

EMPLOYEE HEALTH MONITORING USING DATA MINING OCCUPATIONAL HEALTH EXAMINATIONS

Identification of Patterns and Factors Affecting the Health of Employees Based on Datamining of Occupational Examinations with the Purpose of Promoting Occupational Health

Ehsan Soleimanfar

MSc, Dept. of Industrial Engineering, Faculty of Industrial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

Mehrzad Navabakhsh

* Assistant Professor, Dept. of Industrial Engineering, Faculty of Industrial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran. Corresponding author:

m_navabakhsh@azad.ac.ir

Mohamad Vahid Sebt

Assistant Professor, Dept. of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

Received: 16 January 2019

Accepted: 14 May 2019

DOI: 10.30699/ijhehp.7.3.295

ABSTRACT

Background and Objective: Paying attention to the health of workers as a significant part of the population is important as they play an important role in the development of the society, which also has caught the attention of government officials and World Health Organization (WHO). Based on the rules and regulations of workers in different occupations, each year they must undergo certain medical tests and examinations to ensure they have sufficient health to perform their duties. This study aimed to predict the results of examinations, extraction of knowledge and identifying patterns and agents that affect workers' health.

Materials and Methods: This was a descriptive-analytic study conducted in Tehran among 1267 employees of various occupations who participated in annual examinations of labor medicine in 2017 and 80 variables related to their health and occupational and family background were collected during the examinations. Due to the size and type of data, the C5.0 decision tree method was used to perform data mining and discovery process.

Results: Using the C5.0 decision tree, a model with accuracy of 99.05% was introduced. According to this model, variables with the greatest impact on the health of the employees were identified. Hearing status, especially hearing loss at frequencies of 6000 and 4000 Hz, had the greatest impact on the results of employee health examinations.

Conclusion: According to the extracted patterns and identification of determinants that had the greatest impact on the result of medical examinations, it is possible to control the specified factors to improve the health of workers.

Keywords: Employee's health, Data mining, Decision tree, Periodic examinations

Paper Type: Research Article.

► **Citation (Vancouver):** Soleimanfar E, Navabakhsh M, Sebt MV. Identification of Patterns and Factors Affecting the Health of Employees Based on Datamining of Occupational Examinations with the Purpose of Promoting Occupational Health. *Iran J Health Educ Health Promot. Autumn 2019*;7(3): 295-305. [Persian]

► **Citation (APA):** Soleimanfar E., Navabakhsh M., Sebt MV. (Autumn 2019). Identification of Patterns and Factors Affecting the Health of Employees Based on Datamining of Occupational Examinations with the Purpose of Promoting Occupational Health. *Iranian Journal of Health Education & Health Promotion.*, 7(3), 295-305. [Persian]

شناسایی الگوها و عوامل مؤثر بر سلامت کارکنان مبتنی بر داده‌کاوی معاینات طب کار با هدف ارتقای سلامت شغلی

چکیده

زمینه و هدف: توجه به سلامت کارکنان به عنوان بخش قابل توجهی از جمعیت جامعه که در پیشرفت و توسعه نقش بسزایی دارند، امری اجتناب‌ناپذیر است. این پژوهش با هدف پیش‌بینی نتایج معاینات، به استخراج دانش و شناسایی الگوها و عواملی که بر سلامت کارکنان تأثیر می‌گذارد پرداخته است.

مواد و روش‌ها: پژوهش حاضر از نوع توصیفی-تحلیلی است. جامعه آماری پژوهش شامل ۱۲۶۷ نفر از کارکنان مشاغل مختلفی است که در سال ۱۳۹۶ در شهر تهران برای انجام معاینات سالیانه طب کار ارزیابی و سنجیده شدند و داده‌های ۸۰ متغیر مربوط به سطح سلامت و سوابق شغلی و خانوادگی آن‌ها در مباحث مختلف طی معاینات جمع‌آوری شد. با توجه به حجم و نوع داده‌ها از روش درخت تصمیم C5.0 برای انجام فرایند داده‌کاوی و کشف دانش استفاده شد.

یافته‌ها: با استفاده از درخت تصمیم C5.0 مدلی با دقت ۹۹/۰۵ درصد معرفی شد. بر اساس نتایج این مدل، متغیرهایی که بیشترین تأثیر را بر سلامت کارکنان داشته‌اند، شناسایی شدند. در میان آنها وضعیت شغلی، مخصوصاً ضعف شنوایی در فرکانس‌های ۶۰۰۰ و ۴۰۰۰ هرتز بیشترین تأثیر را بر نتیجه معاینات سلامت کارکنان داشته است. همچنین ۷ الگو از داده‌ها استخراج شد.

نتیجه‌گیری: با توجه به الگوهای استخراج‌شده و شناسایی متغیرهایی که بیشترین تأثیر را در نتیجه معاینات طب کار داشته‌اند می‌توان با کنترل عوامل مشخص‌شده به ارتقای سطح سلامت کارکنان پرداخت.

کلیدواژه: سلامت کارکنان، داده‌کاوی، درخت تصمیم، معاینات دوره‌ای

نوع مقاله: مطالعه پژوهشی.

احسان سلیمان‌فر

گروه مهندسی صنایع، دانشکده مهندسی صنایع، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

مهرزاد نوابخش

* گروه مهندسی صنایع دانشکده مهندسی صنایع، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران. (نویسنده مسئول):

m_navabakhsh@azad.ac.ir

محمد وحید سبط

گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه خوارزمی، تهران، ایران.

◀ **استناد (ونکوور):** سلیمان فر، ا. نوابخش م، سبط م. شناسایی الگوها و عوامل مؤثر بر سلامت کارکنان مبتنی بر داده‌کاوی معاینات طب کار با هدف ارتقای سلامت شغلی. *فصلنامه آموزش بهداشت و ارتقاء سلامت*. پاییز ۱۳۹۸؛ ۷(۳): ۲۹۵-۳۰۵.

◀ **استناد (APA):** سلیمان فر، احسان؛ نوابخش، مهرزاد؛ سبط، محمدوحید (پاییز ۱۳۹۸). شناسایی الگوها و عوامل مؤثر بر سلامت کارکنان مبتنی بر داده‌کاوی معاینات طب کار با هدف ارتقای سلامت شغلی. *فصلنامه آموزش بهداشت و ارتقاء سلامت*. ۷(۳): ۲۹۵-۳۰۵.

تاریخ دریافت: ۱۳۹۷/۱۰/۲۶

تاریخ پذیرش: ۱۳۹۸/۰۲/۲۴

مقدمه

یکی از مهم‌ترین عوامل پیشرفت و توسعه پایدار اجتماعی و اقتصادی، در نظر گرفتن و اهمیت به سلامت نیروی کار است. نیروی کار بخش بزرگی از جمعیت جامعه را تشکیل می‌دهد و توجه به سلامت کارکنان، تأثیر بسزایی در افزایش سطح سلامت کل جامعه دارد. طب کار شاخه‌ای از دانش پزشکی است که وظیفه حفظ و ارتقای سلامت کارکنان در سازمان‌های مختلف اعم از صنعتی و خدماتی را بر عهده دارد (۱).

معاینات دوره‌ای طب کار با هدف کشف زودهنگام بیماری‌های ناشی از کار اجرا می‌شود تا از زیان‌های جبران‌ناپذیر پیشرفت بیماری تا مرحله غیر قابل درمان و در نهایت از کارافتادگی کارکنان جلوگیری شود (۲). معاینات طب کار نوعی برنامه معاینه پزشکی اجباری برای تمام کارکنان است که در فواصل زمانی تعیین شده با ضوابط قانونی انجام می‌شود که شامل معاینه‌های دوره‌ای عمومی پزشکی و معاینات پزشکی خاص مشاغل مشخص می‌شود که بسته به نوع شغل بررسی می‌شود. به طور کلی دو روش عمده برای انجام معاینات پزشکی کارکنان وجود دارد که عبارتند از: اول، آزمایش‌های مربوط به کارکنان جدید (قبل از استخدام یا قبل از شروع کار) و دوم، آزمایشات دوره‌ای که در فواصل منظم انجام می‌شود و پس از استخدام و کار انجام می‌گیرد (۳).

استفاده از فناوری اطلاعات در سازمان‌های مراقبت‌های بهداشتی به منظور پاسخ به نیازهای پزشکی و کمک در تصمیم‌گیری‌های روزانه پزشکان روزبه‌روز در حال افزایش است. ابزارهای داده‌کاوی می‌توانند محدودیت‌های انسانی مانند ذهن یا خطاهای ناشی از خستگی را جبران کنند و در فرایند تصمیم‌گیری مفید واقع شوند (۴، ۵). پیشرفت‌های فناوری باعث افزایش قابل‌ملاحظه‌ای در حجم داده‌های پزشکی شده است (۶)، اما مجموعه داده‌های خام ارزشی بالایی ندارند و باید با تجزیه و تحلیل، این داده‌ها را به دانش و اطلاعات قابل استفاده تبدیل کرد (۷). امروزه این موضوع که الگوریتم‌های داده‌کاوی می‌توانند در همه زمینه‌های علوم زیست پزشکی قابل استفاده باشند حقیقتی انکارناپذیر است؛ از جمله می‌توان به استفاده

در زمینه‌های بیوانفورماتیک، انفورماتیک تصویربرداری، فناوری اطلاعات بالینی و اطلاع‌رسانی عمومی سلامت اشاره کرد (۸).

علم داده‌کاوی همچنین می‌تواند وابستگی‌های بین داده‌های خام را کشف کند. این کشف برای پزشکان و علم پزشکی ارزشمند و قابل استفاده است. در واقع علم داده‌کاوی می‌تواند به عنوان ابزاری معتبر، حساس و قابل اعتماد برای کشف الگوها و روابط بین داده‌ها استفاده شود (۹). کاهش هزینه‌ها در تشخیص بیماری، دقت بالا و همچنین کاهش منابع انسانی از عمده فواید به کارگیری علم داده‌کاوی در مسائل پزشکی است (۱۰، ۱۱).

درخت تصمیم‌گیری از الگوریتم‌های محبوب طبقه‌بندی در پزشکی و زمینه‌های دیگر است (۱۲، ۱۳). این روش نوعی نمایش گرافیکی از روابط بین داده‌ها در پایگاه داده است که برای طبقه‌بندی داده‌ها استفاده می‌شود. درخت تصمیم‌گیری بیشتر در طبقه‌بندی و پیش‌بینی استفاده می‌شود و از آن می‌توان به عنوان روشی ساده و قدرتمند برای کشف دانش نهفته استفاده کرد (۱۴). همچنین پیاده‌سازی آن آسان و تفسیر نتایج آن راحت است. درخت تصمیم قادر است از روابط موجود در مجموعه داده‌ها توصیفات قابل درکی را برای انسان تولید کند و پیش‌بینی خود را در قالب قوانینی ارائه دهد که از نظر پارامترهای آماری ارزش مناسبی دارند (۱۵).

پژوهش‌های متعددی در حوزه داده‌کاوی و بررسی سلامت کارکنان انجام گرفته است که از این میان می‌توان به این پژوهش‌ها اشاره کرد. Stoia و همکاران (۲۰۱۶) با استفاده از درخت تصمیم به پیش‌بینی مسمومیت ناشی از سرب در محیط کار با بررسی آزمایش‌هایی که از کارگران به عمل آمده بود پرداختند و نشان دادند مدل‌سازی درخت تصمیم در مدیریت ریسک مسمومیت مفید است (۱۶).

Ameri و همکاران (۲۰۱۴) با استفاده از روش‌های داده‌کاوی، عوامل تأثیرگذار بر سکتة قلبی را در بیماران دیابتی بررسی کردند. نتیجه پژوهش آنها نشان داد درخت تصمیم C&R در بین الگوریتم‌های مختلف دسته‌بندی با دقت بیشتری عوامل مؤثر بر سکتة قلبی را شناسایی کرده است (۱۷).

در پژوهش حاضر با استفاده از الگوریتم درخت تصمیم C5.0 عواملی که در سلامت کارکنان و استخراج الگوها از داده‌های موجود با هدف کاهش میزان از کارافتادگی و حفظ سطح سلامت مطلوب کارکنان بیشترین تأثیر داشتند شناسایی شده است.

مواد و روش‌ها

این پژوهش از نوع توصیفی-تحلیلی است که بر روی ۱۲۶۷ نفر از کارکنان مشاغل مختلف شهر تهران در سال ۱۳۹۶ انجام شده است. پژوهش حاضر حاصل تحلیل داده‌های به دست آمده از معاینات طب کار است. یکی از وظایف مراکز طب کار در مقابل کارفرمایانی که کارکنان خود را برای انجام معاینات دوره اعزام می‌کنند، ارائه آمار و تحلیل داده‌های مربوط به وضعیت سلامت کارکنان آن سازمان است. بر این اساس داده‌ها با رعایت مسائل اخلاق پژوهش و محرمانه ماندن اطلاعات فردی کارکنان با استفاده از روش‌های داده‌کاوی تجزیه و تحلیل شد. داده‌های این تحقیق شامل ۸۰ متغیر است. این متغیرها در ۶ گروه «سابقه شخصی، خانوادگی و پزشکی»، «داده‌های عوامل زیان‌آور در محیط کار»، «داده‌های معاینات بالینی»، «داده‌های مربوط به آزمایش شمارش کامل سلول‌های خونی (Complete blood count)»، «داده‌های مربوط به آزمایش‌های زیست شیمی خون (Biochemistry)» و «داده‌های مربوط به معاینات پاراکلینیکی» تقسیم‌بندی شده‌اند. در جدول ۱ متغیرهای موجود در هر یک از گروه‌ها آورده شده است.

یکی از روش‌های متداول انجام پژوهش‌های داده‌کاوی استفاده از روش فرآیندهای استاندارد متقابل صنعت برای داده‌کاوی Cross Industry Process For Data Mining است (۱۱). این روش مراحل اجرایی مختلفی دارد که به صورت زیر در پژوهش حاضر اجرا شده است:

مرحله اول: شناخت محیط مطالعه شده

در این مرحله برای شناخت نسبت به موضوع مطالعه شده، مصاحبه‌هایی با پزشکان و دیگر عوامل اجرایی معاینات طب کار انجام گرفت. همچنین مقادیر مجاز هر کدام از متغیرهای موجود در این پژوهش

Ameri و همکاران (۲۰۱۳) در پژوهشی دیگر که روی داده‌های بیماران دیابتی انجام دادند، با استفاده از درخت تصمیم C5.0 و شبکه‌های عصبی، احتمال بروز عوارض میکروواسکولار و ماکروواسکولار را بررسی کردند و دقت بالای درخت تصمیم گیری C5.0 را در نتایج پژوهش نشان دادند (۱۸).

Choi و همکاران (۲۰۱۴) در مطالعه‌ای که در زمینه نقش حلال‌های صنعتی و فلزات سنگین بر روی شنوایی کارگران بود، به این نتیجه رسیدند که وجود فلزات سنگین و یا حلال‌های آلی در محیط کار ممکن است تأثیر سر و صدا را در از دست دادن شنوایی در محل کار تشدید کند (۱۹).

در یک مطالعه دیگر Yang و همکاران (۲۰۱۵) در یک شرکت بزرگ تولید نیکل در کشور چین با استفاده از روش‌های آماری و استفاده از متغیرهایی از قبیل میزان قند خون، چربی‌ها و غلظت فلزات در ادرار به بررسی وضعیت سلامت کارکنان به منظور اثبات رابطه بین تماس با فلزات سنگین و ابتلا به دیابت یا پیش‌دیابت پرداختند (۲۰).

Wong و همکاران (۲۰۱۱) نیز با مطالعه روی پرسشنامه‌های سلامت شغلی ۱۳ کشور آسیا و اقیانوس آرام به این نتایج رسیدند که قوانین کار در کشورهای آسیا و اقیانوس آرام بسیار متنوع است. مشکلات رایج آنها شامل کمبود نیروی انسانی، هزینه‌های بالا و آگاهی‌نداشتن کارکنان بوده است (۲۱).

Shafizadeh و همکاران (۲۰۱۷) در پژوهشی به منظور پیشگیری و شناسایی به موقع بیماری‌های شغلی، مطالعه‌ای بر روی کارکنان بیمارستان نزاجا اصفهان انجام دادند که نتایج این مطالعه نشان داد میان گروه‌بندی بر اساس سن، جنسیت و شغل کارکنان با بیماری‌های شغلی ارتباط معناداری وجود دارد (۲۲).

در مطالعه‌ای دیگر که Rafeemanesh و همکاران (۲۰۱۳) به منظور تشخیص زودرس بیماری‌های شاغلان انجام دادند، نتایج حاکی از آن بود که معاینات بدو استخدام و معاینات دوره‌ای در تشخیص زودهنگام بیماری افرادی که از بیماری خود مطلع نبودند اهمیت دارد (۲۳).

جدول ۱. متغیرهای موجود در پژوهش

ردیف	دسته بندی معاینات	متغیرهای موجود
۱	سابقه شخصی، خانوادگی و پزشکی	سن (ماه)، سطح تحصیلات، سابقه کار (ماه)، جنسیت، سابقه بیماری، تغییر علائم بیماری در محیط کار، داشتن علائم بیماری مشابه با همکاران، تغییر علائم بیماری در مرخصی، حساسیت به دارو، سابقه بستری در بیمارستان، سابقه عمل جراحی، سابقه سرطان یا بیماری مزمن در فامیل، مصرف داروی خاص، مصرف سیگار در زمان حال، سابقه مصرف سیگار در گذشته، انجام ورزش یا سرگرمی در اوقات فراغت، داشتن حادثه شغلی، غیبت بیش از ۳ روز در محیط کار به خاطر بیماری، نزدیکی منزل به مراکز صنعتی، سابقه معرفی به کمیسیون پزشکی
۲	عوامل زیان‌آور در محیط کار	صدا، ارتعاش، اشعه، گرما، گرد و غبار، بخار فلزات، حلال، اسید و باز، عوامل بیولوژیک، نشستن یا ایستادن طولانی، حمل بار سنگین، وضعیت نامناسب بدن، نوبت کاری
۳	معاینات بالینی	وزن (کیلوگرم)، فشار خون (Min & Max)، قد (سانتی‌متر)، تعداد ضربان قلب در دقیقه
۴	آزمایش شمارش کامل سلول‌های خونی Complete blood count	گلبول سفید (white blood cell) ، گلبول قرمز (red blood cell) ، هموگلوبین (Hemoglobin) ، هماتوکریت (Hematocrit) ، پلاکت (platelet)
۵	آزمایش‌های زیست‌شیمی خون (Biochemistry)	سرعت رسوب گلبول قرمز (Erythrocyte Sedimentation Rate) ، قند خون (fasting blood glucose) ، اوره خون (Urea)، کراتینین (Creatinine)، اسید اوریک (Uric Acid)، کلسترول (Cholesterol) ، تری‌گلیسیرید (Triglyceride) ، لیپوپروتئین با چگالی بالا (High Density Lipoprotein) ، لیپوپروتئین با چگالی کم (low-density lipoprotein)، آنزیم‌های کبدی (Serum glutamic- oxaloacetic transaminase & Serum Glutamic-Pyruvic Transaminase)
۶	معاینات پاراکلینیکی	شنوایی گوش‌ها در فرکانس ۵۰۰، ۱۰۰۰، ۲۰۰۰، ۳۰۰۰، ۴۰۰۰، ۶۰۰۰، ۸۰۰۰ نتیجه کلی شنوایی سنجی گوش‌ها، وضعیت اسپیرومتری، متقارن بودن چشم‌ها، دید دور چشم‌ها، دید نزدیک چشم‌ها، تست کوررنگی، تشخیص اجزا، تشخیص عمق بینایی
۷	نتیجه کلی معاینات	نتیجه کلی معاینات (قبول یا مردود (دارای محدودیت کاری))

مشخص و در پایگاه داده بر اساس مقادیر مجاز هر متغیر برچسب‌گذاری و طبقه‌بندی نرمال یا غیر نرمال بودن هر یک از رکوردها انجام شد.

مرحله دوم: آماده‌سازی داده‌ها

مبحث آماده‌سازی و پیش‌پردازش داده‌ها یکی از مهم‌ترین بخش‌های داده‌کاوی و استخراج دانش به شمار می‌رود. در این مرحله برای

سهولت در کار، مقادیر متغیرهایی که به صورت کیفی و اسمی بودند کدگذاری شدند. برای رعایت مسائل اخلاق پژوهش، داده‌هایی که مربوط به هویت افراد مطالعه‌شده بودند شامل نام و نام خانوادگی، محل سکونت و شماره تماس حذف شد. همچنین متغیر نشستن یا ایستادن طولانی مدت به دلیل اینکه در تمامی کارکنانی که در این پژوهش مطالعه شدند همه‌گیر بود و واردکردن این متغیر در فرایند مدل‌سازی بی‌ثمر تلقی شد، از مجموعه داده‌ها حذف شد. تعداد ۷ رکورد

یکی از متداول‌ترین روش‌های داده‌کاوی و استخراج الگو، مخصوصاً در حوزه پزشکی، استفاده از روش درخت تصمیم است (۱۲، ۱۳). این روش نوعی ابزار پشتیبان تصمیم‌گیری و یادگیری ماشین است که داده‌ها را بر اساس متغیرهای چندگانه طبقه‌بندی می‌کند (۲۴، ۲۵). درخت تصمیم با استفاده از سلسله انشعاب از متغیرهای مهم و اساسی به ایجاد تمایز واضح در دسته‌های تشخیصی می‌پردازد (۲۶)؛ بدین صورت که هر برگ یا گره پایانی زیرمجموعه‌ای از داده‌هایی را توصیف می‌کند که برای ساخت مدل استفاده شده است. هر گره در

یک درخت تصمیم، نماینده یک صفت در یک موضوع طبقه‌بندی شده است و هر شاخه نماینده مقداری است که یک گره می‌تواند اختیار کند و در واقع اشتراک ویژگی‌ها را نشان می‌دهد (۲۷). در این پژوهش از الگوریتم C5.0 در نرم‌افزار IBM SPSS Modeler نسخه ۱۸ برای ساخت مدل استفاده شده است. روش مدل‌سازی بدین صورت بود که در ابتدا داده‌ها به دو قسمت آموزشی (۷۰ درصد) و آزمایشی (۳۰ درصد) تقسیم شدند. در حقیقت الگوریتم با استفاده از داده‌های آموزشی مدل را می‌سازد و با استفاده از داده‌های آزمایشی مدل را ارزیابی می‌کند. همچنین از شاخص جینی (Gini Index) به عنوان معیار تقسیم شاخه‌ها استفاده شده است و با استفاده از این معیار درخت به زیر شاخه‌ها شکسته می‌شود. برای استخراج الگو، مسیر از ریشه تا برگ دنبال می‌شود و به صورت شرطی (اگر و آنگاه) بیان می‌شود. همچنین در محیط نرم‌افزار حداقل میزان اطمینان (Confidence) برای الگوهای ایجاد شده ۵۰ درصد تعیین شده است. متغیر نتیجه کلی معاینات به عنوان متغیر پیش‌بینی شونده انتخاب شده است.

مرحله چهارم: ارزیابی مدل

پژوهش‌های داده‌کاوی برای اثبات میزان کارایی مدل ساخته‌شده نیازمند روش‌هایی برای سنجش و ارزیابی نتایج مدل هستند. در این تحقیق از شاخص‌های ارزیابی Accuracy، Precision و Recall استفاده شده است. به کمک این شاخص‌ها می‌توان تشخیص داد که کدام روش رده‌بندی صحت و دقت بهتری نسبت به روش‌های دیگر دارد و همچنین مدل ایجادشده تا چه اندازه قابلیت پیش‌بینی دارد. برای به‌دست‌آوردن مقادیر معیارهای ارزیابی در این پژوهش از ماتریس اغتشاش (Confusion Matrix) استفاده شده است.

جدول ۲. ماتریس اغتشاش پیش‌بینی نتیجه کلی معاینات طب کار

		مقادیر واقعی	
		دسته مثبت	دسته منفی
مقادیر پیش‌بینی	دسته مثبت	TP	FP
	دسته منفی	FN	TN

(True Negatives) TN: بیانگر تعداد رکوردهایی است که رده واقعی آنها منفی بوده و الگوریتم رده‌بندی، آنها را منفی تشخیص داده است.

(False Positives) FP: بیانگر تعداد رکوردهایی است که رده واقعی آنها منفی بوده و الگوریتم رده‌بندی، آنها را به اشتباه مثبت تشخیص داده است.

(False Negatives) FN: بیانگر تعداد رکوردهایی است که رده واقعی آنها مثبت بوده و الگوریتم رده‌بندی، آنها را به اشتباه منفی تشخیص داده است.

(True Positives) TP: بیانگر تعداد رکوردهایی است که رده واقعی آنها مثبت بوده و الگوریتم رده‌بندی، آنها را مثبت تشخیص داده است.

با توجه به ماتریس اغتشاش جدول ۲، معیارهای ارزیابی ذکر شده در بالا به صورت زیر محاسبه می‌شوند.

$$\text{precision} = \frac{TP}{TP + FP} \quad \text{فرمول ۱:}$$

$$\text{Recall} = \frac{TP}{FN + TP} \quad \text{فرمول ۲:}$$

$$\text{Accuracy} = \frac{TN + TP}{n} \quad \text{فرمول ۳:}$$

Precision بیانگر مواردی است که به عنوان مردود یا دارای محدودیت شغلی پیش‌بینی شده‌اند و در واقعیت نیز همین طور هستند. Recall بیانگر توانایی روش در کاهش نرخ منفی کاذب (FN) در مواردی است که مدل فرد را در دسته مربوطه پیش‌بینی می‌کند. Accuracy از نسبت تعداد نمونه‌هایی که درست در دسته مدنظر تشخیص داده شده‌اند به تعداد کل رکوردها (n) به دست می‌آید. در این پژوهش به منظور انتخاب روش مناسب مدل‌سازی و مقایسه دقت مدل‌ها به جز الگوریتم C5.0، از الگوریتم‌های رایج دیگر درخت تصمیم مانند CHAID، C&R و QUEST نیز استفاده شده است. الگوریتم‌های CHAID و QUEST بر اساس معیار P-value

کار داشته‌اند، به همراه میزان تأثیرشان در جدول ۵ آورده شده است.

جدول ۵. متغیرهای دارای بیشترین تأثیر بر نتیجه معاینات

ردیف	نام متغیر	وزن اهمیت
۱	وضعیت شنوایی گوش چپ	۰/۱۸
۲	سطح شنوایی گوش چپ در فرکانس ۶۰۰۰	۰/۱۶
۳	سابقه بستری در بیمارستان	۰/۱۰
۴	سطح شنوایی گوش راست در فرکانس ۶۰۰۰	۰/۰۹
۵	سطح شنوایی گوش چپ در فرکانس ۴۰۰۰	۰/۰۹
۶	مصرف سیگار در زمان حال	۰/۰۸
۷	سطح شنوایی گوش راست در فرکانس ۴۰۰۰	۰/۰۷
۸	لیپوپروتئین با چگالی بالا (HDL Chol)	۰/۰۵
۹	گلبول قرمز (RBC)	۰/۰۵
۱۰	اوره خون (Urea)	۰/۰۵
۱۱	کراتینین (Creatinine)	۰/۰۵
۱۲	قرارداشتن در معرض بخار فلزات	۰/۰۵

یکی از مزایای استفاده از درخت تصمیم، ارائه مجموعه الگوهای است که در پیش‌بینی متغیر هدف و یا به عنوان یک سیستم کمکی برای تصمیم‌گیری پزشکی می‌تواند کاربرد داشته باشد. بر اساس خروجی مدل ارائه‌شده، الگوهای به‌دست آمده با نرم‌افزار مورد کارشناسی صاحب‌نظران قرار گرفته است و ۷ الگو به صورت قوانین «اگر و آنگاه» از داده‌های مطالعه‌شده استخراج شده که به شرح زیر است:

- اگر سطح کارمند تحصیلات کاردانی باشد، سابقه بستری در بیمارستان داشته باشد، سطح واکنش شنوایی گوش راست در فرکانس ۴۰۰۰ هرتز بالاتر از ۵۰ باشد، سطح واکنش شنوایی گوش چپ کارمند در فرکانس ۶۰۰۰ هرتز بالاتر از ۳۵ باشد، کلسترول مفید خون (HDL) بالای ۳۲ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۹۲/۳ درصد مردود است.
- اگر کارمند سیگار مصرف کند، سطح واکنش شنوایی گوش راست در فرکانس ۶۰۰۰ هرتز بالاتر از ۴۵ باشد، سطح واکنش شنوایی گوش چپ در فرکانس ۴۰۰۰ بالاتر از ۳۵ باشد، سطح اوره خون بالاتر از ۲۵ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۷۰ درصد مردود است.

مدل‌سازی را انجام می‌دهند. در الگوریتم C&R از روش رگرسیون برای مدل‌سازی و ساخت درخت تصمیم استفاده می‌شود. در جدول ۳ مقایسه بین ویژگی‌های ارزیابی الگوریتم‌های مختلف درخت تصمیم آورده شده است.

جدول ۳. معیارهای ارزیابی مدل ساخته‌شده (اعداد بر حسب درصد است.)

نوع درخت تصمیم	Recall	Precision	Accuracy
C5.0 (مدل پیشنهادی)	۸۷/۵۰	۹۰/۷۴	۹۹/۰۵
CHAID	۵۳/۵۷	۷۶/۹۲	۹۷/۲۴
C & R	۶۶/۰۷	۷۱/۱۵	۹۷/۳۲
QUEST	۶۷/۸۵	۵۷/۵۷	۹۶/۳۷

یافته‌ها

در هر پژوهش داده‌کاوی مبحث ارزیابی مدل و محاسبه نتایج معیارهای ارزیابی موضوعی مهم و اساسی تلقی می‌شود. در پژوهش حاضر به منظور تعیین بهترین الگوریتم برای ایجاد مدل پیش‌بینی نتایج معاینات طب کار، انواع مختلف الگوریتم‌های درخت تصمیم ارزیابی شد. بر اساس نتایج ارزیابی بین الگوریتم‌های مختلف که در جدول شماره ۴ نیز آورده شده، نتایج حاکی از آن است الگوریتم درخت تصمیم C5.0 بهترین عملکرد را در پیش‌بینی نتیجه معاینات طب کار دارد. اطلاعات جمعیت‌شناختی افراد شرکت‌کننده در این پژوهش در جدول ۴ آورده شده است.

جدول ۴. اطلاعات جمعیت‌شناختی افراد شرکت‌کننده در پژوهش

سن	جنسیت	شغل	سطح تحصیلات
۲۲ تا ۳۰ سال:		کارمند دفتری و مشاغل ستادی:	
۲۶۲ نفر		۵۶ نفر	زیر دیپلم: ۶۸ نفر
۳۰ تا ۴۵ سال:		تکنسین کارگاه‌های	دیپلم: ۲۰۳ نفر
۵۲۳ نفر	مرد: ۱۲۴۷ نفر	صنعتی: ۹۸۲ نفر	فوق دیپلم: ۵۴۱ نفر
۴۵ تا ۶۰ سال:	زن: ۲۰ نفر	راننده: ۴۹ نفر	لیسانس و فوق
۴۶۹ نفر		کارگر خدماتی و سایر مشاغل: ۱۸۰ نفر	لیسانس: ۴۵۵ نفر
۶۰ تا ۷۵ سال:			
۱۳ نفر			

بر اساس مدل ایجادشده با استفاده از الگوریتم درخت تصمیم C5.0 متغیرهایی که بیشترین تأثیر را در پیش‌بینی نتیجه معاینات طب

۳. اگر شنوایی گوش راست کارمند مشکل حاد داشته باشد، سطح واکنش شنوایی گوش چپ در فرکانس ۴۰۰۰ هرتز بالاتر از ۳۰ باشد، سابقه حادثه شغلی داشته باشد، سطح واکنش شنوایی گوش چپ در فرکانس ۱۰۰۰ هرتز بالاتر از ۱۵ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۷۶/۹ درصد مردود است.
۴. اگر شنوایی گوش راست کارمند مشکل حاد داشته باشد، سابقه حادثه شغلی داشته باشد، سطح واکنش شنوایی گوش چپ در فرکانس ۴۰۰۰ هرتز بالاتر از ۳۰ باشد، سطح پلاکت خون بالای ۲۱۹ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۱۰۰ درصد مردود است.
۵. اگر شنوایی گوش راست مشکل حاد داشته باشد، سطح واکنش شنوایی گوش چپ کارمند در فرکانس ۴۰۰۰ هرتز بالاتر از ۵۴/۵ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۵۴/۱۲ درصد مردود است.
۶. اگر کارمند در معرض بخار فلزات نباشد، سیگار نکشد، سطح گلبول قرمز خون کمتر از ۵/۹۲ باشد، سطح واکنش شنوایی گوش راست در فرکانس ۴۰۰۰ هرتز زیر ۵۰ باشد، سطح اوره خون بالاتر از ۲۵ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۹۹/۴ درصد قبول است.
۷. اگر گوش راست کارمند مشکل شنوایی متوسط داشته باشد، سطح شنوایی گوش چپ در فرکانس ۴۰۰۰ هرتز بالاتر از ۴۰ باشد، آنگاه وضعیت سلامت کارمند به احتمال ۷۲/۲ درصد قبول است.

بحث

پژوهش حاضر با هدف ارتقای سلامت شغلی کارکنان انجام گرفته است. در این پژوهش که بر روی داده‌های به‌دست‌آمده از معاینات دوره‌ای طب کار انجام شده است، داده‌های مربوط به ۸۰ متغیر مربوط به معاینات طب کار با استفاده از الگوریتم درخت تصمیم C5.0 بررسی شده است. خروجی حاصل از این پژوهش، شناسایی متغیرهای دارای بیشترین تأثیر بر سلامت کارکنان و همچنین

استخراج الگوهای موجود در پایگاه داده بوده است. نتایج به‌دست‌آمده از این پژوهش نشان می‌دهد وضعیت سلامت دستگاه شنوایی، نقش مهمی در نتیجه معاینات سلامت کارکنان دارد. همچنین در معاینه شنوایی سنجی، سطح شنوایی در فرکانس‌های بالا مخصوصاً ۴۰۰۰ و ۶۰۰۰ هرتز از بقیه فرکانس‌ها مهم‌تر بوده است. در پژوهشی که Pourzarea و همکاران (۲۰۱۶) روی کارکنان کارخانه لاستیک‌سازی انجام داده‌اند نیز به اهمیت سنجش شنوایی کارکنان در فرکانس‌های بالا تأکید شده است (۲۷). Rachiotis و همکاران (۲۰۰۶) مطالعه‌ای بر روی شنوایی سنجی کارکنان صنایع تولید قطعات الکترونیکی انجام داده‌اند و نتایج مشابهی با پژوهش حاضر در کاهش سطح شنوایی کارکنان در فرکانس ۴۰۰۰ هرتز داشته‌اند (۲۸). همچنین نتایج تحقیق حاضر با پژوهشی که Loukzadeh و همکاران (۲۰۱۳) در بررسی وضعیت سلامت شنوایی لکوموتورانان داشته‌اند همسو است. نتایج پژوهش آنها بیانگر افزایش آستانه شنوایی در فرکانس‌های بالا به‌خصوص در ۶۰۰۰ هرتز است که با نتایج پژوهش حاضر منطبق است (۲۹). بر اساس نتایج پژوهش حاضر، وضعیت شنوایی گوش چپ نسبت به گوش راست اهمیت بیشتری دارد. Ghorbani و همکاران (۲۰۰۶) در مطالعه‌ای که روی رابطه افت شنوایی با میزان سر و صدا و سابقه کار داشته‌اند، به این نتیجه رسیده‌اند که افت شنوایی در گوش چپ بیشتر رخ داده است (۳۰). Berjis و همکاران (۲۰۱۹) در مطالعه‌ای که به منظور بررسی افت شنوایی در رانندگان وسایل نقلیه سنگین داشتند، به این نتیجه رسیدند که گوش چپ بیشتر در معرض آسیب قرار دارد (۳۱). Bakhshian و همکاران (۲۰۱۸) نیز در بررسی افت شنوایی کارگران جوشکار، به این نتیجه رسیده‌اند که افت شنوایی بیشتر در گوش سمت چپ و در فرکانس ۶۰۰۰ رخ داده است که با پژوهش حاضر همخوانی دارد (۳۲).

بر اساس الگوهای به‌دست‌آمده در پژوهش پیش رو، بین مصرف سیگار و کاهش شنوایی ارتباط وجود دارد. Palmer و همکاران (۲۰۰۴) در پژوهشی که به منظور بررسی ارتباط بین مصرف سیگار و مشکلات شنوایی انجام داده‌اند، به این نتیجه رسیده‌اند

بوده است. بر این اساس محافظت از دستگاه شنوایی مخصوصاً در مشاغلی که عامل زیان‌آور سر و صدا در آن از حد مجاز بیشتر است، باید مورد توجه قرار گیرد. همچنین از الگوهای به‌دست‌آمده در این پژوهش می‌توان به عنوان ابزاری کمکی برای تصمیم‌گیری پزشکان طب کار به منظور تشخیص سلامت کارکنان برای انجام وظایف شغلی‌شان استفاده کرد.

تقدیر و تشکر

مقاله حاضر از پایان‌نامه کارشناسی ارشد با کد شناسایی ۱۴۱۴۰۳۰۹۵۲۰۱۸ و با حمایت معاونت پژوهش و فناوری دانشگاه آزاد اسلامی واحد تهران جنوب انجام شده است. بدین وسیله از همه افرادی که در این پژوهش همکاری داشته‌اند تشکر و قدردانی می‌شود.

تعارض در منافع

سهم تمام نویسندگان در این مطالعه یکسان است و هیچ‌گونه تضاد منافی وجود ندارد.

که مصرف سیگار می‌تواند بر شنوایی کارکنان تأثیر منفی بگذارد (۳۳). Farhadian و همکاران (۲۰۱۴) در مطالعه خود ریسک عوامل مؤثر بر افت شنوایی کارکنان یک صنعت فلزی را ارزیابی کرده‌اند و از نظر رابطه مصرف سیگار و افت شنوایی کارکنان به نتایج مشابهی با نتایج پژوهش حاضر دست یافتند (۳۴).

نتیجه‌گیری

با توجه به اینکه افراد شاغل، قسمت قابل توجهی از زمان خود را در طول شبانه روز در محیط کار سپری می‌کنند، تلاش برای ارتقای سلامت شغلی امری اجتناب‌ناپذیر تلقی می‌شود. برای ارتقای سلامت شغلی اولین گام شناسایی و اولویت‌بندی عواملی است که می‌تواند بر سلامت کارکنان تأثیر بگذارد. در پژوهش حاضر با استفاده از داده‌های معاینات طب کار، به شناسایی عوامل مؤثر بر سلامت شغلی پرداخته شده است. بر اساس نتایج به‌دست‌آمده، عملکرد دستگاه شنوایی مخصوصاً وضعیت شنوایی گوش چپ، مهم‌ترین عامل در تشخیص سلامت کارکنان در معاینات طب کار

References

- World Health Organization. Occupational Health: A Manual for Primary Health Care Workers [Internet]. 2002. Available from: <http://www.who.int/iris/handle/10665/200733>
- Howe HF. Organization and Operation of an Occupational Health Program. *Journal of Occupational and Environmental Medicine*. 1975; 17(8):528-540. https://journals.lww.com/joem/citation/1975/08000/organization_and_operation_of_an_occupational.14.aspx
- Park J, Kim Y, Kim KS. Remodeling of the Occupational Medical Examination Program in South Korea. *International Archives of Occupational and Environmental Health*. 1999; 72(6):411-417. <https://doi.org/10.1007/s004200050394> PMID:10473842
- Milovic, B. Prediction and Decision Making in Health Care Using Data Mining. *International Journal of Public Health Science*. 2012; 1(2):69-78. <https://doi.org/10.11591/ijphs.v1i2.1380>
- Candelieri A, Dolce G, Riganello F, Sannita WG. Data Mining in Neurology. In: Funatsu K, editor. *Knowledge-Oriented Applications in Data Mining*. London: InTech; 2011. <https://doi.org/10.5772/13966>
- Peyman N, Rezai-Rad M, Tehrani H, Gholian-Aval M, Vahedian-Shahroodi M, Miri HH. Digital Media-based Health Intervention on the promotion of Women's physical activity: a quasi-experimental study. *BMC public health*. 2018;18(1):134. <https://doi.org/10.1186/s12889-018-5025-5> PMID:29334970 PMCid:PMC5769504
- Al Jarullah AA. Decision Tree Discovery for the Diagnosis of Type II Diabetes. Paper presented at 2011 International Conference on Innovations in Information Technology. 25 Apr 2011; Abu Dhabi, United Arab Emirates. <https://doi.org/10.1109/INNOVATIONS.2011.5893838>
- Bellazzi R, Ferrazzi F, Sacchi L. Predictive Data Mining in Clinical Medicine: A Focus on Selected Methods and Applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2011; 1(5):416-30. <https://doi.org/10.1002/widm.23>
- Khajehei M, Etemady F. Data Mining and Medical Research Studies. Paper presented at 2010 Second International Conference on Computational Intelligence, Modelling and Simulation. 28-30 September 2010; Washington DC, USA. <https://doi.org/10.1109/CIMSiM.2010.24>

10. Jayalakshmi T, Santhakumaran A. A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks. Paper presented at 2010 International Conference on Data Storage and Data Engineering (DSDE). 9-10 February 2010; Bangalore, India.
<https://doi.org/10.1109/DSDE.2010.58>
11. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. Amsterdam: Elsevier; 2011. <https://www.elsevier.com/books/data-mining-concepts-and-techniques/han/978-0-12-381479-1>
12. Bellazzi R, Zupan B. Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. International Journal of Medical Informatics. 2008; 77(2):81-97.
<https://doi.org/10.1016/j.ijmedinf.2006.11.006> PMID:17188928
13. Delen D, Walker G, Kadam A. Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods. Artificial Intelligence in Medicine. 2005; 34(2):113-27.
<https://doi.org/10.1016/j.artmed.2004.07.002> PMID:15894176
14. Khan FS, Anwer RM, Torgersson O, Falkman G. Data Mining in Oral Medicine Using Decision Trees. World Academy of Science, Engineering and Technology. 2008; 37(1):225-230. <https://doi.org/10.5281/zenodo.1335258>
15. Soni J, Ansari U, Sharma D, Soni S. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications. 2011; 17(8):43-8.
<https://doi.org/10.5120/2237-2860>
16. Stoia M, Kurtanek Z, Oancea S. Reliability of a Decision-Tree Model in Predicting Occupational Lead Poisoning in a Group of Highly Exposed Workers. American Journal of Industrial Medicine. 2016; 59(7):575-582.
<https://doi.org/10.1002/ajim.22589> PMID:27219678
17. Ameri H, Alizadeh S, Barzegari A. Identification of Influencing Factors for Heart Attack in Diabetic Patients Using C & R Algorithm. Daneshvar Medicine. 2014; 21(112):71-82.
18. Ameri H, Alizade S, Barzegari A. Knowledge Extraction of Diabetics' Data by Decision Tree Method. Journal of Health Management. 2013; 16(53):58-72. <http://jha.iuums.ac.ir/article-1-1351-en.html>
19. Choi YH, Kim K. Noise-Induced Hearing Loss in Korean Workers: Co-Exposure to Organic Solvents and Heavy Metals in Nationwide Industries. PloS One. 2014; 9(5):e97538. <https://doi.org/10.1371/journal.pone.0097538> PMID:24870407 PMCID:PMC4037174
20. Yang AM, Cheng N, Pu HQ, Liu SM, Li JS, Bassig BA, et al. Metal Exposure and Risk of Diabetes and Prediabetes Among Chinese Occupational Workers. Biomedical and Environmental Sciences. 2015; 28(12):875-883. <https://doi.org/10.3967/bes2015.121>
21. Wong TW, Wong AH. A Review of Statutory Medical Examinations in Asian-Pacific Countries. American Journal of Industrial Medicine. 2011; 54(1):78-88.
<https://doi.org/10.1002/ajim.20897> PMID:20862702
22. ShafiZadeh S, Rezayee E, Hasanpour M. Evaluation of Occupational Health Monitoring Results Separated With Gender, Age and Occupational Groups in NEZAJA Hospital, Isfahan. Journal of Nurse and Physician Within War. 2017; 4(13):169-177. <http://npwjm.ajaums.ac.ir/article-1-378-fa.html>
23. Rafeemanesh E, Lotfi H, Taheri R, Rahimpour F. The Role of Occupational Examinations in Early Diagnosis of Diseases of Workers. Iran Occupational Health. 2013; 10 (3):12-19. <http://ioh.iuums.ac.ir/article-1-717-en.html>
24. Stiglic G, Kocbek S, Pernek I, Kokol P. Comprehensive Decision Tree Models in Bioinformatics. PloS One. 2012; 7(3):e33812. <https://doi.org/10.1371/journal.pone.0033812> PMID:22479449 PMCID:PMC3316502
25. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics. 2006; 15(3):651-674.
<https://doi.org/10.1198/106186006X133933>
26. Phua L, Smith K, Gayler R. A Comprehensive Survey of Data Mining-Based Fraud Detection Research. Clayton School of Information Technology. 2015; 77(2):81-97.
<https://doi.org/10.1016/j.chb.2012.01.002>
27. Pourzarea G, Attarchi M, Valirad F, Mohammadi S. The Effect of Simultaneous Exposure to Organic Solvents and Noise on High Frequency Hearing Loss in Tire Manufacturing Company Workers. Tibbi-i Kar Journal. 2016; 8(2):72-80. <http://tkj.ssu.ac.ir/article-1-621-fa.html>
28. Rachiotis G, Alexopoulos C, Drivas, S. Occupational Exposure to Noise, and Hearing Function Among Electro Production Workers. Auris Nasus Larynx. 2006; 33(4):381-385.
<https://doi.org/10.1016/j.anl.2006.03.008> PMID:16704911
29. Loukazadeh Z, Zare Mirakabad Z, Mahnam M, Hashemi S. Evaluation of Hearing Status of Yazd Railway's Train Drivers

- in 2011. Tibbi-i Kar Journal. 2013; 5(1):8-15. <http://tkj.ssu.ac.ir/article-1-312-en.html>
30. Ghorbani Shahna F. Noise Induced Hearing Loss and Its Relationship With Dose and Exposure Length. The Journal of Qazvin University of Medical Sciences. 2006; 10(1):84-88. <http://journal.qums.ac.ir/article-1-194-fa.html>
31. Berjis N, Soheilipoor S, Poorabdian S, Akbari S. Evaluating the Relative Frequency of Hearing Loss on Heavy Vehicles Drivers. Journal of Isfahan Medical School. 2019; 28(120): 1471-1476. <http://jjims.mui.ac.ir/index.php/jjims/article/view/821>
32. Bakhshian Shahrababaki S, Nasirpour F, Saber Siahpous S. Hearing Loss Assessment of Welders Who Referred to Health Center of Fardis. Aristotle University Medical Journal. 2018; 7(1):9-14. URL: <http://aums.abzums.ac.ir/article-1-715-en.html>
<https://doi.org/10.29252/aums.7.1.9>
33. Palmer KT, Griffin MJ, Syddall HE, Coggon D. Cigarette Smoking, Occupational Exposure to Noise, and Self Reported Hearing Difficulties. Occupational and Environmental Medicine. 2004; 61(4):340-4. <https://doi.org/10.1136/oem.2003.009183>
PMid:15031392 PMCID:PMC1740747
34. Farhadian M, Aliabadi M, Shahidi R. Risk Assessment of Influence Factors on Occupational Hearing Loss in Noise-Exposed Workers in Typical Metal Industry. Journal of Occupational Hygiene Engineering. 2014; 1(3):37-44. <http://johe.umsha.ac.ir/article-1-79-en.html>