International Journal of Industrial Ergonomics 52 (2016) 92–99

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

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

Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders

A. Suárez Sánchez [a, *], F.J. Iglesias-Rodríguez [a], P. Riesgo Fernández [a], F.J. de Cos Juez [b]

[a] Department of Business Administration, School of Mining, Energy and Materials Engineering of Oviedo, University of Oviedo, 33004 Oviedo, Spain
[b] Department of Mining Exploitation, School of Mining, Energy and Materials Engineering of Oviedo, University of Oviedo, 33004 Oviedo, Spain

ARTICLE INFO

Article history:
Received 28 August 2014
Received in revised form
15 September 2015
Accepted 18 September 2015
Available online 1 October 2015

Keywords:
Work-related musculoskeletal disorders
Risk factor assessment
K-nearest neighbor technique
Categorical variable

ABSTRACT

The *K*-nearest neighbor (KNN) technique is a widely used classifier because of its simplicity and high efficiency. We adapted this technique and applied it with success to predict work-related musculoskeletal disorders. Among the general working population, the algorithm was able to identify workers that had reported work-related musculoskeletal complaints in the last twelve months. According to the model that was developed, poor lighting conditions, exposure to vibrations, an uncomfortable chair and a high mental demand are the factors that have the strongest influence on the development of this type of health problem.

Relevance to industry:

The approach described in this paper allows the KNN technique to be implemented for the prediction of musculoskeletal disorders among the general working population. The model is able to overcome the limitations of other traditional statistical learning techniques to predict this type of disorder with accuracy and effectiveness. The results may be used as a decision support tool for the prioritization of resources dedicated to ergonomic intervention programs.

1. Introduction

Throughout the world, musculoskeletal symptoms and disorders are common amongst the working population. This type of work-related health problem can result in serious social impacts on both individuals and communities and represents a considerable economic burden to employers, employees and society as a whole (Buckle and Devereux, 2002; Hanson et al., 2006; Dunning et al., 2010). Bhattacharya (2014) estimated the direct costs of MSDs in the USA to be $1.5 billion in 2007. According to recent EU figures (European Agency for Safety and Health at Work, 2010), 24.7% of European workers complain of backache, 22.8% suffer from muscular pains and 45.5% report working in painful or tiring positions.

Work-related musculoskeletal disorders (WMSD) are usually defined as diverse types of injuries, such as pain in muscles, nerves, tendons, ligaments, bones, joints, spinal disks, and cartilage, among others, that result from traumatizing the body in the workplace by demanding more than it is prepared for. It is difficult to compare data from studies of musculoskeletal disorders due to the differences in the terminologies used to define this type of complaint. For example, Violante et al. (2000) highlight the importance of concept definition in this subject and the necessity of clearly distinguishing between the terms musculoskeletal disorders and musculoskeletal symptoms. In the literature, diverse terms are used to describe musculoskeletal pain, injury, symptoms, trouble, discomfort and diseases or disorders (Scuffham et al., 2010). Given the importance of musculoskeletal symptoms and disorders, a large branch of ergonomics focuses on the study of this type of problem. Knowledge of the main factors associated with them is essential to minimize their occurrence.

Recently, diverse works have been published on the prevalence and/or incidence of musculoskeletal problems in different occupational groups, such as computer operators and office workers (Jensen et al., 2002; Robertson et al., 2009; Luttmann et al., 2010; Choobineh et al., 2011; Collins and O'Sullivan, 2015), health care workers (Daraiseh et al., 2010; Jaworek et al., 2010; Scuffham et al., 2010; Callison and Nussbaum, 2012; Cavanagh et al., 2012; Lin et al., 2012; Long et al., 2012; Miranda et al., 2014), manufacturing and assembly line workers (Ferguson et al., 2012; Yu et al., 2012; Gerr et al., 2014a, b, Meyers et al., 2014; Warren et al., 2015, Cote et al.,

* Corresponding author.
E-mail address: suarezana@uniovi.es (A. Suárez Sánchez).

http://dx.doi.org/10.1016/j.ergon.2015.09.012
0169-8141/

2014; Harris-Adamson et al., 2014; Garg et al., 2014a, b, Kapellusch et al., 2014) and even sign language interpreters (Fischer and Woodcock, 2012). These studies usually focus on the risk factors connected to the tasks and activities of the studied profession.

In contrast, a few studies have focused on specific groups of factors and have tried to establish their associations with WMSDs. For example, Widanarko et al. (2011) described the influence of individual factors, such as gender and age, on WMSDs whereas Eatough et al. (2012) focused on the effects of psychosocial factors. Meanwhile Mariscal Saldaria et al. (2012) observed that prevention-related information decreases the likelihood of employees suffering from musculoskeletal problems.

The implementation of artificial intelligence and advanced data mining techniques to ergonomics is not new. For example, Asensio-Cuesta et al. (2010) successfully developed a new approach to obtain neural network models to classify the risk of low back disorders that were presented by workers who had certain lifting jobs involving manual material handling. Meanwhile, the *K*-nearest neighbors (KNN) algorithm is a data mining technique that has been widely used in pattern recognition and classification (Jiawei and Micheline, 2006). The aim of the present research is to develop a holistic model using the KNN technique to identify factors (individual factors, working conditions, and workplace design, among others) that are more strongly associated with the prevalence of occupational WMSDs among the general working population. From this model, it is possible to define a musculoskeletal risk profile. The results may be used as a decision support tool for the prioritization of resources dedicated to ergonomic intervention programs.

2. Materials and methods

2.1. Data set

The data set used in this work captures the results of the Sixth National Survey on Working Conditions (VI Encuesta Nacional de Condiciones de Trabajo). This survey was published in 2007 by the Spanish National Institute for Safety and Hygiene at Work (Instituto Nacional de Seguridad e Higiene en el Trabajo, INSHT), a subsidiary body of the Spanish Ministry of Labour and Immigration (Ministerio de Trabajo e Inmigración, MTIN). The aim of the survey was to provide an overview of the health and safety conditions in Spanish workplaces. Some of the survey's specific objectives included:

- To identify which workplace factors affect workers' health and the extent to which workers are exposed to them;
- To identify existing occupational health and safety management structures and to assess their activities according to the practical measures undertaken; and
- To identify trends in working conditions in the Spanish labor market.

The working population that was eligible to take part in the survey consisted of 18,518,444 workers from the entire Spanish territory, employed across the whole range of economic activities.

Fieldwork was carried out between December 2006 and April 2007, with personal interviews conducted in the homes of 11,054 (5917 male and 5137 female) workers who responded to a questionnaire consisting of 78 items. The sampling procedure was multistage, stratified cluster sampling, with a random selection of primary and secondary sampling units, and workers were selected by random routes. For a confidence level of 95.5% and $P = Q$, the error for the overall sample was ±0.95%.

The National Survey on Working Conditions has been carried out periodically in Spain since 1987. To guarantee its validity, the content of the questionnaire is defined and improved in every edition by the experts of the National Institute for Safety and Hygiene at Work (Instituto Nacional de Seguridad e Higiene en el Trabajo, 2007) and is guided by the published conceptual framework and a literature review.

The topics of the 78 items in the questionnaire ranged from working conditions to health damage and also included questions on the occupational health and safety resources of the company. The items were structured in the following groups:

- Labor relation and type of contract.
- Information from the company and the working center.
- Type of work and seniority.
- Thermal environment.
- Physical agents.
- Chemical and biological agents.
- Safety hazards and conditions.
- Workplace design, physical demand and psychosocial factors.
- Health and safety management and resources.
- Working hours.
- Health and safety activities.
- Violent behavior at work.
- Work-related accidents and health damage.

The reliability of the questionnaire was measured by Cronbach's Alpha coefficient, which is the most common measure of internal consistency (Cronbach, 1951; Frost et al., 2007) and is especially appropriate for the type of scales used in this study. The test was carried out for each group of items, and the calculated coefficients ranged from 0.70 to 0.93. The standard threshold for adequate reliability for the use of measures for group comparisons is 0.70 (Frost et al., 2007). The internal consistency was especially good for "Workplace design" (0.81) and "Psychosocial factors" (0.79).

In the case of self-reported health damage, the interviewed worker was asked to mention up to eight work-related diseases and symptoms that, from their point of view, they were suffering from as a result of their work. These diseases and symptoms could be selected from a list of 29 that were offered by the interviewer. The list included six issues that can be classified as self-reported musculoskeletal symptoms and/or disorders:

- Neck pain.
- Back pain.
- Slipped disc (due to the heterogeneity of the sample, this colloquial term was included in the questionnaire to allow the worker to self-report any of the disorders associated to the general concept).
- Upper limb pain, including pain of the shoulder, arm, elbow and forearm (excluding wrist, hand and finger pain).
- Wrist, hand and finger pain.
- Lower limb pain, including pain of the hip, thigh, knee, lower leg, ankle and feet.

Again, Cronbach's Alpha coefficients were used to determine the internal reliability of the outcome measures and ranged from 0.72 to 0.81.

To develop a model to predict the global prevalence of WMSDs, a new binary variable (WMSD_BIN) was created, which can be considered to be the target variable, and it condenses the information recorded in the group of variables concerning self-reported health damage. This target variable identifies any worker who, in the previous year, suffered from any of the above-mentioned musculoskeletal symptoms.

Of course, all musculoskeletal disorders have different etiologies, but many studies have observed associations between