DOI: 10.48305/jhsr.v19i2.1499 Published by Vesnu Publications

Journal of Health System Research Vol. 19, No. 2, Summer

Received: 19.12.2022 Accepted: 14.05.2023 Published: 06.07.2023

Designing a Support System for Predicting the Survival of Patients with Melanoma Based on Data Mining Algorithms

Farinaz Sanaei[1], Seyed Abdollah Amin Mousavi[2], Abbas Toloie-Eshlaghy[3], Ali Rajabzadeh-Ghatari[4]

Original Article

Abstract

Background: Melanoma is one of the most commonly diagnosed cancers and the second cause of cancer-related death among people. This disease is the rarest and most malignant type of skin cancer. In advanced conditions, it has the ability to spread to internal organs and can lead to death. In Iran, for several years, significant data about melanoma have been collected either manually or electronically, due to its prevalence and the high costs it leaves on the country's healthcare system, but despite these valuable data, the health system is still unaware of the high potential of data mining in predicting the survival of patients with melanoma. Therefore, the aim of this study was to design an intelligent system to predict the survival of these patients.

Methods: This study was practical in terms of nature and descriptive-analytical and retrospective in terms of method. The research population consisted of patients with melanoma cancer from the database of the National Cancer Research Center affiliated to Shahid Beheshti University, located in Tajrish Martyrs Hospital, Tehran, Iran (between 2007 and 2012), who were followed up for 5 years (n = 4118). SPSS and Weka software were used to design the support system for melanoma cancer survival prediction. The final model for predicting melanoma cancer survival was selected based on the evaluation indices of data mining algorithms.

Findings: Neural network algorithms, simple Bayes, Bayesian network (BN) and combination of decision tree with simple Bayes, logistic regression, J48, and ID3 were selected as the used models of the country's database. Based on the findings, the neural network performed better with a value of 0.97 in terms of accuracy and 91.03 in terms of features.

Conclusion: The performance of the neural network in all evaluation indices was statistically higher than other selected algorithms. Therefore, this algorithm was chosen as a support system for predicting melanoma cancer survival.

Keywords: Data mining; Prediction; Melanoma; Disease survival; Neural Networks; Support system

Citation: Sanaei F, Mousavi SAA, Toloie-Eshlaghy A, Rajabzadeh-Ghatari A. **Designing a Support System for Predicting the Survival of Patients with Melanoma Based on Data Mining Algorithms.** J Health Syst Res 2023; 19(2): 160-65.

1- PhD Student, Department of Information Technology Management, School of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran
2- Assistant Professor, Department of Industrial Management, School of Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran
3- Professor, Department of Information Technology Management, School of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran
4- Professor, Department of Industrial Management, School of Management and Economics, Tarbiat Modares University, Tehran, Iran
Corresponding Author: Seyed Abdolla Amin Mousavi; Assistant Professor, Department of Industrial Management, School of Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran; Email: a.mousavi@iauctb.ac.ir

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI: 10.48305/jhsr.v19i2.1499 Published by Vesnu Publications

طراحی سیستم پشتیبان برای پیش‌بینی بقای بیماران مبتلا به سرطان ملانوم مبتنی بر الگوریتم‌های داده‌کاوی

فریناز صناعی[۱]، سید عبداله امین موسوی[۲]، عباس طلوعی اشلقی[۳]، علی رجب‌زاده قطری[۴]

مقاله پژوهشی

چکیده

مقدمه: ملانوم از جمله شایع‌ترین سرطان‌های تشخیصی و دومین علت مرگ ناشی از سرطان در میان افراد است. این بیماری، نادرترین و بدخیم‌ترین نوع سرطان پوست به شمار می‌رود و در شرایط پیشرفته، توانایی انتشار به ارگان‌های داخلی را دارد و می‌تواند منجر به مرگ شود. چندین سال است که در ایران داده‌های قابل توجهی درباره ملانوم چه به صورت دستی و چه به شکل الکترونیکی، به علت شیوع آن و هزینه‌های بالایی که بر سیستم بهداشت و درمان کشور به جای می‌گذارد، جمع‌آوری شده، اما با وجود این داده‌های ارزشمند، سیستم بهداشتی هنوز از پتانسیل بالای داده‌کاوی در پیش‌بینی بقای بیماران مبتلا به ملانوم غافل مانده است. پژوهش حاضر با هدف طراحی سیستم هوشمند برای پیش‌بینی بقای بیماران مبتلا به ملانوم انجام شد.

روش‌ها: این مطالعه از نظر ماهیت، کاربردی و از نظر روش، توصیفی- تحلیلی و گذشته‌نگر بود. جامعه تحقیق را بیماران مبتلا به سرطان ملانوم پایگاه داده مرکز تحقیقات کشوری سرطان وابسته به دانشگاه شهید بهشتی، واقع در بیمارستان شهدای تجریش (بین سال‌های ۱۳۸۷ تا ۱۳۹۲) که تا ۵ سال مورد پیگیری قرار گرفته بودند (۴۱۱۸ نفر)، تشکیل داد. برای طراحی سیستم پشتیبان پیش‌بینی بقای سرطان ملانوم، از نرم‌افزارهای SPSS و Weka استفاده شد. مدل نهایی پیش‌بینی بقای سرطان ملانوم بر اساس شاخص‌های ارزیابی الگوریتم‌های داده‌کاوی انتخاب گردید.

یافته‌ها: الگوریتم‌های شبکه عصبی، Bayes ساده، Bayesian network (BN) و ترکیب درخت تصمیم‌گیری با Bayes ساده، رگرسیون لجستیک، J48 و ID3 به عنوان مدل‌های استفاده شده پایگاه داده کشور انتخاب شدند. بر اساس یافته‌ها، شبکه عصبی با مقدار ۰/۹۷ از لحاظ دقت و ۹۱/۰۳ از لحاظ ویژگی، عملکرد بهتری داشت.

نتیجه‌گیری: عملکرد شبکه عصبی در همه شاخص‌های ارزیابی از لحاظ آماری نسبت به سایر الگوریتم‌های منتخب بالاتر بود. بنابراین، الگوریتم شبکه عصبی به عنوان سیستم پشتیبان پیش‌بینی بقای سرطان ملانوم انتخاب گردید.

واژه‌های کلیدی: داده‌کاوی؛ پیش‌بینی؛ ملانوم؛ بقای بیماری؛ شبکه‌های عصبی؛ سیستم پشتیبان

ارجاع: صناعی فریناز، موسوی سید عبداله امین، طلوعی اشلقی عباس، رجب‌زاده قطری علی. **طراحی سیستم پشتیبان برای پیش‌بینی بقای بیماران مبتلا به سرطان ملانوم مبتنی بر الگوریتم‌های داده‌کاوی.** مجله تحقیقات نظام سلامت ۱۴۰۲؛ ۱۹ (۲): ۱۶۵-۱۶۰

دریافت مقاله: ۱۴۰۱/۹/۲۸ پذیرش مقاله: ۱۴۰۲/۲/۲۴ تاریخ چاپ: ۱۴۰۲/۴/۱۵

مقدمه

سرطان، یکی از مهم‌ترین بیماری‌های غیر واگیر می‌باشد که زندگی انسان را تحت تأثیر قرار داده و اثر بزرگی بر کلیه امور زندگی گذاشته است. چنانچه این رشد غیر طبیعی سلول‌ها مهار نشود، منجر به مرگ می‌گردد (۱). بر اساس آخرین آمار سازمان بهداشت جهانی (World Health Organization یا WHO)، در جوامع پیشرفته، ۱۸ تا ۲۲ درصد مردم و یا به عبارت دیگر، از هر ۵ نفر ۱ نفر در طول عمر خود به سرطان مبتلا می‌شود (۲). طبق برآوردهای انجمن سرطان آمریکا در سال ۲۰۲۲، حدود ۹۹۷۸۰ نفر به بیماری ملانوم مبتلا شدند و حدود ۷۶۵۰ نفر در اثر این بیماری جان خود را از دست دادند.

در حال حاضر با توجه به کشنده بودن سرطان، این بیماری در ایران عامل مهم مرگ و میر و آسیب‌های اجتماعی است. در نتیجه، بهداشت سرطان، تصمیم‌گیری مدیریت هوشمند و ایجاد سیستم تصمیم‌یار از مهم‌ترین اولویت‌های نظام سلامت کشور به شمار می‌رود. در صنعت سلامت، پیشگیری، غربالگری و درمان اولیه سرطان از اولویت بالایی برخوردار است (۳).

اگرچه ملانوم بدخیم پیشرفته اغلب غیر قابل درمان است، با این وجود با تشخیص زودهنگام در بسیاری از موارد در مراحل اولیه و اغلب قبل از مرحله متاستاز، می‌توان آن را درمان کرد و به میزان قابل توجهی مرگ و میر ناشی از این سرطان را کاهش داد (۵، ۴). پیشرفت فن‌آوری آسیب‌شناسی پوستی، نقش قابل توجهی در تشخیص و افزایش نرخ بقا در این بیماری دارد (۶). به عنوان مثال، بیماران مبتلا به ملانوم کمتر یا برابر با ۰/۷۵ میلی‌متر ضخامت، پیش‌آگهی

۱- دانشجوی دکتری تخصصی، گروه مدیریت فن‌آوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۲- استادیار، گروه مدیریت صنعتی، دانشکده مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران
۳- استاد، گروه مدیریت فن‌آوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۴- استاد، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
نویسنده مسؤول: سید عبداله امین موسوی؛ استادیار، گروه مدیریت صنعتی، دانشکده مدیریت، واحد تهران مرکز، دانشگاه آزاد اسلامی، تهران، ایران
Email: a.mousavi@iauctb.ac.ir

خوب و بیشتر از ۹۳ درصد آن‌ها میزان بقای ۵ ساله دارند (۷). بنابراین، تشخیص زودهنگام برای کاهش مرگ و میر مربوط به ملانوم بسیار مهم است (۸). نتایج تحقیقات اخیر نشان داده است که احتمال ابتلا به سرطان‌های دیگر در افرادی که به ملانوم مبتلا می‌شوند، بیشتر می‌باشد (۹). اگر این پیش‌بینی با دقت قابل قبولی همراه باشد، در درازمدت باعث مدیریت بهتر سرطان‌ها، طول عمر بیماران و کاهش چشمگیر هزینه‌ها می‌شود (۱۰). پیش‌بینی بقای ملانوم امری حیاتی در حوزه پزشکی است (۱۱).

با مرور پژوهش‌های انجام شده، میزان بقای بیماران مبتلا به ملانوم در ایران ۲۸/۵ درصد تخمین زده شده است (۱۲). در مطالعه دیگری، میانگین میزان بقای بیماران ۷۵ درصد به دست آمد که در زنان بیشتر از مردان بود (۱۳). میزان بقای بیماران مبتلا به ملانوما بر حسب سن، جنسیت، تعداد، میتوز، ضخامت تومور، میزان ارتشاح لنفوسیت، محل تومور و مرحله بر اساس آزمون‌های آماری، ۶/۱۲ ± ۶۱/۰۹ ماه به دست آمد (۱۴). در خصوص دستگاه تشخیص هوشمند ملانوما با استفاده از پردازش تصویر و الگوریتم یادگیری Bayes، دقت تشخیص صحیح برای تصاویر ثابت، ۹۵ درصد و برای پویش بلادرنگ، ۵۰ درصد مشاهده گردید (۱۵). برای دسته‌بندی داده‌های سرطان پستان با الگوریتم‌های Bayes ساده، J48 و شبکه عصبی، صحت گزارش شده ۹۴/۹ درصد بود (۱۶). اپیدمی ملانوم به صورت واقعی و بقا به طور کلی مستقل از ضخامت در حال بهبود می‌باشد، اما بهبودی در بقا توسط آن با اقلیت‌های خاص و زیرگروه‌های ندولار و آکرال تجربه نشده است (۱۷). در پیش‌بینی سرطان پوست با استفاده از داده‌کاوی، صحت Learning from Examples based on Rough Sets (LFRS)، VF15 و شبکه عصبی به ترتیب ۹۳/۴ و ۹۶/۲ و ۹۸ درصد به دست آمد (۱۸).

در تحقیق حاضر برای اولین بار در کشور با بررسی الگوریتم‌های داده‌کاوی بر پایگاه داده کل کشور ایران، سیستم تصمیم‌یار بقای ملانوم با بالاترین دقت طراحی گردید تا سیاست‌گذاران بهداشتی، کشور را در کنترل هرچه بهتر مرگ و میر ملانوم، افزایش کیفیت زندگی، امید به زندگی بیماران و مدیریت هوشمند سلامت یاری نماید. نتایج و الگوهای حاصل از به کارگیری الگوریتم‌های داده‌کاوی می‌تواند راهکار مؤثری در حیطه تشخیص و پیش‌بینی ابتلا به این بیماری، کاهش خطای انسانی و افزایش دقت تشخیص باشد. بنابراین، این سیستم می‌تواند به عنوان راهنمای پزشک در پیش‌بینی بقای بیماران مبتلا به سرطان ملانوم، گام مؤثری جهت پیشبرد راهبردهای هوشمندانه صنعت سلامت و بهداشت بردارد.

با توجه به این که دامنه تحقیقاتی پژوهش حاضر نسبت به مطالعات پیشین گسترده‌تر است و مقایسه الگوریتم‌های داده‌کاوی پشتیبانی‌کننده از تصمیم‌گیری بالینی در قالب یک مدل معرفی شده است، سیستم پشتیبانی برای پیش‌بینی بقای بیماران مبتلا به ملانوم در تحقیق حاضر از لحاظ قدرت، صحت و پیش‌بینی، دقیق‌ترین مدل می‌باشد و می‌تواند به عنوان سیستم پشتیبان در پیش‌بینی بقا یا فوت بیماران مبتلا به سرطان ملانوم کمک مؤثری نماید.

روش‌ها

این پژوهش از نظر ماهیت، کاربردی و از نظر هدف، به روش توصیفی- تحلیلی گذشته‌نگر و از نظر زمان به صورت مقطعی انجام شد. داده‌های مورد استفاده، بیماران مبتلا به ملانوم بین سال‌های ۱۳۸۷ تا ۱۳۹۱ بودند که در سال ۱۳۹۸ به تفکیک هر سال، در ایران بررسی شدند.

تعداد کل بیماران مورد بررسی، ۷۶۹۰ نفر بود. بیمارانی که پس از تشخیص ملانوم، کمتر از ۵ سال پیگیری شده بودند، حذف شدند. همچنین، بیمارانی که قبل از رسیدن به پنجمین سال پیگیری بیماری فوت کرده بودند و علت فوت آن‌ها غیر از ملانوم بود، حذف شدند که در نهایت، تعداد به ۴۱۱۸ نفر رسیدند. کل نمونه به دو دسته آموزش و آزمایش تقسیم‌بندی گردید. بر این اساس، از نمونه‌گیری تصادفی و با تخصیص ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای آزمون سیستم‌های طبقه‌بندی استفاده و تعداد دفعات تکرار آموزش- تست، ۱۰ مرتبه تعیین گردید. برای پیاده‌سازی سیستم پشتیبان پیش‌بینی بقای سرطان ملانوم، از نرم‌افزارهای SPSS نسخه ۲۴ (version 24, IBM Corporation, Armonk, NY) و Weka نسخه 3.6.9 استفاده شد. مدل پیش‌بینی بقای سرطان ملانوم بر اساس شاخص‌های ارزیابی الگوریتم‌های داده‌کاوی انتخاب شد.

جدول ۱ دربرگیرنده متغیرهای مهم در پیش‌بینی بقای ملانوم و توضیحات مربوط به مقادیر متغیرها می‌باشد و بر اساس بررسی پژوهش‌های بقای بیماران مبتلا به ملانوم و نظرخواهی از متخصصان مجرب جراحی پوست ایجاد شده است. برای رعایت مسایل اخلاقی، برخی مشخصات بیماران (متغیرها) محرمانه ماند. خصوصیات متغیرها با توجه به کدگذاری بین‌المللی سرطان‌ها (ICD-O یا International Classification of Diseases for Oncology) طبق توصیه WHO صورت گرفت تا بتوان اطلاعات را در مقیاس‌های منطقه‌ای، ملی و بین‌المللی مقایسه نمود. افزایش جمع‌آوری و ثبت روزافزون داده‌های پیچیده در دنیای پزشکی، دلیل استفاده از فرایند داده‌کاوی از پایگاه داده است.

برای بررسی دقیق‌تر و جزیی‌تر و همچنین، اجرای عملیات پیش‌پردازش داده‌ها همچون مدیریت مقادیر از دست رفته و پرت، مجموعه داده کل کشور وارد نرم‌افزار SPSS شد. متغیرهای پیش‌بینی‌کننده در این مجموعه شامل جنسیت، ارتشاح لنفوسیتی، وضعیت تأهل، درجه بیماری، محل تومور، مرحله سرطان، رفتار، اشعه درمانی، سن، ضخامت تومور و تعداد میتوز با توجه به ICD-O کدگذاری شد و همراه با سایر اطلاعات دموگرافیک بیماران (سن، جنسیت، محل سکونت) مورد تجزیه و تحلیل قرار گرفت.

توزیع متغیرهای پیش‌بینی‌کننده، متغیرهای پیوسته پیش‌بینی‌کننده پیوسته و متغیر وابسته در جداول ۴-۲ ارایه شده است.

مجموعه داده ملانوم که در این مرحله استفاده شد، بانک اطلاعاتی داده‌های کشوری ملانوم وابسته به دانشگاه شهید بهشتی بود که پیش‌پردازش یا شناخت داده‌ها در این فایل امکان‌پذیر نبود. بنابراین، برای شناخت اولیه، داده‌ها وارد نرم‌افزار SPSS گردید.

داده‌های ۷۶۹۰ نفر بر طبق هدف مطالعه حاضر که بررسی بیماران از سال ۱۳۸۷ تا ۱۳۹۲ بود، جمع‌آوری گردید. مجموعه داده کل سرطان پوست کشور در سال ۱۳۸۷، ۷۹۸ نفر؛ در سال ۱۳۸۸، ۸۲۰ نفر؛ در سال ۱۳۸۹، ۷۳۶ نفر؛ در سال ۱۳۹۰، ۸۲۲ نفر و در سال ۱۳۹۱، ۹۴۲ نفر گزارش شد. سپس مجموعه داده برای شناسایی داده‌های مفقود شده، ناصحیح و ناهمسان وارد نرم‌افزار SPSS گردید و با استفاده از دستور ذیل مورد بررسی قرار گرفت. متغیر وابسته، متغیری از نوع دوتایی بود که صفر و ۱ به ترتیب عدم بقا و بقا را نشان می‌داد. برای تعیین متغیر وابسته، مجموعه داده از نرم‌افزار SPSS به نرم‌افزار Excel نسخه ۲۰۱۶ وارد شد.

متغیر «کدگذاری مجدد زمان بقا» از چهار کاراکتر تشکیل می‌شود که دو نویسه اول نشانگر سال پیگیری و دو نویسه بعدی نشانگر ماه پیگیری می‌باشد.

جدول ۱. داده‌های ملانوم برای استخراج متغیرهای مورد نیاز از پرونده‌های بیماران کل کشور

متغیر	مقادیر متغیر
سن	بر اساس سن یادداشت شده در پرونده
ضخامت تومور	بر اساس ضخامت تومور اندازه‌گیری شده و درج در پرونده: I = کمتر یا مساوی ۱ میلی‌متر، II = بین ۱/۰۱ تا ۲ میلی‌متر، III = بین ۲/۰۱ تا ٤ میلی‌متر، IV = بزرگ‌تر یا مساوی ٤ میلی‌متر
جنسیت	مرد = ۱، زن = ۲
ارتشاح لنفوسیتی	۱ = خفیف، ۲ = متوسط و ۳ = شدید
وضعیت تأهل	۱ = مجرد، ۲ = متأهل، ۳ = دوجنسی و ۹ = نامشخص
درجه بیماری	۱ = درجه اول، ۲ = درجه دو، ۳ = درجه سه، ٤ = درجه چهار، ۹ = تشخیص داده نشده
محل تومور	صفر = پوست لب، ۱ = پلک چشم، ۲ = گوش خارجی، ۳ = بقیه جاهای صورت، ٤ = سر و گردن، ٥ = تنه، ٦ = اندام فوقانی و شانه، ۷ = اندام تحتانی و لگن، ۸ = بینابین نواحی، ۹ = تشخیص داده نشده
مرحله بیماری (سرطان)	IA = ۱ (۱)، IB = ۲ (۱)، IIA = ۳ (۲)، IIB = ٤ (۲)، IIC = ٥ (۲)، IIIA = ٦ (۳)، IIIB = ۷ (۳)، IIIC = ۸ (۳)، IV = ۹ (٤)
رفتار	صفر = خوش‌خیم، ۱ = نامشخص ازلحاظ خوش‌خیمی یا بدخیمی، ۲ = کارسینوم درجا، ۳ = بدخیم اولیه، ٦ = بدخیم ثانویه (متاستاتیک)، ۹ = نامشخص از لحاظ اولیه یا ثانویه
اشعه درمانی	۱ = بله، ۲ = خیر

جدول ۲. متغیرهای پیش‌بینی‌کننده در مجموعه داده‌های ملانوم مرکز تحقیقات کشوری سرطان

اسامی متغیر گروهی	تعداد مقادیر منحصر به فرد
جنسیت	۲
ارتشاح لنفوسیتی	۳
وضعیت تأهل	٤
درجه بیماری	٥
محل تومور	۱۰
مرحله سرطان	۹
رفتار	٦
اشعه درمانی	۲

ابتدا هر یک از اعداد در ستون جداگانه‌ای در نرم‌افزار Excel تقسیم بر صد شد و سپس با به کارگیری دستور ذیل، متغیر وابسته ایجاد گردید.

```
If STR > = 5 year and VSR = alive then
Record = 1
Else if STR < 5 year and COD = melanom then
Record = 0
Else
رکورد را نادیده بگیر
End if
```

رکوردهایی که در هیچ‌یک از گروه‌های بقا و عدم بقا قرار نگرفته بود، حذف شد و تعداد رکوردها به ۴۱۱۸ رکورد رسید. اگر میزان مقادیر مفقود شده مجموعه داده‌ها کمتر از ۱ درصد باشد، در فرایند استخراج دانش خللی ایجاد نمی‌کند. چنانچه بین ۱ تا ۵ درصد قابل مدیریت، بین ۵ تا ۱۵ درصد مستلزم اعمال روش‌های پیچیده برای مدیریت و بیشتر از ۱۵ درصد ممکن است به صورت جدی بر هر نوع تفسیر دانشی تأثیرگذار باشد. بنابراین، برای مدیریت این مقادیر، به جای حذف آن‌ها که باعث از دست رفتن اطلاعات ارزشمندی می‌شود، از خطمشی‌های خاصی استفاده شد.

جدول ۳. متغیرهای پیوسته پیش‌بینی‌کننده در مجموعه داده‌های ملانوم مرکز تحقیقات کشوری سرطان

اسامی متغیر پیوسته	میانگین ± انحراف معیار	بازه (محدوده)
سن	۵۹/۲۵ ± ۱۹/۳۳	۱-۱۰۱
ضخامت تومور	۲/۳۹ ± ۲/۲٤	۰/۰۱-۹/۰۵
تعداد میتوز	۲/۰٦ ± ۱/۲۰	۱-۱۳

البته لازم به ذکر است که برای اعمال خطمشی‌های خاص، باید مقادیر مفقود شده به صورت کاملاً تصادفی رخ داده باشند؛ بدین معنی که فقدان مقادیر، هیچ وابستگی به خود مقدار یا مقادیر متغیرها نداشته باشد. به عبارت دیگر، احتمال این که یک متغیر مقادیر مفقوده داشته باشد، در همه رکوردها مشابه باشد.

جدول ۴. توزیع متغیر وابسته پایگاه داده‌های کشوری

دسته	تعداد رکورد (درصد)
صفر = عدم بقا	۱٤۰٤ (۳٤/۱)
۱ = بقا	۲۷۱٤ (٦٥/۹)
جمع کل	٤۱۱۸ (۱۰۰)

یافته‌ها

بررسی نتایج داده‌های ۷ الگوریتم داده‌کاوی (در نرم‌افزار Weka) مورد استفاده جهت بررسی بقای افراد مبتلا به سرطان ملانوم نشان داد که از میان این الگوریتم‌ها، طبق الگوریتم شبکه عصبی از لحاظ صحت، حساسیت و ویژگی بالاتر از دیگر الگوریتم‌ها بود (جدول ۵).

جدول ۵. حساسیت، صحت، ویژگی و دقت الگوریتم‌های پیش‌بین بقای بیماران مبتلا به سرطان ملانوم در پایگاه داده‌های کشوری

مدل	صحت (درصد)	حساسیت (درصد)	ویژگی (درصد)	دقت (درصد)	ROC (درصد)
BN	۹۰/۸۰	۹۳/۱٤	۸٦/۱۸	۹۰/۸۰	۹٥/٤۰
J48	۹۳/۲۰	۹۷/۱٦	۸٥/٤۷	۹۳/۲۰	۹۳/۷۰
شبکه عصبی	۹۳/٥۰	۹۷/۷۹	۹۱/۰۳	۹۷/۰۰	۹٦/۱۰
رگرسیون لجستیک	۹۲/۷۰	۹٦/۳٥	۸٥/٥٤	۹۲/۷۰	۹٥/٦۰
ترکیب درخت تصمیم و Bayes ساده	۹۰/۳۰	۹۰/۲۳	۹۰/٤٥	۹۰/۷۰	۹٥/۷۰
Bayes ساده	۹۱/۳۰	۹۳/٦٥	۸۹/٤٥	۹۱/۳۰	۹٥/۷۰
ID3	۹۱/۳۰	۹۳/٦۲	۸٦/۷٥	۹۱/۲۸	۹۱/۳۰

BN: Bayesian network; ROC: Receiver operating characteristic

شاخص Kappa (عملکرد) برای الگوریتم‌های استفاده شده در پژوهش در پایگاه‌های کشوری در جدول ۶ ارایه شده است.

جدول ۶. عملکرد Kappa الگوریتم‌های پایگاه داده‌های کشور

الگوریتم	Kappa
شبکه عصبی	۰/۹٥۲
J48	۰/۸٤٤
رگرسیون لجستیک	۰/۸۳۳
BN	۰/۸۰٤
Bayes ساده	۰/۸۰٦
ID3	۰/۸۰٥

BN: Bayesian network

در حالت کلی، الگوریتم‌های اجرا شده در پایگاه داده‌های کشوری از لحاظ عملکرد در شاخص‌های ارزیابی نسبت به یکدیگر تفاوت معنی‌داری را نشان دادند (χ^2 = ۰/۰۰۱، P < ۰/۰۰۱).

به منظور مقایسه دو به دوی مدل‌های (الگوریتم‌های) اجرا شده در پایگاه داده‌های کشوری با یکدیگر از آزمون Wilcoxon استفاده شد. تفاوت معنی‌داری بین مدل‌ها با یکدیگر در سطح معنی‌داری ۰/۰۰۷ گزارش گردید. همچنین، نتایج تصحیح Bonferroni (لازم به ذکر است در این مورد نیاز است برای تصحیح نرخ خطا از روش Bonferroni استفاده شود) نشان داد که مدل‌های شبکه عصبی، Bayes ساده، BN و ترکیب درخت تصمیم‌گیری با Bayes ساده با ID3، J48 و رگرسیون لجستیک در شاخص‌های ارزیابی تفاوت معنی‌داری با یکدیگر داشتند؛ در حالی که عملکرد الگوریتم‌های شبکه عصبی، Bayes ساده و BN و ترکیب درخت تصمیم‌گیری با Bayes ساده با یکدیگر متفاوت نبودند.

در مقام مقایسه این چهار الگوریتم، شبکه عصبی علاوه بر شاخص‌های ارزیابی، از نظر سطح زیر منحنی (۰/۸۷۹) و شاخص ترجیح یافته (۰/۹۵۲) از تمامی الگوریتم‌ها مناسب‌تر بود. در پایگاه داده کل کشور، الگوریتم شبکه عصبی از لحاظ شاخص‌های ارزیابی شده با بالاترین صحت (۹۳/۵ درصد)، دقت (۹۷/۰۰ درصد)، حساسیت (۹۷/۷۹ درصد) و ویژگی (۹۱/۰۳ درصد)، نسبت به سایر الگوریتم‌ها برتری داشت.

بحث

هدف از انجام پژوهش حاضر، طراحی سیستم پشتیبان برای پیش‌بینی بقای بیماران مبتلا به ملانوم بود که با استفاده از داده‌کاوی، هوش مصنوعی و بررسی پایگاه داده‌های واقعی کل کشور انجام شد. الگوریتم شبکه عصبی با توجه به شاخص‌های ارزیابی، به عنوان سیستم پشتیبان معرفی گردید. از مقایسه الگوریتم‌های مختلف بررسی شده، مدل پیش‌بینی بقای ملانوم در مطالعه حاضر به دست آمد. در پایگاه داده‌های کل کشور، به ۱۱ متغیر اساسی اشاره شد که در مقایسه با تحقیق انجام شده در خارج از کشور، متغیر «نژاد» به علت تعدد نژاد، تفاوت مدت زمان بقای ملانوم در نژادهای مختلف و رابطه آن با زمان بقا (۱۹) با نتایج بررسی حاضر هم‌راستا نبود. نتایج پژوهش‌های انجام شده در زمینه میزان بقای ملانوما بر حسب سن (۱۲)، جنسیت (۱۲)، محل تومور (۱۴، ۱۲)، مرحله بیماری (۱۴)، دستگاه تشخیص هوشمند ملانوما (۱۵) و دقت تشخیص (۱۵)، با یافته‌های مطالعه حاضر همخوانی داشت، اما در خصوص روش انجام کار (۱۳)، معیارهای ارزیابی (۱۳) و عوامل مؤثر دیگر همچون سن (۱۴)، درجه بیماری (۱۴)، تعداد میتوز (۱۴)، استفاده از سیستم‌های هوشمند (۱۴)، داده‌کاوی و معیارهای ارزیابی (۱۴)، نتایج متفاوتی با یافته‌های تحقیق حاضر به دست آمد. در پژوهشی که به مقایسه سه الگوریتم متفاوت Bayes ساده، J48 و شبکه عصبی برای دسته‌بندی داده‌های سرطان پرداخت (۱۶)، دقت عملکرد الگوریتم شبکه عصبی و نحوه ارزیابی آن با یافته‌های مطالعه حاضر هم‌راستا، اما در معیارهای دیگر ارزیابی همچون حساسیت و ویژگی، همسو نبود.

نقص در ثبت پرونده‌های مورد بررسی برای استخراج متغیرهای مربوط به پایگاه داده‌های کشوری که متعلق به بازه زمانی مشخص شده می‌باشد، از محدودیت‌های انجام تحقیق حاضر بود. پیشنهاد می‌شود پژوهشگران عملکرد مدل ایجاد در پایگاه داده بومی را با پایگاه‌های داده منتخب سایت‌های مختلف مقایسه کنند و با ادغام چند پایگاه داده ثبت اطلاعات بیماران ملانوم، تعداد رکوردها را افزایش دهند و از متغیرهای کاربردی‌تری استفاده نمایند.

نتیجه‌گیری

پیشنهاد می‌شود جهت سیستم پشتیبان تصمیم برای بقای بیماران مبتلا به ملانوم، از شبکه عصبی استفاده گردد. این مدل در همه شاخص‌های ارزیابی تفاوت معنی‌داری نسبت به سایر مدل‌ها داشت و از لحاظ قدرت و پایایی، عملکرد بهتری را از خود نشان داد. از آن‌جایی که الگوریتم‌های اجرا شده در

پایگاه داده‌های کشور، مقادیر فوتی را بهتر پیش‌بینی می‌کنند، در صورت عدم وجود متخصصان پوست، الگوریتم شبکه عصبی به عنوان تصمیم‌یار هوشمند، جایگزین مناسبی برای آن‌ها خواهد بود.

تشکر و قدردانی

بدین وسیله نویسندگان از تمام افرادی که در انجام این مطالعه همکاری نمودند، تشکر و قدردانی به عمل می‌آورند.

References

1. Dolfe R, Matinzadeh K. Investigating skin cancer with unsupervised learning. Stockholm, Sweden: KTH Royal Institute of Technology; 2019.
2. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al. SEER Cancer Statistics Review, 1975-2014. Bethesda, MD: National Cancer Institute; 2018.
3. Saad H, Nagarur N. Data mining techniques in predicting breast cancer. J Applied Sci 2020; 20(4): 124-33.
4. Do T, Hoang T, Pomponiu V, Zhou Y, Chen Z, Cheung M, et al. Accessible melanoma detection using smartphones and mobile image analysis. IEEE Trans Multimedia 2018; 20(10): 2849-64.
5. Habif TP, Dinulos JG, Chapman MS, Zug KA. Skin disease: Diagnosis and treatment. Philadelphia, PA: Elsevier Mosby; 2017.
6. Dorj UO, Lee KK, Choi JY, Lee M. The skin cancer classification using deep convolutional neural network. Multimed Tools Appl 2018; 77(8): 9909-24.
7. Joshi A, Khosravy M, Gupta N. Machine learning for predictive analysis. Proceedings of ICTIS 2020. Singapore: Springer; 2021.
8. Saritas MM, Yasar A. Performance analysis of ANN and Naive Bayes classification algorithm for data classification. International Journal of Intelligent Systems and Applications in Engineering 2019; 7(2): 88-91.
9. Chen S, Han C, Miao X, Li X, Yin C, Zou J, et al. Targeting MC1R depalmitoylation to prevent melanomagenesis in redheads. Nat Commun 2019; 10(1): 877.
10. Lobo S, Pallavi MS. Predicting protein in cancer diagnosis using effective classification and feature selection technique. Proceedings of the 2018 International Conference on Communication and Signal Processing (ICCSP); 2018 Apr 3-8; Chennai, India.
11. Burd CE, Liu W, Huynh MV, Waqas MA, Gillahan JE, Clark KS, et al. Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. Cancer Discov 2014; 4(12): 1418-29.
12. Esmaili N, Mohagegi MA, Safayi Naraghi Z, Emami Razavi SZ, Chavoshi M, Emami Razavi SH. Five-year survival rate of cutaneous malignant melanoma in Iran. Res J Biol Sci 2008; 3(11): 1261-4.
13. Safaii Naraghi Z, Bahadori M, Ehsani AH, Robati RM, Ghiasi M, Nozan Z. Evaluation of primary cutaneous malignant melanoma according to Breslow and Clarke pathological indices. Tehran Univ Med J 2006; 64(5): 79-86. [In Persian].
14. Ebrahimzade Ardakani M, Akhavan Tafti M, Dehghanizade N. Survival analysis in patients with malignant melanoma in Yazd. Dermatol Cosmet 2014; 5(3): 124-30. [In Persian].
15. Vahdani N, Haghi pour S, Shams Asr A. Intelligent and real-time diagnosis of melanoma skin cancer by using image processing and bayesian learning algorith [MSc Thesis]. Tabriz, Iran: Seraj Higher Education Institute; 2019. [In Persian].
16. Soria D, Garibaldi JM, Biganzoli E, Ellis IO. A comparison of three different methods for classification of breast cancer data. Proceedings of the 2008 Seventh International Conference on Machine Learning and Applications; 2008 Dec 11-13; San Diego, CA, USA.
17. Shaikh WR, Dusza SW, Weinstock MA, Oliveria SA, Geller AC, Halpern AC. Melanoma thickness and survival trends in the United States, 1989 to 2009. J Natl Cancer Inst 2016; 108(1).
18. Beenish H, Fahad M. Skin cancer prediction using data mining and its techniques: A Review. Proceedings of the 2020 International Conference on Computing and Information Technology (ICCIT-1441); 2020 Sep 9-10; Tabuk, Saudi Arabia.
19. Gamba CS, Clarke CA, Keegan TH, Tao L, Swetter SM. Melanoma survival disadvantage in young, non-Hispanic white males compared with females. JAMA Dermatol 2013; 149(8): 912-20.