Applied Ergonomics 105 (2022) 103847

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Work-related overexertion injuries in cleaning occupations: An exploration of the factors to predict the days of absence by means of machine learning methodologies

Aroa González Fuentes [a], Nélida M. Busto Serrano [b], Fernando Sánchez Lasheras [c,d], Gregorio Fidalgo Valverde [d,e], Ana Suárez Sánchez [d,e,*]

[a] School of Mining, Energy and Materials Engineering of Oviedo, University of Oviedo, Spain
[b] Labor and Social Security Inspectorate. Ministry of Labor and Social Economy, Spain
[c] Department of Mathematics, University of Oviedo, 33007, Oviedo, Spain
[d] Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), University of Oviedo, 33004, Oviedo, Spain
[e] Department of Business Management, University of Oviedo, 33004, Oviedo, Spain

ARTICLE INFO

Keywords:
Work-related overexertion injuries
Absenteeism
Musculoskeletal disorders (MSD)
Cleaning sector
Machine learning

ABSTRACT

The special characteristics of the cleaning industry have an important impact on the health and safety of its workforce. Making use of data from more than 79,000 occupational accidents, the aim of the present research is to use machine learning techniques to develop a model to predict incapacity for work (expressed in days of absence) due to work-related overexertion injuries among service sector cleaners in Spain. The severity of accidents caused by overexertion depends on several factors that can be classified into the following categories: injury typology, individual factors, employment conditions, accident circumstances and health and safety management and standards in the company.

1. Introduction

The cleaning industry has suffered from invisibility for many years, but it is now beginning to receive well-deserved recognition due to the advent of COVID-19. Cleaning is a necessity in every home and in all types of work environment. The institutional and industrial cleaning industry provides essential products and services to clean and maintain a healthy indoor environment for commercial establishments of all sizes and types, including schools, hospitals, day care centers, food service operations, office complexes and other similar establishments.

Regarding the recent context in the cleaning industry, the global pandemic has produced worse unemployment rates due to the decreasing demand of domestic maids and housekeepers, and also janitors and cleaners in several workplaces. On the other hand, the demand for cleaning in other industries, for instance the health sector, has soared. In fact, some governmental statistics have highlighted this phenomenon, which has occurred all around the world (U.S. Bureau of labor statistics, 2021).

In Europe, the cleaning sector employs more than 4 million people and has produced incomes of 120 billion euro. Accordingly, employment has increased by 15%. Among European countries, France, Germany, Italy, Spain and the United Kingdom account for 70% of the overall turnover. The European cleaning industry has particular employment conditions. Indeed, 3.7 million people work part-time, while 2.3 million people are full-time workers (European Cleaning and Facility Services Industry, 2020).

There are different work positions in the cleaning sector such as janitors, custodians, cleaners, housekeepers and maids, among others. Their activity usually involves exposure to different kinds of risks, such as those associated to the physical environment in which cleaning services are performed (Occupational Safety and Health Administration, 2021). Most of them are also affected by ergonomic-related risk factors that are strongly associated with the prevalence of Musculoskeletal Disorders (MSD) (Lim et al., 2022; Graves et al., 2002; Luenda et al., 2008; Chang et al., 2012; Wami et al., 2019).

MSD are characterized by persistent discomfort, aches or pain, and they frequently cause mobility limitations that restrict the function and productivity of the affected individual and involve nerves, tendons,

* Corresponding author. Department of Business Management, University of Oviedo, 33004, Oviedo, Spain.
E-mail addresses: UO212081@uniovi.es (A. González Fuentes), nelida.busto@mites.gob.es (N.M. Busto Serrano), sanchezfernando@uniovi.es (F. Sánchez Lasheras), gfidalgo@uniovi.es (G. Fidalgo Valverde), suarezana@uniovi.es (A. Suárez Sánchez).

https://doi.org/10.1016/j.apergo.2022.103847
Received 14 March 2022; Received in revised form 5 July 2022; Accepted 7 July 2022
Available online 30 July 2022
0003-6870/

cartilages, ligaments, joints and muscles (World Health Organization, 2021). Several studies show the janitorial workers' vulnerability to upper limb MSD, especially when they perform repetitive tasks such as sweeping, vacuuming, mopping, or scrubbing, heavier tasks like handling furniture or waste management, or maintain awkward postures (Kumar and Kumar, 2008; Johansonsson and Ljunggren, 1989; Melese et al., 2020, Choi and Chin, 2018).

Physical work overload seems to increase absenteeism rate by 2.7 times (Schwartz et al., 2021). Work overload can be classified into two categories: physical or psychological overload. In fact, several research has shown the impact that both of them have on the occurrence of human error, stress manifestation and lower job satisfaction levels which, in turn, may be related to work-related injury rates (Christian et al., 2009; Yurko et al., 2010; Elo et al., 2003; Littman et al., 2006).

In addition to physical and chemical risk factors, workers in janitorial services face a host of social and economic factors that have been associated with a higher risk of work-related injuries and illnesses. They are often women, over 40 years of age, and the workforce is made up of a large proportion of immigrants. Furthermore, janitors and cleaners are usually low-wage workers, and the profession is considered low status, which limits their bargaining power to demand safer work (Smith and Anderson, 2017).

The recent advances in machine learning and numerical simulation make it possible to work with large datasets to describe relations between variables and to forecast different types of phenomena. There are a number of studies that apply machine learning techniques to ergonomics (Suárez Sánchez et al., 2016; Artime Ríos et al., 2019, 2020; Busto Serrano et al., 2020). Besides, González Fuentes et al. (2020) proposed a model to predict the number of days absent from work due to health-related leave among workers in the energy sector, according to different risk factors.

Making use of data from more than 79,000 occupational accidents, the aim of the present research is to use machine learning techniques to develop a model to predict the incapacity for work (expressed in days of absence) due to work-related overexertion injuries among service sector cleaners in Spain. This makes it possible to identify the most relevant factors (personal, ergonomic, work-environment, employment conditions, etc.) that affect the severity of this kind of injuries. The resulting model can be applied to other data sets to predict the severity of potential injuries based on information on the key factors. This means it can be used as an ergonomic risk assessment method for cleaning occupations.

2. Material and methods

2.1. Data set

As in many countries in the world, recording and notification of occupational injuries is mandatory in Spain. The introduction, in 2003, of the statewide Electronic Declaration of Accidents at Work (Delt@), together with some specific regional records, led to the creation of Work Accident Statistics - Estadística de Accidentes de Trabajo (Ministerio de Trabajo y Economía Social, 2021). For each occupational injury causing at least one day of absence from work, this database records information regarding the personal and professional characteristics of the injured worker, the characteristics of their job, company and work center, the circumstances associated to the injury and the nature and consequences of the damage to their health.

The datasets are stored by the Ministry of Labor and Social Economy. General access to the main statistics is possible through the web portal of the Ministry (Ministerio de Trabajo y Economía Social, 2020, https://www.mites.gob.es/estadisticas/eat/welcome.htm). Upon request, the duly anonymized microdata are provided free of charge for research purposes.

The original database codifies 58 variables that can be classified into the following topics:

- Personal and professional characteristics of the worker injured.
- Information about the employer and the enterprise or establishment.
- Information about the place of occurrence.
- Information about the accident and its sequence.
- Information about the injury.
- Information about the temporary incapacity for work.

With the exception of military occupations, all categories of the National Classification of Economic Activities (CNAE) and the National Classification of Occupations (CNO) are covered. The cases included in the database are those work-related accidents suffered by the group of workers who are affiliated to the Spanish Social Security scheme. This research employs the microdata of the Work Accident Statistics from 2009 to 2019, which initially included more than 6.5 million accidents and recurrences.

According to the aims of the work, the sample was reduced using the following criteria:

- Main occupation of the worker: only those classified as "Cleaning staff for offices, hotels and other similar establishments" were selected (398,712 cases).
- Mode of injury: only those classified as "Physical overexertion on the musculoskeletal system" were selected (145,209 cases).
- Status in employment: only employees were considered; self-employed workers were eliminated.
- Social Security regime: workers under special Social Security regimes were excluded.
- Location of the place of the accident: commuting accidents were excluded; traffic accidents were excluded as well.
- Whether it was the worker's usual work: all those cases where the worker was not performing their usual work were excluded.
- Place of occurrence: industrial locations and locations associated with the primary sector were excluded.
- Type of work-work process: only cases classified under the headings "Services, health care, assistance to people" and "Cleaning of premises, of machines" were included.
- Deviation leading to the accident: cases beyond the scope of the study were excluded; only those related to body movements were included.
- Part of the body injured: injuries to the neck, back, trunk, upper limb and lower limb were included.

The final sample consisted of 79,265 occupational accidents classified as overexertion, occurring between 2009 and 2019 among Social Security-insured service sector cleaners.

Before implementing the machine learning techniques, a number of independent variables were modified to provide the information in a more suitable way. Many of the categorical variables were transformed into binary variables. There is a variable in the database that records the severity of the injury as assessed by the health-care professional. However, the duration of the incapacity for work is an indirect indicator of severity which has important implications in the economic and human costs of the injury for both the worker and the company. Thus, the target dependent variable selected for the development of the models was the incapacity for work expressed in calendar days of absence.

2.2. Software and hardware

All the calculus in the present research was performed in a server with Linux Ubuntu 7.5.0 distribution, 64 GM RAM memory and an Intel® Core™ i7-7700K CPU with 4.20 GHz. All machine learning models were trained with the help of the R-project software version 3.6.3. The libraries earth (Milborrow, 2021), e1071 (Meyer et al., 2021), metrics (Hamner and Frasco, 2018), and tictoc (Izrailev, 2021) were employed.

2.3. Multivariate adaptive regression splines

Multivariate adaptive regression splines (MARS) is a multivariate methodology that can be considered as a generalization of the recursive partitioning regression model. The MARS model can be expressed as (Friedman, 1991):

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i \beta(x_i)$$

Where y_t represents the response variable for each $t \in R$ and β_i are the model parameters for the x_i variables and $i = 1, \dots, k$. The intercept coefficient is represented by β_0 and $\beta(x_i)$ are the base functions. $\beta(x_i)$ can be expressed as $\beta(x_i) = max(0, x_i - c)$ where c is a threshold value.

MARS generates different threshold values for each variable (García Nieto et al., 2013). The resulting MARS model is the sum and product of base functions that are optimized with the help of a forward/backward stepwise algorithm. In a first stage, a model is created with the help of a forward stepwise algorithm, and this model is usually formed by a large number of base functions; afterwards, with the help of a backward stepwise algorithm, all those terms that contribute less to the model fitness are removed.

The variables' importance in a MARS model can be tested with the help of any of the following methods that are described below (Alonso-Fernández et al., 2013):

Residual Sum of Squares (RSS). This coefficient represents the sum of squares of residuals. It can be defined as the difference of actual values minus predicted ones. Their formula is as follows:

$$RSS = \sum_{i=1}^{n} (y_t - \hat{y}_t)^2$$

Where y_t are the actual values and $\hat{y}_t$ their forecasts.

Generic cross validation (GCV). This is an approach of the cross-validation method by the leave-one-out formula:

$$GCV = \frac{RSS}{\left(1 - \frac{M}{n}\right)^2}$$

Where M is the number of parameters of the model which depends on the number of terms and the threshold values employed minus 2.

Nsubsets: A criterion that assesses the importance of the variables, taking into account the number of models each variable takes part in.

The MARS methodology is well-known, and its performance has been tested in different applications like, for example, the forecasting of the bone mineral density in post-menopausal women (De Cos Juez et al., 2009) or the study of the presence of cyanotoxins in reservoirs (García Nieto et al., 2011).

2.4. Support vector machines

Support vector machines (SVM) is a well-known machine learning methodology originally developed in the 1990s as a classification algorithm (Cortes and Vapnik, 1995) and later, extended as a regression method.

SVM is a very convenient technique for learning features in a high dimension characteristics space (Rosado et al., 2013). One of the main properties of SVM regression models is that its training can be configured in such a way that the training not only focuses on the error minimization, but also on the improvement of other characteristics if such an error goes below a certain value. In summary, the objective function of a SVM regression model is to minimize the $l2-$ *norm* of its coefficients instead of just performing as a pure error minimization.

This means that the error term is embedded in the constraints where the absolute error is fixed to be an under certain margin, called ε. Please note that the ε value can be adjusted in order to obtain the required model accuracy.

What has been explained above means that the objective function is expressed as (Casteleiro Roca et al., 2017):

$$min \frac{1}{2}\omega^2 + C \sum_{i=1}^{n} |\xi_i|$$

subject to:

$$|y_i - \omega_i x_i| \leq \varepsilon + |\xi_i|$$

In the two equations presented above, ξ_i represents the deviation and C is a hyperparameter that can be modified, but taking into account that as C raises, the tolerance to those points out of the ε band increases.

As in the case of MARS, SVM is also a well-known methodology that has been tested in both classification and regression problems. Two interesting examples of successful applications of this methodology were the training and validation of a survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers (Rosado et al., 2013) and the forecasting of the higher heating value in biomass torrefaction (García Nieto et al., 2019).

3. Results

3.1. Characteristics of the sample

As stated above, the final sample consisted of 79,265 occupational accidents. Most of the injured workers (87.3%) were women and 8198 of them (which represents 10.3% of the sample) were immigrants. Their average age was 45.9 years, 42.0 among men and 46.4 among women. The average seniority of the workers was 5.5 years. Regarding their status in employment, the majority of the injured workers (69.1%) had an indefinite contract, but almost half of them (46.4%) were part-time; most of them (90.6%) were employees in the private sector.

Only 3.2% of the accidents occurred out of the worker's usual workplace. The highest number occurred on Mondays, and the frequency decreased throughout the week, especially during weekends. Regarding the hour of work, the largest number of injuries occurred during the second hour of work and this number decreased throughout the rest of the shift. Only 6.6% occurred at night. In 23.1% of the cases, no risk assessment had been performed.

The most frequent deviations were uncoordinated movements (31.9%), inappropriate lifts (28.3%), inappropriate pushes or pulls (12.0%) and inappropriate twists or turns (10.6%). The part of the body most frequently injured (38.5%) was the back, followed by the upper limbs (32.5%), the lower limbs (17.9%), the neck (7.2%) and the trunk (3.9%). The number of days of absence due to the injury (target variable) ranged from 1 to 625, with an average of 28.39 and a standard deviation of 41.96.

3.2. Machine learning models

Fig. 1 shows a flowchart of the methodology applied in the present research. In a first step, the data set is split randomly into two subsets, one with 80% of information for model training and another with 20% for model validation. Afterwards, a MARS and a SVM model that both make use of the training data set are trained. The performance of both models is measured with the help of the validation data set and, also, another MARS model is trained that makes use of the training data set employing only those variables considered to be important by the previous MARS model. The performance of this new MARS model is also measured with the help of the validation data set. This process is repeated 5 times making use of different training and validation data sets.

In a first step a MARS model was trained. The use of a MARS model allowed us to find out which variables were relevant in order to predict the days of absence and also the relative importance of those variables.

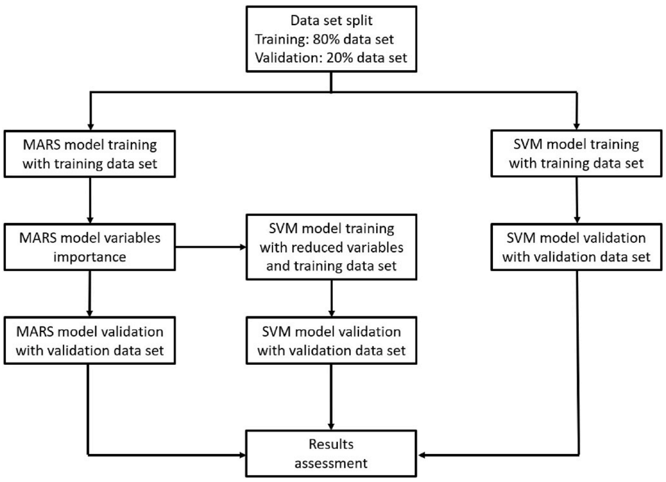

Fig. 1. Workflow of the methodology applied in the present research.

As an example, Table 1 presents those variables that were chosen as important by one of the MARS models trained, showing their relative importance with the help of the nsubsets, GCV and RSS metrics.

After the training of the MARS models, two different SVM models were also trained by making use of the same training data set and validated with the same validation data set as the MARS model. The first of the two SVM models made use of all the variables available in the data set while the second one employed only those variables that were found to be relevant by the MARS model. This second model was called "SVM reduced variables".

Please also remember that each of the sets of models (MARS, SVM and SVM reduced variables) was trained by making use of 80% of the available individuals selected in a random mode while the other 20% were employed for model validation. Please note that each set of models (MARS, SVM and SVM reduced variables) employed a different training and validation data set.

The performance of all the models was tested with the help of the RMSE, MAE, MAPE and Pearson's correlation coefficients. The results are shown in Table 2, where also the time required for the computation of each model can be seen. As may be observed in this table, the performance of the SVM model is higher than that of the MARS models, because in all of them the RMSE, MAE and MAPE values are lower, while in the case of the r-squared coefficient, their values are higher. This means that the SVM model has a greater ability than the MARS to predict the duration of leave. Unfortunately, as can also be observed in Table 2, the computation time is about 15 times higher. The third kind of model whose results are presented in Table 2 is the one called SVM reduced variables. In this case, the performance is quite similar to the SVM model, but in all cases slightly worse. In the case of RMSE, the values are 14% higher, while MAE is 2.8% higher, MAPE values are almost identical and r-squared performance is reduced by about 1.2%. Please also note that computational times are almost 6% higher.

Fig. 2 shows the residuals distribution of the MARS, SVM and SVM reduced variables models. Please note that the residuals are the difference of the real leave duration minus the forecasted value obtained by each of the models. As can be observed, the distribution of residuals of the SVM and SVM reduced variables models are narrower, which means a lower dispersion of the difference between the real and predicted values of the duration of leave. These results are in line with those shown in Table 2. In the case of the MARS model, the average value of a residual was 0.001, with a standard deviation of 40.924 and a median of −10.403, while for the SVM the average value was of −1.288 with a standard deviation of 8.831 and a median of −4.184. In the case of the SVM reduced set of variables the average was −1.499 with a standard deviation of 8.137 and a median value of −4.183.

Fig. 3 presents some of the splines that form part of the MARS models. These Figures not only show which variables are relevant but also explain how they influence the target variable (incapacity for work in terms of the number of days of absence due to the injury). For example, the first spline (a) of Fig. 3 shows the contribution of variable "Edad_del_trabajador_accidentado_el_dia_del_accidente" (age of the worker) to the prediction of the value of the target variable: the number of days of absence significantly increase with the age of the worker.

To be able to understand what is behind the work-related injuries studied and their potential consequences for the workers' health status,

Table 1
Importance of variables according to the results obtained with one of the MARS models (nsubsets, GCV, RSS).

	nsubsets	gcv	rss
Extremidad_superior	58	100	100
Extremidad_inferior	57	80.1	82
Edad_del_trabajador_accidentado_el_dia_del_accidente	56	69.8	72.8
Sexo_8	55	58.7	63.2
Indefinido	53	52.7	57.9
Viernes	53	52.7	57.9
Agente06_OtrasHerramientas_manuales_limpiar	53	52.7	57.9
Plantilla_del_centro_32	52	50	55.5
Lugar_habitual	52	50	55.5
Desviacion_75	52	50	55.5
Antiguedad_en_el_puesto_de_trabajo_en_meses_15	51	47.3	53.2
Agente04_herramientas_manuales_limpiar	51	47.3	53.2
Agente14_humanos	50	44.7	51
Agente08_material_limpieza	45	36.4	43.6
Situacion_profesional_12	41	34.7	41.6
Desviacion_72	41	34.7	41.6
Actividad_Manipulacion	40	33.3	40.4
Desviacion_70	40	33.3	40.4
TipoTrabajo_Asistencia	39	32.1	39.3
Nacionales	38	30.9	38.1
Actividad_Maquinas	38	30.9	38.1
Agente07_maquinas_limpiar	35	26.9	34.7
Jueves	32	23.2	31.4
TipoTrabajo_Limpieza	28	19.7	28.1
Agente09_carros	25	17.2	25.6
Hora_de_trabajo_72	22	16.2	24.1
Evaluacion_de_riesgos_74	18	14.1	21.5
Agente05_herramientas_mecanicas_limpiar	17	13.4	20.7
Desviacion_74	15	11.8	19
Agente11_cargas	13	10.4	17.3
TC	10	8	14.6

Table 2
Performance of the MARS, SVM and SVM reduced set of variables models (RMSE; root mean squared error, MAE: mean absolute error, MAPE: mean absolute percentage error; r-squared: squared value of the Pearson's correlation coefficient).

MARS				
RMSE	MAE	MAPE	R-squared	time (sec.)
40.8108	22.37458	1.8685%	0.04383504	430.213
41.22899	22.57842	1.8805%	0.04723708	456.953
40.71752	22.31591	1.8590%	0.04761595	444.088
40.88696	22.40823	1.8685%	0.04644732	431.134
41.32255	22.5796	1.8813%	0.04638609	458.654
SVM				
RMSE	**MAE**	**MAPE**	**R-squared**	**time (sec.)**
8.030453	4.333495	0.4109%	0.9676599	6497.292
8.148747	4.373064	0.4157%	0.9673812	7574.861
7.923442	4.319536	0.4100%	0.9685324	7256.092
7.483582	4.30675	0.4122%	0.9727428	6565.644
7.511101	4.369971	0.4166%	0.9730518	7472.346
SVM reduced set of variables				
RMSE	**MAE**	**MAPE**	**R-squared**	**time (sec.)**
8.954931	4.410451	0.4090%	0.9575009	7159.141
9.086879	4.485919	0.4169%	0.9570749	7409.129
8.95468	4.439549	0.4090%	0.9517834	7160.403
9.449801	4.544427	0.4127%	0.9521157	9467.453
8.254001	4.430831	0.4154%	0.9653515	6170.361

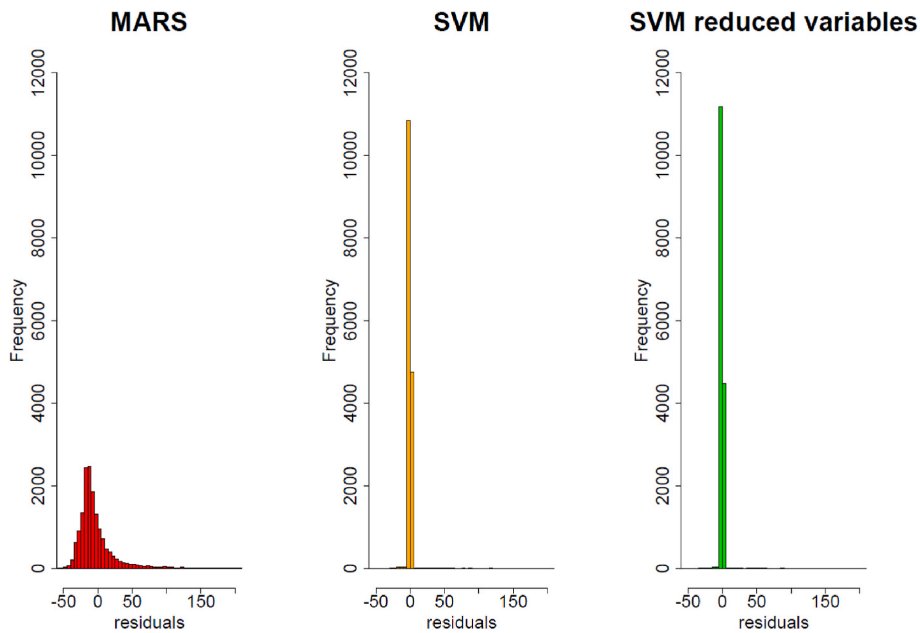

Fig. 2. Residuals distribution of MARS SVM and SVM reduced variables models.

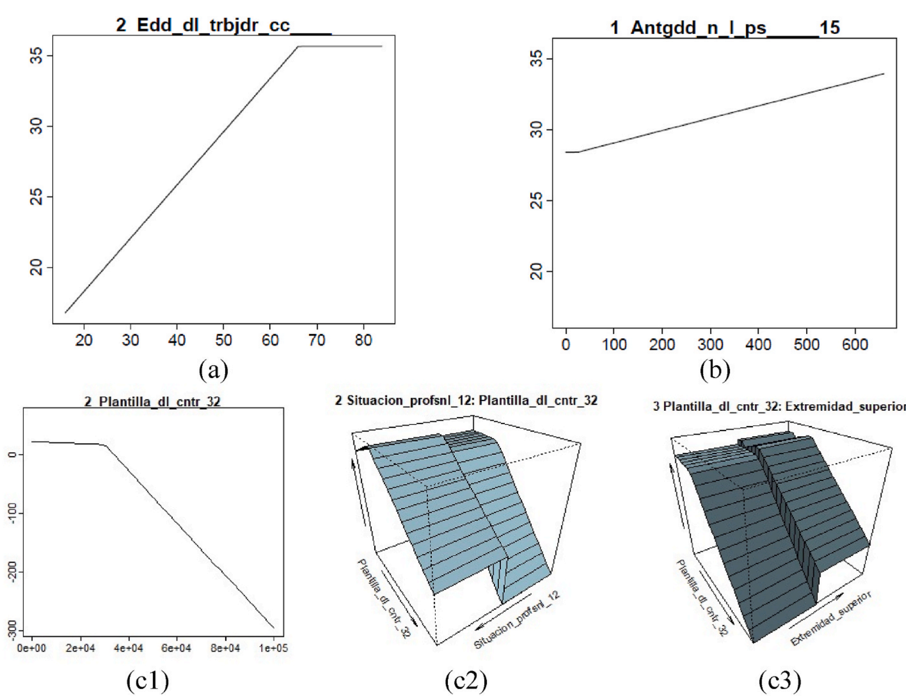

Fig. 3. Some relevant variables according to the MARS models, and how they influence the target variable (number of days of absence due to the injury): (a) Age; (b) Seniority; (c) Size of the establishment.

the variables identified by the MARS models have been sorted according the following categories:

- Injury typology.
- Individual factors.
- Employment conditions.
- Accident circumstances.
- Health and Safety management and company standards.

Table 3 summarizes the 20 most important variables, which appear in all (or four) of the MARS models, together with their definition. The variables are classified into the different categories. Within each category, they are arranged in order of importance according to their average position in the set of models.

As the day of the week when the injury occurred was defined using seven binary variables, Fig. 4 helps to show the behavior of such item by summarizing both the number of accidents and the average severity of the injury versus the day of the week.

4. Discussion

As shown in the Results section, the use of SVM models, when

Table 3
Summary of the 20 most important variables in the MARS models.

Variable	Definition
Injury typology	
Extremidad_superior	Part of the body injured: upper limb
Extremidad_inferior	Part of the body injured: lower limb
Individual factors	
Edad_del_trabajador_accidentado_el_dia_del_accidente	Age of the injured worker
Antiguedad_en_el_puesto_de_trabajo_en_meses_15	Seniority of the worker (in months)
Sexo_8	Sex
Employment conditions	
Indefinido	Type of contract: fixed-term or indefinite contract
Nacionales	Nationality: country national or migrant worker
Situacion_profesional_12	Employment situation: whether private or public employee
Accident circumstances	
Viernes	Day of the week: Friday
Agente06_OtrasHerramientas_manuales_limpiar	Material item or agency associated with the injury: powered hand tools for waxing, lubricating, washing, cleaning
Hora_de_trabajo_72	Hours worked in the activity when the accident occurred
Lugar_habitual	Type of location of the accident: the worker's usual workplace or another place
Agente14_humanos	Material item or agency associated with the injury: human beings
Desviacion_72	Deviation: inappropriate pushes or pulls
Agente04_herramientas_manuales_limpiar	Material item or agency associated with the injury: non-powered hand tools for waxing, lubricating, washing, cleaning
Jueves	Day of the week: Thursday
Desviacion_74	Deviation: inappropriate twists or turns
Desviacion_70	Deviation: other movements of the body as a result of or with physical exertion
Health and safety management and company standards	
Plantilla_del_centro_32	Size of establishment (number of workers)
Evaluacion_de_riesgos_74	Existence of risks assessment

compared with MARS, improves the performance of the model considerably. One of the main disadvantages of SVM models is the lack of model explainability. In other words, while in a MARS model equation it is easy to find out what the relationship is among variables, SVM is a black box model. This means that the relationship among variables is difficult to explain. Taking into account the good performance of the SVM model but its lack of explainability, a second kind of SVM models were trained that made use only of those variables considered important by the MARS model. As the performance obtained was quite close to the one achieved by the SVM trained with all the variables, those models are the preferred ones, as they achieve a good prediction level by only making use of some variables whose relationship can be explained with

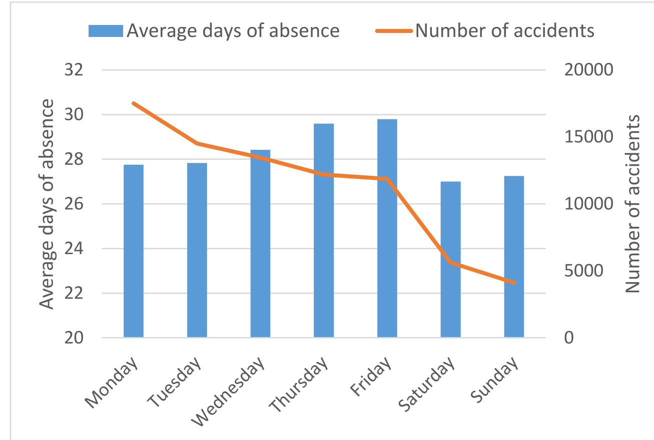

Fig. 4. Evolution of the average days of absence and of the number of accidents along the week.

the help of a MARS model. Moreover, from the authors' point of view, a remarkable issue is the fact that the proposed methodology can be employed with any dabase that makes use of continuous independent and dependent variables and it is not only limited to the context of health and safety studies.

The developed MARS models show the relevance of some specific factors on the impact of the injury, in terms of days of absence of the worker. As stated above, this means they can be used as an ergonomic risk assessment method specifically developed for cleaning occupations.

As stated above (see Table 3), the variables identified by the MARS models have been sorted according the five different categories.

4.1. Injury typology

The location of the injury is an important factor in the prediction of long-term sick leave, considering that MSD are the leading cause of absenteeism among European workers (Bevan, 2015).

Both upper- and lower-limb injuries have been identified by the MARS models as being among the most influential factors in sick leave duration. The average number of days of absence associated with an injury of the upper limb (35.9) and of the lower limb (31.9) are significantly higher than the average of the sample (28.4). However, there is unanimity on identifying the injury of the upper-limb MSD as a sick leave aggravation cause. In fact, this finding is in tune with many studies that highlight the relevance of upper limb MSD in the context of employment (Da Costa et al., 2015).

Furthermore, talking about cleaning workers, it seems evident that their task content involves continuous repetitive movements and awkward postures, among other ergonomic risk factors that have an impact on upper-limb MSD prevalence (Naik and Khan, 2020).

Unfortunately, there is no general guideline as to how to successfully develop an ergonomic intervention suitable to any workplace. Consequently, many studies and their conclusions must be analyzed and applied carefully (Sundstrup et al., 2020). In this sense, working tasks and workplace heterogeneity in the cleaning sector are key in understanding that ergonomic programs must be designed after an exhaustive working condition assessment in order to ensure their suitability to the case study.

4.2. Individual factors

Age and sex take very relevant positions in the MARS models. Both of them are workers' characteristics with an important impact on MSD prevalence proved by many studies (Lee et al., 2018; Ramos Vieira et al., 2015; Shariat et al., 2018; Macpherson et al., 2018; Palmer and Goodson, 2015).

As occupational statistics in Spain show, women tend to work preferably in the service sector, whereas men are more likely to work in the productive sector. The sample used in this study confirms the overwhelming feminization of the cleaning sector in Spain, and the average leave duration in this sample is significantly higher among women (28.9 days) than among men (24.7 days).

Beyond women's widespread presence in this sector, gender features as a major factor affecting MSD prevalence, placing women in the center of the issue. Therefore, any occupational ergonomics intervention must consider biological and social differences behind sex and gender to ensure success in the achievement of its objectives.

Focusing on the next variable, a profound impact of ageing on sick leave duration has been shown. The MARS models (Fig. 3(a)) have detected a strong and direct relation between the number of days of absence and the age of the injured worker.

Indeed, the European Union sees the ageing of the working population as an emerging issue and a challenge to be dealt with, to ensure appropriate health and safety standards and to maintain the competitiveness of their labor markets (European Commission, 2015).

Most health and safety standards and policies were originally developed years ago, focusing on a working population that did not correspond to the current one in terms of the average age of the workers. Therefore, it is critical to make the required efforts to adapt health and safety management to the present circumstances, taking into account the needs of an ageing working population.

Another variable included in this category is the seniority of the worker. As shown in Fig. 3(b), the predicted number of days of absence increases with the seniority of the worker. Its behavior is consistently similar to that of the age of the worker, but further research would be necessary to identify what part of this influence can be attributed to the long-term cumulative exposure to the work-related risk factors.

4.3. Employment conditions

Three relevant variables are included in this category:

- Type of contract: fixed-term or indefinite contract.
- Employment situation: working for the private or public sector.
- Nationality: country national or migrant worker.

It is believed that precarious employment conditions have a negative impact on health and safety standards, so it is to be expected that fixed-term contract, migrant and private sector workers were more susceptible to injuries, including those related to MSD. However, the current study detected that indefinite contract (29.0 average days of absence vs 26.9 in the fixed-term contract workers), country national (29.0 average days of absence vs 23.0 in migrant workers) and public sector workers (31.6 average days of absence vs 28.1 in private sector workers) are more likely to suffer from this kind of damage to their health.

In any case, this could be explained by the effect that job instability has on the underreporting of work-related accidents and the shortening of sick leaves. Thus, a plausible theory could be that this is caused by workers being afraid of losing their jobs.

This hypothesis is in line with the idea developed by some studies (Martinez Aso, 2013) suggesting that migrant workers are more exposed to accidents at work than country national workers but, in parallel, they ignore their rights, the legal rules that are applied and the procedures to communicate with the public authorities responsible for health and safety matters (Moyce and Schenker, 2018; Schwartz et al., 2019).

In fact, it has been detected that many of the perceived barriers to reporting occupational injuries can be eliminated by informing workers of their right to report injuries and helping them to recognize an occupational injury (Green et al., 2019).

For the same reason, workers that have been contracted on a fixed-term basis could be underreporting work accidents as well, since it is not true that they suffer fewer accidents; on the contrary, there are studies that claim there is a contractual effect that increases their probability (Guadalupe, 2003).

As for the public workers' situation, they actually represent the summum of job stability, so the results of this study can be explained in the same way as both of the previous variables.

4.4. Accident circumstances

With regard to this category, the day of the week seems to be one of the factors with the strongest influence on the duration of the sick leave. In fact, although the number of accidents decreases throughout the week, accidents that happened on Thursdays and Fridays tend to be more serious (see Fig. 4). The relevance of this phenomenon has already been studied (Amiri et al., 2014; Fontaneda et al., 2019), with a particular focus on road transport (Sussman and Coplen, 2000). Everything seems to point to the effects of fatigue as the main cause, so that the overall accumulation of physical exposure (Veerasammy et al., 2022) throughout the week leads to more severe injuries in the final days. This is consistent with other studies that have detected a persistence of fatigue from Tuesday to Friday (Yung et al., 2014).

Another interesting accident circumstance to mention is the negative potential effect that working in third-party facilities has on the duration of leave (30.0 average days of absence versus 28.3 average days in accidents occurred in the worker's usual workplace). The Spanish authorities have already identified this as a risk factor; in fact, there is a Spanish Royal Decree that establishes how to deal with health and safety management in case of several companies and workers attending the same workplace (Ministerio de Trabajo y Asuntos Sociales, 2004). Also, there is a specific Spanish Law for the construction sector that develops the issue. Regarding the cleaning sector, this aspect has to be of major concern as many times tasks are performed on the customers' premises.

Finally, in this category, several factors have appeared related to tools, equipment and the kinds of tasks performed by workers that have a direct impact on the duration of sick leave. For example, injuries associated with the use of powered hand tools for waxing, lubricating, washing and cleaning lead to significantly longer leave (30.3 average days of absence versus 28.4 average days in other cases). Vibrations, deficient design and control difficulties have already been highlighted as risk factors when operating these types of tools (Kumar and Kumar, 2008).

Again, this seems to highlight the relevance of ergonomic design and intervention with regard to materials and procedures to prevent overstrain and heavy workloads.

4.5. Health and safety management and company standards

Variables such as the size of establishment and the existence of risk assessment seem to point to how the strength of the management system affects health and safety standards and accident rates.

Fig. 3(c) describe the influence of the size of the establishment in the predicted variable according to the MARS models. The duration of the leave generally decreases when the size of the establishment increases (Fig. 3(c1)): health and safety management standards tend to improve with the size of the company. Nevertheless, the combined effect of the size of the establishment with other variables is particularly interesting. The decrease in the duration of the leave is less pronounced among public sector employees (Fig. 3(c2)), which is consistent with the previous comments regarding job stability. At the same time, the decrease is less pronounced in the case of injuries of the upper limb (Fig. 3(c3)), and this would suggest that this type of injuries has more to do with material than with managerial conditions.

Numerous studies have focused on the effectiveness of health and safety management (Robson et al., 2007); in fact, this is one of the main objectives of safety science researchers. Now, a question could be raised concerning whether some particularities of the cleaning sector should be considered in the design and development of health and safety management systems. In this sense, studies focused on the cleaning industry (Goldenhar et al., 1999) have found dissatisfaction on the part of the workers and carelessness from the owners in relation to health and safety management and standards.

In addition, there is a particular subsector within the cleaning industry whose tasks and, most especially, work organization system, has spurred interest among public authorities in Spain, including courts and labour inspectorates, due to its precarious working conditions and low health and safety standards: the hotel housekeeping staff. A literature review suggests that this issue has gone beyond Spanish borders (Rawan Nimri et al., 2020).

4.6. Limitations

This study has some limitations. First of all, it must be taken into account that, as all the machine learning models, those proposed in the present research require of large amounts of data for their training and that, in case new information is available, systems should be trained again making use of the whole data set, as the machine learning models cannot learn in an incremental way. The present research made use of two machine learning methodologies: MARS and SVM. MARS is proposed as a complement to SVM due to the good performance of the resulting SVM model when variable selection is made with the help of a MARS model. Please note that the variable selection is possible as MARS is not a black box model in the same way as SVM. Although the computational cost of MARS is higher than that required for other regression models, due to the advances of computers in the last years, it is not a problem in the context of the present research. It is also worth mentioning that MARS regression is performed in a stepwise greedy manner. It means that only the best basis function given the current model is either included or removed from the model in each step.

Second, our model has been developed considering all the injuries identified as “overexertion” in the Spanish accident reporting system. MSD of different body parts sometimes have unique features; further research performing independent models for each particular injury may provide more specific findings on the influence of certain factors.

Third, the database used does not include independent variables containing detailed information on the presence of other factors previously identified in the literature as important in the development of MSD, such as the existence of repetitiveness, awkward postures, load-carrying, etc.; on the contrary, these have needed to be deducted from the circumstances of the accident. Something similar can be said of variables regarding the previous health status of the workers. Including some extra items concerning these aspects in the accident reporting system, and thus in the resulting database, would lead to a more accurate ergonomic analysis.

5. Conclusions

The SVM models showed a better predictive capability than the MARS models. This research made use of both of them to obtain a high-performance model wherein the relationships between the variables could be explained. For future research, it would be of interest to explore the performance of deep learning methodologies in this context.

The severity of accidents caused by overexertion depends on several factors that can be classified into the following categories: injury typology, personal factors, employment conditions, accident circumstances and health and safety management and standards in the company. In any case, those factors are interrelated, and it is hard to establish where the influence of each variable starts and ends.

In relation to labour conditions (type of contract, job situation and migrant workers), there seems to be a contradiction between the model developed (where, apparently, the severity of the injuries is lower in more unstable jobs) and previous studies. In contrast, the results of this work can be coherently explained by the division of the Spanish labor market into two levels of workers based on their job stability perception: workers in more unstable jobs might be reluctant to report injuries for fear of losing their jobs. This aspect may guide the authorities on the way to working for the reinforcement of labor rights and the protection of the most vulnerable workforce.

Regarding personal factors of the workers, ageing and the gender gap seem to be another great challenge for public agents, as their impact on increasing absenteeism and lowering workers’ well-being has been demonstrated time and again.

Ergonomics is now again at the centre of the proposed solution. The type of injury and the circumstances that surround the accident can be fought with risk assessment, ergonomic intervention and a proper design of equipment and procedures.

The identification of critical features of the cleaning industry that have a major impact on injury rates is key in ensuring an efficient design of the health and safety management system. Likewise, ergonomic intervention should adapt to the actual working conditions and the particular circumstances of this sector. In this sense, hotel housekeeping staff are a clear example that needs further research.

Declaration of competing interest

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

Acknowledgements

The authors wish to acknowledge the Spanish Ministry of Labor and Social Economy (Ministerio de Trabajo y Economía Social), for providing us with the dataset used in this work.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alonso Fernández, J.R., Díaz Muñiz, C., García Nieto, P.J., de Cos Juez, F.J., Sánchez Lasheras, F., Roqueñí, M.N., 2013. Forecasting the cyanotoxins presence in fresh waters: a new model based on genetic algorithms combined with the MARS technique. Ecol. Eng. 53, 68–78. https://doi.org/10.1016/j.ecoleng.2012.12.015.

Amiri, M., Ardeshir, A., Fazel Zarandi, M.H., 2014. Risk-based analysis of construction accidents in Iran during 2007-2011-meta analyze study. Iran. J. Public Health 43 (4), 507–522. http://www.ncbi.nlm.nih.gov/pmc/articles/pmc4433733/.

Artime Ríos, E.M., Sánchez Lasheras, F., Suárez Sánchez, A., Iglesias-Rodríguez, F.J., Seguí Crespo, M.D.M., 2019. Prediction of computer vision syndrome in health personnel by means of genetic algorithms and binary regression trees. Sensors 12 (19), 2800. https://doi.org/10.3390/s19122800.

Artime Ríos, E.M., Suárez Sánchez, A., Sánchez Lasheras, F., del Seguí Crespo, M.M., 2020. Genetic algorithm based on support vector machines for computer vision syndrome classification in health personnel. Neural Comput. Appl. 32 (5), 1239–1248. https://doi.org/10.1007/s00521-018-3581-3.

Bevan, S., 2015. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract. Res. Clin. Rheumatol. 29, 356–373. https://doi.org/10.1016/j.berh.2015.08.002 (¡Error! Referencia de hipervínculo no válida).

Busto Serrano, N., Suárez Sánchez, A., Sánchez Lasheras, F., Iglesias-Rodríguez, F.J., Fidalgo Valverde, G., 2020. Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). Appl. Ergon. 82, 102981 https://doi.org/10.1016/j.apergo.2019.102981.

Casteleiro Roca, J.L., Jove, E., Sánchez Lasheras, F., Méndez Pérez, J.A., Calvo Rolle, J.L., de Cos Juez, F.J., 2017. Power cell SOC modelling for intelligent virtual sensor implementation. J. Sens. https://doi.org/10.1155/2017/9640546, 2017, ID 9640546.

Chang, J.H., De Wu, J., Liu, C.Y., Hsu, D.J., 2012. Prevalence of musculoskeletal disorders and ergonomic assessments of cleaners. Am. J. Ind. Med. 55 (7), 593–604. https://doi.org/10.1002/ajim.22064.

Choi, A., Chin, G., 2018. Effects of the center of mass of a stick vacuum cleaner on the muscle activities of the upper extremity during floor vacuuming. Appl. Ergon. 70, 1–5. https://doi.org/10.1016/j.apergo.2018.02.001.

Christian, M.S., Bradley, J.C., Wallace, J.C., Burke, M.J., 2009. Workplace safety: a meta-analysis of roles of person and situation factors. J. Appl. Psychol. 94 (5), 1103–1127. https://psycnet.apa.org/doi/10.1037/a0016172.

Cortes, C., Vapnik, V.N., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. https://doi:10.1007/BF00994018.

Da Costa, J.T., Baptista, J.S., Vaz, M., 2015. Incidence and prevalence of upper-limb work related musculoskeletal disorders: a systematic review. Work 51 (4), 635–644. https://doi.org/10.3233/wor-152032.

de Cos Juez, F.J., Sánchez Lasheras, F., García Nieto, P.J., Suárez Suárez, M.A., 2009. A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. Int. J. Comput. Math. 86 (10–11), 1878–1887. https://doi.org/10.1080/00207160902783557.

Elo, A.L., Leppänen, A., Jahkola, A., 2003. Validity of a single-item measure of stress symptoms. Scand. J. Work. Environ. Health 29 (6), 444–451. https://doi.org/10.5271/sjweh.752.

European Cleaning and Facility Services Industry, 2020. The Claning Industry in Europe. EFCI'S Report 2020 (79863609229-39). accessed November 2021. https://www.efci.eu/wp-content/uploads/flipbooks/6/.

European Commission, 2015. The 2015 Ageing Report: Economic and Budgetary Projections for the 28 EU Member States (2013-2060).

Fontaneda, I., Camino López, M.A., González Alcántara, O.J., Ritzel, D.O., 2019. Gender differences in lost work days due to occupational accidents. Saf. Sci. 114, 23–29. https://doi.org/10.1016/j.ssci.2018.12.027.

Friedman, J., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–67. https://doi.org/10.1214/aos/1176347963.

García Nieto, P.J., Sánchez Lasheras, F., de Cos Juez, F.J., Alonso Fernández, J.R., 2011. Study of cyanotoxins presence from experimental cyanobacteria concentrations using a new data mining methodology based on multivariate adaptive regression splines in Trasona reservoir (Northern Spain). J. Hazad. Mater. 195, 414–421. https://doi.org/10.1016/j.jhazmat.2011.08.061.

García Nieto, P.J., García–Gonzalo, E., Sánchez Lasheras, F., Paredes–Sánchez, J.P., Riesgo Fernández, P., 2019. Forecast of the higher heating value in biomass torrefaction by means of machine learning techniques. J. Comput. Appl. Math. 357, 284–301. https://doi.org/10.1016/j.cam.2019.03.009.

Goldenhar, L.M., Ruder, A.M., Ewers, L.M., Earnest, S., Haag, W.M., Petersen, M.R., 1999. Concerns of the dry-cleaning industry: a qualitative investigation of labor and management. Am. J. Ind. Med. 35 (2), 112–123. https://doi.org/10.1002/(sici)1097-0274(199902)35:2%3C112::aid-ajim2%3E3.0.co;2-u.

González Fuentes, A., Busto Serrano, N.M., Sánchez Lasheras, F., Fidalgo Valverde, G., Suárez Sánchez, A., 2020. Prediction of health-related leave days among workers in the energy sector by means of genetic algorithms. Energies 13 (10), 2475. https://doi.org/10.3390/en13102475.

Graves, J.R., Way, K., Riley, D., Lawton, C., Morris, L., 2002. Development of risk filter and risk assessment worksheets for HSE guidance-"Upper limb disorders in the workplace". Appl. Ergon. 35 (5), 475–484. https://doi.org/10.1016/j.apergo.2004.03.011.

Green, D.R., Goodwin, S., Kim, H., Ryan, A.D., McGrovern, P.M., Churc, T.R., Schwartz, A., Arauz, R.F., 2019. Knowledge of work-related injury reporting and perceived barriers among janitors. J. Saf. Res. 69, 1–10. https://doi.org/10.1016/j.jsr.2019.01.003.

Guadalupe, M., 2003. The hidden costs of fixed term contracts: the impact on work accidents. Lab. Econ. 10 (3), 339–357. https://doi.org/10.1016/S0927-5371(02)00136-7 (¡Error! Referencia de hipervínculo no válida).

Hamner, B., Frasco, M., 2018. Metrics: Evaluation Metrics for Machine Learning (R Package Version, 0.1.4. https://CRAN.R-project.org/package=Metrics.

Izrailev, S., 2021. Tictoc: Functions for Timing R Scripts, as Well as Implementations of Stack and List Structures. R package version 1.0.1. https://CRAN.R-project.org/package=tictoc.

Johansonsson, S.E., Ljunggren, G., 1989. Perceived exertion during a self-imposed pace of work for a group of cleaners. Appl. Ergon. 20 (4), 307–312. https://doi.org/10.1016/0003-6870(89)90196-8.

Kumar, R., Kumar, S., 2008. Musculoskeletal risk factors in cleaning occupation – a literature review. Int. J. Ind. Ergon. 38, 158–170. https://doi.org/10.1016/j.ergon.2006.04.004.

Lee, S., Hwan Park, M., Jeong, B.Y., 2018. Gender differences in public office workers' satisfaction, subjective symptoms and musculoskeletal complaints in workplace and office environments. Int. J. Occup. Saf. Ergon. 24 (2), 165–170. https://doi.org/10.1080/10803548.2016.1272959.

Lim, M.C., Lukman, K.A., Giloi, N., Lim, J.F., Avoi, R., Syed Abdul Rahim, S.S., Jeffree, M.S., 2022. Prevalnece of upper limb musculoskeletal disorders and its associated risk factors among janitorial workers: a cross-sectional study. Ann. Med. Surg. 73 https://doi.org/10.1016/j.amsu.2021.103201.

Littman, A.J., White, E., Satia, J.A., Bowen, D.J., Kristal, A.R., 2006. Reliability and validity of 2 single-item measures of psychosocial stress. Epidemiology 17 (4), 398–403. https://doi.org/10.1097/01.ede.0000219721.89552.51.

Luenda, E.C., Loomis, D., Demissie, Z., 2008. Occupational hazards experience by cleaning workers and janitors: a review of the epidemiologic literature. Work 34 (1), 105–116. https://doi.org/10.3233/wor-2009-0907.

Macpherson, R.A., Lane, T.J., Collie, A., 2018. Age, sex, and the changing disability burden of compensated work-related musculoskeletal disorders in Canada and Australia. BMC Publ. Health 18, 758. https://doi.org/10.1186/s12889-018-5590-7.

Martinez Aso, M., 2013. La eficacia de la protección del derecho a la seguridad y salud en el trabajo de los trabajadores extranjeros. Universitat de Girona. Departament de Dret Privat. http://hdl.handle.net/10803/124038.

Melese, H., Gerbreyesus, T., Alamer, A., Berhe, A., 2020. Prevalence and associated factors of musculoskeletal disorders among cleaners working at Mekelle University, Ethiopia. J. Pain Res. 13, 2239–2246, 10.2147%2FJPR.S263319.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2021. e1071: Misc Functions of the Department of Statistics. Probability Theory Group (R package version 1.7-9). https://CRAN.R-roject.org/package=e1071.

Milborrow, S., 2021. Earth: Multivariate Adaptive Regression Splines. R package version 5.3.1. https://CRAN.R-project.org/package=earth.

Ministerio de Trabajo y Asuntos Sociales, 2004. Real Decreto 171/2004, de 30 de enero, por el que se desarrolla el artículo 24 de la Ley 31/1995, de 8 de noviembre, de Prevención de Riesgos Laborales, en materia de coordinación de actividades empresariales. https://www.boe.es/eli/es/rd/2004/01/30/171/con.

Ministerio de Trabajo y Economía Social, 2021. Secretaría General Técnica. Subdirección General de Estadística y Análisis Sociolaboral. Base de datos. ACCIDENTES DE TRABAJO.

Ministerio de Trabajo y Economía Social, 2020. Secretaría general técnica. Subdirección general de Estadística y análisis sociolaboral. Estadística de Accidentes de Trabajo, (accessed March 2022). https://www.mites.gob.es/estadisticas/eat/welcome.htm.

Moyce, S.C., Schenker, M., 2018. Migrant workers and their occupational health and safety. Annu. Rev. Publ. Health 39, 351–365 (¡Error! Referencia de hipervínculo no válida).

Naik, G., Khan, M.R., 2020. Prevalence of MSDs and postural risk assessment in floor mopping activity through subjective and objective measures. Safety and Health at Work 11, 80–87. https://doi.org/10.1016/j.shaw.2019.12.005.

Nimri, R., Kensbock, S., Bailey, J., Jennings, G., Patiar, A., 2020. Realizing dignity in housekeeping work: evidence of live star hotels. J. Hum. Resour. Hospit. Tourism 19 (3), 368–387. https://doi.org/10.1080/15332845.2020.1737770.

Occupational Safety and Health Administration, 2021. United Sates Department of Labor. https://www.osha.gov/cleaning-industry (accessed October 2021).

Palmer, K.T., Goodson, N., 2015. Ageing, musculoskeletal health and work. Best Pract. Res. Clin. Rheumatol. 29 (3), 391–404. https://doi.org/10.1016/j.berh.2015.03.004.

Ramos Vieira, E., Venturoso Gongora Buckeridge Serra, M., Brentini de Almeida, L., Vieira Villela, W., Domingos Scalon, J., Veiga Quemelo, P.R., 2015. Symptoms and risks for musculoskeletal disorders among male and female footwear industry workers. Int. J. Ind. Ergon. 48, 110–116. https://doi.org/10.1016/j.ergon.2015.05.001.

Robson, L.S., Clarke, J.A., Cullen, K., Bielecky, A., Severin, C., Bigelow, P.L., Irvin, E., Culyer, A., Mahood, Q., 2007. The effectiveness of occupational health and safety management system interventions: a systematic review. Saf. Sci. 45 (3), 329–353. https://doi.org/10.1016/j.ssci.2006.07.003.

Rosado, P., Lequerica Fernández, P., Villallaín, L., Peña, I., Sánchez Lasheras, F., de Vicente, J.C., 2013. Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines. Expert Syst. Appl. 40, 4770–4776. https://doi.org/10.1016/j.eswa.2013.02.032.

Schwartz, A., Goodwin, S., Albin, T., Kim, H., Ryan, A.D., Churc, T.R., Green, D.R., McGrovern, P.M., Erdman, A.E., Arauz, R.F., 2021. Janitors' mental workload, psychocosial factors, physical fitness, anda injury: the SWEEP study. Int. J. Ind. Ergon. 83, 103132 https://doi.org/10.1016/j.ergon.2021.103132.

Schwartz, A., Goodwin, S., Kim, H., Ryan, A.D., Churc, T.R., Albin, T., McGrovern, P.M., Erdman, A.E., Green, D.R., Arauz, R.F., 2019. Janitor ergonomics and injuries in the safe workload ergonomic exposure project (SWEEP) study. Appl. Ergon. 81, 102847 https://doi.org/10.1016/j.apergo.2019.102874.

Shariat, A., Cardoso, J.R., Cleland, J.A., Danee, M., Ansari, N.N., Kargarfard, M., Mohd Tamrin, S.B., 2018. Prevalence rate of neck, shoulder and lower back pain in association with age, body mass index and gender among Malaysian office workers. Work 60 (2), 191–199. https://doi.org/10.3233/WOR-182738.

Smith, C.K., Anderson, N.J., 2017. Work-related injuries among commercial janitors in Washington State, comparisons by gender. J. Saf. Res. 62, 199–207.

Suárez Sánchez, A., Iglesias-Rodríguez, F.J., Riesgo Fernández, P., de Cos Juez, F.J., 2016. Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. Int. J. Ind. Ergon. 52, 92–99. https://doi.org/10.1016/j.ergon.2015.09.012.

Sundstrup, E., Seeberg, K.G.V., Bengtsen, E., 2020. A systematic review of workplace interventions to rehabilitate musculoskeletal disorders among employees with physical demanding work. Occup. Rehabil. 30, 588–612. https://doi.org/10.1007/s10926-020-09879-x. ¡Error! Referencia de hipervínculo no válida.

Sussman, D., Coplen, M., 2000. Fatigue and alertness in the United States railroad industry part I: the nature of the problem. Transport. Res. F Traffic Psychol. Behav. 3 (4), 211–220. https://doi.org/10.1016/S1369-8478(01)00005-5.

U.S. Bureau of labor statistics, 2021. https://www.bls.gov/ooh/building-and-grounds-cleaning/janitors-and-building-cleaners.htm (accessed October 2021).

Veerasammy, S., Davidson, J.B., Fischer, S.L., 2022. Multi-task exposure assessment to infer musculoskeletal disorder risk: a scoping review of injury causation theories and tools available to assess exposures. Appl. Ergon. 102, 103766 https://doi.org/10.1016/j.apergo.2022.103766.

Wami, S.D., Dessie, A., Chercos, D.H., 2019. The impact of work-related risk factors on the development of neck and upper limb pain among low wage hotel housekeepers in Gondar town, Northwest Ethiopia: institution-based cross-sectional study. Environ. Health Prev. Med. 24, 1–10. https://doi.org/10.1186/s12199-019-0779-7.

World Health Organization (WHO), 2021. Musculoskeletal conditions. last accessed January 2022. https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions.

Yung, M., Bigelow, P.L., Hastings, D.M., Wells, R.P., 2014. Detecting within- and between-day manifestations of neuromuscular fatigue at work: an exploratory study. Ergonomics 57 (10), 1562–1573. https://doi.org/10.1080/00140139.2014.934299.

Yurko, Y.Y., Scerbo, M.W., Prabhu, A.S., Acker, C.E., Stefanidis, D., 2010. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX Tool. Simulat. Healthc. J. Soc. Med. Simulat. 5 (5), 267–271. https://doi.org/10.1097/sih.0b013e3181e3f329.