Iranian Journal of Ergonomics

Iran J Ergon, 2024; 11(4): 261-271

https://journal.iehfs.ir/

Original Article

Prediction of Musculoskeletal Disorders Based on People's Demographic Information Using Artificial Intelligence Methods and the Cornell Musculoskeletal Discomfort Questionnaire

Mousa Nazari[1], **Arezoo Sammak Amani**[2,*], **Mohammad Amin Mououdi**[2], **Mohammad Mahdi Alyan Nezhadi**[1]

[1] Department of Computer Science, University of Science and Technology of Mazandaran, Behshahr, Iran
[2] Department of Occupational Health, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran

Article History:
Received: 16/12/2023
Revised: 15/02/2024
Accepted: 17/02/2024
ePublished: 19/03/2024

***Corresponding author:** Arezoo Sammak Amani, Department of Occupational Health, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran.
Email: arezoo.sam76@yahoo.com

Abstract

Objectives: Work-related musculoskeletal Disorders (WMSDs) are the most significant challenges in both developing and developed countries, affecting the majority of individuals throughout their lives. Considering the detrimental effects of musculoskeletal disorders on the productivity and general health of employees, this research utilizes the Cornell Musculoskeletal Disorder Questionnaire (CMDQ) to develop an intelligent model for assessing and predicting the levels of musculoskeletal disorders.

Methods: In this descriptive-analytical study, 810 employees from five organizations (in four occupational categories, including administrative, technical, production, and services) completed the CMDQ voluntarily. After collecting the questionnaire and performing relevant statistical analyses, data normalization and clustering based on the K-Means method were used to determine levels of musculoskeletal disorders. Finally, the multilayer perceptron artificial neural network was trained to predict the levels of musculoskeletal disorders; moreover, the criteria of precision, accuracy, recall, and F1-score were used to evaluate the proposed model.

Results: The performance of the proposed model in predicting the levels of musculoskeletal disorders is presented in two scenarios (use and non-use of the Synthetic Minority Oversampling Technique (SMOTE) method) based on the evaluation criteria provided. The accuracy, precision, recall, and F1-score values were 0.724, 0.709, 0.756, and 0.720, respectively. The appropriate accuracy and precision in the proposed model indicate its capability to identify the levels of musculoskeletal disorders in individuals and help healthcare professionals take necessary measures to prevent and predict them.

Conclusion: This study employs the CMDQ questionnaire and artificial intelligence to analyze musculoskeletal disorders in the workplace. The proposed model demonstrates significant accuracy and precision compared to similar studies. The results indicate that this model can be utilized to identify and predict musculoskeletal disorders in organizational employees, offering the potential to expedite the identification process and reduce costs.

Keywords: Artificial intelligence, CMDQ, Neural network, WMSD

Copyright © 2024 Iranian Journal of Ergonomics. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, provided the original work is properly cite

Extended Abstract

Background and Objective

Work-related musculoskeletal disorders (WMSDs) are one of the serious problems in developing and advanced countries, threatening individuals throughout their lives. According to the World Health Organization, musculoskeletal disorders are not only limited to old age but can affect people across the lifespan. One of the questionnaires used to assess the prevalence of musculoskeletal disorders is the Male and female versions of the Cornell Musculoskeletal Discomfort Questionnaire (CMDQ) for standing workers developed by Professor Alan Hedge. Machine learning can be used to predict and classify musculoskeletal disorders. Supervised and unsupervised learning are examples of two different types of machine learning approaches. Considering the harmful effects of musculoskeletal disorders on the productivity and general health of employees, the present study aimed to provide an intelligent model to determine and predict the level of skeletal-muscular disorders with the help of demographic information using the CMDQ and machine learning algorithms, such as k-means clustering method, Synthetic Minority Oversampling Technique (SMOTE) for labeling and the learning method with perceptron neural network monitoring.

Materials and Methods

In this descriptive-analytical study, 810 employees of five organizations with four occupational classes(administrative, technical, production, and service) voluntarily completed the CMDQ to assess musculoskeletal disorders. The proposed method to determine and predict the level of musculoskeletal disorders consists of five main steps. After collecting questionnaire data, descriptive statistics related to data evaluation and verification were carried out in the first stage. Thereafter, pre-processing operations, data normalization, and calculation of missing values in the data sets were performed. In the following, the K-means clustering algorithm is used as one of the most famous clustering and unsupervised learning methods for clustering and labeling the severity and level of musculoskeletal disorders. Due to the nature of the categories identified in the clustering stage by the expert and the imbalance in different categories, the SMOTE technique was employed to manage imbalanced data. Finally, the multilayer perceptron artificial neural network was trained as a supervised learning method to predict the level of musculoskeletal disorders. The criteria of precision, accuracy, recall, and F1-score were used to evaluate the proposed model.

Results

Clustering was performed on the score collected from the questionnaires using the K-Means algorithm. The elbow method was used to determine the appropriate number of clusters. Based on the Elbow method and the investigations carried out, the data were clustered into five categories (no problem, with low problem, medium, high, and critical) based on the K-Means method. The performance results of the proposed model in predicting the level of musculoskeletal disorders in two modes (use and non-use of the SMOTE method) are presented based on evaluation criteria. The accuracy, precision, recall, and F1-score values were 0.724, 0.709, 0.756, and 0.720, respectively. The performance results of the proposed model in predicting the level of musculoskeletal disorders in two modes (use and non-use of the SMOTE method) are presented based on evaluation criteria.

The results pointed to a significant improvement in the performance of the proposed model using the SMOTE method. Moreover, the efficiency improvement rate of the proposed model using the SMOTE method was 0.184 based on accuracy criteria and a maximum of 0.329 based on accuracy criteria, compared to the non-use of this method. The appropriate accuracy and precision values in the proposed model demonstrate its ability to identify the level of skeletal-muscular disorders in people and assist health professionals in identifying and taking effective measures to prevent and predict them.

Discussion

Due to the different nature of the columns in the CMDQ, it is preferable to use a method to cluster the data instead of categorizing the data only based on the final scores. Since, in this case, the difference between the severity and frequency of musculoskeletal disorders is also considered. Nonetheless, in a study by Akbari et al., musculoskeletal disorders were assigned to 10 identical categories only based on the final score of the questionnaire, and it needs revision and correction since, in addition to the final score, the score of body parts should also be taken into consideration when interpreting and making decisions in this regard. The SMOTE method can predict the level of musculoskeletal disorders with 72.4% accuracy. In this study, a score of ≥612.5 was determined as a critical point. In a study by Moodi et al., a score above 367 was determined as a critical point by artificial intelligence (machine learning); nonetheless, no reference was made to the accuracy and precision of the method used in the research (Gaussian Mixture Method (GMM)). This difference can be ascribed to the accuracy and precision of the k-means method compared to the GMM method.

Conclusion

According to the purpose envisaged in this research for the proposed model, two applications can be expressed. The first application includes identifying the musculoskeletal disorders of employees of new organizations, and the mere use of people's demographic information accelerates this identification process. Considering the reduction of the required time of the process, which can lead to the reduction of the associated cost, it could be envisaged that more willingness and cooperation from organizations and their employees should be expected.

Please cite this article as follows: Nazari M, Sammak Amani A, Mououdi M A, Alyan Nezhadi M M. Prediction of Musculoskeletal Disorders Based on People's Demographic Information Using Artificial Intelligence Methods and the Cornell Musculoskeletal Discomfort Questionnaire . Iran J Ergon. 2024; 11(4): 261-271.

مجله ارگونومی
دوره ۱۱، شماره ۴، زمستان ۱۴۰۲، صفحات: ۲۷۱-۲۶۱
https://journal.iehfs.ir
UMSHA Press

مقاله پژوهشی

پیش‌بینی اختلالات اسکلتی‌عضلانی بر اساس اطلاعات دموگرافیک افراد به کمک روش‌های هوش مصنوعی و پرسش‌نامه‌ی CMDQ

موسی نظری[۱]، آرزو سماک امانی[۲،*]، محمد امین موعودی[۲]، محمدمهدی علیان‌نژادی[۱]

[۱] گروه علوم کامپیوتر، دانشکده‌ی علوم، دانشگاه علم و فناوری مازندران، مازندران، ایران
[۲] گروه مهندسی بهداشت حرفه‌ای، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی‌درمانی مازندران، مازندران، ایران

تاریخ دریافت مقاله: ۱۴۰۲/۰۹/۲۵
تاریخ داوری مقاله: ۱۴۰۲/۱۱/۲۶
تاریخ پذیرش مقاله: ۱۴۰۲/۱۱/۲۸
تاریخ انتشار مقاله: ۱۴۰۲/۱۲/۲۹

تمامی حقوق نشر برای دانشگاه علوم پزشکی همدان محفوظ است.

* **نویسنده مسئول:** آرزو سماک امانی، گروه مهندسی بهداشت حرفه‌ای، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی‌درمانی مازندران، مازندران، ایران.
ایمیل: arezoo.sam76@yahoo.com

چکیده

اهداف: اختلالات اسکلتی‌عضلانی مرتبط با کار (WMSDs) از معضلات مهم در کشورهای درحال‌توسعه و پیشرفته است و اکثر افراد در طول زندگی خود، با آن روبه‌رو هستند. با توجه به اثرهای زیان‌آور اختلالات اسکلتی‌عضلانی در بهره‌وری و سلامت عمومی کارکنان، این پژوهش با استفاده از پرسش‌نامه‌ی اختلالات اسکلتی‌عضلانی کرنل (CMDQ) به‌منظور ارائه‌ی مدلی هوشمند برای تعیین سطح و پیش‌بینی اختلالات اسکلتی‌عضلانی انجام شد.

روش کار: در این مطالعه‌ی توصیفی‌تحلیلی، ۸۱۰ نفر از کارکنان پنج سازمان با چهار طبقه‌ی شغلی اداری، فنی، تولید و خدمات، داوطلبانه برای ارزیابی اختلالات اسکلتی‌عضلانی، پرسش‌نامه‌ی اختلالات اسکلتی‌عضلانی کرنل (CMDQ) را تکمیل کردند. پس از جمع‌آوری داده‌های پرسش‌نامه‌ای و انجام تحلیل‌های آماری مرتبط، از نرمال‌سازی داده‌ها و خوشه‌بندی بر اساس روش K-Means برای تعیین سطوح اختلالات اسکلتی‌عضلانی استفاده شد. در نهایت، شبکه‌ی عصبی مصنوعی پرسپترون چندلایه برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی آموزش داده شد و معیارهای دقت، صحت، Recall و F1-score برای ارزیابی مدل پیشنهادی به کار گرفته شدند.

یافته‌ها: نتایج عملکرد مدل پیشنهادی در پیش‌بینی سطح اختلالات اسکلتی‌عضلانی در دو حالت استفاده و عدم استفاده از روش SMOTE بر اساس معیارهای ارزیابی ارائه شده است. مقادیر صحت، دقت، Recall و F1-score به‌ترتیب، برابر با ۰/۷۲۴، ۰/۷۰۹، ۰/۷۵۶ و ۰/۷۲۰ به دست آمد. مقدار مناسب صحت و دقت در مدل پیشنهادی نشان‌دهنده‌ی قابلیت آن در شناسایی سطح اختلالات اسکلتی‌عضلانی افراد و کمک به متخصصان حوزه‌ی بهداشت در شناسایی و اقدامات لازم برای پیشگیری و پیش‌بینی آن‌ها است.

نتیجه‌گیری: این مطالعه با استفاده از پرسش‌نامه‌ی CMDQ و روش‌های هوش مصنوعی به تحلیل اختلالات اسکلتی‌عضلانی در محیط کار پرداخته است. مدل پیشنهادی در مقایسه با مطالعات مشابه، دارای دقت و صحت قابل توجهی است. نتایج نشان دادند که از این مدل می‌توان برای شناسایی و پیش‌بینی اختلالات اسکلتی‌عضلانی در کارکنان سازمان‌ها با امکان تسریع فرایند شناسایی و کاهش هزینه‌ها بهره برد.

کلید واژه‌ها: اختلالات اسکلتی‌عضلانی، CMDQ، هوش مصنوعی، شبکه‌ی عصبی

استناد: نظری، موسی؛ سماک امانی، آرزو؛ موعودی، محمد امین؛ علیان نژادی، محمد مهدی. پیش‌بینی اختلالات اسکلتی‌عضلانی بر اساس اطلاعات دموگرافیک افراد به کمک روش‌های هوش مصنوعی و پرسش‌نامه‌ی CMDQ . مجله ارگونومی، زمستان ۱۴۰۲؛۱۱(۴): ۲۷۱-۲۶۱.

مقدمه

اختلالات اسکلتی‌عضلانی مرتبط با کار از جمله مشکلات عمده‌ای است که کشورهای درحال‌توسعه و پیشرفته با آن روبه‌رو هستند و سالانه، هزینه‌های زیادی را بر جوامع تحمیل می‌کند [۱]. بر اساس گزارش سازمان بهداشت جهانی، اختلالات اسکلتی‌عضلانی دومین

عامل ایجادکننده‌ی ناتوانی و کمردرد و به‌صورت کلی، یکی از علل ناتوانی در سطح جهان است. بر اساس اعلام سازمان بهداشت جهانی، اختلالات اسکلتی‌عضلانی تنها مرتبط با افزایش سن نیست، بلکه افراد در طول زندگی خود، ممکن است با آن روبه‌رو شوند. در دنیا از هر پنج نفر، یک نفر اختلالات اسکلتی‌عضلانی را تجربه می‌کند [۲].

گزارش سازمان کار امریکا نشان می‌دهد که در میان هر دو بزرگ‌سال امریکایی، یک نفر به نوعی از اختلالات اسکلتی‌عضلانی دچار می‌شود. سازمان کار امریکا در سال ۲۰۱۱، هزینه‌ی این نوع اختلالات را ۲۱۳میلیارد دلار و در سال ۲۰۱۵، هزینه‌ی درمان آن را ۸۷۴میلیارد دلار اعلام کرده است [۳].

بزرگ‌ترین مطالعه‌ی بهداشت شغلی در اروپا را مؤسسه‌ی تجاری اتحادیه‌ی اروپا (ETUI) انجام داد و مشخص شد که ۴۶ درصد از کارگران اروپایی دچار کمردرد هستند و ۴۳ درصد از آن‌ها از درد در نواحی شانه، گردن و عضلات اندام فوقانی رنج می‌برند [۴].

پرسش‌نامه‌ی ناراحتی اسکلتی‌عضلانی کرنل (CMDQ)، نسخه‌ی مرتبط با فعالیت ایستاده (خانم‌ها و آقایان)، ابزاری مناسب برای جمع‌آوری اطلاعات ناراحتی‌های اسکلتی‌عضلانی است. این پرسش‌نامه را پروفسور آلن هِج (Alan Hedge) و همکاران (۱۹۹۹) در ۱۲ قسمت بدن، شامل اندام‌های گردن، شانه‌ی چپ و راست، قسمت فوقانی پشت، قسمت فوقانی بازوی راست و چپ، قسمت تحتانی پشت، ساعد راست و چپ، مچ چپ و راست، باسن، ران چپ و راست، زانوی چپ و راست و قسمت تحتانی پای راست و چپ، کف پای راست و چپ و در ۳ بخش فراوانی ناراحتی، شدت ناراحتی و تأثیر در توان کاری تنظیم کرده‌اند. امتیاز به‌دست‌آمده برای هریک از اندام‌ها بین ۰ تا ۹۰ خواهد بود. امتیاز کلی برابر با حاصل‌ضرب امتیاز فراوانی ناراحتی (هرگز=۰، ۱ تا ۲ بار در هفته=۱/۵، ۳ تا ۴ بار در هفته=۳/۵، یک بار در روز=۵ و چندین بار در روز=۱۰)، امتیاز شدت ناراحتی (۱، ۲، ۳) و امتیاز تأثیر در توان کاری (۱، ۲، ۳) خواهد بود [۵].

مطالعه‌ای را موعودی و همکاران برای ارزیابی اختلالات اسکلتی‌عضلانی در ۲۵۶ نفر از کارکنان اداری شرکت ملی گاز ایران منطقه‌ی ۹ عملیات انتقال گاز با استفاده از پرسش‌نامه‌ی CMDQ انجام دادند. بیشترین ناراحتی در اندام‌های فوقانی، کمر (۱۱/۸۵ درصد) و کمترین بار ناراحتی در ساعد راست (۱/۳۳ درصد) وجود داشت. امتیاز ۳۶۷ به بالا را الگوریتم مدل ترکیبی گوسین (GMM) ناحیه‌ی بحرانی تعیین کرد [۶].

همیشه این پرسش مطرح است که چه کسانی در معرض خطر بیشتری قرار دارند و در نهایت، چگونه می‌توان این اختلالات را پیش‌بینی و از آن‌ها پیشگیری کرد. در سالیان اخیر، محققان از الگوریتم‌های یادگیری ماشین (Machine Learning) به‌عنوان اصلی‌ترین ابزار پیش‌بینی در علوم و صنایع مختلف استفاده کرده‌اند [۷]. اختلالات اسکلتی‌عضلانی مشکلی مربوط به سلامت و شایع است که می‌تواند بر کیفیت زندگی افراد تأثیر بگذارد. یادگیری ماشین می‌تواند برای پیش‌بینی و دسته‌بندی اختلالات اسکلتی‌عضلانی به کار رود.

مطالعه‌ای را گومز برای پیش‌بینی اختلالات اسکلتی‌عضلانی درباره‌ی ۱۷۴ نفر از کارگران صنعت فراوری گوشت با استفاده از مدل‌های یادگیری ماشین انجام داده است. شایع‌ترین قسمت‌های درگیر در بدن شانه‌ها با شیوع ۴۹/۴ درصد، پشت با شیوع ۴۷/۱ درصد، دست‌ها و مچ با شیوع ۳۱/۶ درصد و گردن با شیوع ۲۹/۳ درصد تعیین شد. از مدل‌های طبقه‌بندی مبتنی بر رگرسیون لجستیک و درخت‌های تصمیم برای پیش‌بینی اختلالات اسکلتی‌عضلانی بر اساس ۲۰ متغیر شامل ۴ متغیر مواجهه‌ی فیزیکی، ۶ متغیر روانی‌اجتماعی و ۱۰ متغیر فردی در این پژوهش استفاده شده است [۸].

چاندانا و پال خطر اختلال کمر در ۲۳۵ نفر از کارکنان در مشاغل صنعتی را بر اساس الگوریتم ماشین بردار پشتیبان (Support vector machine) طبقه‌بندی کردند. داده‌ها در دو دسته‌ی اختلالات کم و پرخطر بر اساس ۵ متغیر مستقل و ۱ متغیر وابسته با دقت ۷۷ درصد دسته‌بندی شده‌اند [۹].

روش‌های یادگیری ماشین یکی از زیرشاخه‌های اصلی هوش مصنوعی هستند که قادرند بدون دانستن روابط پنهان بین داده‌ها، آن‌ها را یاد بگیرند. یادگیری تحت نظارت و بدون نظارت نمونه‌هایی از دو نوع مختلف رویکرد مدل یادگیری ماشین هستند. آن‌ها در نحوه‌ی آموزش مدل‌ها و شرایط داده‌های آموزشی مورد نیاز متفاوت هستند [۱۰].

یادگیری نظارت‌شده نوعی رویکرد یادگیری ماشینی است که با استفاده از مجموعه‌داده‌های برچسب‌گذاری‌شده تعریف می‌شود. این مجموعه‌داده‌ها برای آموزش الگوریتم‌ها به‌منظور طبقه‌بندی داده‌ها یا پیش‌بینی دقیق نتایج طراحی شده‌اند. با استفاده از ورودی‌ها و خروجی‌های برچسب‌دار، مدل می‌تواند دقت خود را اندازه‌گیری کند و در طول زمان یاد بگیرد. از طرف دیگر، در رویکرد یادگیری بدون نظارت، آموزش مدل‌ها مبتنی بر داده‌های آموزشی خام و بدون برچسب است. از این روش اغلب برای شناسایی الگوها و روندها در مجموعه‌داده‌های خام یا خوشه‌بندی داده‌های مشابه در تعداد معینی از خوشه‌ها استفاده می‌شود [۱۱].

این پژوهش با استفاده از پرسش‌نامه‌ی اختلالات اسکلتی‌عضلانی کرنل (CMDQ)، تحلیل آمار توصیفی و الگوریتم‌های یادگیری ماشین برای کارکنان پنج سازمان انجام شده است. هدف اصلی این پژوهش ارائه‌ی مدلی هوشمند برای تعیین سطح و پیش‌بینی اختلالات اسکلتی‌عضلانی بر اساس اطلاعات دموگرافیک افراد است. روش پیشنهادی شامل پنج مرحله‌ی اصلی است. در مرحله‌ی نخست، تحلیل‌های آماری مرتبط با ارزیابی و صحت‌سنجی داده انجام می‌شود. سپس، عملیات پیش‌پردازش، نرمال‌سازی داده‌ها و محاسبه‌ی مقادیر گم‌شده در داده‌ها انجام می‌شود. در ادامه، از الگوریتم خوشه‌بندی K-means [۱۲] به‌عنوان یکی از روش‌های مشهور خوشه‌بندی و یادگیری بدون نظارت برای خوشه‌بندی و برچسب‌گذاری میزان شدت و سطح اختلالات اسکلتی‌عضلانی استفاده شده است. با توجه به ماهیت دسته‌های مشخص‌شده در مرحله‌ی خوشه‌بندی توسط خبره

و عدم توازن در دسته‌های مختلف، تکنیک SMOTE [۱۳] برای مدیریت عدم تعادل داده‌ها به کار گرفته شده است. در نهایت، شبکه‌ی عصبی مصنوعی برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی آموزش داده و ارائه شده است.

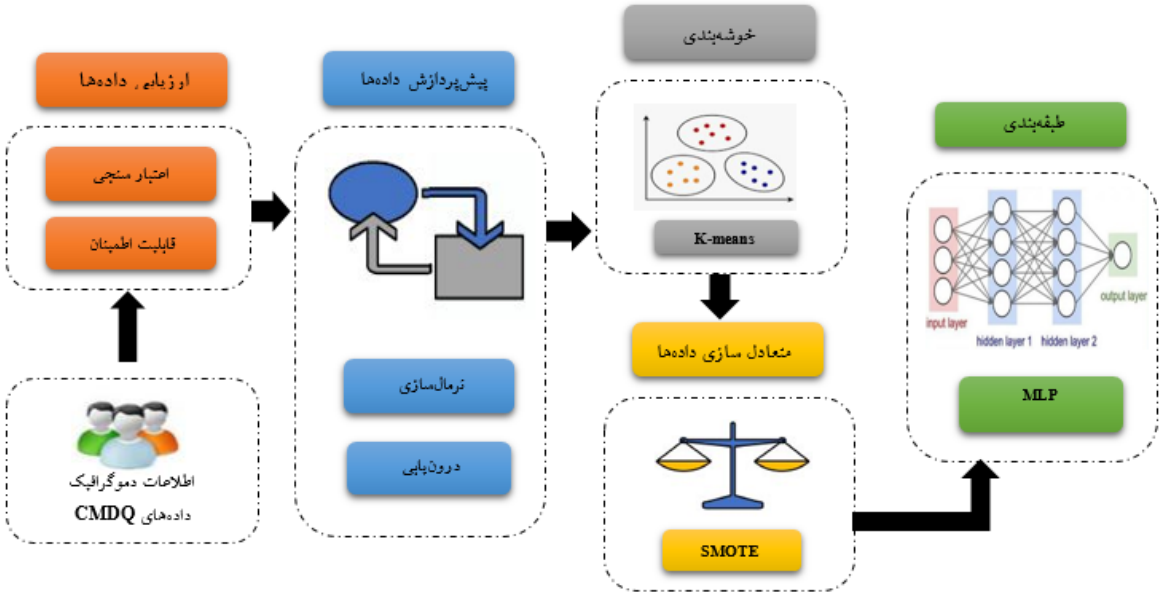
روش کار

مدل پیشنهادی برای پیش‌بینی سطح و دسته‌ی اختلالات اسکلتی‌عضلانی شامل پنج مرحله‌ی آزمون و صحت‌سنجی آماری، پیش‌پردازش، خوشه‌بندی، متعادل‌سازی داده‌ها و دسته‌بندی (Classification) است. شکل ۱ چهارچوب کلی مدل پیشنهادی است.

جمع‌آوری و ارزیابی داده

در این مطالعه‌ی توصیفی‌تحلیلی، ۸۱۰ نفر از کارکنان پنج سازمان (سه سازمان دولتی و دو سازمان خصوصی)، با چهار طبقه‌ی شغلی اداری، فنی، تولید و خدمات، داوطلبانه برای ارزیابی سطح و پیش‌بینی اختلالات اسکلتی‌عضلانی انتخاب شدند. ابتدا، هدف از انجام این بررسی برای افراد مورد مطالعه توضیح داده شد. سپس، افراد پرسش‌نامه‌ی CMDQ را تکمیل کردند. این پرسش‌نامه در سه بخش فراوانی ناراحتی، شدت ناراحتی و تأثیر در توان کاری در هفته ی کاری گذشته تنظیم شد. همچنین، پرسش‌نامه دارای نقشه‌ی بدن است و ۱۲ عضو بدن را که در مجموع، ۲۰ قسمت از بدن است، تحلیل می‌کند. ویژگی‌های دموگرافیک نظیر نوع شغل، سابقه‌ی کار، سن، جنسیت، میزان تحصیلات، نوع تأهل و BMI به ابتدای پرسش‌نامه افزوده شد.

در تحلیل روایی پرسش‌نامه برای قسمت شدت درد و ناراحتی، از ضریب هم‌بستگی Kappa (بین ۰/۹۶۰-۰/۸۲۸) و برای قسمت فراوانی درد و ناراحتی، از ضریب هم‌بستگی Spearman (بین ۰/۹۴۱ – ۰/۸۳۶) استفاده شد. در تحلیل پایایی پرسش‌نامه، ضریب آلفای کرونباخ برابر با ۰/۹۸۶ محاسبه شد. ضریب آلفای کرونباخ در هر سه بخش فراوانی ناراحتی، شدت ناراحتی و تأثیر ناراحتی در توان کاری به‌ترتیب، برابر با ۰/۹۵۵، ۰/۹۶۱ و ۰/۹۶۹ به دست آمد [۱۴]. در ۲۰ درصد از افراد، نمونه‌برداری تکرار شد.



شکل۱: چهارچوب کلی روش پیشنهادی

پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها به فرایند آماده‌سازی داده‌های خام به‌منظور افزایش سازگاری داده‌ها و همچنین، افزایش اطمینان و عملکرد بهتر الگوریتم‌های یادگیری ماشین اطلاق می‌شود. در این پژوهش، پُر کردن مقادیر گم‌شده (Missing value) و نرمال‌سازی داده‌ها بر داده‌های خام اعمال شده است. برای بازیابی اطلاعات گم‌شده در داده‌های خام، از روش درون‌یابی با رابطه‌ی زیر استفاده شده است [۱۵].

$$f(x_i)=\begin{cases} x_i & if\, x_i \neq NaN \\ 0 & if\ \ x_i = NaN\ and\ (x_{i+1}=NaN\ or\ x_{i-1}=NaN\ ) \\ (x_{i+1}+x_{i-1})/2 & if\ \ x_i = NaN\ and\ \ x_{i+1}\neq NaN\ and\ x_{i-1}\neq NaN \end{cases} \quad (۱)$$

در رابطه‌ی فوق، مقصود از x_i و NaN به‌ترتیب، ویژگی داده‌ی ورودی و داده‌ی گم‌شده است.

پس از پر کردن مقادیر گم‌شده، از مقیاس‌بندی Min – Max برای نرمال‌سازی و قرارگیری داده‌ها در محدوده‌ی صفر تا یک استفاده شده است. مقیاس‌بندی Min – Max برای نرمال‌سازی داده‌ها بر اساس رابطه‌ی زیر به کار گرفته می‌شود [۱۷,۱۶]:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (۲)$$

در رابطه‌ی فوق، x' مقدار نرمال‌شده است.

خوشه‌بندی

فرایند یافتن داده‌های مشابه در داده‌های بدون برچسب و دسته‌بندی آن‌ها خوشه‌بندی نامیده می‌شود. در طی دهه‌های اخیر، الگوریتم‌ها و روش‌های مختلف خوشه‌بندی را محققان ارائه کرده‌اند [۱۸]. یکی از روش‌های پرکاربرد و درعین‌حال ساده‌ی خوشه‌بندی الگوریتم K-Means است. الگوریتم K-Means روشی مبتنی بر مرکز است، به‌طوری که با محاسبه و تعیین مراکز، داده‌ها به خوشه‌ای تعلق پیدا می‌کنند که با مرکز آن خوشه فاصله‌ی کمتری داشته باشند [۱۲]. شکل ۲ شبه‌کد مربوط به این الگوریتم را نشان می‌دهد.

مهم‌ترین چالش در الگوریتم K-means یافتن تعداد مناسب خوشه (K) است. در این پژوهش، از روش Elbow (آرنج) برای تعیین مقدار مناسب K در الگوریتم K-means استفاده شده است. در روش Elbow، مقادیر مجموع مربعی درون‌خوشه‌ای (WCSS) برای مقادیر مختلف K نمایش داده می‌شود و مقدار K در محل شکست (آرنج) نمودار، مقدار بهینه و مناسب تعداد خوشه‌ها در نظر گرفته می‌شود (شکل ۳) [۱۲].

متعادل‌سازی داده‌ها

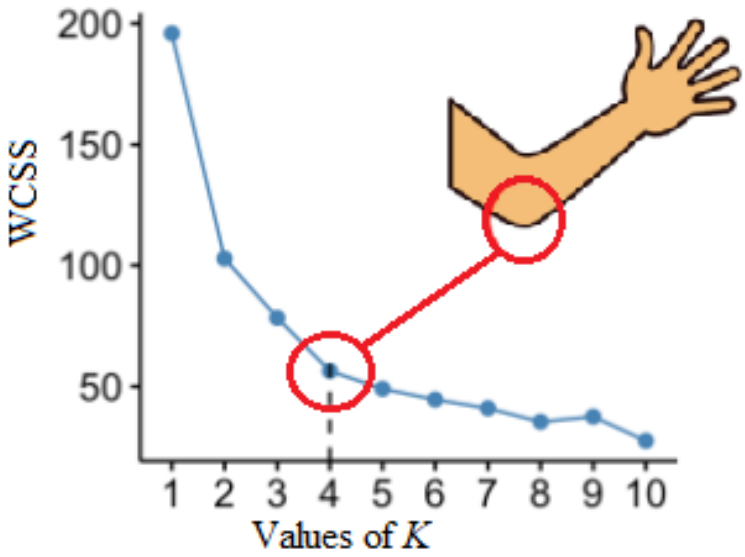
نکته‌ی قابل توجه در مورد نتایج خوشه‌بندی عدم تعادل و توازن در تعداد اعضای هر خوشه است. مدل‌های یادگیری با ناظر در مواجهه با چنین داده‌هایی، دچار مشکلاتی همچون سوگیری (Bias) به دسته‌ی اکثریت و کاهش دقت پیش‌بینی برای دسته‌های اقلیت می‌شوند [۱۹].

رویکرد اصلی در مدل‌های یادگیری ماشین برای پرداختن به مجموعه‌داده‌های نامتعادل، نمونه‌برداری بیش از حد از دسته‌ی اقلیت است. نمونه‌های جدید بر اساس ترکیبی از نمونه‌های موجود ایجاد می‌شوند. این نوعی افزایش داده برای کلاس اقلیت است و تکنیک نمونه‌برداری بیش از حد اقلیت مصنوعی یا به اختصار، SMOTE نامیده می‌شود [۲۰].

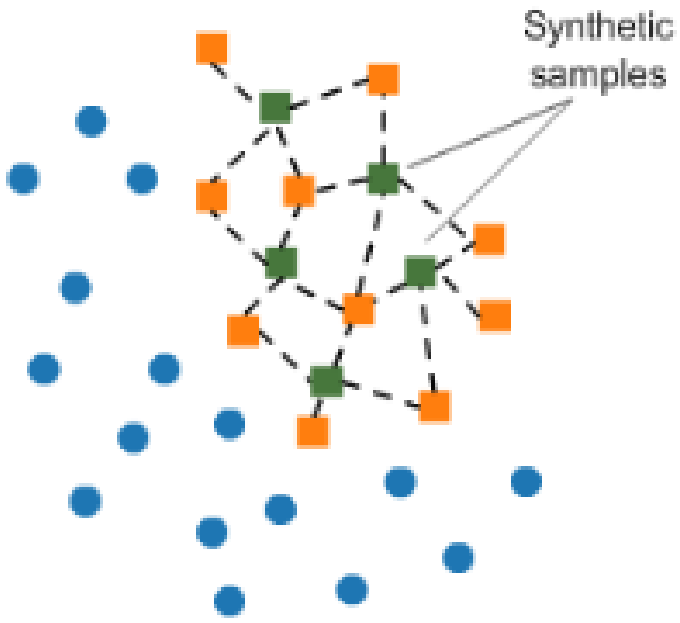
ایده‌ی اصلی این الگوریتم ایجاد نمونه‌های کلاس اقلیت جدید با برداشتن گام‌های کوچک از یکی از نمونه‌های کلاس اقلیت به یکی از k همسایگان نزدیک آن در فضای ویژگی است. الگوریتم با انتخاب تصادفی یکی از k همسایه‌های نزدیک و سپس، اضافه کردن اغتشاشی کوچک به بردار ویژگی بین نمونه‌ی اصلی و همسایه‌ی انتخاب‌شده، نمونه‌ای جدید ایجاد می‌کند. نمونه‌های ایجادشده شبیه نمونه‌های کلاس اقلیت در فضای ویژگی است، اما کپی دقیقی از هیچ‌یک از نمونه‌های موجود نیست (شکل ۴).

Input:
Data = $\{t_1, t_2, \cdots, t_N\}$
K
Output:
Membership matrix
K-means Algorithm:
assign initial values for $c_1, c_2, \cdots, c_K$ randomly
repeat
assign each item t_i to the clusters which has the closest mean;
calculate new mean for each cluster;
until convergence criteria is met;

شکل ۲: شبه‌کد الگوریتم خوشه‌بندی K-Means



شکل ۳: روش Elbow برای تعیین مقدار مناسب تعداد خوشه در الگوریتم K-Means



شکل ۴: نحوه‌ی عملکرد الگوریتم SMOTE

طبقه‌بندی

در سالیان اخیر، مدل‌های بسیار زیادی از الگوریتم‌های نظارت‌شده را محققان ارائه کرده و توسعه داده‌اند [۲۱]. در این پژوهش نیز روش‌های متعدد یادگیری نظارت‌شده مانند درخت تصمیم، جنگل تصادفی و موارد مشابه برای مرحله‌ی آخر مدل پیشنهادی بررسی شده است و در نهایت، شبکه‌ی عصبی مصنوعی پرسپترون چندلایه (MLP) انتخاب شده است. MLP از یک سری لایه‌ی پنهان تشکیل شده است که هر لایه از واحدهایی پردازشی به نام نورون تشکیل شده است. واحدهای پردازشی در هر لایه از داده‌های ورودی خود یاد می‌گیرند و داده‌های پردازش‌شده را به لایه‌ی بعدی منتقل می‌کنند [۲۲].

تعداد نورون‌ها در لایه‌ی ورودی برابر با تعداد ویژگی داده‌ها است و تعداد نورون‌های لایه‌ی خروجی با تعداد دسته‌های (کلاس‌های) مسئله مرتبط است. همچنین، تعداد نورون‌های لایه(های) مخفی، پارامتر مدل محسوب می‌شود و خبره یا همان طراح مدل در حین فاز آموزش، برای رسیدن به دقت بالاتر، آن را تعیین می‌کند [۲۳]. وزن‌ها پارامترهای اصلی MLP هستند که روابط بین ورودی‌ها و خروجی‌ها را تعیین می‌کنند. محققان الگوریتم‌های مختلفی را برای آموزش MLP توسعه و ارائه داده‌اند. الگوریتم‌های گرادیان نزولی و پس‌انتشار روش‌های مهم آموزش برای MLP محسوب می‌شوند [۲۴]. در این پژوهش، از یکی از مشتقات الگوریتم گرادیان نزولی برای آموزش وزن‌های مدل پیشنهادی استفاده شده است.

یافته‌ها

از تعداد ۸۱۰ پرسشنامه‌ی تکمیل شده، تعداد ۸۰۴ پرسشنامه وارد آنالیز آماری شدند. ازاین تعداد نمونه، ۵۹۸ نفر مرد و ۲۰۶ نفر زن هستند. جدول ۱ اطلاعات دموگرافیک افراد را نشان می‌دهد.

در این مرحله از پژوهش، خوشه‌بندی امتیازهای جمع آوری شده از پرسشنامه‌ها به کمک الگوریتم K-Means صورت گرفته است. برای انجام خوشه‌بندی در ابتدا، نمونه‌هایی که مقدار صفر داشتند، جدا و حذف شدند. بر اساس تحلیل به دست آمده از روش Elbow و بررسی‌های صورت گرفته، داده‌ها در چهار دسته بر اساس روش K-Means خوشه‌بندی شده‌اند (جدول ۲).

جدول ۱: اطلاعات دموگرافیک

	جنسیت (تعداد/ درصد)		طبقه‌بندی شغلی (تعداد/ درصد)				BMI (میانگین، S.D)	سن (میانگین، S.D)	وضعیت تأهل (تعداد/ درصد)			سابقه‌ی کار (میانگین، S.D)
	مرد	زن	اداری	فنی	تولید	خدمات			متأهل	مجرد	مطلقه	
سازمان	۱۷۹	۴۹	۱۶۰	۶۵	۰	۳	۲۶/۸۳	۴۰/۴۶	۲۱۴	۱۳	۱	۱۵/۴۹
دولتی A	٪۷۸/۵	٪۲۱/۵	٪۷۰/۱	٪۲۸/۵	٪۰/۰	٪۱/۴	۴/۱۱	۸/۱۳	٪۹۳/۸	٪۵/۷	٪۰/۴	۷/۸۹
سازمان	۲۲۹	۲۷	۱۵۶	۵۴	۱	۴۵	۲۷/۲۷	۳۸/۳۴	۲۳۶	۲۰	۰	۱۲/۴۹
دولتی B	٪۸۹/۵	٪۱۰/۵	٪۶۰/۹	٪۲۱/۱	٪۰/۴	٪۱۷/۶	۴/۷۳	۶/۱۶	٪۹۲/۲	٪۷/۸	٪۰/۰	۶/۵۸
سازمان	۵۳	۱۱۰	۱۵۷	۰	۰	۶	۲۶/۶۴	۳۸/۳۷	۱۴۴	۱۴	۵	۱۳/۷۱
دولتی C	٪۳۲/۵	٪۶۷/۴	٪۹۶/۳	٪۰/۰	٪۰/۰	٪۳/۷	۴/۰۲	۷/۹۹	٪۸۸/۳	٪۸/۵۸	٪۳/۰۶	۷/۹۴

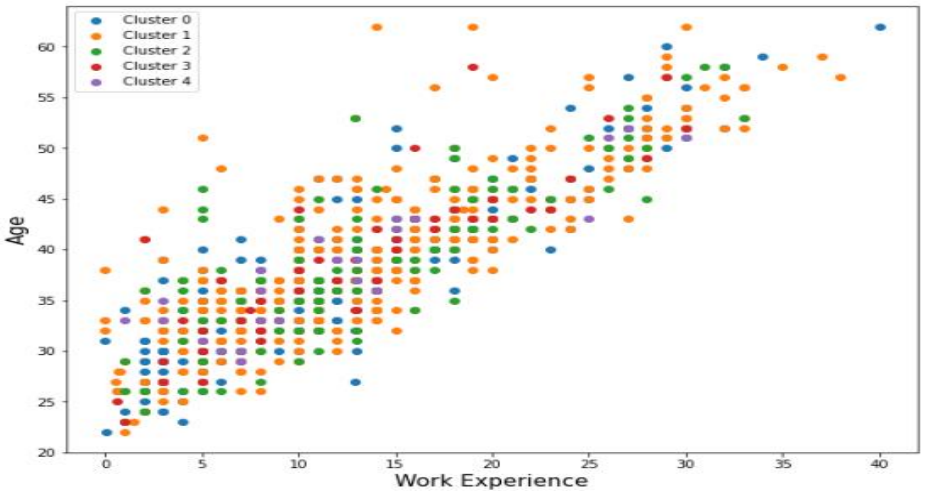
ادامه جدول ۱

سازمان	۱۱۰	۷	۱۸	۱۵	۶۴	۲۰	۲۵/۹۲	۳۵/۳۴	۱۰۰	۱۷	۰	۹/۹۷
خصوصی A	٪۹۴/۰	٪۶/۰	٪۱۵/۴	٪۱۲/۸	٪۵۴/۷	٪۱۷/۱	۴/۶۸	۸/۴۳	٪۸۵/۵	٪۱۴/۵	٪۰/۰	۷/۷۵
سازمان	۲۷	۱۳	۵	۷	۲۳	۵	۲۶/۴۰	۳۵/۲۸	۲۸	۱۲	۰	۸/۵۸
خصوصیB	٪۶۷/۵	٪۳۲/۵	٪۱۲/۵	٪۱۷/۵	٪۵۷/۵	٪۱۲/۵	۳/۱۱	۶/۸۲	٪۷۰	٪۳۰	٪۰/۰	۶/۱۷
کل	۵۹۸	۲۰۶	۴۹۶	۱۴۱	۸۸	۷۹	۲۶/۷۸	۳۸/۳۷	۷۲۲	۷۶	۶	۱۳/۰۲
	٪۷۴/۳	٪۲۵/۶	٪۶۱/۶۹	٪۱۷/۵۳	٪۱۱	٪۹/۹	۴/۳۶	۷/۹۹	٪۸۹/۸	٪۹/۴۵	٪۰/۷۴	۷/۹۴

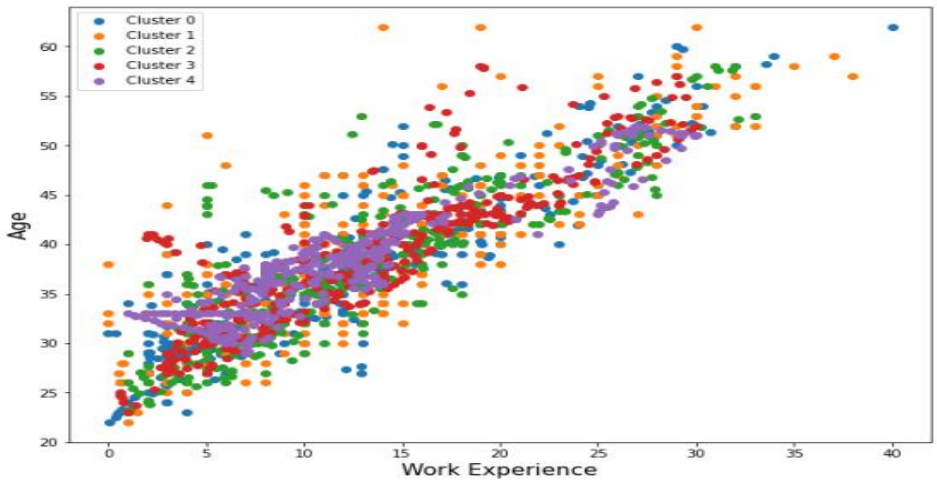
جدول ۲: نتایج حاصل از خوشه‌بندی

رنگ محدوده	تفسیر	امتیاز CMDQ	تعداد	درصد
سفید	بدون مشکل	۰	۲۰۸	۲۵/۸۷
سبز	مشکل کم	۱-۹۸/۵	۴۰۶	۵۰/۵۰
زرد	مشکل متوسط	۹۹-۲۷۹/۵	۱۱۶	۱۴/۴۳
نارنجی	مشکل زیاد	۵۸۳/۲۸۴-۵	۵۰	۶/۲۱
قرمز	ناحیه‌ی بحرانی	۶۱۲/۵ به بالا	۲۴	۲/۶۱

با توجه به نتایج، نکته‌ی قابل توجه و مشهود در داده‌ها عدم توازن و تعادل در تعداد اعضا و نمونه‌های مرتبط با هر خوشه است (جدول ۲). برای رفع این مسئله و ایجاد توازن در داده‌های مرتبط با هر دسته، از روش SMOTE استفاده شده است. شکل‌های ۵ و ۶ به‌ترتیب، داده‌ها را بر اساس دو ویژگی دموگرافیک سن و سابقه‌ی کاری افراد برای قبل و بعد از اعمال روش SMOTE نشان می‌دهند.



شکل ۵: نمایش داده‌ها قبل از اعمال الگوریتم SMOTE



شکل ۶: نمایش داده‌ها بعد از اعمال الگوریتم SMOTE

```
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 dense_40 (Dense)             (None, 512)               5120

 activation_24 (Activation)   (None, 512)               0

 dropout_24 (Dropout)         (None, 512)               0

 dense_41 (Dense)             (None, 256)               131328

 activation_25 (Activation)   (None, 256)               0

 dropout_25 (Dropout)         (None, 256)               0

 dense_42 (Dense)             (None, 128)               32896

 dense_43 (Dense)             (None, 5)                 645

=================================================================
Total params: 169,989
Trainable params: 169,989
Non-trainable params: 0
_________________________________________________________________
```

شکل ۷: معماری شبکه‌ی عصبی

تمامی پیاده‌سازی‌ها اعم از پیش‌پردازش، خوشه‌بندی، الگوریتم SMOTE و در نهایت، MLP در محیط برنامه‌نویسی پایتون (Python) در رایانه‌ای با سیستم‌عامل ویندوز ۱۰، پردازنده‌ی Intel(R) Core(TM) i7-6500U با فرکانس (GHz) ۲/۵۶ و حافظه‌ی اصلی (RAM) (Gbyte) ۸ انجام شده است. همچنین، برای پیاده‌سازی شبکه‌ی عصبی MLP، از کتابخانه‌ی تنسورفلو (TensorFlow) استفاده شده است [۲۵،۲۶]. شکل ۷ معماری شبکه‌ی عصبی MLP مورد استفاده در این پژوهش برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی را نشان می‌دهد. در نهایت، برای آموزش وزن‌های شبکه‌ی عصبیMLP، روش بهینه‌سازی مبتنی بر گرادیان نزولی با عنوان RMSprop از کتابخانه‌ی تنسورفلو به کار گرفته شده است.

در این پژوهش، از چهار معیار صحت، دقت، Recall و F1-score به‌منظور سنجش و ارزیابی مدل پیشنهادی برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی استفاده شده است. معیار صحت بیانگر نسبت پیش‌بینی‌های درست به کل پیش‌بینی‌ها است.

با توجه به اینکه در این پژوهش با داده‌های نامتعادل روبه‌رو هستیم، از سه معیار دیگر برای سنجش و ارزیابی دقیق‌تر مدل پیشنهادی استفاده شده است. دقت و Recall دو معیار ارزیابی اندازه‌گیری دقت و جامعیت مدل پیش‌بینی هستند. دقت نسبت نمونه‌های درست واقعی تشخیص‌داده‌شده توسط مدل به کل نمونه‌های درست پیش‌بینی‌شده را نشان می‌دهد. معیار Recall نسبت نمونه‌های درست واقعی پیش‌بینی‌شده به کل نمونه‌های مثبت واقعی است. در آخر، معیار F1-score ترکیبی از دو معیار دقت و Recall برای سنجش صحت پیش‌بینی صحیح مدل در کل مجموعه‌داده‌ها است [۲۷].

جدول ۳ نتایج به‌دست‌آمده از مدل پیشنهادی برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی بر اساس معیارهای فوق را نشان می‌دهد. نتایج عملکرد مدل پیشنهادی در پیش‌بینی سطح اختلالات اسکلتی‌عضلانی در دو حالت استفاده و عدم استفاده از روش SMOTE بر اساس معیارهای ارزیابی ارائه شده است. نتایج نشان‌دهنده‌ی بهبود قابل توجه عملکرد مدل پیشنهادی در استفاده از روش SMOTE است. همچنین، میزان بهبود کارایی مدل پیشنهادی در استفاده از روش SMOTE برابر با ۰/۱۸۴ بر اساس معیار صحت و حداکثر برابر با ۰/۳۲۹ بر اساس معیار دقت، نسبت به عدم استفاده از آن است. سطح مناسب صحت و دقت در مدل پیشنهادی نشان‌دهنده‌ی قابلیت آن در شناسایی سطح اختلالات اسکلتی‌عضلانی افراد و کمک به متخصصان حوزه‌ی بهداشت در شناسایی و انجام اقدامات لازم برای پیشگیری از آن‌ها و درمان آن‌ها است.

جدول ۳: نتایج مدل پیشنهادی برای پیش‌بینی سطح اختلالات اسکلتی‌عضلانی

معیار ارزیابی	عدم استفاده از روش SMOTE	استفاده از روش SMOTE
صحت	۰/۵۴۰	۰/۷۲۴
دقت	۰/۳۸۰	۰/۷۰۹
Recall	۰/۵۴۰	۰/۷۵۶
F1-score	۰/۴۲۱	۰/۷۲۰

بحث

با توجه به اثرات زیان‌آور اختلالات اسکلتی‌عضلانی در بهره‌وری و سلامت عمومی کارکنان، بررسی اختلالات اسکلتی‌عضلانی در محیط کار از اهمیت ویژه‌ای برخوردار است. مطالعه‌ی حاضر با استفاده از پرسش‌نامه‌ی اختلالات اسکلتی‌عضلانی کرنل (CMDQ) برای ارائه‌ی مدلی هوشمند به منظور تعیین سطح و پیش‌بینی اختلالات اسکلتی‌عضلانی انجام شد. بعد از حذف داده‌های غیرقابل استفاده، ۸۰۴ داده وارد تحلیل آماری شدند و بعد از عملیات نرمال‌سازی داده‌ها، خوشه‌بندی داده‌ها با استفاده از روش K-Means انجام شد. موعودی و همکاران در مطالعه‌ای مشابه، از روش GMM برای

خوشه‌بندی داده‌ها استفاده کردند و به خوشه‌بندی پنج گانه‌ای برای داده‌های مشابه با مطالعه‌ی حاضر دست یافتند. از مزیت‌های مطالعه‌ی حاضر نسبت به مطالعه‌ی موعودی و همکاران، می‌توان به توانایی پیش‌بینی سطح اختلالات اسکلتی‌عضلانی توسط شبکه‌ی عصبی مصنوعی اشاره کرد. با توجه به ماهیت متفاوت ستون‌ها در پرسشنامه‌ی CMDQ، استفاده از روشی برای خوشه‌بندی داده‌ها به جای دسته‌بندی داده‌ها تنها بر اساس نمرات نهایی ارجحیت دارد؛ زیرا در این حالت، افتراق بین شدت و فراوانی اختلالات اسکلتی-عضلانی نیز در نظر گرفته می‌شود، اما در مطالعه‌ی اکبری و همکاران [۲۸]، ناراحتی‌های اسکلتی‌عضلانی تنها بر اساس نمره‌ی نهایی پرسش‌نامه، به ۱۰ دسته‌ی یکسان تقسیم شده است که به نظر می-رسد باید بازبینی و اصلاح شود. به عنوان مثال، اگر فردی صرفاً در دو ناحیه‌ی کمر و ساق پا دچار مشکل جدی باشد و فراوانی درد و شدت ناراحتی و تأثیر در توان کاری برای این فرد به ترتیب، ۱۰ و ۳ و ۳ باشد، طبق جدول، با امتیاز ۱۸۰ دارای مشکل جزئی است و در ناحیه‌ی سبز قرار می‌گیرد، درصورتی که چنین فردی در مطالعه‌ی فعلی با مشکل متوسط در ناحیه‌ی زردرنگ قرار می گیرد. میرمحمدی و همکاران [۲۹] نیز در مطالعه‌ی خود، به چنین موضوعی اشاره کردند، اما به خلاف مطالعه‌ی حاضر، روشی برای دسته‌بندی داده‌ها ارائه نکردند؛ لذا مطالعه‌ی حاضر از این جهت دارای مزیت است. روش SMOTE با صحت ۷۲/۴ درصد امکان پیش‌بینی سطح اختلالات اسکلتی‌عضلانی را دارد. در این پژوهش، امتیاز ۶۱۲/۵ به بالا ناحیه‌ی بحرانی در نظر گرفته شد. در مطالعه-ی موعودی و همکاران [۶]، امتیاز ناحیه‌ی بحرانی ۳۶۷ به بالا تعیین شد، اما اشاره‌ای به صحت و دقت روش به کارگرفته‌شده در پژوهش (GMM) نشد. این تفاوت را می‌توان مربوط به صحت و دقت روش K-Means در مقایسه با روش GMM دانست.

نتیجه‌گیری

با توجه به هدف متصورشده در این پژوهش، برای مدل پیشنهادی دو کاربرد قابل بیان است. کاربرد اول شامل شناسایی اختلالات اسکلتی عضلانی کارکنان سازمان‌های جدید است که با توجه به استفاده ازمطالعات دموگرافیک افراد به تنهایی، به تسریع فرایند شناسایی سطح اختلالات منجر می‌شود.

با توجه به کاهش زمان مورد نیاز فرایند که می‌تواند به کاهش هزینه‌ی مرتبط با آن منجر شود، انتظار رغبت و همکاری بیشتر سازمان‌ها و کارکنان آنها قابل تصور است.

همچنین، با توجه به تغییر اطلاعات دموگرافیک افراد در طول زمان، مانند سن، وزن و سابقه‌ی کاری، قابلیت پیش‌بینی سطوح مختلف اختلالات اسکلتی عضلانی در افراد توسط متخصصان، کاربرد ثانویهی مدل پیشنهادی است.

بر مبنای مطالعه‌ی انجام شده، تحقیق مشابهی که با پرسشنامه‌ی CMDQ و هوش مصنوعی انجام شده باشد، یافت نشد. با توجه به اینکه موضوع مورد پژوهش و کاربرد هوش مصنوعی در این زمینه جزء موضوعات نوین است، خلأ مطالعاتی در این زمینه وجود دارد و پیشنهاد می‌شود که مطالعات آتی با رویکردی پیشبینانه و پیشگیرانه در زمینه‌ی اختلالات اسکلتی عضلانی صورت گیرد.

تشکر و قدردانی

از تمام کسانی که در انجام این پژوهش ما را یاری دادند، کمال تشکر را داریم.

تضاد منافع

نویسندگان این مقاله هیچ‌گونه تعارضی در منافع ندارند.

سهم نویسندگان

تمامی نویسندگان به یک اندازه در آماده سازی این مقاله مشارکت کرده‌اند.

ملاحظات اخلاقی

از تمامی شرکت‌کنندگان در این پژوهش رضایت‌نامه‌ی کتبی گرفته شده است.

حمایت مالی

محققان پشتیبان مالی این تحقیق بوده‌اند.

REFERENCES

1. Coury HJ. The effects of production changes on the musculoskeletal disorders in Brazil and South America. Int J Ind Ergon. 2000; 25(1):103-4.
2. World Health Organization. Musculoskeletal conditions. Web Site;2019. [Link]
3. United States bone and joint initiative: Prevalence, societal economic cost. 3rd ed. Illinois: The burden of musculoskeletal disease in the United States.2016. [Link]
4. European Trade Union Institute (ETUI). Musculoskeletal disorders.2013.[Link]
5. Hedge A, Morimoto S, McCrobie D. Effects of keyboard tray geometry on upper body posture and comfort. Ergonomics. 1999;42(10):1333-49. [DOI: 10.1080/001401399184983] [PMID]
6. Mououdi M A, Sammak Amani A, Ghezelbash K, Ghahari M, Kebriyaee Nasab T. Musculoskeletal Disorders (MSDS) in the Administrative Staff of the National Iranian Gas Transmission Company-District 9 (NIGTC-D9). [In Persian]. Iran South Med J. 2022;25(3):250-60. [DOI: 10.52547/ismj.25.3.250]
7. Shinde PP, Shah S. A review of machine learning and deep learning applications. In2018 Fourth international conference on computing communication control and automation (ICCUBEA) 2018: 1-6. [DOI: 10.1109/ICCUBEA.2018.8697857]
8. Gomes, Mervyn Prediction of work-related musculoskeletal discomfort in the meat processing industry using statistical models. Int J Ind Ergon. 2020;75:102876. [DOI:10.1016/j.ergon.2019.102876]
9. Chandna, Pankaj. Pal, Mahesh. Infinite ensemble of support vector machines for prediction of musculoskeletal disorders risk. Int J Appl Sci Eng.2011; 3.6: 71-7. [DOI: 10.4314/ijest.v3i6.6]
10. Mahesh B, Machine learning algorithms-a review. Int j sci

res. 2020;9(1):381-6. [DOI : 10.21275/ART20203995]
11. Ayodele T. Types of machine learning algorithms, New advances in machine learning. 3th ed by Zhang, Y. InTech;2010:19-48. [DOI: 10.5772/9385]
12. Ikotun, A M., Ezugwu A E, Abualigah L, Abuhaija B, and Heming J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Inf Sci.2022;622(C):178-210. [DOI:10.1016/j.ins.2022.11.139]
13. Wongvorachan, T, He S, Bulut O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Inf.2023; 14(1):54. [DOI:10.3390/info14010054]
14. Afifehzadeh-Kashani H, Choobineh A, Bakand S, et al. Validity and Reliability Farsi Version Cornell Musculoskeletal Discomfort Questionnaire (CMDQ).[In Persian]. Iran Occup Health .2011; 7(4): 10.
15. Zheng, Z, Yang, Y, Niu X, Dai H N, Zhou Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Trans Industr Inform. 2018;14(4):1606–15. [DOI:10.1109/TII.2017.2785963]
16. Nam H, Kim HE. Batch-instance normalization for adaptively style-invariant neural networks. Advances in Neural Information Processing Systems. 2018;31. [DOI:10.48550/arxiv.1805.07925]
17. Pandey A K, Jain A. Comparative analysis of KNN algorithm using various normalization techniques. Int J Comput Netw Inf Secur. 2017; 9:36–42. [DOI: 10.5815/IJCNIS.2017.11.04]
18. Golalipour K, Akbari E, Hamidi S, Lee M, and Enayatifar R. From clustering to clustering ensemble selection: A review. Engineering Applications of Artificial Intelligence.2021;104: 104388. [DOI:10.1016/j.engappai.2021.104388]
19. Aguiar G, Krawczyk B, Cano A. A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework. Mach Learn . 2023:1-79. [DOI: 10.1007/s10994-023-06353-6]
20. Elreedy D, Atiya, A.F. & Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn .2023:1-21. [DOI:10.1007/s10994-022-06296-4]
21. Gasparetto A, Marcuzzo M, Zangari A, and Albarelli A. A survey on text classification algorithms: From text to predictions. Information .2022;13(2):83. [DOI:10.3390/info13020083]
22. Desai Meha, and Shah M. An anatomization on breast cancer detection and diagnosis employing multilayer perceptron neural network (MLP) and Convolutional neural network (CNN). Clinical eHealth .2021;4:1-11. [DOI:10.1016/j.ceh.2020.11.002]
23. Egbueri J C, and Agbasi K C. Performances of MLR, RBF-NN, and MLP-NN in the evaluation and prediction of water resources quality for irrigation purposes under two modeling scenarios. Geocarto Int .2022;37(26):14399-431. [DOI:10.1080/10106049.2022.2087758]
24. Tao P, Cheng J, and Chen L. Brain-inspired chaotic backpropagation for MLP. Neural Net.2022;155(C):1-3. [DOI:10.1016/j.neunet.2022.08.004]
25. Pang, B., Nijkamp, E. and Wu, Y.N., 2020. Deep learning with tensorflow: A review. J Educ Behav Stat.2020;45(2): 227-48. [DOI:10.3102/10769986198727]
26. Singh P, Manure A, Singh P and Manure A. Introduction to tensorflow 2.0. Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python. 2020 1-24. [DOI: 10.1007/978-1-4842-5558-2-1]
27. Grandini M, Bagli E, and Visani G. Metrics for multi-class classification: an overview. ArXiv.2020; abs/2008.05756. [DOI: 10.48550/arXiv.2008.05756]
28. Akbari J, Kazemi M, Mazareie A, Moradirad R, Razavi A. The Ergonomic assessment of exposure to risk factors that cause musculoskeletal disorders in Office workers by using ROSA.[In Persian]. J Ilam Uni Med Sci. 2017; 25(2) :8-17. [DOI: 10.29252/sjimu.25.2.8]
29. Mirmohammadi S T, Gook O, Mousavinasab SN, Mahmoodi Sharafe H. Investigating the Prevalence of Musculoskeletal Disorders in Melli Bank Staff and Determining Its Relationship with Office Tension in North Khorasan Province in 2019.[In Persian]. Iran J Ergon. 2020; 7(4):31-9. [DOI: 10.30699/jergon.7.4.31]