

شناسایی علایم و بیماری‌های زمینه‌ای مرتبط با بیماری کووید-۱۹ و پیش‌بینی وضعیت مرگ با کمک رگرسیون لوژستیک و شبکه‌های عصبی مصنوعی: یک رویکرد داده‌کاوی

نسرین تلخی^۱، نوشین اکبری شارک^۱، زهرا رجب زاده^۱، مریم سالاری^۲، سید مسعود ساداتی^۳، محمد تقی شاکری^۴

۱- کارشناسی ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد، مشهد، ایران

۲- استادیار آمار زیستی، مرکز مدیریت و فناوری اطلاعات، دانشگاه علوم پزشکی مشهد، مشهد، ایران

۳- کارشناسی ارشد مدیریت فناوری اطلاعات، مرکز آمار و مدیریت فناوری اطلاعات، بیمارستان امام رضا، دانشگاه علوم پزشکی مشهد، مشهد، ایران

۴- استاد آمار زیستی، مرکز تحقیقات عوامل اجتماعی تعیین کننده سلامت، دانشگاه علوم پزشکی مشهد، مشهد، ایران

چکیده

مقدمه و اهداف: شیوع و نرخ مرگ‌ومیر بالای بیماری کووید-۱۹، علائم، اطلاعات جمعیت شناختی و بیماری‌های زمینه‌ای مؤثر در پیش‌بینی مرگ ناشی از آن را ضروری می‌سازد. لذا در این مطالعه قصد داریم به پیش‌بینی رفتار مرگ‌ومیر ناشی از کووید-۱۹ در استان خراسان رضوی بپردازیم.

روش کار: در این مطالعه داده‌های کامل ۴۷۴۶۰ نفر از بیماران بستری در بیمارستان‌های استان خراسان رضوی از ۴ اسفند ۱۳۹۸ تا ۲۱ شهریور ۱۴۰۰ جمع‌آوری شد. برای تشخیص بازماندگان و غیر بازماندگان ناشی از کووید-۱۹ روش شبکه‌های عصبی و رگرسیون لوژستیک و برای مقایسه دو مدل از حساسیت، ویژگی، صحت پیش‌بینی و سطح زیر منحنی مشخصه عملکرد استفاده گردید.

یافته‌ها: کاهش سطح هوشیاری، سرفه، درصد اکسیژن خون کمتر از ۹۳٪، سن، سرطان، بیماری‌های مزمن کلیه، تب داشتن، سردرد داشتن، سیگاری بودن، و بیماری‌های مزمن خون به‌عنوان ده عامل مهم‌تر در پیش‌بینی مرگ شناسایی شدند. صحت مدل شبکه عصبی و رگرسیون لوژستیک به ترتیب برابر ۸۹/۹۰٪ و ۸۳/۶۷٪ درصد، همچنین حساسیت، ویژگی و سطح زیر منحنی راک در دو مدل به ترتیب (۷۶/۱۴٪، ۶۸/۹۴٪)، (۹۱/۹۹٪، ۸۵/۳۰٪) و (۷۷/۱۴٪، ۶۸/۹۸٪) بود.

نتیجه‌گیری: یافته‌های ما اهمیت برخی اطلاعات جمعیت شناختی، بیماری‌های زمینه‌ای و علائم بالینی را ارائه کرد. همچنین، مدل شبکه عصبی می‌تواند مرگ را با دقت بیشتری نسبت به مدل رگرسیون لوژستیک پیش‌بینی کند. با این حال، تحقیقات پزشکی در این زمینه با به‌کارگیری سایر روش‌های یادگیری ماشین و قدرت بالای آن‌ها، نتایج کامل‌کننده‌ای به دنبال خواهد داشت.

اطلاعات مقاله

تاریخ دریافت

۱۴۰۱/۰۴/۱۲

تاریخ پذیرش

۱۴۰۱/۰۹/۱۹

نویسنده رابط

محمد تقی شاکری

ایمیل نویسنده رابط

ShakeriMI@mums.ac.ir

نشانی نویسنده رابط

مرکز تحقیقات عوامل اجتماعی تعیین

کننده سلامت، دانشگاه علوم پزشکی

مشهد، مشهد، ایران

واژگان کلیدی: یادگیری ماشین،

کووید-۱۹، شبکه‌های عصبی،

داده‌کاوی، رگرسیون لوژستیک

مقدمه

کووید-۱۹ در نقاط مختلف جهان، اولین موارد قطعی کووید-۱۹ در ایران در ۱ اسفند ۱۳۹۸ اعلام شد (۴).

این ویروس می‌تواند بین ۶ ساعت تا ۷ روز بسته به نوع سطح، بر روی سطوح زنده بماند و به سرعت از انسان به انسان منتقل شود (۵). علائم شایع این ویروس شامل خستگی، ضعف عمومی، مشکل در تنفس، درد قفسه سینه، گلودرد، تب، دیسترس حاد تنفسی، درد عضلانی و غیره است در حالیکه برخی از افراد آلوده هیچ علامتی ندارند (۵، ۶). میزان مرگ‌ومیر در سالمندان و گروه‌های پرخطر شامل افراد مبتلا به بیماری‌های قلبی عروقی،

در آذرماه ۱۳۹۸، یک ویروس کرونای جدید برای اولین بار در شهر ووهان، استان هوبئی چین ظاهر شد (۱). شیوع بالای آن باعث شد که ویروس به‌سرعت در سراسر جهان گسترش یابد و به یک بیماری همه‌گیر تبدیل شود و در نتیجه زندگی روزمره مردم را تحت تأثیر قرار دهد (۲). گزارش‌های اخیر سازمان بهداشت جهانی (WHO) نشان می‌دهد که این ویروس در همه کشورها گسترش یافته و زندگی شخصی، اقتصاد، صنعت و غیره را تحت تأثیر قرار داده است (۳). به دنبال همه‌گیری بیماری

سطوح شیوع کم، متوسط و بالا تقسیم شدند. مرگومیر به علت کووید-۱۹ به جنسیت، نژاد/قومیت و محیط بیرون بستگی داشت (۱۹). با توجه به اینکه روش‌های یادگیری ماشین از جمله شبکه عصبی دارای نتایج امیدوارکننده‌ای در تحقیقات پزشکی بوده است، لذا در این تحقیق این نوع الگوریتم نیز مورد استفاده قرار گرفته است. شبکه عصبی برای تشخیص، طبقه‌بندی و پیش‌بینی استفاده می‌شود و یکی از مزیت‌های آن استفاده برای مجموعه داده‌های بزرگ و بررسی ارتباطات غیرخطی و اثرات متقابل پیچیده بین عوامل است (۲۰). این مدل بر پایه پیش‌فرض‌ها نیست و قادر به تشخیص ارتباطاتی است که برخی از مدل‌های آماری مانند رگرسیون چندگانه قادر به تشخیص آن نمی‌باشند (۲۱). رگرسیون لوژستیک نیز یکی از روش‌های متداول در طبقه‌بندی و پیش‌بینی در مطالعات با پیامدهای دوسطحی یا چند سطحی است (۲۲)، که با توجه به قابلیت برآورد احتمال با استفاده از فرمول لوژستیک به‌طور گسترده در مطالعات اپیدمیولوژی مورد استفاده قرار می‌گیرد (۲۳). با توجه به تفاوت‌های ذاتی این دو مدل، مقایسه عملکردهای آن‌ها و برتری آن‌ها در داده‌های متفاوت مورد توجه است. لذا در پژوهش حاضر باهدف شناسایی عوامل مرتبط با وضعیت مرگ افراد و بررسی اهمیت ویژگی‌های مؤثر بر مرگ از دو مدل رگرسیون لوژستیک و تکنیک شبکه‌های عصبی مصنوعی که از جمله روش‌های یادگیری ماشین است استفاده می‌شود و درنهایت به مقایسه عملکرد آن‌ها می‌پردازیم.

روش کار

طراحی مطالعه و داده‌های پژوهش

داده‌های مورد استفاده در این طرح مربوط به بیماران مبتلا به کووید-۱۹ که در بیمارستان‌های علوم پزشکی در سطح استان خراسان رضوی پرونده پزشکی تشکیل داده‌اند می‌باشد. اطلاعات ثبت‌شده این بیماران از ۴ اسفند ۱۳۹۸ تا ۲۱ شهریور ۱۴۰۰ از سامانه وزارتی مدیریت اورژانس بیش بیمارستانی و حوادث دانشگاه با رعایت موازین اخلاقی دریافت شد. جهت محرمانه ماندن اطلاعات بیماران، اطلاعات بدون نام افراد و تنها با شماره پرونده مورد دسترسی و تجزیه و تحلیل قرار گرفت. لازم به ذکر است که این سامانه از ابتدا برای ثبت اطلاعات ستاد هدایت و بیماران ارجاعی بین بیمارستان‌های کشور طراحی شده است. با شیوع کووید-۱۹ فرم ثبت مشخصات بیماران مبتلا به سندروم حاد تنفسی در این سامانه طراحی و کلیه بیمارستان‌های سطح کشور موظف به تکمیل اطلاعات بیماران محتمل و قطعی کووید-۱۹ در آن شدند. فرم ثبت اطلاعات بیماران شامل مشخصات جمعیت‌شناختی، اطلاعات هویتی، بیماری‌های زمینه‌ای، علائم

دیابت، بیماری مزمن تنفسی و فشارخون به‌طور قابل‌توجهی بیشتر از افراد سالم است (۷).

کووید-۱۹ به چهار نوع خفیف، متوسط، شدید و بحرانی تقسیم می‌شود (۸). درحالی‌که اکثر افراد آلوده علائم خفیفی دارند، برخی افراد ممکن است به‌سرعت بمیرند. بسیاری از افراد آلوده به کووید-۱۹ بدون علامت هستند، اما می‌توانند ویروس را به دیگران منتقل کنند. این امر تشخیص بیماری را بر اساس علائم بالینی دشوار می‌کند (۹). تشخیص زودهنگام ویروس مسیر انتقال را قطع می‌کند و درمان به‌موقع عوارض و گسترش آن را کاهش می‌دهد که نکاتی کلیدی برای کنترل بیماری هستند (۸).

با توجه به اینکه دولت‌ها مداخلات و راهکارهایی مانند حفظ فاصله اجتماعی، پوشیدن ماسک، ماندن در خانه، عدم تجمع در اماکن عمومی و غیره را برای کاهش روند همه‌گیری به اجرا گذاشتند (۱۰)، با این حال، پس از گذشت حدود دو سال و انجام مداخلات مختلف برای مقابله با بیماری همه‌گیر، این بیماری همچنان باعث مرگ بسیاری از افراد می‌شود (۱۱). با توجه به شیوع گسترده ویروس، امکان ترکیب مجدد ژنوم‌ها و ایجاد یک کروناویروس جدید وجود دارد. بنابراین، این بیماری عفونی احتمالاً به‌صورت دوره‌ای در انسان ظاهر می‌شود (۱). با توجه به اینکه تعداد مرگومیر به علت ابتلا به این بیماری رو به افزایش است و نتایج قطعی در مورد این بیماری وجود ندارد، لذا تحقیقات بیشتری در مورد عواملی که می‌توانند باعث تعجیل در مرگومیر شوند، احساس می‌شود.

در شرایط فعلی، تحلیلگران داده با کمک مدل‌های آماری، ریاضی و یادگیری ماشین می‌توانند نقش مهمی در پیش‌بینی و تشخیص مرگ ایفا کنند و این اطلاعات را در اختیار دولت‌ها و سیاستمداران قرار دهند. امروزه روش‌های آماری و الگوریتم‌های یادگیری ماشین به‌طور گسترده‌ای در زمینه پزشکی استفاده می‌شود و نتایج موفقیت‌آمیزی داشته است (۲، ۱۲). از سوی دیگر، هوش مصنوعی و رگرسیون لوژستیک برای تشخیص مسائل در بسیاری از مطالعات استفاده شده است (۹، ۱۳-۱۷).

در یک مطالعه سیستماتیک و مروری در ایران با استفاده از رگرسیون لوژستیک، تأثیر قابل توجه افزایش سن، سیگار کشیدن، فشارخون بالا، دیابت، بیماری قلبی عروقی، بیماری تنفسی، بیماری کلیوی و بدخیمی با مرگومیر همراه بود. تعداد گلوبول‌های سفید بالاتر، نوتروفیل‌ها، فریتین، پلاکت پایین، و تعداد لنفوسیت‌ها همگی با مرگومیر مرتبط بودند (۱۸). در مطالعه مروری دیگر، ۳۱۲۵ شهرستان آلوده در ایالات متحده آمریکا با استفاده از تحلیل خوشه‌ای به سه کلاس مربوط به

$$\text{Logit}[\pi(X)] = \log \left[\frac{\pi(X)}{1 - \pi(X)} \right] = \alpha + \beta X$$

شبکه‌های عصبی مصنوعی یکی از تکنیک‌ها در حوزه هوش مصنوعی و از دهه ۱۹۹۰ دارای کاربردهای متعدد در تشخیص پزشکی بوده است. مدل‌های شبکه عصبی مصنوعی سعی می‌کنند ارتباطی بین داده‌های ورودی و خروجی پیدا کنند. به عبارت دیگر، آنها سعی می‌کنند مشابه مغز انسان عمل کنند. همچنین، شبکه‌های عصبی مصنوعی ابزارهای هوشمند پیشرفته‌ای برای محققان و زمینه‌های تشخیص پزشکی مانند سرطان، بیماری‌های قلبی، هیپاتیت و توسعه دارو هستند (۲۷) یکی از مراحل اصلی و حائز اهمیت در داده‌کاوی، مرحله پیش‌پردازش داده‌ها است. هرگونه پاک‌سازی، کدگذاری مجدد، حذف یا جایگزین کردن داده‌های از دست‌رفته، نرمال‌سازی ورودی‌ها، انتخاب ورودی‌های کامل با مقادیر معتبر و همچنین تعیین مجموعه داده‌های آموزشی و آزمایشی، همه بخشی از مرحله آماده‌سازی هستند.

در مرحله پیش‌پردازش، داده‌های نادرست، نامرتب، ناقص و داده‌های گمشده حذف شدند و در نهایت یک مجموعه داده کامل با ۳۲ متغیر و ۴۷۴۶۰ نمونه یا مشاهده باقی ماند. در فرآیند ساخت مدل، ۹۰٪ نمونه‌ها به صورت تصادفی به داده‌های آموزشی و ۱۰٪ باقی‌مانده داده‌ها به مجموعه داده‌های آزمایشی جهت ارزیابی عملکرد مدل اختصاص داده شد. در مرحله مدل‌سازی، مدل بر روی مجموعه داده آموزشی ساخته می‌شود (وزن‌های مناسب و بهینه‌ی شبکه که بین متغیرهای ورودی و نرون‌های لایه میانی و هدف ارتباط برقرار می‌کنند، به دست می‌آیند) و با اعمال ورودی‌ها یا متغیرهای پیش‌گو، مقادیر هدف را پیش‌بینی می‌کند. به عبارت دیگر تحت شرایط زیر تابع هدف پیش‌بینی می‌شود:

$$Y_i = \begin{cases} 0 & \text{if } \sum w_i X_i > 0 \\ 1 & \text{if } \sum w_i X_i \leq 0 \end{cases}$$

برای شناسایی متغیرهای مهم و تأثیرگذار از رگرسیون لوژیستیک با رویکرد پس‌رونده استفاده شد و در نهایت متغیرهای معنادار و مهم جهت به دست آوردن یک مدل ساده و قابل تفسیر در مدل باقی ماندند.

فرآیند بهینه‌سازی شبکه عصبی با استفاده از روش اعتبارسنجی متقاطع تکرارشونده (Repeated 10-fold cross-validation) به تعداد پنج مرتبه تکرار، اجرا شد. در مرحله ارزیابی مدل، عملکرد مدل یا طبقه‌بندی‌کننده توسط ماتریس درهم‌آمیختگی و شاخص‌های مرتبط با آن مانند صحت،

بیماری، تاریخ مراجعه، نتایج آزمون پی‌سی‌آر، نتیجه سی‌تی‌اسکن ریه، تاریخ ترخیص یا فوت و نام بخش بستری است. کاربران بیمارستان‌ها به محض ورود بیمار و تشخیص پزشک به ابتلای قطعی یا حتمی بیمار به کووید-۱۹، اطلاعات فوق را در هنگام وقوع به‌طور کامل در سامانه MCMC ثبت می‌کنند. این سامانه از طرف مرکز مدیریت حوادث و فوریت‌های پزشکی که زیرمجموعه معاونت درمان است، پایش می‌شود. اطلاعات این سامانه در خصوص بیماری کووید-۱۹ یکی از اصلی‌ترین منابع تحلیل و رصد وضعیت پاندمی در ایران است. متغیرهای سن (Age)، جنسیت (Sex)، تب (Fever)، سرفه (Cough)، درد عضلانی (Muscular pain)، ناراحتی تنفسی (Distress)، کاهش سطح هوشیاری (Conscious)، کاهش حس بویایی (Decreased sense of smell)، کاهش حس چشایی (Decreased sense of taste)، تشنج (Convulsions)، سردرد داشتن (Headache)، سرگیجه (Confusion)، درد قفسه سینه (Chest pain)، التهاب پوست (Skin inflammation)، معده درد (Stomachache)، تهوع (Nausea)، استفراغ (Vomit)، اسهال (Diarrhea)، بی‌اشتهایی (Anorexia)، وضعیت سیگار کشیدن (Smoking status)، مصرف موارد مخدر (Drug use)، PO2، سرطان (Cancer)، بیماری‌های مزمن کبدی (Liver disease)، دیابت (Diabetes)، بیماری‌های مزمن خون (Blood diseases)، نقص ایمنی (Immunodeficiency)، بیماری‌های قلبی (Heart disease)، بیماری‌های مزمن کلیوی (Kidney disease)، آسم (Asthma)، اختلالات عصبی مزمن (Neurological disease)، فشارخون بالا (HTN) به‌عنوان متغیرهای مستقل و وضعیت مرگ به علت کووید-۱۹ به‌عنوان متغیر وابسته در نظر گرفته شد.

مدل‌سازی

رگرسیون لوژیستیک یکی از روش‌های آماری است که برای حل مسائل شناسایی الگو و طبقه‌بندی مورد استفاده قرار می‌گیرد. این روش در علوم پزشکی و زیستی به‌طور متداول مورد استفاده قرار گرفته است (۲۴). یکی از دلایل محبوبیت مدل رگرسیون لوژیستیک در مقایسه با سایر روش‌ها در بین محققان پزشکی این است که ضریب شیب به توان رسیده در مدل رگرسیون لوژیستیک را می‌توان به‌عنوان نسبت شانس (OR) تفسیر کرد (۲۵، ۲۶). در رگرسیون لوژیستیک متغیر پاسخ (Y) یک متغیر دوحالته است و متغیرهای مستقل X و $\pi(X)$ شانس موفقیت را با توجه به مقدار خاصی از X نشان می‌دهد. بنابراین رگرسیون لوژیستیک می‌تواند به‌صورت معادله زیر نمایش داده شود:

یافته‌ها

برای انجام کلیه تحلیل‌های آماری و روش شبکه‌های عصبی در این پژوهش از نرم‌افزار R نسخه ۴.۱.۱ استفاده شد. آمار توصیفی برای علائم مشاهده‌شده، بیماری‌های زمینه‌ای و سایر ویژگی‌های بیماران مبتلا به کووید-۱۹ در جدول ۱ گزارش شده است. از تعداد ۴۷۴۶۰ بیمار مبتلا به کووید-۱۹ تعداد ۴۲۹۷۶ نفر زنده ماندند و ۴۴۸۴ نفر فوت کردند. انحراف معیار \pm میانگین مربوط به سن در افراد بازمانده و غیر بازمانده به ترتیب برابر $51/91 \pm 22/06$ و $67/94 \pm 17/17$ سال بود. متغیر سن بر اساس اطلاعات موجود در مورد دسته‌های سنی طبق سازمان جهانی بهداشت (WHO) طبقه‌بندی شد، سپس برای بررسی ارتباط بین وضعیت مرگ و متغیرهای کیفی از آزمون کای اسکور استفاده شد. سطح معنی‌داری ۰/۰۵ در نظر گرفته شد. توصیف سایر ویژگی‌های جامعه مورد مطالعه را می‌توان در جدول ۱ مشاهده کرد.

حساسیت، ویژگی، دقت و سطح زیر منحنی راک سنجیده می‌شود. این شاخص‌ها به صورت زیر تعریف می‌شوند:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \text{Specificity} = \frac{TN}{TN+FP} \quad \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

که در آن TP، FN، FP و TN به ترتیب مثبت واقعی، منفی کاذب، مثبت کاذب و منفی واقعی هستند (۲۸-۳۰). در پایان، از نمودار آبنبات‌چوبی برای به تصویر کشیدن اهمیت متغیرها استفاده شد.

جدول شماره ۱- آمار توصیفی برای ویژگی‌ها و علائم جمعیت مورد مطالعه

متغیر	سطح	بازمانده (n=۴۲۹۷۶) n (%)	غیر بازمانده (n=۴۴۸۴) n (%)	آماره آزمون (p-value)
سن (انحراف معیار \pm میانگین)		۵۱/۹۱ \pm ۲۲/۰۶	۶۷/۹۴ \pm ۱۷/۱۷	-۶۱/۳۹ (<۰/۰۰۱)
سن	سن ≤ 18	۱۲۶۷۶ (۲۹/۵)	۷۰۳ (۸/۳)	۳۸۴۷/۶۹ (<۰/۰۰۱)
	۱۸ $<$ سن < 65	۱۳۵۹۵ (۳۱/۶)	۲۲۴۶ (۲۶/۵)	
	۶۵ \leq سن < 75	۶۰۵۱ (۱۴/۱)	۱۹۱۳ (۲۲/۵)	
	سن ≥ 75	۷۵۲۳ (۱۷/۵)	۳۴۹۵ (۴۱/۲)	
جنسیت (مرد)		۲۱۴۲۶ (۴۹/۹)	۴۸۰۶ (۵۶/۶)	۱۳۰/۷۹ (<۰/۰۰۱)
داشتن تب (بلی)		۱۶۲۸۰ (۳۷/۹)	۲۷۳۸ (۳۲/۳)	۹۵/۶۷ (<۰/۰۰۱)
داشتن سرفه (بلی)		۱۷۲۸۲ (۴۰/۲)	۲۴۸۵ (۲۹/۳)	۳۵۷/۳۱ (<۰/۰۰۱)
داشتن درد عضلانی (بلی)		۸۸۵۴ (۲۰/۶)	۹۸۰ (۱۱/۶)	۳۷۵/۴۹ (<۰/۰۰۱)
دیسترس حاد تنفسی (بلی)		۲۲۴۵۱ (۵۲/۲)	۶۳۵۰ (۷۴/۸)	۱۴۶۹/۲۴ (<۰/۰۰۱)
کاهش سطح هوشیاری (بلی)		۲۳۶۷ (۵/۵)	۲۲۲۷ (۲۶/۲)	۳۷۴۹/۱۷ (<۰/۰۰۱)
کاهش حس بویایی (بلی)		۵۱۱ (۱/۲)	۲۸ (۰/۳)	۵۰/۴۴ (<۰/۰۰۱)
کاهش حس چشایی (بلی)		۲۳۰ (۰/۵)	۲۳ (۰/۳)	۱۰/۱۰ (۰/۰۰۶)
داشتن تشنج (بلی)		۳۱۴ (۰/۷)	۴۹ (۰/۶)	۲/۳۷ (۰/۷۰۰)
داشتن سردرد (بلی)		۳۴۷۱ (۸/۱)	۱۷۶ (۲/۱)	۳۸۷/۶۳ (<۰/۰۰۱)
داشتن سرگیجه (بلی)		۹۸۴ (۲/۳)	۱۱۴ (۱/۳)	۳۰/۳۶ (<۰/۰۰۱)
داشتن درد در قفسه سینه (بلی)		۱۵۸۱ (۳/۷)	۱۹۲ (۲/۳)	۴۲/۶۸ (<۰/۰۰۱)
داشتن التهاب پوستی (بلی)		۶۵ (۰/۲)	۶ (۰/۱)	۳/۳۳ (۰/۸۷۲)
داشتن درد در معده (بلی)		۱۲۷۶ (۳/۰)	۱۱۷ (۱/۴)	۶۸/۰۱ (<۰/۰۰۱)
داشتن حالت تهوع (بلی)		۲۷۱۳ (۶/۳)	۳۳۲ (۳/۹)	۷۳/۲۸ (<۰/۰۰۱)
داشتن استفراغ (بلی)		۱۹۴۸ (۴/۵)	۲۰۸ (۲/۵)	۷۶/۴۴ (<۰/۰۰۱)
داشتن اسهال (بلی)		۱۶۸۳ (۳/۹)	۱۵۱ (۱/۸)	۹۴/۰۸ (۰/۰۰۸)
بی‌اشتهایی بودن (بلی)		۳۱۵۸ (۷/۳)	۶۲۰ (۷/۳)	۰/۱۷ (۰/۸۹۱)
وضعیت سیگاری بودن (بلی)		۹۸۹ (۲/۳)	۲۵۴ (۳/۰)	۱۴/۴۱ (<۰/۰۰۱)
مصرف مواد مخدر (بلی)		۱۴۳۶ (۳/۳)	۴۵۶ (۵/۴)	۸۲/۷۲ (<۰/۰۰۱)
کاهش سطح اکسیژن خون (بلی)		۲۰۰۰۶ (۴۶/۶)	۷۱۹۶ (۸۴/۸)	۴۱۵۸/۶۱ (<۰/۰۰۱)

۳۷۰/۳۱ (<۰/۰۰۱)	۳۷۳ (۴/۴)	۵۷۱ (۱/۳)	ابتلا به سرطان (بلی)
۴۲/۳۹ (<۰/۰۰۱)	۹۷ (۱/۱)	۲۲۸ (۰/۵)	بیماری‌های مزمن کبد (بلی)
۵۲۷/۸۳ (<۰/۰۰۱)	۱۶۹۰ (۱۹/۹)	۴۶۹۴ (۱۰/۹)	ابتلا به دیابت (بلی)
۴۴/۸۵ (<۰/۰۰۱)	۸۵ (۱/۰)	۱۸۴ (۰/۴)	بیماری‌های مزمن خون (بلی)
۸/۳۵ (۰/۰۰۲)	۲۳ (۰/۳)	۵۸ (۰/۱)	داشتن نقص ایمنی (بلی)
۶۴۷/۵۳ (<۰/۰۰۱)	۱۶۱۵ (۱۹/۰)	۴۰۹۹ (۹/۵)	بیماری‌های مزمن قلب (بلی)
۲۴۸/۸۳ (<۰/۰۰۱)	۳۵۴ (۴/۲)	۶۶۹ (۱/۶)	بیماری‌های مزمن کلیه (بلی)
۱/۱۱ (۰/۲۹۲)	۱۸۶ (۲/۲)	۸۶۶ (۲/۰)	ابتلا به آسم (بلی)
۱۷۳/۴۲ (<۰/۰۰۱)	۲۹۱ (۳/۴)	۵۹۸ (۱/۴)	اختلالات عصبی مزمن (بلی)
۵۰۷/۸۲ (<۰/۰۰۱)	۲۲۵۲ (۲۶/۵)	۶۹۹۱ (۱۶/۳)	ابتلا به فشارخون (بلی)

نام متغیر (سطح غیرمرجع) در جدول نشان داده شده است. آزمون یو من ویتنی و یلکاکسون استفاده شده است.

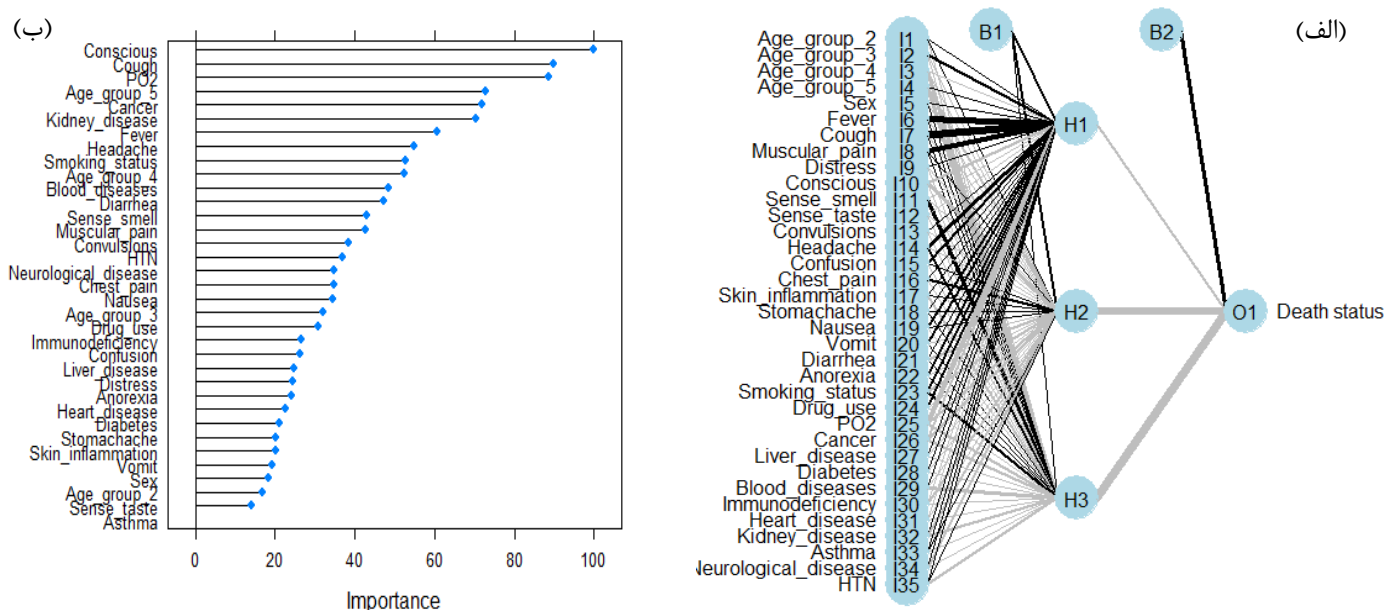
همان‌طور که جدول ۲ نشان می‌دهد، دقت مدل شبکه عصبی مصنوعی برابر ۹۰/۳۰٪ و حساسیت طبقه‌بندی کننده بر روی داده‌های آزمایش ۷۷/۶۶٪ در مرحله آموزش بود، اما این مقادیر در مرحله آزمایش به ترتیب برابر ۸۹/۹۰٪ و ۷۶/۱۴٪ به دست آمد. تجزیه و تحلیل مدل شبکه عصبی مصنوعی نشان داد که این مدل با توانایی بالا، امکان طبقه‌بندی را فراهم کرد. اهمیت متغیرهای مستقل در مدل شبکه عصبی مصنوعی نیز محاسبه و این مقادیر در نمودار ۱ (ب) نشان داده شده است. ده تا از مهم‌ترین عوامل پیش‌بینی کننده مرگ شامل کاهش سطح هوشیاری، داشتن سرفه، سطح اکسیژن کمتر از ۹۳ درصد، سن بالای ۷۵ سال، سرطان، بیماری‌های مزمن کلیه، تب داشتن، سردرد داشتن، سیگاری بودن، و بیماری‌های مزمن خون بود. اهمیت سایر عوامل نیز در نمودار ۱ (ب) قابل مشاهده است.

مدل شبکه عصبی با ورود همه متغیرهای پیش‌بین مورد مطالعه آموزش داده شد. در طول ساخت مدل شبکه عصبی، برای به دست آوردن بهترین عملکرد، مدل بهینه شد. پارامترهای تنظیم شامل size (تعداد نرون های لایه پنهان) و weight decay (پارامتر کاهش وزن) هستند. محدوده تنظیم {۱، ۰.۲، ۰.۱، ۰.۰۵، ۰.۰۱} و {۰/۵، ۰/۲، ۰/۱} به ترتیب برای انتخاب مقادیر بهینه ابر پارامترهای size و weight decay در نظر گرفته شد. مقدار بهینه برای ابر پارامتر weight decay، ۰/۵ و برای size برابر ۳ و تعداد یک‌لایه پنهان به دست آمد. در فرآیند بهینه‌سازی، تعداد تکرارها برای همگرایی مدل ۱۰۰ و تابع فعالیت (Activation function) در لایه پنهان و لایه خروجی، سیگموئید در نظر گرفته شده است. نتایج و ارزیابی شبکه عصبی مصنوعی در جدول ۲ نشان داده شده است و همچنین معماری شبکه برازش یافته نیز در نمودار ۱ (الف) قابل نمایش است.

جدول شماره ۲- ارزیابی عملکرد مدل شبکه عصبی و رگرسیون لوژستیک

مدل	شبکه عصبی		رگرسیون لوژستیک	
	آموزش	آزمایش	آموزش	آزمایش
صحت (/)	۹۰/۳۰	۸۹/۹۰	۸۴/۳۶	۸۳/۶۷
حساسیت (/)	۷۷/۶۶	۷۶/۱۴	۶۹/۳۴	۶۸/۹۴
ویژگی (/)	۹۲/۰۵	۹۱/۹۹	۸۵/۸۹	۸۵/۳۰
سطح زیر منحنی راک (/)	۷۷/۶۵	۷۷/۱۴	۶۹/۳۳	۶۸/۹۸
دقت (/)	۵۹/۰۵	۵۷/۵۳	۴۸/۷۸	۴۶/۲۸

تعداد داده‌های آموزش: ۴۲۷۱۴ و تعداد داده‌های آزمایش: ۴۷۴۶



نمودار شماره ۱- (الف) ساختار و معماری شبکه عصبی برازش یافته، (ب) نمایش میزان اهمیت متغیرها بر اساس مدل شبکه عصبی

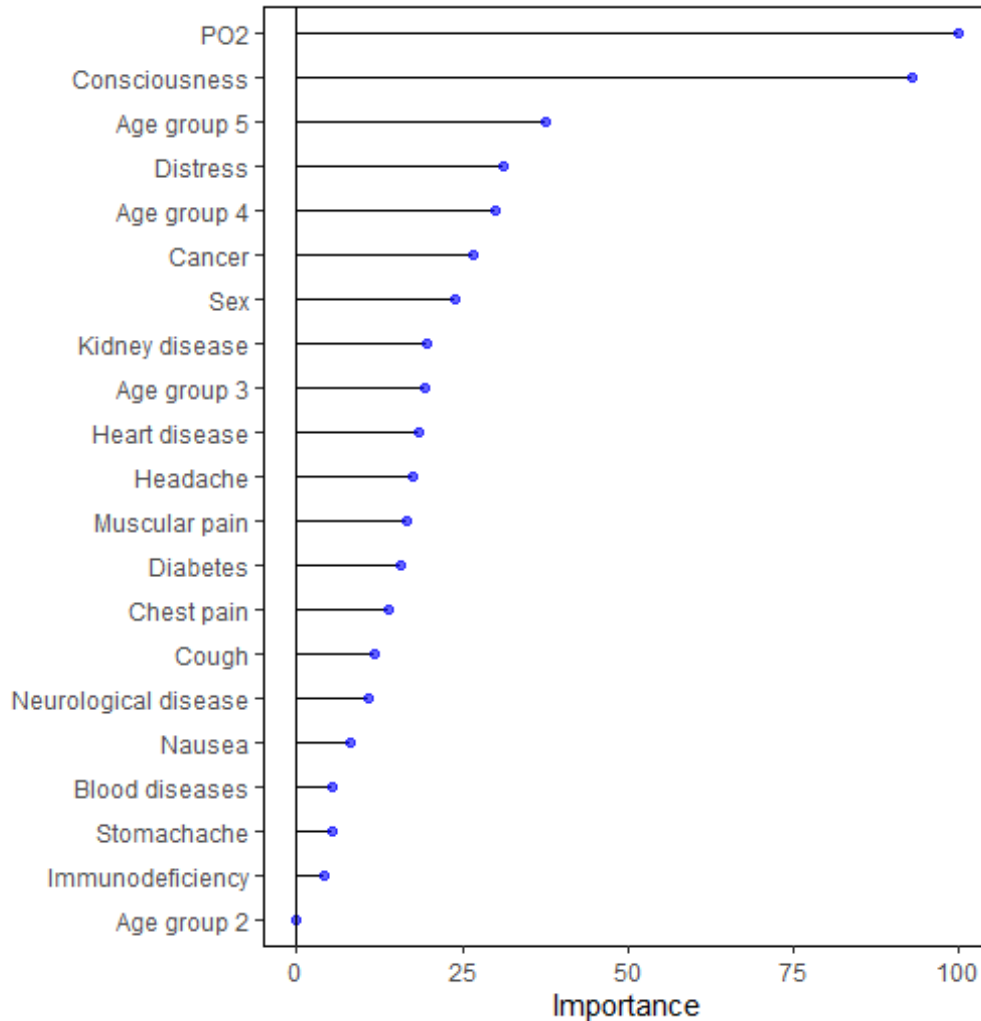
و کاهش سطح اکسیژن خون (کمتر از ۹۳٪) ($OR=۳/۹۷$) دارای نسبت شانس قابل ملاحظه‌ای بودند. به عبارت دیگر، شانس رخ دادن مرگ به علت کووید-۱۹ در افراد بین ۶۵ تا ۷۵ سال، $۳/۶۲$ ، در افراد بزرگتر از ۷۵ سال، $۴/۷۸$ برابر است. علاوه بر سن، کاهش سطح هوشیاری و کاهش سطح اکسیژن خون به ترتیب به اندازه $۳/۹۵$ و $۳/۹۷$ برابر شانس رخداد مرگ را افزایش می‌دهند. اهمیت متغیرها در نمودار ۲ نیز نشان داده شده است.

متغیرهای معنادار در مدل رگرسیون لوژیستیک در جدول ۳ نشان داده شده است. شاخص‌های ارزیابی این مدل نیز در جدول ۲ گزارش شده است. مدل شبکه عصبی نسبت به رگرسیون لوژیستیک توانایی بالاتری در پیش‌بینی سالم ماندن یا مرگ افراد با استفاده از متغیرهای موجود را دارد. متغیرهای مهم در این مدل تا حد زیادی با شبکه عصبی همخوان بود به طوریکه متغیرهای سن بین ۶۵ تا ۷۵ ($OR=۳/۶۲$) و همچنین بزرگتر از ۷۵ سال ($OR=۴/۷۸$)، کاهش سطح هوشیاری ($OR=۳/۹۵$)

جدول شماره ۳- نتایج رگرسیون لوژیستیک

متغیر	بر آورد	خطای معیار	آماره Z	p-value	(فاصله اطمینان ۹۵ درصد) نسبت شانس
عرض از مبدأ	-۴/۱۳	۰/۱۱	-۳۸/۷۰	< ۰/۰۰۱	-
۴۵ ≤ سن	۰/۱۰	۰/۱۱	۰/۹۳	۰/۳۵	۱/۱۱ (۰/۹۰, ۱/۳۸)
۶۵ ≤ سن	۰/۸۵	۰/۱۱	۷/۹۹	< ۰/۰۰۱	۲/۳۴ (۱/۹۱, ۲/۸۹)
۶۵ < سن < ۷۵	۱/۲۹	۰/۱۱	۱۱/۹۶	< ۰/۰۰۱	۳/۶۲ (۲/۹۵, ۴/۴۹)
سن ≥ ۷۵	۱/۵۶	۰/۱۱	۱۴/۷۷	< ۰/۰۰۱	۴/۷۸ (۳/۹۰, ۵/۹۱)
جنسیت (مرد)	۰/۲۷	۰/۰۳	۹/۷۰	< ۰/۰۰۱	۱/۳۲ (۱/۲۵, ۱/۳۹)
داشتن سرفه (بلی)	-۰/۱۶	۰/۰۳	-۵/۲۸	< ۰/۰۰۱	۰/۸۵ (۰/۸۰, ۰/۹۰)
داشتن درد عضلانی (بلی)	-۰/۲۹	۰/۰۴	-۷/۰۵	< ۰/۰۰۱	۰/۷۴ (۰/۶۹, ۰/۸۱)
دیسترس حاد تنفسی (بلی)	۰/۴۱	۰/۰۳	۱۲/۳۴	< ۰/۰۰۱	۱/۵۱ (۱/۴۱, ۱/۶۱)
کاهش سطح هوشیاری (بلی)	۱/۳۷	۰/۰۴	۳۵/۰۲	< ۰/۰۰۱	۳/۹۵ (۳/۹۰, ۴/۲۶)
داشتن سردرد (بلی)	-۰/۶۵	۰/۰۹	-۷/۳۸	< ۰/۰۰۱	۰/۵۲ (۰/۴۴, ۰/۶۲)
داشتن درد در قفسه سینه (بلی)	-۰/۵۳	۰/۰۹	-۵/۹۸	< ۰/۰۰۱	۰/۵۹ (۰/۴۹, ۰/۷۰)
داشتن درد در معده (بلی)	-۰/۳۳	۰/۱۲	-۲/۸۸	< ۰/۰۰۱	۰/۷۲ (۰/۵۷, ۰/۹۰)
داشتن حالت تهوع (بلی)	-۰/۲۷	۰/۰۷	-۳/۹۵	< ۰/۰۰۱	۰/۷۶ (۰/۶۶, ۰/۸۷)
کاهش سطح اکسیژن خون (بلی)	۱/۳۸	۰/۰۴	۳۷/۵۹	< ۰/۰۰۱	۳/۹۷ (۳/۷۰, ۴/۲۷)
ابتلا به سرطان (بلی)	۰/۸۹	۰/۰۸	۱۰/۶۸	< ۰/۰۰۱	۲/۴۳ (۲/۰۶, ۲/۸۶)

۱/۲۹(۱/۲۰, ۱/۳۹)	< ۰/۰۰۱	۶/۷۱	۰/۰۴	۰/۲۵	ابتلا به دیابت (بلی)
۱/۶۱(۱/۱۶, ۲/۲۲)	< ۰/۰۰۱	۲/۸۸	۰/۱۷	۰/۴۸	بیماری‌های مزمن خون (بلی)
۲/۰۶(۱/۱۳, ۳/۶۵)	< ۰/۰۰۱	۲/۴۳	۰/۳۰	۰/۷۲	داشتن نقص ایمنی (بلی)
۱/۳۵(۱/۲۵, ۱/۴۶)	< ۰/۰۰۱	۷/۶۹	۰/۰۴	۰/۳۰	بیماری‌های مزمن قلب (بلی)
۱/۹۴(۱/۶۵, ۲/۲۸)	< ۰/۰۰۱	۸/۱۶	۰/۰۸	۰/۶۶	بیماری‌های مزمن کلیه (بلی)
۱/۵۵(۱/۳۰, ۱/۸۴)	< ۰/۰۰۱	۴/۹۵	۰/۰۹	۰/۴۴	اختلالات عصبی مزمن (بلی)



نمودار شماره ۲- نمایش میزان اهمیت متغیرها بر اساس مدل رگرسیون لوزستیک

بحث

سندرم حاد تنفسی کرونا ویروس ۲ (SARS-CoV-2) نوع جدیدی از ویروس کرونا است. اگرچه میزان مرگ‌ومیر آن بیشتر از آنفولانزا است، اما به نظر می‌رسد این میزان کمتر از سارس و سندرم تنفسی خاورمیانه (MERS) باشد (۳۱). با توجه به نرخ بالای مرگ‌ومیر کووید-۱۹، پیش‌بینی مرگ‌ومیر بیماران بر اساس ویژگی‌های بالینی و ویژگی‌های جمعیتی می‌تواند برای مدیران و پزشکان بیمارستان مفید باشد، به‌ویژه زمانی که بیمارستان توسط بیماران تسخیر شده باشد. الگوریتم‌های یادگیری ماشین در پیش‌بینی بسیاری از مشکلات، به‌ویژه در

داده‌های پیچیده و گسترش بیماری، ارتباط قابل توجهی دارند

(۳۲).

عملکرد مدل‌ها با حساسیت، ویژگی، مساحت زیر منحنی و دقت ارزیابی شد. تمامی شاخص‌های ارزیابی در مدل شبکه عصبی بالاتر از مدل رگرسیون لوزستیک بود. شایع‌ترین و قابل‌توجه‌ترین بیماری‌های زمینه‌ای بیماری‌های مزمن کلیوی و اختلالات مزمن عصبی بودند. اگرچه تب یکی از رایج‌ترین علائم کووید-۱۹ است، اما در مدل‌های پیش‌بینی کننده ما قابل اهمیت نبود.

محمدی و همکاران در ایران از شبکه‌های عصبی مصنوعی و رگرسیون لوزستیک برای تشخیص بیماران COVID-19

کلیوی و بیماری های قلبی بر خطر مرگ در بیماران مبتلا به COVID-19 تأثیرگذار است. آنها اذعان کردند که عوامل مرتبط با مرگومیر یافت شده در این تحقیق می تواند به شناسایی بیماران مبتلا به کووید-۱۹ که در معرض خطر بیشتر پیش آگهی ضعیف هستند کمک کند. نظارت بر این عوامل می تواند به هشدارهای اولیه برای مداخلات مناسب کمک کند (۳۳).

کاندی و همکاران از رگرسیون لوژستیک و رویکردهای طبقه بندی کننده ساده بیزی چندجمله ای برای شناسایی COVID-19 و طبقه بندی گزارش های بالینی به چهار کلاس مختلف COVID، SARS، ARDS، و هر دوی COVID و ARDS استفاده کردند. کلمه SARS مخفف عبارت severe acute respiratory syndrome و ARDS مخفف عبارت acute respiratory distress syndrome می باشند. بر اساس نتایج آنها، رویکردهای یادگیری ماشینی با دقت ۹۶.۲ درصد بالاترین عملکرد را داشتند (۳۴). وانگ و همکاران روند اپیدمی COVID-19 را با استفاده از مدل لوژستیک و روش های یادگیری ماشینی پیش بینی کردند (۳۵). شعبان و همکاران باهدف شناسایی عفونت COVID-19 در ۲۷۹ مورد با پیاده سازی موتور استنتاج فازی و شبکه عصبی عمیق به بررسی این بیماری پرداختند (۳۶).

از محدودیت های این مطالعه، می توان به ناقص بودن اطلاعات و گمشدگی در داده ها، کاهش نمونه ها و عدم مشارکت تمام نمونه های موجود در مدل سازی اشاره کرد. محدودیت دیگر این است که اگرچه در این مطالعه نیز برتری مدل شبکه عصبی نتیجه شد ولی بیان این نکته نیز مهم است که تفسیر مستقیم و احتمالاتی برای متغیرهای موجود در شبکه عصبی همانند آنچه در رگرسیون لوژستیک وجود دارد امکان پذیر نیست.

نتیجه گیری

یافته های ما در مقاله حاضر، اهمیت برخی اطلاعات جمعیت شناختی، بیماری های زمینه ای و علائم بالینی را در پیش بینی بازماندگان و غیر بازماندگان مبتلا به کووید-۱۹ ارائه کرد. مدل شبکه عصبی با توان بالاتری نسبت به مدل لوژستیک نتایج مورد اعتماد و پیش بینی های دقیق تری ارائه نمود. باین حال، تحقیقات پزشکی در این زمینه با به کارگیری سایر روش های یادگیری ماشینی و قدرت بالای آن ها، نتایج کامل کننده ای به دنبال خواهد داشت. عوامل مرتبط با مرگومیر یافت شده در این تحقیق می تواند به شناسایی بیماران مبتلا به کووید-۱۹ که در معرض خطر بیشتر پیش آگهی ضعیف هستند کمک کند. نظارت بر این عوامل می تواند به هشدارهای اولیه برای مداخلات مناسب کمک کند. علاوه بر این، با ارائه این مقاله، عمل

استفاده کردند. آنها دریافتند که مدل شبکه های عصبی مصنوعی از دقت بالاتری نسبت به مدل رگرسیون لوژستیک برخوردار است و هر دو مدل توانایی بالایی در تشخیص عفونت COVID-19 دارند. آنها تب، سرفه، تنگی نفس، خستگی، لرز و سردرد را به عنوان علائم شایع در موارد تأیید شده یافتند، درحالی که تنها متغیرهای سرفه و سردرد با مطالعه ما مطابقت داشتند. شایع ترین بیماری های زمینه ای عبارت بودند از: فشارخون، دیابت، بیماری انسدادی مزمن ریه و بیماری عروق کرونر قلب درحالی که در مطالعه ما شایع ترین بیماری ها عبارت بودند از: سرطان، دیابت، بیماری کلیوی، بیماری قلبی و بیماری خون (۹).

مطالعه مروری تیان و همکاران در زمینه شناسایی عوامل خطر مرتبط با مرگومیر ناشی از کووید ۱۹ در ۳۱۲۵ شهرستان ایالات متحده نشان داد مرگومیر COVID-19 به جنسیت، نژاد/قومیت و محیط بیرون بستگی دارد. افزایش آگاهی از تأثیر این عوامل مهم ممکن است به تصمیم گیرندگان، مقامات بهداشت عمومی و عموم مردم کمک کند تا خطر بیماری همه گیر، به ویژه در کاهش مرگومیر COVID-19 را کنترل کنند (۱۹). در مطالعه حاضر تأثیر جنسیت بر مرگ میر بررسی و تأثیرگذار شناخته شد و با نتایج مطالعه مروری تیان و همکاران همخوان است.

در مطالعه تیلور و همکاران، تأثیر قابل توجه افزایش سن، سیگار کشیدن، فشارخون بالا، دیابت، بیماری قلبی عروقی، بیماری تنفسی، بیماری کلیوی و بدخیمی با مرگومیر همراه بود. نمره ارزیابی نارسایی متوالی ارگان بالاتر و فیزیولوژی حاد و ارزیابی سلامت مزمن و نیاز به تهویه مکانیکی در هنگام پذیرش با مرگومیر همراه بود. تعداد گلبول های سفید بالاتر، نوتروفیل ها، فریتین، پلاکت پایین، و تعداد لنفوسیت ها همگی با مرگومیر مرتبط بودند. در نتیجه، افزایش سن، بیماری های زمینه ای، شدت بیماری بر اساس دستگاه های امتیازدهی معتبر، و پاسخ میزبان به بیماری با مرگومیر همراه بود. درحالی که جنسیت مرد و افزایش BMI تأثیرگذار نبود (۱۸). در مطالعه پیش رو، همانند مطالعه تیلور و همکاران متغیر سن، به خصوص سنین بالا، سیگاری بودن، فشارخون بالا، بیماری های بدخیم مانند سرطان، ابتلا به دیابت، بیماری های مزمن قلبی، بیماری تنفسی، بیماری های مزمن کلیوی با مرگومیر همراه و تأثیرگذار شناخته شد، همخوان بود.

علاوه بر این، نتایج ما با مطالعه مروری سپندی و همکاران نیز همخوان بود. این مطالعه مروری با استفاده از فراتحلیل در ایران نشان داد که متغیرهایی نظیر سن (در افراد مسن تر)، جنسیت و بیماری های همراه مانند دیابت شیرین، فشارخون بالا، اختلالات

و کاربرد دیگری از روش‌های یادگیری ماشین را ارائه و امیدواریم نتایج ما به جامعه پزشکی کمک کند.

References

- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England journal of medicine*. 2020.
- Rabie AH, Saleh AI, Mansour NA. A Covid-19's integrated herd immunity (CIHI) based on classifying people vulnerability. *Computers in Biology and Medicine*. 2022;140:105112.
- Brem A, Viardot E, Nylund PA. Implications of the coronavirus (COVID-19) outbreak for innovation: Which technologies will improve our lives? *Technological Forecasting and Social Change*. 2021;163:120451.
- Abdi M. Coronavirus disease 2019 (COVID-19) outbreak in Iran: Actions and problems. *Infection Control & Hospital Epidemiology*. 2020;41(6):754-5.
- Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*. 2020;139:110050.
- Pontoh RS, Z S, Hidayat Y, Aldella R, Jiwani NM, Sukono. Covid-19 Modelling in South Korea using A Time Series Approach. *International Journal of Advanced Science and Technology*. 2020;29(7):1620 - 32.
- Mohammadzadeh rostami F, Nasr Esfahani BN, Ahadi AM, Shalibeik S. A Review of Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Iran-J-Med-Microbiol*. 2020;14(2):154-61.
- Wu Z, Huang Z, Wu Y, Jin Y, Wang Y, Zhao H, et al. Risk stratification for mortality in cardiovascular disease survivors: A survival conditional inference tree analysis. *Nutrition, Metabolism & Cardiovascular Diseases*. 2021;31(2):420-8.
- Mohammadi F, Pourzamani H, Karimi H, Mohammadi M, Mohammadi M, Ardalani N, et al. Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran. *Biomedical journal*. 2021.
- Belhadi A, Kamble S, Jabbour CJC, Gunasekaran A, Ndubisi NO, Venkatesh M. Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries. *Technological Forecasting and Social Change*. 2021;163:120447.
- Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, solitons, and fractals*. 2021;142:110512.
- Srinivas M, Lin YY, Liao HYM, editors. Deep dictionary learning for fine-grained image classification. 2017 IEEE International Conference on Image Processing (ICIP); 2017 17-20 Sept. 2017.
- Ing EB, Miller NR, Nguyen A, Su W, Bursztyn LL, Poole M, et al. Neural network and logistic regression diagnostic prediction models for giant cell arteritis: development and validation. *Clinical Ophthalmology (Auckland, NZ)*. 2019;13:421.
- Shawarib MZA, Latif AEA, Al-Zatmah BEE-D, Abu-Naser SS. Breast Cancer Diagnosis and Survival Prediction Using JNN. *International Journal of Engineering and Information Systems (IJEAIS)*. 2020;4(10).
- Gour M, Jain S. Stacked convolutional neural network for diagnosis of covid-19 disease from x-ray images. *arXiv preprint arXiv:200613817*. 2020.
- Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of personalized medicine*. 2020;10(2):21.
- Selvi RT, Muthulakshmi I. An optimal artificial neural network based big data application for heart disease diagnosis and classification model. *Journal of Ambient Intelligence and Humanized Computing*. 2021;12(6):6129-39.
- Taylor EH, Marson EJ, Elhadi M, Macleod KDM, Yu YC, Davids R, et al. Factors associated with mortality in patients with COVID-19 admitted to intensive care: a systematic review and meta-analysis. *Anaesthesia*. 2021;76(9):1224-32.
- Tian T, Zhang J, Hu L, Jiang Y, Duan C, Li Z, et al. Risk factors associated with mortality of COVID-19 in 3125 counties of the United States. *Infectious diseases of poverty*. 2021;10(1):3.
- Hagan MT, Demuth HB, Beale M. *Neural network design*: PWS Publishing Co.; 1997.
- Bourdès V, Bonnevey S, Lisboa P, Defrance R, Pérol D, Chabaud S, et al. Comparison of artificial neural network with logistic regression as classification models for variable selection for prediction of breast cancer patient outcomes. *Advances in Artificial Neural Systems*. 2010;2010.
- Ohno-Machado L. Modeling medical prognosis: survival analysis techniques. *Journal of biomedical informatics*. 2001;34(6):428-39.
- Kayri M, Çokluk Ö. Using multinomial logistic regression analysis in artificial neural network: An application. *Ozean Journal of Applied Sciences*. 2010;3(2):259-68.
- Swaminathan S. *Logistic Regression — Detailed Overview*. 2018.
- Schober P, Vetter TR. Logistic Regression in Medical Research. *Anesth Analg*. 2021;132(2):365-6.
- Bagley SC, White H, Golomb BA. Logistic regression in the medical literature:: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*. 2001;54(10):979-85.
- Moein S. *Artificial Neural Network for Medical Diagnosis. Medical Diagnosis Using Artificial Neural Networks*. Hershey, PA, USA: IGI Global; 2014. p. 85-94.
- Nellore SB. Various performance measures in Binary classification-An Overview of ROC study. *IJISSET-International Journal of Innovative Science, Engineering & Technology*. 2015;2.
- Tong Z, Liu Y, Ma H, Zhang J, Lin B, Bao X, et al. Development, validation and comparison of artificial neural network models and logistic regression models predicting survival of unresectable pancreatic cancer. *Frontiers in bioengineering and biotechnology*. 2020;8:196.
- Fawcett T. An introduction to ROC analysis. *Pattern recognition letters*. 2006;27(8):861-74.
- Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: are they closely related? *Clinical Microbiology and Infection*. 2020;26(6):729-34.
- Shawaqfah M, Almomani F. Forecast of the Outbreak of COVID-19 Using Artificial Neural Network: Case Study Qatar, Spain, and Italy. *Results in Physics*. 2021:104484.

33. Sepandi M, Taghdir M, Alimohamadi Y, Afrashteh S, Hosamirudsari H. Factors Associated with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis. Iranian journal of public health. 2020;49(7):1211-21.

34. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Din MMU. Machine learning based approaches for detecting COVID-19 using clinical text data. International Journal of Information Technology. 2020;12(3):731-9.

35. Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons & Fractals. 2020;139:110058.

36. Shaban WM, Rabie AH, Saleh AI, Abo-Elvoud M. Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network. Applied Soft Computing. 2021;99:106906.

Tehran University of
Medical Sciences

Iranian Epidemiological Association

Original Article

Identification Symptoms and Underlying Diseases Related to COVID-19 And Prediction of Death Status Using Artificial Neural Network and Logistic Regression: A Data Mining Approach

Nasrin Talkhi¹, Nooshin Akbari sharak¹, Zahra Rajabzadeh¹, Maryam Salari², Seyed Masoud Sadati³, Mohammad Taghi Shakeri⁴

1- MSc of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

2- Assistant Professor in Biostatistics, Expert Management and Information Technology, Mashhad University of Medical Sciences, Mashhad, Iran

3- MSc of IT management, Center of Statistics and Information Technology Management, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

4- Professor in Biostatistics, Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

Article Information

Received

3 July 2022

Accepted

10 December 2022

Corresponding author

Mohammad Taghi Shakeri

Corresponding author E-mail

ShakeriMT@mums.ac.ir

Keywords:

Machine learning, COVID-19, Neural networks, Data mining, Logistic regression

Abstract

Background and Objectives: Due to the high prevalence of COVID-19 disease and its high mortality rate, it is necessary to identify the symptoms, demographic information and underlying diseases that effectively predict COVID-19 death. Therefore, in this study, we aimed to predict the mortality behavior due to COVID-19 in Khorasan Razavi province.

Methods: This study collected data from 51,460 patients admitted to the hospitals of Khorasan Razavi province from 25 March 2017 to 12 September 2014. Logistic regression and Neural network methods, including machine learning methods, were used to identify survivors and non-survivors caused by COVID-19.

Results: Decreased consciousness, cough, PO2 level less than 93%, age, cancer, chronic kidney diseases, fever, headache, smoking status, and chronic blood diseases are the most important predictors of death. The accuracy of the artificial neural network model was 89.90% in the test phase. Also, the sensitivity, specificity and area under the rock curve in this model are equal to 76.14%, 91.99% and 77.65%, respectively.

Conclusion: Our findings highlight the importance of some demographic information, underlying diseases, and clinical signs in predicting survivors and non-survivors of COVID-19. Also, the neural network model provided high accuracy in prediction. However, medical research in this field will lead to complementary results by using other methods of machine learning and their high power.

Copyright © 2022 The Authors. Published by Tehran University of Medical Sciences.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.