	16,494	16.4	16,302	16.3	173	20.3	5	17.9	14	20.3
>30 years	1,548	1.5	1,524	1.5	20	2.3	0	0.0	4	5.8

N number of accidents; % percentage of accidents

3.1.4 Circumstances variables

Regarding the accident location, 90.3% of light accidents occurred at the usual workplace, while 4.3% were moving between workplaces and 5.3% were working at other workplaces. On the other hand, for serious, very serious, and fatal accidents, this proportion changes considerably. In the usual workplace, 73.2% were serious, 42.9% very serious, and 52.2% fatal accidents, while 8.6% of serious, 28.6% of very serious, and 26.1% of fatal occurred when workers moved between areas. In other workplaces occurred 18.2% of the serious, 28.6% of the very serious, and 21.7% of fatal accidents.

Table 7 shows the distribution of the accidents by deviation and severity. Deviation describes the abnormal occurrence that has adversely interfered with the normal process of work performance and has led to the occurrence or origin of the accident. The "Loss of machine control" deviation followed by "Involuntary body movement" presents the highest number of light accidents (26.2% and 25.9% respectively). Regarding the serious, very serious and fatal accidents the main deviation is "Falls involving people" (30.6%

Table 7 Accidents in the subsector of manufacturing fabricated metal products for construction by deviation and severity between 2009 and 2022

Deviation	Total		Light		Serious		Very Serious		Fatal	
	N	TAR(%)	*N*	LAR(%)	*N*	SAR(%)	*N*	VSAR(%)	*N*	FAR(%)
	100,805	100	99,856	100	852	100	28	100	69	100
Electricity, explosion, fire	934	0.9	917	0.9	13	1.5	1	3.6	3	4.3
Falling object	5,135	5.1	5,116	5.1	19	2.2	0	0.0	0	0.0
Fall, slide	9,919	9.8	9,751	9.8	153	18.0	5	17.9	10	14.5
Loss of machine control	26,373	26.2	26,131	26.2	222	26.1	5	17.0	14	20.3
Falls involving people	10,149	10.1	9,859	9.9	261	30.6	11	39.3	18	26.1
Voluntary body movement	17,687	17.5	17,598	17.6	86	10.1	0	0.0	3	4.3
Involuntary body movement	25,925	25.7	25,879	25.9	44	5.2	2	7.1	0	0.0
Shock or jolting action	409	0.4	401	0.4	7	0.8	1	3.6	0	0.0
No information	4,275	4.2	4,204	4.2	47	5.5	3	10.7	21	30.4

N number of accidents; % percentage of accidents

LAR, 39.3% VSAR and 26.1% FAR) followed by "Loss of machine control" and "Fall, slide". Between the three deviations, they account for 74.7% of serious accidents, 74.2% of very serious, and 60.9% of fatal accidents. Also noteworthy is the 30.4% of fatal accidents within the 'Others' deviation, usually due to a lack of information.

3.1.5 Consequences variables

Table 8 shows the distribution of the accidents by injury and severity. "Wounds, superficial injuries" together with "Dislocations, sprains and strains" were the main injuries caused by light accidents (46.8% and 36.1% respectively). Regarding serious accidents, "Crushed bones" was the highest injury with 48.2% of the total. Very serious accidents caused injuries mainly in "Concussions and internal injuries" (32.1%), "Multiple lesions" (28.6%), and "Crushed bones" (25.0%), Finally, the death of workers were caused mainly by "Multiple lesions" (39.1%), "Hearth attacks, strokes and other nontraumatic diseases" (30.4%), and "Concussions and internal injuries" (23.2%). "Multiple lesions" refer to cases where the victim suffers two or more types of injuries of similar severity, which may explain how multiple injuries sustained after an accident can led to death. The group "Heart attacks, strokes, and other nontraumatic diseases" includes strictly natural causes resulting from a specific health condition, such as infarction, stroke, ictus, fainting, sudden low blood pressure, retinal detachment, etc.

3.2 Knowledge modelling

For the fuzzy cluster analysis (Fuzzy C-Means), a subset of variables was selected based on theoretical, technical and statistical criteria. The selection process was designed to ensure compatibility with the algorithm, the discriminatory capacity between observations, and the interpretability of the results. The Fuzzy C-Means algorithm is grounded in the principle of minimising the Euclidean distance between points and cluster centres, with the weighting determined by fuzzy membership degrees [26]. Consequently, it necessitates continuous numerical variables or variables converted to that format as input, and its performance is adversely impacted by the presence of poorly coded or highly correlated categorical variables [27]. For the development of the models, accident records with missing data in any of the variables were eliminated, working with a sample of 100,206 records.

The variables were then analysed to see if there was a correlation between them (Fig. 3). Following the variable selection criteria described above, the following variables were selected: Gender, Company Staff, Day of Week, Time of the Day and Accident Location.

Length of service could estimate work experience level. Despite its moderate correlation with Age, it was retained as the two variables capture different effects (accumulated experience versus stage of life). Accident Location acts as a proxy for the organisational context in which the accident occurred, which could introduce redundancy. The Day of the Week has been converted to a circular format using trigonometric functions. This transformation preserves the cyclical nature of the variable and allows for adequate representation in metric spaces [28]. Time of the Day captures intra-daily temporal patterns. Although age (AGE) is not highly correlated with the others ($\rho \approx 0.32$ with Length of service) and has discriminatory capacity within the population, it has been discarded due to severe imbalance in the data, with very little data for women. Time of the working day was excluded due to its high correlation with Time

Table 8 Accidents in the subsector of manufacturing fabricated metal products for construction by injury and severity between 2009 and 2022

Injury	Total		Light		Serious		Very Serious		Fatal	
	N	TAR(%)	*N*	LAR(%)	*N*	SAR(%)	*N*	VSAR(%)	*N*	FAR(%)
	100,805	100	99,856	100	852	100	28	100	69	100
No information	470	0.5	465	0.5	5	0.6	0	0.0	0	0.0
Wounds, superficial injuries	46,864	46.5	46,748	46.8	116	13.6	0	0.0	0	0.0
Crushed bones	7,910	7.8	7,491	7.5	411	48.2	7	25.0	1	1.4
Dislocations, sprains, and strains	36,043	35.8	36,007	36.1	36	4.2	0	0.0	0	0.0
Amputations	395	0.4	320	0.3	73	8.6	1	3.6	1	1.4
Concussions and internal injuries	5,055	5.0	4,971	5.0	59	6.9	9	32.1	16	23.2
Burns	1,224	1.2	1,208	1.2	16	1.9	0	0.0	0	0.0
Poisonings and infections	119	0.1	117	0.1	2	0.2	0	0.0	0	0.0
Drowning and asphyxiation	88	0.1	86	0.1	1	0.1	0	0.0	1	1.4
Effects of noise. Vibration and pressure	153	0.2	153	0.2	0	0.0	0	0.0	0	0.0
Extreme temperature Effects	378	0.4	372	0.4	3	0.4	1	3.6	2	2.9
Multiple lesions	784	0.8	652	0.7	97	11.4	8	28.6	27	39.1
Heart attacks, strokes, and other nontraumatic diseases	94	0.1	48	0.0	23	2.7	2	7.1	21	30.4
Other injuries	1,228	1.2	1,218	1.2	10	1.2	0	0.0	0	0.0

N number of accidents; % percentage of accidents

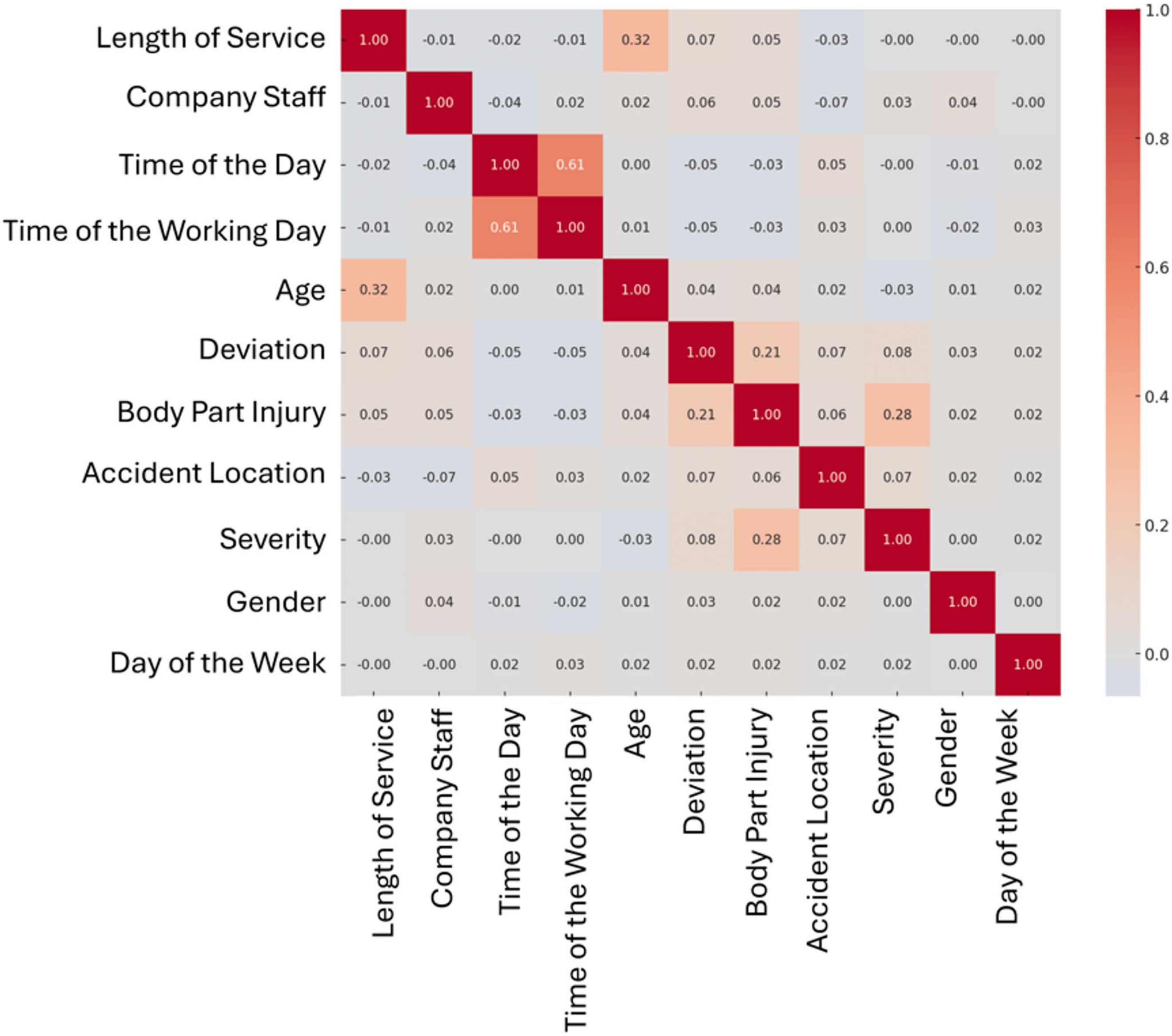

Fig. 3 Correlation matrix of variables (Coefficient of Spearman for categorical variables)

of Day ($\rho \approx 0.61$). However, categorical variables such as Severity, Deviation, Injury and Body Part Injured were excluded due to their nominal, non-ordinal nature and the complexity that their one-hot encoding would introduce into a distance-based algorithm [29]. These variables could be incorporated in a subsequent phase, as elements of interpretation or external validation of the clusters obtained. This baseline set provides a balanced representation of individual characteristics (age, gender), work characteristics (seniority, size of the centre), and temporal characteristics (time, day), which is particularly relevant for generating diffuse risk or behaviour profiles.

Regarding the treatment of categorical variables included in the model, the Accident Location variable was coded into three levels (usual workplace, moving between workplaces, and other workplace) using the one-hot coding method. The Day of the week variable can be considered as an ordinal categorical. To simplify its coding, circular coding was used, resulting in two variables: Day1 and Day2. Finally, the severity variable was grouped into two levels (Non-critical and Critical). The Critical level includes fatal, very serious, and serious accidents. On the other hand, the No-Critical level includes light accidents.

The unsupervised model was created using the Fuzzy C-Means (FCM) algorithm. A partition of 70% of the data was made, considering the distribution of the data according to the severity variable. Given that the dataset is highly unbalanced, the SMOTE algorithm was applied to increase

Table 9 Cluster centroid values and cluster descriptions

Cluster	Length of service	Company Staff	Time of the Day (h)	Age	Other Workplace	Usual Workplace	Between Workplaces	Day	Predicted Severity
Noise Cluster	91.96	144.15	9:44	42	39.20%	37.50%	23.30%	Tuesday (Day1 = 0.377 Day2 = 0.635)	Critical or No-Critical
Cluster 0	55.58	39.29	12:07	40	0%	100%	0%	Tuesday (Day1 = 0.954 Day2 = 0.203)	No Critical
Cluster 1	55,21	35.12	12:00	39	0%	100%	0%	Monday (Day1 = 0.196 Day2 = 0.903)	Critical

Table 10 Confusion matrix for the semi-supervised model. Recall, precision, and F1-score for the training and the validation samples

Sample	Observed	Predicted		RECALL	PRECISION	F1-SCORE
		Critical	No Critical			
Training	**Critical**	42,423	27,476	60.69%	51.75%	0.559
	No Critical	39,549	30,350	43.42%	52.49%	0.475
	Overall	50.00%	50.00%	52.06%	52.12%	0.517
Validation	**Critical**	182	103	63.86%	1.06%	0.021
	No Critical	16,994	12,963	43.27%	99.21%	0.603
	Overall	0.94%	99.06%	23.44%	50.22%	0.196

the number of minority cases (corresponding to the Critical level of the severity variable) and obtain a balanced sample. The fuzzy cluster grouping was performed with an average fuzzification coefficient (m = 2) and a delta parameter of 0.5 as the convergence criterion. The clustering was considered for the formation of two clusters, allowing samples with a lower degree of membership in either of the two clusters to be included in a noise cluster. Table 9 shows the composition of the centroids for the three clusters, and the description of each of the clusters, as well as their relationship with the severity variable.

The semi-supervised model was created using the relationship between the samples in each cluster and the severity variable. The samples from the Noise cluster were included in Cluster 1 or Cluster 0 according to their degree of belonging to one cluster or the other (approximately 30% of observations have this cluster as dominant, which allows this percentage of data to be recovered according to its proximity to cluster 1 or cluster 0), then Cluster 1 and Cluster 2 were assigned a Critical level. The assigned levels were critical if they classified most samples at this level and a No-Critical level if they did not. Using this criterion, clusters 1 and 0 were classified as Critical and No-Critical, respectively. This enabled each sample to be classified according to its cluster membership and severity level. Table 10 shows the number of samples that were correctly classified as Critical or No Critical. Examining the values reveals that, while the model does allow samples with critical severity to be detected (RECALL close to 61%), it is not very accurate in its classification.

3.3 Insight validation

Thirty per cent of the samples were used to validate the semi-supervised model. In this case, no balancing technique was employed, and a severely imbalanced sample (few critical cases) was used. Table 10 shows the obtained results. Although the model can correctly detect 65% of the critical cases in the sample (182 out of 285), it also detects many false positives, reducing its accuracy. On the other hand, if the model characterises a sample as No-Critical, the probability that it is No-Critical is high.

Considering the critical case detection capability of the developed model and the characteristics of the samples associated with each cluster, the preventive measures described in Table 11 can be recommended.

Table 11 Preventive measures for each cluster

Cluster	Practical Recommendations
Noise Cluster	Logistics and move planning Training in uncontrolled environments Contractor management Early schedule control
Cluster 0	Reinforcement of breaks and rotation Routine safety checks Safe working protocols at fixed workstations Preventive leadership
Cluster 1	Weekly kick-off meetings Progressive task planning Motivational reinforcement Initial equipment checks

4 Discussion

The findings reveal that, whilst the absolute figures and accident mechanisms fluctuate according to the level of development, the metal sector consistently occupies the top position in terms of relative risk when benchmarked against other productive sectors (Table 1). This phenomenon has been observed in both industrialised economies, where significant injuries and amputations resulting from machinery are prevalent [25], and in low-wage countries, where annual prevalence rates occasionally exceed 50% of the workforce [21]. A recent global meta-analysis estimates an average annual prevalence of 55% in the iron and steel industries, with consistent risks across all income levels [21]. Data from the World Steel Association shows dozens of fatal accidents and significant rates in the steel industry each year, confirming that risks persist even in environments with advanced safety policies [30]. A comparison with Spanish data will facilitate an assessment of the extent to which global trends are replicated at the national level, and whether current prevention policies are aligned with the challenges identified internationally.

In Spain, the subsector of manufacturing of fabricated metal products for construction has a great economic and employment importance within the metal sector, and this importance also has negative consequences, such as the occupational accident rate. It represents 30% of total accidents, 33% of serious and very serious accidents and 21% of fatal accidents in the metal sector.

Mondays present the highest rate (not by far) of total and light accidents, but not of fatal accidents, so it could be talked of light "Monday effect"[16]. Another noteworthy aspect is that accidents at weekends are drastically reduced, which, together with the reduction in the number of accidents after the eighth hour of the working day, indicates that overtime is not common in the sub-sector.

During the daytime, 40.9% of the accidents occurred after a meal break, results similar as other studies in the metal sector and in other sectors [11, 16]. About the time of the working day, between the start of the working day and the following three hours, up 59.8% of total accidents occur. It is somewhat contradictory because workers should be fresher because they have been working fewer hours. Some authors suggest that this could be due to failures in the routine tasks at the start of the working day [31]– [32].

The study reflects that the accidents occurred mainly in men, results expected by the characteristics of workers in the sector and in line with previous studies [10]– [11]. The age ranges with the highest accident rates are between 30 and 39, and between 40 and 49, results according to previous studies in metal sector but with difference regarding the fatal accidents, 40 and 49 years present more cases than the 50 and 59 age range, highest at the whole metal sector.

Small companies, the most common due to the characteristics of the Spanish economy, have the highest number of accidents, both light as serious and fatal. Accidents involving workers with less than three months' experience in the company account for approximately 20% of the total, serious, very serious, and fatal accidents. When workers move between workplaces or carry out their work in other workplaces, the number of serious, very serious, and fatal accidents increases considerably in comparison with the proportion of accidents in the usual workplace. These three aspects are significant for the design of training and information actions both by the prevention services of the companies as well as by the prevention services of the companies.

The risk assessments of workstations in subsector 251 should be reviewed to try to avoid deviations associated with "Loss of machine control", "Falls involving people" and "Fall, slide", which are the most frequent causes of fatal accidents.

Analysis of the centroids reveals significant differences in the operational and risk profiles of the three clusters. The Noise Cluster comprises larger companies with longer-serving employees and a higher frequency of accidents in the morning and on days close to the start of the week. The distribution of accidents by location is heterogeneous. This cluster has the highest ratio of critical to non-critical accidents of the three clusters (64.4% of accidents are critical). Cluster 0 comprises medium-sized companies whose employees have less experience. It comprises accidents that mainly occur at midday between Tuesday and Wednesday. All accidents occur in the usual workplace. This cluster contains a lower proportion of critical accidents than non-critical accidents (42.8% of cases are critical). Cluster 1 has a similar profile to Cluster 0 in terms of size and seniority. However, accidents in this cluster are concentrated on Mondays at midday in the usual workplace. This cluster also has a slightly higher percentage of critical cases (44.3%). Cluster analysis groups accidents into different categories to propose appropriate preventive measures. It also reveals the latent structure presented by the descriptive analysis.

The validation sample is very unfavourable, as most of the cases are in the noise cluster and the data is highly imbalanced if we compared with the training sample (data of the noise cluster in validation sample is approximately 24% higher than in the training sample). However, the model's performance with the validation sample is similar to that with the training sample, suggesting that it does not have serious overfitting problems.

The main limitation of this study is the lack of worker-specific data in accident reports, such as education level, experience, and health and safety training. This information could facilitate a more detailed analysis of cases and contribute to the development of more effective prevention policies within companies, as well as inform the strategies of health and safety organisations. The limitations of the semi-supervised model used are typical of this type of model and are also derived from working with unbalanced data and categorical variables. While the model has low accuracy, it can detect critical cases. This is important for risk prevention, since it is better to incorrectly assign high criticality to a non-critical case than to fail to detect a critical case. The model's accuracy could be improved by incorporating an additional context variable alongside the cluster membership information in a supervised model (logistic regression or neural network).

5 Conclusions

The article discussed the severity and the main factors of the accidents of the subsector of manufacturing fabricated metal products for construction (subsector 251) between 2009 and 2022. It represents 30% of total accidents, 33% of serious and very serious accidents and 21% of fatal accidents in the metal sector.

The first four hours of the working day, the "meal breaks", and a light "Monday effect" have the highest accident rates among the variables time of the working day, the daytime, and the day of the week. These data require improved control by managers and supervisors during these temporary periods.

Falls involving people, falls, slips, and loss of machine control are the deviations that cause the highest occupational mortality. Wounds, superficial injuries, along with dislocations, sprains, and strains, are the main injuries resulting from minor accidents, while heart attacks, strokes, and other non-traumatic diseases, followed by multiple lesions, are the main causes of fatal accidents.

A non-supervised model has been developed that groups together accidents recorded in previous profiles for possible preventive action. The proposed semi-supervised model can detect critical cases (182 out of 285 critical cases in the sample were correctly classified) and can be used as an initial screening tool.

Twelve preventive measures have been proposed for the different groups of accidents, which have been classified according to severity, time of day, age of worker, experience and company size. General training actions must be carried out focus mainly in men, middle-aged workers and small companies, and specific training actions must be developed in work outside the usual workplace and for workers who will be incorporated into new companies.

To improve the model's precision and ensure its continued use, it would be advisable to study the possibility of incorporating the semi-supervised model's information alongside another variable (e.g. deviation, injury or body part injury) into a supervised model, updating the model's data monthly with company data.

Acknowledgements This paper is based on the ongoing research of the Working Group on Risk Engineering in Manufacturing (REM) of the Manufacturing Engineering Society.

Funding statement Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding for open access charge: CRUE-Universitat Politècnica de València

Data availability The data used for the research were requested from the Spanish Ministry of Labour and Social Economy. The authors do not have permission to share data.

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. International Labour Office (ILO) (2023) Safety and health at work [Online]. Available from: http://www.ilo.org/global/topics/safety-and-health-at-work/lang%2D-en/index.htm. Accessed June 2025
2. Peng Y, Zhang S, Wu P (2015) Factors influencing workplace accident costs of Building projects. Saf Sci 72:97–104. https://doi.org/10.1016/j.ssci.2014.08.008
3. Forteza FJ, Carretero-Gomez JM, Sese A (2017) Occupational risks, accidents on sites and economic performance of construction firms. Saf Sci 94:61–76. https://doi.org/10.1016/j.ssci.2017.01.003
4. Darda 'u Rafindadi A, Shafiq N, Othman I, Ibrahim A, Aliyu MM, Mikic´ M, Alarifi H (2023) Data mining of the essential causes of different types of fatal construction accidents. Heliyon 9:e13389. https://doi.org/10.1016/j.heliyon.2023.e13389
5. Perotti S, Russo MC (2018) Work-related fatal injuries in Brescia County (Northern Italy), 1982 to 2015: a forensic analysis. J Forensic Leg Med 58:122–125. https://doi.org/10.1016/j.jflm.2018.06.002
6. Anyfantis ID, Leka S, Reniers G, Boustras G (2021) Employers´ perceived importance and the use (or non-use) of workplace risk assessment in micro-sized and small enterprises in Europe with

focus on Cyprus. Saf Sci 139:105256. https://doi.org/10.1016/j.ssci.2021.105256
7. Cieslewicz W, Araszkiewicz K, Sikora P (2019) Accident rate as a measure of safety assessment in Polish civil engineering. Safety 5:77. https://doi.org/10.3390/safety5040077
8. Nenonen S (2011) Fatal workplace accidents in outsourced operations in the manufacturing industry. Saf Sci 49:1394–1403. https://doi.org/10.1016/j.ssci.2011.06.004
9. Hedlund FH (2013) Recorded fatal and permanently disabling injuries in South African manufacturing industry - overview, analysis and reflection. Saf Sci 55:149–159. https://doi.org/10.1016/j.ssci.2013.01.005
10. Díaz Aramburu C (2022) Informe de siniestralidad en el sector del metal. Instituto Nacional de Seguridad e Higiene en el Trabajo, Available on: https://www.insst.es/documents/94886/96082/Informe+de+siniestalidad+en+el+sector+del+metal.pdf/1894c648-cf42-fe7d-2e00-6301e0024edc?t=1655192187441. Accessed Oct 2024
11. Fuentes-Bargues JL, Sánchez-Lite A, González-Gaya C, Rosales-Prieto VF, Reniers G (2022) A study of situational circumstances related to Spain´s occupational accident rates in the metal sector from 2009 to 2019. Saf Sci 150:105700. https://doi.org/10.1016/j.ssci.2022.105700
12. CNAE 09 National Classification of Economic Activities in Spain. Available on: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177032&menu=ultiDatos&idp=1254735976614
13. Ministry of Labour and Social Affairs of Spain (MLSAS) (2002) Orden TAS/2926/2002, de 19 de noviembre, Por La Que se Establecen Nuevos modelos Para La notificación de Los accidentes de Trabajo y se posibilita Su transmisión Por Procedimiento electrónico. Boletín Oficial Del Estado núm 279:21
14. Camino-López MA, Ritzel DO, Fontaneda I, González-Alcantara OJ (2008) Construction industry accidents in Spain. J Saf Res 39:497–507. https://doi.org/10.1016/j.jsr.2008.07.006
15. López-Arquillos A, Rubio-Romero JC, Gibb A (2012) Analysis of construction accidents in Spain, 2003–2008. J Saf Res 43:381–388. https://doi.org/10.1016/j.jsr.2012.07.005
16. Suárez-Cebador M, Rubio-Romero JC, Carrillo-Castrillo JA, López-Arquillos A (2015) A decade of occupational accidents in Andalusian (Spain) public universities. Saf Sci 80:23–32. https://doi.org/10.1016/j.ssci.2015.07.008
17. IBM Corp, Released (2024) IBM SPSS statistics for Windows, version 30.0. IBM Corp, Armonk, NY
18. Released KNIMEAG (2025) KNIME analytics Platform, version 5.4.3, Talacker 50 8001 Zürich Switzerland
19. International Labour Organization (ILO). Vietnams machinery and equipment industry. A systemic sector assessment (2023) Available on: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40ed_emp/%40emp_ent/%40ifp_seed/documents/publication/wcms_892986.pdf
20. National Institute of Occupational & Environmental Health of Vietnam (2020) Available on: https://mhlw-grants.niph.go.jp/system/files/report_pdf/32121%20OSH%20report%20of%20Vietnam-buntan1-6.pdf
21. Shabani S, Moen BE, Deressa W, Mamuya SH (2025) Prevalence of occupational injuries among workers in the iron and steel industries in Tanzania. Ann Glob Health 91:1. https://doi.org/10.5334/aogh.4503
22. Xu Q, Xu K (2020) Statistical analysis of prediction of fatal accidents in the metallurgical industry in China. Int J Environ Res Public Health 17(11):3790. https://doi.org/10.3390/ijerph17113790
23. Zhang X, Wang Z, Li T (2010) The current status of occupational health in China. Environ Health Prev Med 15(5):263–270. https://doi.org/10.1007/s12199-010-0145-2
24. U.S. Bureau of Labor Statistics (2024) Incidence rates of nonfatal occupational injuries and illnesses by industry and case types, 2023. https://www.bls.gov/web/osh/table-1-industry-rates-national.htm. For more information, visit U.S. Bureau of Labor Statistics.
25. Eurostat (2024) Accidents at work – statistics on causes and circumstances 2022. https://ec.europa.eu/eurostat/statisticsexplained/index.php/Accidents_at_work_-_statistics_on_causes_and_circumstances
26. Bezdek JC (1981) Models for pattern recognition. Pattern recognition with fuzzy objective function Algorithms. Advanced applications in pattern recognition. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-0450-1_1
27. Ruspini EH, Bezdek JC, Keller JM (2019) Fuzzy clustering: a historical perspective. IEEE Comput Intell Mag 14(1):45–55. https://doi.org/10.1109/MCI.2018.2881643
28. Guo C, Berkhahn F (2016) Entity embeddings of categorical variables. ArXiv Preprint. Available on: arXiv:1604.06737. Accessed Aug 2025
29. Zhang T, Ramakrishnan R, Livny M (1996) BIRCH: an efficient data clustering method for very large databases. ACM Sigmod Rec 25(2):103–114
30. World Steel Association (2024) Safety and health in the steel industry – Data report 2024. https://worldsteel.org/media/publications/safety-and-health-in-the-steel-industry-data-report-2024/
31. Zermane A, Zahirasri M, Baharundin MR, Yusoff HM (2020) Analysis of the contributing factors for fatal accidents due to falls from heights in Malaysia and the USA. Sci Technol 28(S1):15–36
32. Huang X, Hinze J (2003) Analysis of construction worker fall accidents. J Constr Eng Manag 129(3):262–271. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:3(262)