

مقاله پژوهشی

ترياز بیماران مبتلا به کووید-۱۹: به کارگیری روش یادگیری گروهی برای تحلیل عوامل خطر و پیش‌بینی خطر مرگ ناشی از ابتلا

ندا سادات^۱، شراره رستم نیاکان کلهری^{۲،۳*}، شهرزاد درویشی^۴، جمیله کیانی^۱، فرهاد عباسی^۵، بتول امیری^۱، عرفان جوانمردی^۱، صفیه دانشی^۱

^۱ واحد توسعه تحقیقات بالینی، بیمارستان شهدای خلیج فارس، دانشگاه علوم پزشکی بوشهر، بوشهر، ایران

^۲ گروه مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران

^۳ موسسه تحقیقات انفورماتیک PLRI، دانشکده پزشکی هانوفر، دانشگاه فنی Braunschweig، آلمان

^۴ بیمارستان امام خمینی (ره)، دانشگاه علوم پزشکی بوشهر، بوشهر، ایران

^۵ گروه عفونی، دانشکده پزشکی، دانشگاه علوم پزشکی بوشهر، بوشهر، ایران

* نویسنده مسئول: شراره رستم نیاکان کلهری: گروه مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران. ایمیل:

sh-rniakank@tums.ac.ir

دریافت: ۱۴۰۲/۰۲/۲۷ پذیرش: ۱۴۰۲/۰۳/۲۸

چکیده

مقدمه: شناسایی زودهنگام بیماران پرخطر کووید-۱۹ با استفاده از داده‌های غیر آزمایشگاهی هنگام پذیرش، می‌تواند به استفاده موثر منابع محدود مراقبت‌های بهداشتی کمک نموده و با کاهش هزینه و زمان، تصمیم‌گیری بالینی را بهبود بخشد، در نتیجه مرگ بیماران را کاهش دهد.

اهداف: هدف این مطالعه ارائه یک رویکرد هوشمند به منظور تریاز بیماران مبتلا به کووید-۱۹ با تحلیل عوامل موثر و پیش‌بینی خطر مرگ ناشی از این بیماری با استفاده از روش یادگیری گروهی است.

مواد و روش‌ها: مطالعه حاضر از نوع توصیفی-کاربردی بوده که در سال ۱۴۰۰ انجام شده است. داده‌های غیر آزمایشگاهی هنگام پذیرش ۴۵۵۸ بیمار قطعی مبتلا به کووید-۱۹ مراجعه‌کننده به بیمارستان شهدای خلیج فارس بوشهر ثبت شده در سیستم اطلاعات مراقبت پزشکی مورد استفاده قرار گرفت. پس از پیش‌پردازش داده‌ها، عوامل خطر موثر در مرگ این بیماران شناسایی و از نظر اهمیت رتبه‌بندی شد، سپس با به کارگیری روش یادگیری گروهی (voting) مدل پیش‌بینی مرگ ایجاد و عملکرد آن با معیارهای ماتریس آشفتگی (Confusion Matrix) ارزیابی گردید.

نتایج: از مجموع ۵۴۳۳ بیمار، اطلاعات ۴۵۵۸ نفر وارد مطالعه شد. از این تعداد ۲۲۲۲ (۴۵/۵ درصد) زن و ۲۶۶۳ (۵۹/۵ درصد) مرد بودند. میانگین سنی بیماران ۴۷/۶ به دست آمد. از میان همه عوامل مورد بررسی، ۱۷ عامل در مرگ بیماران موثر شناسایی شد. نتایج رده‌بندی آن‌ها نشان داد شش عامل اول در پیش‌بینی مرگ به ترتیب وضعیت اکسیژن خون، سن، وضعیت هوشیاری، تب، سرفه و بدن درد است. مدل طراحی شده به روش یادگیری گروهی، خطر مرگ را با دقت ۰/۸۵، ویژگی ۰/۸۵ و سطح زیر نمودار ROC ۰/۹۱ برآورد نمود.

نتیجه‌گیری: نتایج مطالعه یک راه‌حل کم‌هزینه، سریع و نوآورانه برای شناسایی زودهنگام بیماران مبتلا به کووید-۱۹ در معرض خطر مرگ برای تریاز موثرتر آن‌ها ارائه می‌نماید که می‌تواند با به‌روزرسانی مداوم با داده‌های جدید در تصمیم‌گیری به مدیران سیستم‌های مراقبت بهداشتی به منظور مدیریت منابع و ارائه درمان‌های متناسب با بیماران با خطر مرگ بالا یاری رساند.

واژگان کلیدی: کووید-۱۹، عفونت‌های ویروس کرونا، یادگیری ماشین، عوامل خطر، تریاز

۱. مقدمه

یک بیماری همه‌گیر معرفی گردید (۱، ۲). به طور عمده، این ویروس زمانی منتقل می‌شود که افراد در هوای آلوده از قطرات یا ذرات معلق حاوی ویروس تنفس کنند (۳). علائم این بیماری در افراد متنوع بوده اما معمولاً شامل

کروناویروس سندرم حاد تنفسی (SARS-CoV-2) عضوی از خانواده ویروس کرونا است که باعث ایجاد بیماری کووید-۱۹ می‌شود. این بیماری در یازده مارس سال ۲۰۲۰ توسط سازمان بهداشت جهانی (WHO) به عنوان

تب، سرفه و احتقان بینی، از دست دادن اشتها، خستگی، تنگی نفس، درد عضلانی و از دست دادن طعم و بو است (۴) که در موارد شدید می‌تواند منجر به ایجاد ذات‌الریه، سندرم دیسترس تنفسی حاد، شوک سپتیک و در نهایت مرگ گردد (۵، ۶). معمولاً این علائم بین ۱ تا ۱۴ روز پس از قرار گرفتن در معرض ویروس در افراد ظاهر می‌گردد (۷). تا ۷ می سال ۲۰۲۱، کل موارد تاییدشده پنومونی کووید-۱۹ در کل جهان به حدود ۱۵۷ میلیون ابتلا و ۶ میلیون مورد مرگ رسید. بر اساس آمار دانشگاه جانز هاپکینز نسبت مرگومیر به مورد (Case Fatality Rate) این عفونت در جهان ۱/۰۴ درصد بوده است (۸). افزایش موارد بیماری به دلیل گسترش این ویروس در سراسر جهان باعث شد در بسیاری از کشورها، تعداد بیماران از ظرفیت بیمارستان فراتر رود و در نتیجه این مراکز از نظر تجهیزات درمانی و مراقبتی، کارکنان خط مقدم درمان و سایر موارد با مشکل روبه‌رو شوند (۹، ۱۲). بستری بیمار بدحال در بخش مراقبت‌های ویژه، تست‌های آزمایشگاهی و تصویربرداری و همچنین وسایل محافظتی برای کارکنان هزینه‌های سنگینی را به سیستم مراقبت بهداشتی تحمیل کرد (۱۳). در چنین شرایطی شناسایی زود هنگام بیماران در معرض خطر مرگ و کنترل شدت بیماری در بیمارستان امری ضروری به نظر می‌رسد (۱۳). شناسایی زود هنگام بیماران پرخطر کووید-۱۹ می‌تواند پاسخگویی سیستم‌های مراقبت بهداشتی را تسهیل کند و از مداخله فوری و مراقبت‌های ویژه اطمینان حاصل کند و در نتیجه پیامد بیماری را بهبود بخشد. همچنین می‌تواند به کاهش بار سیستم‌های بهداشتی کمک کرده و آن‌ها را قادر سازد تا با اولویت‌بندی تخصیص منابع محدود در زمان اوج بیماری تصمیم‌گیری را بهینه کنند (۱۴). این موضوع به‌ویژه در کشورهای با منابع محدود ضروری به نظر می‌رسد (۱۵). از جمله راهکارهای موجود برای حل این موضوع استفاده از مدل‌های پیش‌بینی است. این مدل‌ها قادرند بیماران در معرض خطر مرگومیر را شناسایی کنند و حمایت‌های بالینی لازم برای کاهش مرگومیر را در کمترین زمان ممکن فراهم آورند (۱۶، ۱۹). اما پیش‌بینی دقیق پیامدهای بالینی برای بیماران با توجه به طیف ناهمگون تظاهرات بالینی می‌تواند دشوار باشد. این مسئله چالش بزرگی را برای مدیریت بیماران مبتلا به کووید-۱۹ به‌ویژه در کانون‌های بیماری نیازمند به‌تیراژ حجم بالایی از بیماران به وجود آورده است (۲۰، ۲۱). به همین دلیل سیستم‌های مراقبت‌های بهداشتی در سراسر جهان تلاش می‌کنند تا از یادگیری ماشین برای دستیابی به تصمیم‌گیری مناسب از طریق حذف ارزیابی‌های ذهنی پزشکان استفاده کنند (۲۲، ۲۳). یادگیری ماشین

(Machine learning) زیرشاخه‌ای از هوش مصنوعی است که امکان استخراج مدل‌های پیش‌بینی با کیفیت بالا را از مجموعه داده‌های خام عظیم فراهم می‌کند. الگوریتم‌های یادگیری ماشین در زمینه‌های بی‌شماری از کووید-۱۹ مورد استفاده قرار گرفته‌است و تنها محدود به تشخیص شیوع، شناسایی و طبقه‌بندی تصاویر پزشکی کووید-۱۹ پیش‌بینی شدت ریسک نیست (۲۴، ۲۷). از جمله مهم‌ترین کاربردهای آن، پیش‌بینی میزان عفونت، مرگومیر و همچنین ایجاد مدلی برای طبقه‌بندی بیماران بر اساس یافته‌های بالینی است (۲۸، ۲۹).

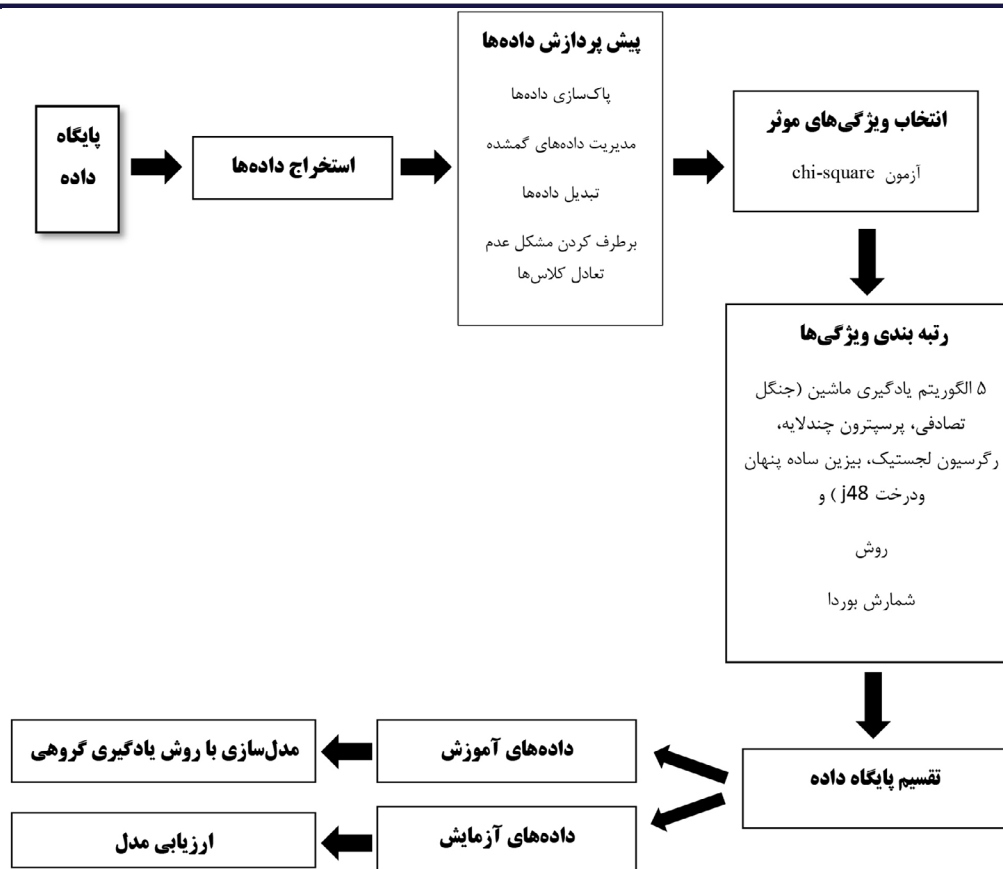
پژوهشگران در نقاط مختلف جهان توانسته‌اند مدل‌های پیش‌بینی مرگ را با روش‌های یادگیری ماشین و داده‌های بیماران و به خصوص داده‌های آزمایشگاهی ایجاد کنند (۲۸، ۳۰-۳۶). به علاوه با توجه به اینکه قومیت و نژاد از عوامل موثر در مرگومیر بیماران کووید-۱۹ است (۲، ۱۱، ۳۷، ۳۹)، این مدل‌ها معمولاً با داده‌های بومی و ملی ایجاد شده‌است. در همین زمینه Ikemura و همکاران در ایالات متحده آمریکا برای پیش‌بینی شانس زنده ماندن بیماران کووید-۱۹ با انواع مختلف داده‌هایی مانند علائم حیاتی، بیومارکرها و بیماری‌های زمینه‌ای مدلی با روش ensemble stacked model ایجاد کردند (۳۰). Karthikeyan و همکاران نیز با استفاده از داده‌های آزمایشگاهی خون بیماران بیمارستانی در چین و به‌کارگیری یادگیری ماشین به پیش‌بینی مرگ بیماران کووید-۱۹ پرداختند (۴۰). در کره جنوبی An و همکاران نیز با استفاده از چند الگوریتم یادگیری ماشین (ماشین بردار پشتیبان، جنگل تصادفی و k نزدیک‌ترین همسایگی) و اطلاعات جمعیت و اجتماعی-شناختی و اطلاعات پزشکی بیماران مدلی را برای پیش‌بینی مرگومیر بیماران کووید-۱۹ توسعه دادند (۴۱). به‌کارگیری مدل‌های توسعه‌یافته بر اساس داده‌های آزمایشگاهی خون بیماران در مرحله تریاژ به سختی امکان‌پذیر است چرا که دسترسی به این داده‌ها با توجه به تعدد بیماران در مراکز مراقبت بهداشتی و محدودیت منابع در هنگام پاندمی دشوار است. از طرف دیگر، در بحث پیش‌بینی و پیش‌آگهی‌های بیماری‌ها، دقت و عملکرد مدل‌های یادگیری ماشین بسیار حائز اهمیت است. بنابراین بهتر است در این‌گونه مسائل، روش‌هایی به کار گرفته شود که این دو مولفه را بهبود بخشند. یکی از این روش‌ها، یادگیری گروهی است که در آن از چندین الگوریتم یادگیری ماشین برای به دست آوردن عملکرد پیش‌بینی بهتر از آن‌چه از الگوریتم‌های یادگیری تشکیل‌دهنده به تنهایی به دست می‌آید، ترکیب می‌شود (۴۲).

۲. اهداف

با توجه به چالش‌های اشاره شده، در این پژوهش با بهره‌گیری از مزایای روش‌های یادگیری گروهی و داده‌های جمعیتی و بالینی غیرآزمایشگاهی هنگام پذیرش بیماران مبتلا به کووید-۱۹ استان بوشهر ابتدا مهم‌ترین عوامل خطر (پیش‌بینی‌کننده مرگ) شناسایی و رتبه‌بندی گردید، سپس مدلی بومی برای پیش‌بینی خطر مرگ این بیماران به منظور شناسایی زود هنگام و تریاژ موثرتر آن‌ها ایجاد شد.

۳. مواد و روش‌ها

مطالعه حاضر از نوع گذشته‌نگر بوده که به‌صورت توصیفی-کاربردی، در سال ۱۴۰۰ و در بیمارستان شهدای خلیج فارس بوشهر به‌عنوان مرکز معین بیماران کووید-۱۹ استان بوشهر با کد اخلاق IR.BPUMS.REC.1399.003 انجام شد. هدف پژوهش تحلیل عوامل خطر و ایجاد یک مدل پیش‌بینی خطر مرگ بیماران کووید-۱۹ با روش یادگیری گروهی بود که در چندین فاز انجام شد. فرایند کلی پژوهش در نمودار زیر ترسیم شده است (تصویر ۱).



تصویر ۱. فرایند کلی انجام پژوهش.

۱.۳. فاز اول: پیش‌پردازش داده‌ها

برای انجام این پژوهش از داده‌های افراد مبتلا قطعی به کووید-۱۹ بر اساس یکی از تست‌های تشخیصی سی‌تی‌اسکن (CT-SCAN) و واکنش زنجیره‌ای پلیمرز (Polymerase Chain Reaction (PCR)) مراجعه‌کننده به بیمارستان شهدای خلیج فارس بوشهر مرکز معین بیماری کووید-۱۹ از ابتدای شروع بیماری (اسفند ۱۳۹۸) تا تاریخ ۱۷ اردیبهشت ۱۴۰۰ شمسی ثبت شده در سیستم مرکز پایش مراقبت پزشکی (Medical care monitoring center) سیستم اطلاعات بیمارستان

استفاده شد. داده‌های موج اول تا چهارم بیماری و سویه آلفا، بتا و گاما با ورود به سیستم مذکور و با قالب اکسل استخراج شد. در این بازه زمانی تعداد ۵۴۳۳ رکورد از بیماران بازاریابی گردید. هر رکورد شامل ۷۲ ویژگی کمی و کیفی از علائم اولیه بالینی هنگام پذیرش، دموگرافیک، اطلاعات مراقبت درمانی (مدت بستری و داروها) در بیمارستان و اطلاعات ترخیصی بیماران بود. در گام اول، پیش‌پردازش داده‌ها در چندین مرحله شامل پاکسازی داده‌ها، مدیریت داده‌های گمشده، تبدیل داده‌ها، متعادل‌سازی کلاس‌ها انجام شد که در ادامه به تفصیل تشریح شده است.

۱.۱.۳. پاک‌سازی داده‌ها

در داده‌های این پژوهش عدم تعادل در نمونه‌های کلاس‌ها دیده شد؛ یعنی کلاس «مرگ» تعداد نمونه‌های بسیار کمتری نسبت به تعداد نمونه‌های کلاس «دیگر انواع ترخیص» داشت. بنابراین ممکن بود الگوریتم‌های یادگیری ماشین به‌خوبی عمل نکند و نوعی سوگیری به سمت کلاس اکثریت به وجود آید. برای حل این مشکل از روش بیش‌نمونه‌گیری اقلیت مصنوعی (Synthetic Minority Oversampling Technique) و نرم‌افزار ۳.۹.۵ Weka استفاده شد. در این روش به صورت مصنوعی تعداد نمونه‌های کلاس اقلیت افزایش می‌یابد و در نتیجه این کار توزیع کلاس‌ها متعادل‌تر می‌گردد.

۲.۳. فاز دوم: شناسایی ویژگی‌های کاندید

۱.۲.۳. انتخاب ویژگی‌های موثر: کاهش ابعاد

به‌منظور شناسایی ویژگی‌های کاندید از آزمون کای-دو استفاده گردید و ویژگی‌های با مقادیر $P\text{-value} < 0.05$ به‌عنوان ویژگی کاندید انتخاب شد.

۲.۲.۳. تعیین اهمیت و رتبه‌بندی ویژگی‌های پیش‌بین (کاندید)

اهمیت ویژگی‌های پیش‌بین با به‌کارگیری پنج الگوریتم درخت تصادفی (Random Tree) RT، پرسپترون چندلایه (Multilayerperceptron) MLP، بیزین ساده پنهان (Hidden Naive Bayes) LR، رگرسیون لجستیک (Logistic Regression) و درخت J48 مشخص و سپس این ویژگی‌ها رتبه‌بندی شدند. هر الگوریتم از مفهوم LOOCV (One-Out Cross-Validation) برای مرتب‌سازی آن‌ها استفاده کرده و یک رتبه‌بندی از این ویژگی‌ها ایجاد نمود. در مرحله بعد از روش شمارش بوردا (Borda Count) برای ترکیب رتبه‌بندی کلی اهمیت n ویژگی پیش‌بین استفاده شد. مقدار Borda Count برای هریک از این ویژگی‌ها برای رتبه‌بندی کلی اهمیت آن‌ها (n ویژگی پیش‌بین) محاسبه شد.

۳.۳. فاز سوم: مدل‌سازی و اعتبارسنجی

مدل‌سازی پژوهش در نرم‌افزار ۳.۹.۵ Weka و با الگوریتم یادگیری گروهی به روش رای‌گیری (voting) انجام شد. یادگیری گروهی، طبقه‌بندی‌های پایه را برای طبقه‌بندی نمونه‌های جدید ترکیب می‌کند. روش رای‌گیری یکی از متداول‌ترین روش‌های یادگیری گروهی است که در آن برای هر نمونه از مجموعه داده‌ها، طبقه‌بندی‌کننده‌های پایه به هر کلاس احتمال می‌دهند که به‌عنوان کلاس نهایی

ابتدا از پارامتر زمان برای پاک‌سازی داده‌ها استفاده شد، به‌طوری‌که رکوردهای بازیابی‌شده در خارج از بازه زمانی با مقادیر نامعتبر و پرت به‌صورت دستی بررسی و در صورت نیاز حذف شدند. سپس برای هر رکورد، ویژگی «وضعیت ترخیص» به‌عنوان کلاس در نظر گرفته شد. کلاس با مقدار «۱» برای مرگ و با مقدار «۰» برای دیگر انواع ترخیص تعیین شد. معیار ورود به مطالعه، مشخص بودن مقدار ویژگی «وضعیت ترخیص» بیمار در رکورد و سن بین ۱ تا ۹۹ سال و معیار خروج از مطالعه، مشخص نبودن وضعیت ترخیص و بدون علامت بودن شخص (به دلیل وابسته بودن ورودی الگوریتم‌ها به علائم و مشخصات دموگرافیک بیماران) بود. پس از آن، رکوردهایی با تعداد مقادیر ویژگی‌های خالی بیش از ۲۰ درصد حذف شدند. با توجه به این‌که هدف شناسایی خطر مرگ بر اساس علائم بیماری در هنگام پذیرش و اطلاعات جمعیتی بیماران بود، لذا داده‌های مربوط به درمان‌ها و داروها از مطالعه خارج و با مشورت پزشک متخصص بیماری‌های عفونی، در نهایت ۴۳ ویژگی انتخاب و وارد مطالعه شدند. از این تعداد، ویژگی‌هایی که در بیش از ۳۰ درصد رکوردهای فاقد مقدار بودند حذف شدند. همچنین با پیشنهاد پزشک متخصص از به‌کارگیری تعدادی دیگر از علائم که در تعداد کمی از بیماران دیده شد (کمتر از یک درصد) صرف نظر گردید. به‌منظور کاهش ابعاد، در ادامه کار، برخی ویژگی‌ها که قابلیت ادغام را داشتند با هم ترکیب و مقادیر آن‌ها در یک ویژگی جدید قرار گرفته و ویژگی‌های قبل حذف شدند.

۲.۱.۳. مدیریت داده‌های گمشده

الگوی گمشدگی داده‌های پژوهش به صورت تصادفی بوده، لذا با به‌کارگیری روش جای‌گذاری چندگانه multiple imputation و با استفاده از رگرسیون لجستیک (Logistic regression) با مقدار imputation ۵، در نرم‌افزار SPSS-22، مقادیر آن‌ها برآورد گردید.

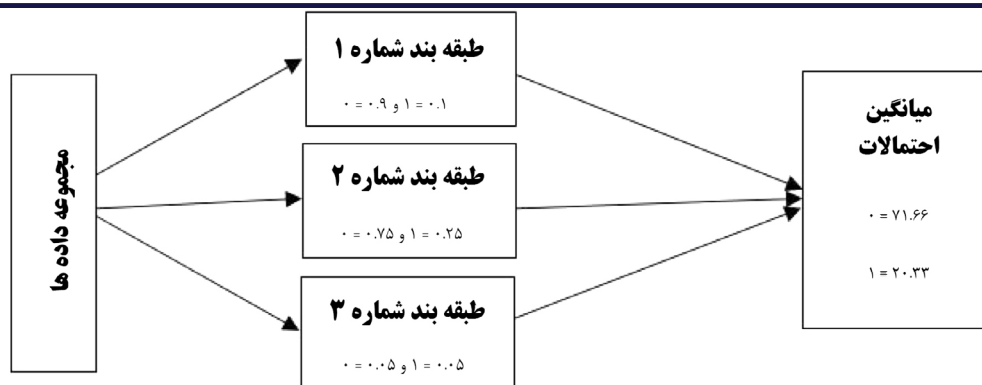
۳.۱.۳. تبدیل داده‌ها

به‌منظور تسهیل فرایند داده‌کاوی و کاهش تنوع داده‌ها و بهره‌وری بیشتر آن‌ها، ویژگی‌هایی با مقادیر متنوع دسته‌بندی گردید. همچنین ویژگی‌های مربوط به وجود یا عدم وجود علائم بیماری در افراد، به ویژگی دوحالتی و سایر ویژگی‌ها گسسته یا پیوسته دیگر نیز به ویژگی طبقه‌ای تبدیل شدند.

۴.۱.۳. متعادل‌سازی کلاس‌ها

آن نمونه تنظیم می‌شود. این احتمالات می‌توانند در یک قانون ترکیبی از الگوریتم رای گیری استفاده شوند. میانگین احتمالات (Average of probabilities) یکی از قوانین ترکیبی در الگوریتم رای گیری است که در آن برای به دست

آوردن احتمال انتخاب هر کلاس به عنوان کلاس نهایی از همه طبقه‌بندی‌کننده‌ها، میانگین گرفته می‌شود و با حداکثر مقدار آن، نتیجه برچسب‌گذاری می‌شود (تصویر ۲).



تصویر ۲. میانگین احتمالاتی در روش رای گیری در الگوریتم یادگیری گروهی.

برای مدل‌سازی به روش یادگیری گروهی، در مرحله اول ابتدا ویژگی‌های پیش‌بین مهم از نتیجه رتبه‌بندی کلی مرحله قبل، در اختیار سه پزشک متخصص بالینی قرار گرفته و تایید شد. این ویژگی‌ها، به‌عنوان متغیرهای ورودی الگوریتم‌های پایه (HNB, j48, LR, RT, MLP)، یادگیری گروهی در نظر گرفته شده و سپس از روش دسته‌بندی‌کننده voting و با استفاده از قانون ترکیب Average of probability voting classier نتایج دسته‌بندی‌های پایه ترکیب شد. در ادامه، داده‌ها به دو مجموعه آموزش (۷۰ درصد) و مجموعه آزمایش (۳۰ درصد) تقسیم و یادگیری با استفاده از داده‌های آموزش انجام شد و در نهایت با داده‌های آزمون مدل تست شد. این روش برای ۱۰ بار و برای تولید ۱۰ مجموعه آموزش و آزمایش تکرار شد (روش ۱۰ fold cross validation).

برای ارزیابی عملکرد مدل معیارهای ارزیابی استاندارد مانند دقت accuracy، حساسیت sensitivity، ویژگی specificity و امتیاز F-score، سطح زیر منحنی و مشخصه عملکرد گیرنده (Receiver operating characteristic) Roc ماتریس آشفتگی (Confusion Matrix) محاسبه شد (جدول ۱). در این ماتریس افرادی که فوت شدند و مدل به‌درستی آن‌ها را تشخیص داده با True Positive (TP)، تعداد افرادی که زنده ماندند اما مدل آن‌ها را فوت تشخیص داده با False positive (FP)، تعداد افرادی که زنده ماندند و مدل آن‌ها را تشخیص داده با True Negative (TN) و تعداد افرادی که زنده ماندند و مدل آن‌ها را فوت کرده تشخیص داده با False Negative (FN) مشخص شده است.

جدول ۱. ماتریس آشفتگی (Confusion Matrix)

متغیرها	مرگ پیش‌بینی شده (+)	بقا پیش‌بینی شده (-)
مرگ واقعی (+)	TP	FN
بقا واقعی (-)	FP	TN

$$\text{فرمول (۱)} \quad Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{فرمول (۲)} \quad Sensitivity = \frac{TP}{TP + FN}$$

$$\text{فرمول (۳)} \quad Specificity = \frac{TN}{TN + FP}$$

$$\text{فرمول (۴)} \quad F - score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

۴. نتایج

تنگی نفس ایجاد شد. در نتیجه، از تعداد ۳۷ ویژگی، ۱۰ ویژگی کم و ۳ ویژگی جدید اضافه شد. در نهایت ۳۰ ویژگی وارد مطالعه گردید.

در مرحله متعادل سازی کلاس ها با روش smote، نمونه های تعداد کلاس اقلیت مرگ چهار برابر (۱۹۲۴) گردید، اما کلاس اکثریت (دیگر انواع ترخیص) ثابت (۴۴/۴) ماند.

نتایج تحلیل توصیفی داده ها نشان داد که میانگین سنی بیماران ۴۷/۶ است. از مجموع ۴۸۸۵ بیمار، تعداد ۲۲۲۲ (۴۵/۵ درصد) نفر زن و ۲۶۶۳ (۵۴/۵ درصد) نفر مرد بودند. از بین ویژگی های مورد بررسی، نتیجه آزمایش مثبت PCR (۸۲/۳ درصد)، سابقه تماس با فرد آلوده (۵۰/۶ درصد) و علائم تنفسی (۴۴ درصد) به ترتیب بیشترین فراوانی را داشتند. کمترین فراوانی نیز متعلق به ویژگی مصرف سیگار (۰/۱ درصد) بود. همچنین از مجموع بیماران مورد بررسی ۹/۸ درصد فوت و بقیه به دیگر انواع ترخیص شدند (جدول ۲).

پس از پیش پردازش داده ها، از مجموع ۵۴۳۳ تعداد ۵۴۸ رکورد حذف و اطلاعات ۴۸۸۵ بیمار وارد مطالعه شد. از این تعداد ۴۸۱ رکورد کلاس «مرگ» و ۴۴۰۴ متعلق به کلاس «دیگر انواع ترخیص» بود. همچنین از میان همه ۷۲ ویژگی (شامل داروها، درمان ها و علائم بیماران)، با توجه به هدف پژوهش، ۴۳ ویژگی مربوط به علائم بیماری و اطلاعات دموگرافیک وارد مطالعه شدند. تعدادی از آن ها که در بیش از ۳۰ درصد رکوردهای فاقد مقدار بودند، حذف شدند که در نهایت ۳۷ ویژگی باقی ماند. برخی ویژگی ها مانند تهوع، استفراغ، اسهال و بی اشتها با هم ادغام و یک ویژگی جدید با عنوان «علائم گوارشی» ایجاد شد. علاوه بر این، ویژگی «سایر بیماری های مزمن» از ترکیب فیلدهای سرطان، بیماری کبدی، بیماری کلیه، ریه و ویژگی «علائم تنفسی» نیز با ادغام ویژگی خس خس سینه و احساس

جدول ۲. معرفی و توزیع فراوانی ویژگی های دموگرافیک و بالینی بیماران کووید-۱۹

ردیف	ویژگی	تعداد (درصد)	ویژگی	تعداد (درصد)
۱	سن	میانگین (۴۷/۶)	بیماری مزمن خونی	ندارد: ۴۸۴۸ (۹۹/۲)، دارد: ۳۷ (۰/۸)
۲	جنس	زن: ۲۲۲۲ (۴۵/۵)، مرد: ۲۶۶۳ (۵۴/۵)	بیماری نقص ایمنی به جز ایدز و سرطان	ندارد: ۴۸۷۳ (۹۹/۸)، دارد: ۱۲ (۰/۲)
۳	سابقه تماس با بیمار آلوده	داشته است: ۲۴۷۲ (۵۰/۶)، نداشته است: ۲۴۱۳ (۴۹/۴)	بارداری	است: ۲۱ (۰/۴)، نیست: ۴۸۶۴ (۹۹/۶)
۴	نتیجه تست pcr	مثبت: ۴۰۲۰ (۸۲/۳)، منفی: ۸۶۵ (۱۷/۷)	دیالیز	ندارد: ۴۳۱۵ (۸۸/۳)، دارد: ۵۷۰ (۱۱/۷)
۵	تب	بیش از ۳۷/۵: ۱۲۳۷ (۲۵/۳)، کمتر از ۳۷/۴: ۳۶۴۸ (۷۴/۷)	آسم	ندارد: ۴۷۶۷ (۹۷/۶)، دارد: ۱۱۸ (۲/۴)
۶	سرفه	ندارد: ۳۳۱۷ (۶۷/۹)، دارد: ۱۵۶۸ (۳۲/۱)	بیماری مزمن ریوی به جز آسم	ندارد: ۴۸۶۶ (۹۹/۶)، دارد: ۱۹ (۰/۴)
۷	بدن درد	ندارد: ۴۱۹۰ (۸۵/۸)، دارد: ۶۹۵ (۱۴/۲)	سایر بیماری مزمن	ندارد: ۴۵۹۵ (۹۵)، دارد: ۲۹۰ (۵)
۸	علائم تنفسی	ندارد: ۲۷۳۵ (۵۶)، دارد: ۲۱۵۰ (۴۴)	هوشیاری	ندارد: ۴۷۷۲ (۹۷/۷)، دارد: ۱۱۳ (۲/۳)
۹	علائم گوارشی	ندارد: ۴۳۸۴ (۸۹/۷)، دارد: ۵۰۱ (۱۰/۳)	سر درد	ندارد: ۴۵۱۷ (۹۲/۵)، دارد: ۳۶۸ (۷/۵)
۱۰	پارزی اندام ها	ندارد: ۴۸۷۶ (۹۹/۸)، دارد: ۹ (۰/۲)	سرگیجه	ندارد: ۴۷۴۲ (۹۷/۱)، دارد: ۱۴۳ (۲/۹)
۱۱	ضایعات پوستی	ندارد: ۴۸۷۷ (۹۹/۸)، دارد: ۸ (۰/۲)	درد قفسه سینه	ندارد: ۴۷۴۱ (۹۷/۱)، دارد: ۱۴۴ (۲/۹)
۱۲	میزان PO2	بیشتر از ۹۳: ۳۷۲۵ (۷۶/۳)، کمتر از ۹۳: ۱۱۶۰ (۲۳/۷)	CT-scan علائم	ندارد: ۳۷۲۵ (۷۳/۸)، دارد: ۱۱۶۰ (۲۶/۲)
۱۳	مصرف سیگار	بله: ۷ (۰/۱)، خیر: ۴۸۷۸ (۹۹/۹)	بیماری قلبی	ندارد: ۴۵۶۷ (۹۳)، دارد: ۳۱۸ (۷)
۱۴	مواد مخدر	بله: ۹ (۰/۲)، خیر: ۴۸۷۶ (۹۹/۸)	فشار خون	ندارد: ۴۲۵۵ (۸۷/۱)، دارد: ۶۳۰ (۱۲/۹)
۱۵	دیابت	ندارد: ۴۳۱۵ (۸۸/۳)، دارد: ۵۷۰ (۱۱/۷)	نوع ترخیص	بقا: ۴۴۰۴ (۹۰/۲)، مرگ: ۴۸۱ (۹/۸)

نظر گرفتن سطح معنی‌داری $p\text{-value} < 0.05$ ، هفده عامل به‌عنوان ویژگی پیش‌بین و کاندید شناسایی شد (جدول ۳).

یکی از مهم‌ترین اهداف این پژوهش تحلیل عوامل موثر بر مرگ‌ومیر بیماران مبتلا به کووید-۱۹ بود. به همین منظور، با انجام آزمون کای-دو (chi-square) و با در

جدول ۳. شناسایی ویژگی‌های کاندید براساس آزمون Chi-Square		
ویژگی	آزمون کای-دو	مقدار P-Value
سن	۱۷۹/۶۲۴	< 0.001
سابقه تماس	۱۰/۰۴۱	0.001
نتیجه تست	۳۸/۷۳۱	< 0.001
تب	۱۱/۵۹۶	0.001
سرفه	۴۳/۸۷۱	< 0.001
بدن درد	۲۶/۴۸	< 0.001
علائم تنفسی	۴۷/۵۷۷	< 0.001
هوشیاری	۳۰۷/۲۸۹	< 0.001
علائم گوارشی	۷/۵۲۵	0.006
سردرد	۲۲/۷۸۶	< 0.001
میزان اکسیژن خون	۶۴۳/۴۸۲	< 0.001
بیماری قلبی	۵۷/۶۷۱	< 0.001
فشار خون	۲۸	< 0.001
بیماری مزمن	۳۵	< 0.001
دیابت	۱۰	< 0.001
بیماری خونی مزمن	۸/۸۰۳	0.003
دیالیز	۴/۶۸۴	0.03

بوردا ترکیب شد و در نهایت رتبه‌بندی کلی عوامل بر اساس اهمیت آن در پیش‌بینی مرگ بیماران مشخص گردید (جدول ۴).

این عوامل با استفاده از هر یک از پنج الگوریتم یادگیری ماشین رده‌بندی و تعیین اهمیت شد. سپس نتایج رده‌بندی حاصل هر یک از الگوریتم‌ها با استفاده از روش شمارش

جدول ۴. رتبه‌بندی ویژگی‌های کاندید براساس اهمیت						
رتبه، ویژگی	جنگل تصادفی، Random (Tree)	پرسپترون چندلایه (Multilayer perceptron)	درخت J۴۸	رگرسیون لجستیک (Logistic Regression)	پنهان ساده (Hidden Naive Bayes)	شمارش بوردا (Borda Count)
۱	وضعیت اکسیژن خون	وضعیت اکسیژن خون	وضعیت اکسیژن خون	وضعیت اکسیژن خون	وضعیت اکسیژن خون	وضعیت اکسیژن خون
۲	سن	هوشیاری	سرفه	بدن درد	سن	سن
۳	هوشیاری	بیماری قلبی	سن	هوشیاری	بدن درد	هوشیاری
۴	سرفه	سرفه	تب	سن	سرفه	تب

۵	سابقه تماس	بدن درد	سابقه تماس	تب	تب	سرفه
۶	علائم تنفسی	فشار خون	علائم تنفسی	علائم گوارشی	سابقه تماس	بدن درد
۷	بیماری قلبی	تب	بیماری قلبی	سابقه تماس	سردرد	بیماری قلبی
۸	سردرد	نتیجه تست	بیماری مزمن	فشار خون	علائم گوارشی	سردرد
۹	تب	سابقه تماس	هوشیاری	سردرد	هوشیاری	علائم تنفسی
۱۰	دیالیز	علائم تنفسی	نتیجه تست	دیابت	بیماری قلبی	علائم گوارشی
۱۱	فشار خون	علائم گوارشی	بدن درد	دیالیز	دیابت	فشار خون
۱۲	بیماری خونی	بیماری مزمن	فشار خون	بیماری خونی	دیالیز	دیابت
۱۳	بدن درد	سردرد	دیابت	بیماری قلبی	بیماری مزمن	بیماری مزمن
۱۴	نتیجه تست	دیابت	سردرد	بیماری مزمن	بیماری مزمن	نتیجه تست
۱۵	علائم گوارشی	دیالیز	علائم گوارشی	نتیجه تست	فشار خون	دیالیز
۱۶	دیابت	بیماری خونی	بیماری خونی	علائم تنفسی	علائم تنفسی	بیماری خونی
۱۷	بیماری مزمن	سن	دیالیز	سرفه	نتیجه تست	سابقه تماس

همان‌طور که در جدول دیده می‌شود وضعیت اکسیژن خون همان‌طور که در جدول دیده می‌شود وضعیت اکسیژن خون بیماران در هنگام پذیرش پراهمیت‌ترین ویژگی و یا عامل خطر مرگ است. با دقت در جدول مشاهده می‌شود در همه پنج الگوریتم پایه، این ویژگی، مهم‌ترین ویژگی بوده و رتبه اول را به دست آورده است. بعد از آن سن، وضعیت هوشیاری، تب، سرفه و بدن درد به ترتیب پنج عامل مهم دیگر در پیش‌بینی مرگ بیماران شناخته شدند. بعد از رتبه‌بندی ویژگی‌ها، در محیط نرم‌افزار Weka ۳.۹.۵

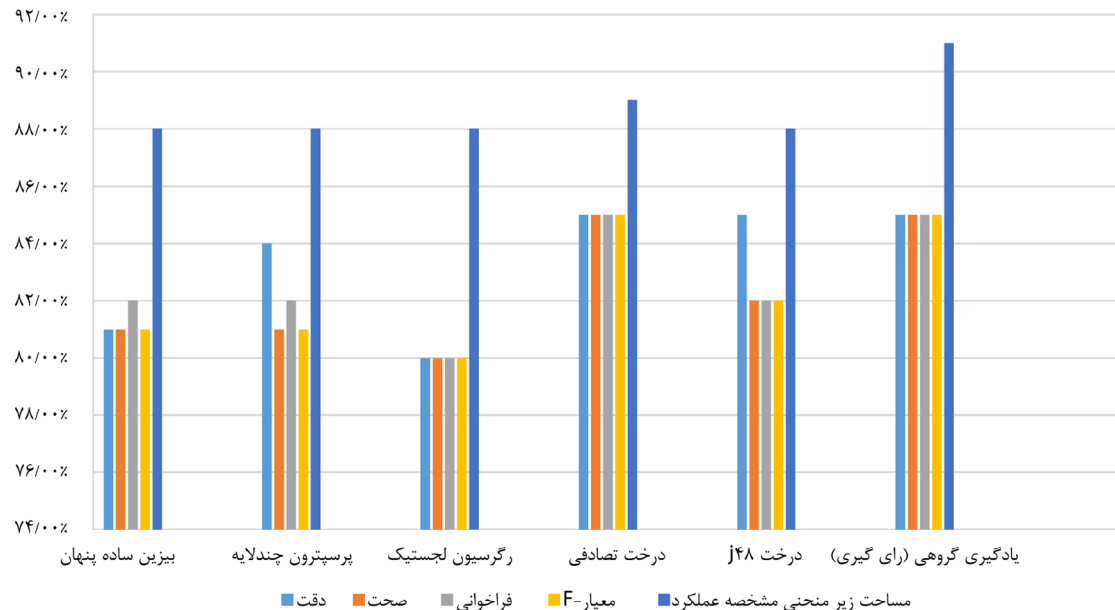
و با استفاده از یادگیری گروهی (Ensemble learning) و در نظر گرفتن پنج الگوریتم یادگیری ماشین به‌عنوان الگوریتم‌های پایه، مدل پیش‌بینی مرگ بیماران ایجاد شد. مدل‌سازی نهایی با ترکیب نتایج حاصل از این الگوریتم‌ها توسط روش یادگیری گروهی انجام شد. اگرچه هدف پژوهش مقایسه عملکرد مدل ایجاد شده با الگوریتم یادگیری گروهی با الگوریتم‌های پایه خود نیست، اما به‌منظور اثبات برتری مدل پیشنهادی پژوهش نسبت به هر یک از الگوریتم‌های پایه تکی، معیارهای ارزیابی عملکرد هر یک از این الگوریتم‌ها جهت مقایسه در جدول ۵ آورده شده است.

جدول ۵. مقایسه معیارهای ارزیابی ۵ الگوریتم یادگیری ماشین و یادگیری گروهی

الگوریتم	درصد مرگ واقعی (TP)	کاذب مرگ (FP)	درصد دقت، Accuracy	درصد صحت، Precision	درصد فراخوانی، Recall	معیار F-measure	درصد، مساحت زیر منحنی تشخیص عملکرد، AUROC
(HNB) بیزین ساده پنهان	۸۱	۲۷	۸۱	۸۱	۸۲	۸۱	۸۸
پرسپترون چندلایه (Multilayer perceptron)	۸۲	۲۸	۸۴	۸۱	۸۲	۸۱	۸۸
رگرسیون لجستیک (Logistic Regression)	۸۰	۳۰	۸۰	۸۰	۸۰	۸۰	۸۸
درخت تصادفی Random Tree	۸۵	۱۹	۸۵	۸۵	۸۵	۸۵	۸۹
درخت J۴۸	۸۵	۲۶	۸۵	۸۲	۸۲	۸۲	۸۸
رای‌گیری (یادگیری گروهی) Ensemble Learning (voting)	۸۵	۲۱	۸۵	۸۵	۸۵	۸۵	۹۱

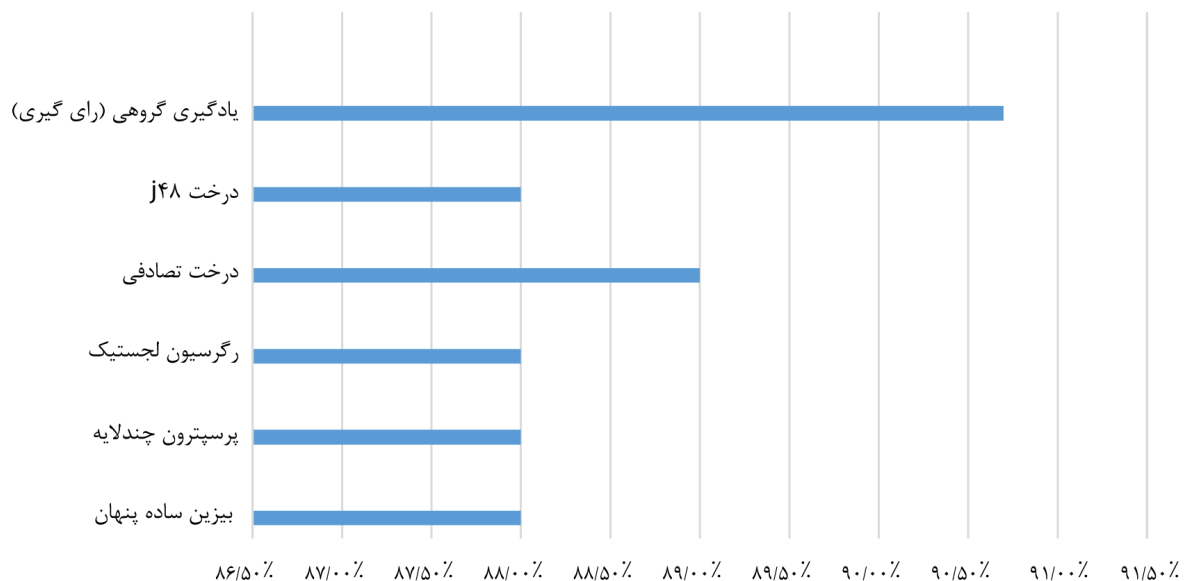
درخت تصادفی (۸۵ درصد) و درخت j۴۸ (۸۵ درصد) با دقت روش یادگیری گروهی (۸۵ درصد) برابر است (شکل ۳) اما مساحت سطح زیر نمودار Roc یادگیری گروهی (۹۱ درصد) از همه الگوریتم‌های پایه خود بیشتر است (شکل ۴). به عبارت دیگر، این روش بهتر از الگوریتم‌های پایه خود، توانسته افراد در معرض خطر مرگ را تشخیص دهد. شکل شماره ۵ نمودار Roc مربوط به روش یادگیری گروهی را نشان می‌دهد.

همان‌طور که در این جدول دیده می‌شود، از بین همه الگوریتم‌های پایه، رگرسیون لجستیک ضعیف‌ترین و درخت تصادفی بهترین عملکرد را داشته است. همچنین سطح زیر نمودار Roc جز در درخت تصادفی (۸۹ درصد)، در سایر الگوریتم‌های پایه با هم برابر و (۸۸ درصد) بوده است. در سطر آخر جدول معیارهای ارزیابی روش پیشنهادی آورده شده است. با توجه در جدول می‌توان دریافت که دقت روش

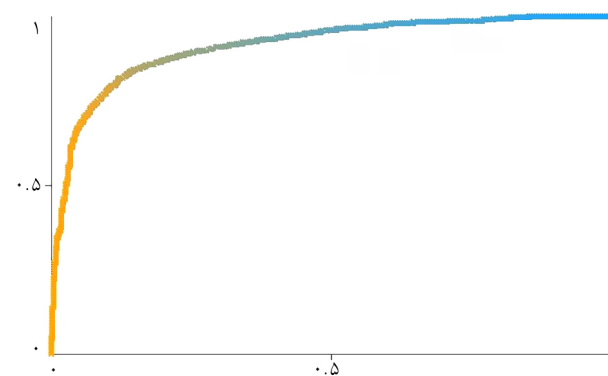


تصویر ۳. مقایسه بصری معیارهای عملکردی الگوریتم‌های یادگیری گروهی و پایه در پیش‌بینی خطر مرگ بیماران کووید-۱۹.

مساحت زیر منحنی Roc الگوریتم یادگیری گروهی و الگوریتم‌های پایه



تصویر ۴. مساحت زیر منحنی Roc الگوریتم یادگیری گروهی و الگوریتم‌های پایه.



تصویر ۴. نمودار ROC یادگیری گروهی با ۱۷ متغیر (AUC-ROC = 0.91)

۵. بحث و نتیجه گیری

این مطالعه یک مدل یادگیری ماشین مبتنی بر یادگیری گروهی را گزارش می‌کند که با هدف تحلیل (شناسایی و رتبه‌بندی) مهم‌ترین عوامل موثر در مرگومیر و ایجاد یک مدل یادگیری ماشین مبتنی بر یادگیری گروهی برای شناسایی بیماران مبتلا به کووید-۱۹ در معرض خطر مرگ به منظور تریاژ موثرتر آن‌ها انجام شد. داده‌های مورد استفاده پژوهش اطلاعات بیماران قطعی مبتلا به کووید-۱۹ مراجعه‌کننده به بیمارستان شهدای خلیج فارس بوشهر در بازه زمانی اسفند ۱۳۹۸ تا اردیبهشت و موج اول تا چهارم بیماری و سویه آلفا و بتا و گاما بود. در این مطالعه، پس از پیش‌پردازش داده‌ها، ۳۰ ویژگی یا عامل وارد مطالعه شد که با استفاده از آزمون کای-دو، ۱۷ عامل تاثیرگذار در مرگ بیماران کووید-۱۹ شناخته شد. با اعمال پنج الگوریتم پایه یادگیری ماشین و به کارگیری روش شمارش بوردا، این عوامل موثر از نظر اهمیت رده‌بندی شد. سپس یک مدل پیش‌بینی مرگ بیماران کووید-۱۹ با استفاده از یادگیری گروهی ایجاد شد و در نهایت عملکرد آن مورد ارزیابی قرار گرفت.

نتایج تحلیل توصیفی داده‌های بیماران حاکی از آن بود که مردان بیش از زنان به بیماری کووید-۱۹ مبتلا شده‌اند. در دو مطالعه مشابه انجام شده توسط جعفری و همکاران و طالبی و همکاران، میزان ابتلا زنان به این بیماری بیش از مردان بود (۴۳، ۴۴) که خلاف یافته‌ی پژوهش حاضر است، اما در دو مطالعه مروری گزارش شد که مردان بیش از زنان به این بیماری دچار شدند (۴۵، ۴۶) که با نتایج مطالعه ما همخوانی داشت. در تبیین این یافته، مطالعاتی وجود دارد که بیان می‌کنند هورمون‌های جنسی نقش محافظتی از کروموزوم X در ایمنی تطبیقی و ذاتی دارند (۴۷). از طرفی دیگر با توجه به این‌که مردان به دلیل کار و فعالیت در اجتماع، تعاملات بیشتری دارند لذا بیشتر در معرض خطر

ابتلا به بیماری‌های واگیردار قرار می‌گیرند (۴۸، ۴۹). در بخشی دیگر از یافته‌ها میانگین سنی بیماران مورد بررسی ۴۷/۶ سال به دست آمد. به عبارتی بیشتر مبتلایان میانسال بودند. در مطالعه مروری Cao و همکاران نیز میانگین سنی بیماران ۴۶/۶۲ گزارش شد. همچنین در دیگر مطالعه مروری انجام شده توسط Rodriguez-Morales این عدد حدود ۵۱/۹۷ بود (۴۵، ۴۶). گزیده کار و همکاران نیز میانگین سنی بیماران را ۴۷ سال بیان کردند. در تبیین این یافته می‌توان گفت معمولاً میانسالان قشر فعال جامعه هستند و مسئولیت‌های بیشتری نظیر پیگیری امور خارج از خانه را بر عهده دارند و از طرف دیگر شیوع بیماری‌های زمینه‌ای در این گروه سنی نیز نسبتاً بالا است. در نتیجه این دو عامل، این افراد بیش از سایر گروه‌های سنی در معرض خطر قرار می‌گیرند.

در بررسی فراوانی علائم و نشانه‌های بیماری، علائم تنفسی (خس‌خس سینه و تنگی نفس)، سرفه و تب شایع‌ترین علائم بودند. این نتایج با چندین مطالعه مشابه همسو است (۵۰، ۵۲). در مطالعه طالبی و همکاران علائم تنگی نفس در بدو ورود بیماران شایع‌ترین علامت بوده و پس از آن سرفه و تب قرار داشت (۴۴).

میزان مرگومیر بیماران در این مطالعه ۹/۸ درصد به دست آمد. در مطالعات مشابه در کشورها و استان‌های دیگر این عدد متفاوت بود. طالبی و همکاران مرگومیر را در یکی از بیمارستان‌های شهر سبزوار ۱۴ درصد بیان کردند (۴۴). دو مطالعه مشابه دیگر در چین این عدد را ۱۴/۱ درصد و ۱۵ درصد گزارش نمودند (۵۰، ۵۲). یکسان نبودن عدد به دست آمده در این مطالعات می‌تواند تفاوت جمعیت مورد مطالعه، ژنتیک، زمان انجام پژوهش و سویه‌های متفاوت بیماری باشد.

در این پژوهش ۱۷ عامل یا ویژگی از ۳۰ عامل مورد بررسی، در پیش‌بینی مرگ بیماران موثر بود. رده‌بندی این ویژگی بر اساس اهمیت نشان داد که وضعیت اکسیژن خونی، سن، وضعیت هوشیاری، تب، سرفه و بدن درد شش عامل اول تاثیرگذار در مرگ بیماران است. مطالعات متعددی نشان دادند که افراد مسن، افراد دارای حداقل مشکلات پزشکی زمینه‌ای/بیماری‌های همراه، مانند بیماری‌های قلبی عروقی، دیابت، فشارخون، بیماری کلیوی و بیماری انسداد ریه بیشتر از دیگران در معرض خطر مرگ این عفونت هستند (۵۳، ۵۵). همچنین نتایج تعدادی از مطالعات اپیدمیولوژی کووید-۱۹ بر اساس اطلاعات جمعیت‌شناختی خاص، نشان داد که سن بالاتر از ۶۵ سال، قومیت، نژاد و جنسیت از دیگر عوامل موثر در مرگومیر افراد است (۲، ۱۱، ۳۷، ۳۹). بنابراین به‌طور کلی می‌توان گفت که عواملی

با روش یادگیری گروهی با سه الگوریتم پایه جنگل تصادفی رگرسیون لجستیک و گرادیان تقویت شده پرداختند. مدل به دست آمده دارای دقت ۹۸/۱۳ درصد و فراخوانی ۹۹ درصد و $AUROC = 0.97$ بود (۶۶).

Kenji و همکاران نیز در پژوهش خود برای پیش‌بینی شانس زنده ماندن بیماران کووید-۱۹ مدلی با روش ensemble stacked model و با استفاده از ۴۸ ویژگی از هر بیمار ایجاد کردند و سطح زیر نمودار $AUROC = 0.907$ بدست آمد (۳۰).

مقایسه معیارهای عملکردی پژوهش حاضر با این مطالعات نشان می‌دهد که پژوهش حاضر دقت و عملکرد پایین‌تری نسبت به آن‌ها داشت. دلیل آن می‌تواند تفاوت نوع الگوریتم ترکیب‌کننده و پایه، در نظر گرفتن داده‌های آزمایشگاهی خون در پیش‌بینی‌کننده‌های آن پژوهش‌ها باشد. با این حال مطالعه Dhruv Patel و همکاران نشان داد که علائم هنگام پذیرش و اطلاعات دموگرافیک بدون در نظر گرفتن فاکتورهای خونی با کاهش اندکی دقت توانایی پیش‌بینی مرگ و میر بیماران را دارد (۳۱) که این خود نیز اعتبار نتایج پژوهش حاضر را تایید می‌کند.

از مهمترین نقاط قوت این پژوهش می‌توان به اعتبار مدل بر اساس نمودار ROC در پیش‌بینی افراد در معرض خطر مرگ بر اساس داده‌های جمعیت شناختی و بالینی و بدون در نظر گرفتن داده‌های آزمایشگاهی خون مورد استفاده در اکثر مطالعات پیشین اشاره کرد. این مدل به ویژه زمانی مفید واقع می‌شود که تصمیمات سریع مورد نیاز باشد اما زمان یا منابع لازم برای به دست آوردن داده‌های نشانگر آزمایشگاهی به موقع در دسترس نیست (۳۱). همچنین در مقایسه با روش‌های آماری محض، نتایج تجربی در مدل‌های آماری و ریاضی در پژوهش‌های قبلی نشان می‌دهد که این روش‌ها در مدیریت حجم زیادی از داده‌ها ناتوان بوده و پیش‌بینی ضعیفی برای این بیماری ارائه کرده‌اند، اما استفاده رویکرد یادگیری ماشین این محدودیت‌ها را نداشته و با تسهیل تصمیم‌گیری پیچیده و جستجوی واقعیت، پتانسیل بالایی در مقابل بحران کووید-۱۹ از خود نشان می‌دهد.

معیارهای ارزیابی عملکرد، قابل اعتماد بودن مدل پیشنهادی را نشان می‌دهد. با این حال برای بهبود مدل و قابلیت تعمیم بیشتر آن، به داده‌های ملی و در یا سطح بین‌المللی نیاز است. پیشنهاد می‌شود در پژوهش‌های آتی با در نظر گرفتن تاثیر سایر ویژگی‌های بالینی و نتایج آزمایشگاهی مهم در مطالعات مشابه و با به کارگیری داده‌های چند مرکزی و همچنین الگوریتم‌های متنوع‌تر،

نظیر سن، مرد بودن، بیماری زمینه‌ای به‌طور قابل توجهی با پیش‌آگهی بدتر مرتبط است (۲۶، ۳۸، ۵۴، ۵۶، ۵۷). تقریباً در همه مطالعات، سالمندی یک عامل پیش‌آگهی قوی است که با نتایج ما هم‌راستا است. بدیهی است که با افزایش سن و تضعیف سیستم ایمنی و یا شرایط پزشکی متعدد دیگر، افراد بیشتر مستعد ابتلا به بیماری بحرانی کووید-۱۹ می‌شوند.

بیماری‌های زمینه‌ای مرتبط با پیش‌آگهی نامناسب شامل فشارخون بالا (۲۶) (۳۸، ۵۸، ۶۰)، دیابت (۲۶) (۳۸، ۵۷، ۵۹)، بیماری ریوی از جمله بیماری مزمن انسدادی ریه و آسم (۳۸)، (۵۴) (۵۸)، بیماری‌های قلبی عروقی (۵۷، ۵۸) (۵۹)، سرطان (۶۱) (۶۲) و بیماری مزمن کلیوی (۳۸، ۶۰، ۶۳) پیش‌تر گزارش شده‌اند که در نتایج پژوهش حاضر برخی از آن‌ها نیز جزو ۱۷ عامل موثر در مرگ بیماران شناخته شده‌اند. مقایسه بین نتایج معیارهای ارزیابی عملکرد روش یادگیری گروهی و الگوریتم‌های پایه خود نشان داد که اگرچه دقت این روش با برخی از الگوریتم‌های پایه خود (درخت تصادفی و درخت $j48$) یکسان (۸۵ درصد) بود، اما با داشتن بیشترین میزان سطح زیر نمودار ROC قدرت بالاتری در شناسایی بیماران با خطر مرگ را دارد. این یافته مورد انتظار بود، زیرا روش‌های یادگیری گروهی با ترکیب نتایج الگوریتم‌های پایه خود عملکرد بهتری نسبت به آن‌ها ارائه می‌کنند.

معیارهای عملکردی مدل ایجادشده در این پژوهش به ترتیب دقت ۸۵ درصد، حساسیت ۸۵ درصد، فراخوانی ۸۵ درصد و سطح زیر نمودار $AUROC = 0.91$ به دست آمد. مطالعات پیشین $AUROC$ را بین ۶۰ درصد تا ۹۰ درصد به دست آوردند که علاوه بر علائم اولیه، بر تست‌های داده‌های آزمایشگاهی خون، تصاویر در زمان تشخیص و یا در طول بیماری متکی بودند (۵، ۱۷، ۲۶، ۶۴، ۶۵).

در زمینه به کارگیری یادگیری گروهی در پیش‌بینی مرگ بیماران، Gao و همکاران در مطالعه خود با به کارگیری روش یادگیری گروهی با چهار الگوریتم پایه شامل رگرسیون لجستیک، ماشین بردار پشتیبان، درخت تصمیم گرادیان تقویت‌شده و شبکه عصبی با استفاده از ۱۴ متغیر یک مدل برای با خطر مرگ و میر استفاده کرده و امکان پیش‌بینی وخامت فیزیولوژیکی و مرگ را تا ۲۰ روز قبل از مرگ ایجاد کردند. مدل در یک گروه اعتبارسنجی داخلی و دو گروه اعتبارسنجی خارجی ارزیابی شد که به ترتیب سطح زیر نمودار Roc ۰/۹۶۲۱، ۰/۹۷۶ و ۰/۹۲۴۶ به دست آمد (۱۴). همچنین Prajoy Podder و همکاران در پژوهش خود به شناسایی بیماران در معرض خطر مرگ با استفاده از ۱۱۱ متغیر مربوط به علائم بالینی و داده‌های آزمایشگاهی خون و

۵. به دلیل در دسترس نبودن نتایج آزمایش‌های بالینی بیمار (ازجمله آزمایش خون و تصاویری سی تی اسکن)، مدل تنها با داده‌های کیفی هنگام پذیرش بیمار ایجاد شد. این کار باعث ایجاد محدودیت استفاده از الگوریتم‌های یادگیری ماشین متنوع‌تر برای مدل‌سازی شد.

تشکر و سپاسگزاری

از مسئولین محترم بیمارستان شهدای خلیج فارس بوشهر و همچنین معاونت محترم تحقیقات و فناوری دانشگاه علوم پزشکی بوشهر که اجرای این مطالعه را تسهیل نمودند، تشکر و قدردانی می‌گردد.

نقش و مشارکت نویسندگان

ن.س.ب و ج.ک: ایده و طراحی مطالعه؛ ن.س.ب و ش.ر.ن.ک: نگارش نسخه اول؛ ب.ا، ف.ع و ش.د: تفسیر نتایج؛ ع.ج و ص.د: جمع‌آوری داده و آنالیز. همه نویسندگان نتایج را بررسی نموده و نسخه نهایی مقاله را تایید نمودند.

حمایت مالی / معنوی

این مطالعه توسط هیچ سازمان یا موسسه ای تامین مالی نشده است.

کد اخلاق:

این مطالعه توسط کمیته اخلاق دانشگاه علوم پزشکی بوشهر با کد اخلاق: (IR.BPUMS.REC1399.003) تایید شده است.

تضاد منافع:

نویسندگان اظهار داشتند که فاقد هرگونه تضاد منافع هستند.

References

1. Organization WH. Coronavirus disease 2019 (COVID-19): situation report, 94. 2020.
2. Wu C, Chen X, Cai Y, Zhou X, Xu S, Huang H, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med. 2020;180(7):934-43. <https://doi.org/https://doi.org/10.1001/jamainternmed.2020.0994>.
3. Hui DS, Azhar EI, Kim Y-J, Memish ZA, Oh M-d, Zumla A. Middle East respiratory syndrome coronavirus: risk factors and determinants of primary, household, and nosocomial transmission. Lancet Infect Dis.

مدل‌هایی با قابلیت اعتماد بالاتر ایجاد شود. همچنین با توجه به شناسایی عوامل پیش‌بینی‌کننده خطر مرگ در این پژوهش پیشنهاد می‌گردد چک‌لیست‌هایی متشکل از این عوامل جهت به‌کارگیری در فرایند تریاژ بیماران ایجاد شود.

نتایج مطالعه یک راه‌حل کم‌هزینه، سریع و نوآورانه برای شناسایی زودهنگام بیماران مبتلا به کووید-۱۹ در معرض خطر مرگ برای تریاژ موثرتر آن‌ها ارائه می‌نماید که می‌تواند با به‌روزرسانی مداوم با داده‌های جدید در تصمیم‌گیری به مدیران سیستم‌های مراقبت بهداشتی به منظور مدیریت منابع و ارائه درمان‌های متناسب با بیماران با خطر مرگ بالا یاری رساند.

۱.۵. محدودیت‌های پژوهش

اگرچه پژوهش حاضر از نتایج خوبی برخوردار بود اما در اجرای آن محدودیت‌هایی وجود داشت. در زیر به برخی از آن‌ها اشاره شده است:

۱. پایگاه داده (مرکز پایش مراقبت پزشکی سیستم اطلاعات بیمارستانی) مورد استفاده پژوهش، تنها شامل اطلاعات بیماران استان بوشهر (تک مرکزی) بود اما با توجه به این‌که که میزان مرگ‌ومیر ناشی از کووید-۱۹ بسته به منطقه ناهمگن است، این موضوع ممکن است تعمیم‌پذیری مدل را تحت تاثیر قرار دهد.

۲. به علت فقدان برخی از داده‌های آزمایشگاهی (خون و سی تی اسکن) بیماران و در نظر نگرفتن همه علائم بیماران در پژوهش حاضر، تعدادی از عوامل تاثیرگذار در مرگ بیماران شناسایی نشد، همین امر باعث شد عملکرد مدل تحت تاثیر منفی قرار گیرد.

۳. پژوهش با داده‌های موج اول تا چهارم بیماری و سویه آلفا، بتا و گاما ویروس انجام شد. لذا با توجه به موج‌های بعدی بیماری و همچنین جهش‌های مختلف ویروس و بالطبع آن رفتارهای متفاوت، نتایج پژوهش نیاز به بازنگری داشته و این ابزار جهت استفاده در آینده باید به‌طور مداوم به‌روزرسانی شود تا از اطمینان بیشتری برخوردار باشد.

۴. در این مطالعه پیش‌بینی مرگ بیماران بر اساس داده‌های هنگام پذیرش انجام شد. با توجه این‌که اقدامات مراقبتی و یا وقایع دیگر در زمان بستری، می‌تواند بر روی سیر بالینی بیمار تاثیرگذار باشد و آن را از احتمال قبلی دور کند، لذا نمی‌توان تنها به داده‌های اولیه هنگام پذیرش اکتفا نمود.

- 2018;18(8):e217-e27.
4. Pardhan S, Vaughan M, Zhang J, Smith L, Chichger H. Sore eyes as the most significant ocular symptom experienced by people with COVID-19: a comparison between pre-COVID-19 and during COVID-19 states. *BMJ Open Ophthalmol.* 2020;5(1):e000632.
5. Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology.* 2020;296(2):E86-E96. <https://doi.org/https://doi.org/10.1148/radiol.2020201433>.
6. Hui DS, Azhar EI, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis.* 2020;91:264-6.
7. Oran DP, Topol EJ. The proportion of SARS-CoV-2 infections that are asymptomatic: a systematic review. *Ann Intern Med.* 2021;174(5):655-62.
8. Lazzerini M, Putoto G. COVID-19 in Italy: momentous decisions and many uncertainties. *Lancet Glob Health.* 2020;8(5):e641-e2.
9. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *Jama.* 2020;323(16):1545-6. <https://doi.org/https://doi.org/10.1001/jama.2020.4031>.
10. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci U S A.* 2020;117(16):9122-6. <https://doi.org/https://doi.org/10.1073/pnas.2004064117>.
11. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents.* 2020;55(3):105924.
12. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: Knowledge in progress. *Diabetes Res Clin Pract.* 2020;162:108142.
13. Kabir MJ, Heidari A, Moeini S, Khatirnamani Z, Kavian Telouri F, Eimery M. Calculation of Direct Medical Costs and Indirect Costs in Patients with Covid-19 Hospitalized in the Intensive Care Unit in Golestan Province. *Manage Strat Health Syst.* 2022;6(4):308-16.
14. Gao Y, Cai G-Y, Fang W, Li H-Y, Wang S-Y, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun.* 2020;11(1):1-10.
15. Organization WH. Maintaining essential health services: operational guidance for the COVID-19 context: interim guidance, 1 June 2020: World Health Organization; 2020 Contract No.: Document Number|.
16. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. *PloS one.* 2020;15(7):e0236618.
17. Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critically ill patients with novel coronavirus disease. *Acad Emerg Med.* 2020;27(6):461-8.
18. Shanbehzadeh M, Orooji A, Kazemi-Arpanahi H. Comparing of data mining techniques for predicting in-hospital mortality among patients with covid-19. *J Biostat Epidemiol.* 2021;7(2):154-73.
19. Josephus BO, Nawir AH, Wijaya E, Moniaga JV, Ohlyver M. Predict mortality in patients infected with COVID-19 virus based on observed characteristics of the patient using logistic regression. *Procedia Comput Sci.* 2021;179:871-7.
20. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. China medical treatment expert group for Covid-19. *N Engl J Med.* 2019;382(18):1708-20.
21. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *Jama.* 2020;323(20):2052-9.
22. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell.* 2020;2(5):283-8.
23. Malki Z, Atlam E-S, Hassanien AE, Dagnew G, Elhosseini MA, Gad I. Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos Solitons Fractals.* 2020;138:110137.
24. Cho A. AI systems aim to sniff out coronavirus outbreaks. *American Association for the Advancement of Science;* 2020.
25. Xiong Z, Wang R, Bai H, Halsey K, Mei J, Li Y, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology.* 2020;296(3):E156-E65. <https://doi.org/https://doi.org/10.1148/radiol.2020201491>.
26. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. *Eur Respir J.* 2020;56(2). <https://doi.org/https://doi.org/10.1183/13993003.01104-2020>.
27. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj.* 2020;369. <https://doi.org/https://doi.org/10.1136/bmj.m1328>.
28. Chowdhury ME, Rahman T, Khandakar A, Al-Madeed S, Zughaier SM, Doi SA, et al. An early warning tool for predicting mortality risk of COVID-19 patients using machine learning. *Cognit Comput.* 2021;1-16. <https://doi.org/https://doi.org/10.1007/s12559-020-09812-7>.
29. Nemati M, Ansary J, Nemati N. Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. *Patterns.* 2020;1(5):100074. <https://doi.org/https://doi.org/10.1016/j.patter.2020.100074>.

- org/10.1016/j.patter.2020.100074.
30. Ikemura K, Bellin E, Yagi Y, Billett H, Saada M, Simone K, et al. Using automated machine learning to predict the mortality of patients with COVID-19: Prediction model development study. *J Med Internet Res*. 2021;23(2):e23458. <https://doi.org/https://doi.org/10.2196/23458>.
31. Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, et al. Machine learning based predictors for COVID-19 disease severity. *Sci Rep*. 2021;11(1):1-7. <https://doi.org/https://doi.org/10.1038/s41598-021-83967-7>.
32. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: Model development and validation. *J Med Internet Res*. 2020;22(11):e24018. <https://doi.org/https://doi.org/10.2196/24018>.
33. Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC Med Inform Decis Mak*. 2022;22(1):1-12.
34. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *MedRxiv*. 2020.
35. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *MedRxiv*. 2020.
36. Shanbehzadeh M, Valinejadi A, Afrah R, Kazemi-Arpanahi H, Orooji A, Kaffashian M. Comparison of machine-learning algorithms efficiency to build a predictive model for mortality risk in COVID-19 hospitalized patients. *Faslnamahi Kumish*. 2022;24(1):128-38.
37. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. 2020;180(8):1081-9. <https://doi.org/https://doi.org/10.1001/jamainternmed.2020.2033>.
38. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *lancet*. 2020;395(10229):1054-62. [https://doi.org/https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/https://doi.org/10.1016/S0140-6736(20)30566-3).
39. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, et al. Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with COVID-19. *Radiology*. 2020;297(1):E197-E206. <https://doi.org/https://doi.org/10.1148/radiol.2020201754>.
40. Karthikeyan A, Garg A, Vinod P, Priyakumar UD. Machine learning based clinical decision support system for early COVID-19 mortality prediction. *Front Public Health*. 2021;9:626697.
41. An C, Lim H, Kim D-W, Chang JH, Choi YJ, Kim SW. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Sci Rep*. 2020;10(1):1-11.
42. Kuncheva LI. Combining pattern classifiers: methods and algorithms. John Wiley & Sons; 2014.
43. Jafari M, Akbari M, Navidkia M, Dashtbin S, Mousavi SF, Heidary M, et al. Comparison of clinical, radiological and laboratory findings in discharged and dead patients with COVID-19. *Vacunas*. 2022.
44. Talebi S, Nematshahi M, Tajabadi A, Khosrogerdi A. Comparison of clinical and epidemiological characteristics of deceased and recovered patients with COVID-19 in Sabzevar, Iran. *Journal Mil Med*. 2020;22(6):509-16.
45. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis*. 2020;34:101623.
46. Cao Y, Liu X, Xiong L, Cai K. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2: a systematic review and meta-analysis. *J Med Virol*. 2020;92(9):1449-59.
47. Hassan SA, Sheikh FN, Jamal S, Ezech JK, Akhtar A. Coronavirus (COVID-19): a review of clinical features, diagnosis, and treatment. *Cureus*. 2020;12(3).
48. Alipio M, Pregoner JD. Epidemiological Characteristics of An Outbreak of Coronavirus Disease 2019 in the Philippines (April 3, 2020). Available at SSRN 3568934.
49. Hooper GL. Health Seeking in Men: A Concept Analysis. *Urol Nurs*. 2016;36(4).
50. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *lancet*. 2020;395(10223):497-506.
51. Ki M. Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. *Epidemiol Health*. 2020;42.
52. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *bmj*. 2020;368.
53. Liu Y, Mao B, Liang S, Yang J. Association between age and clinical characteristics and outcomes of COVID-19. *Eur Respir J*;55(5):2001112. <https://doi.org/https://doi.org/10.1183/13993003.01112-2020>.
54. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: cardiovascular disease, drug therapy, and mortality in Covid-19. *N Engl J Med*. DOI: 10.1056/NEJ-Moa2007621. Mass Medical Soc; 2020.
55. Bajgain KT, Badal S, Bajgain BB, Santana MJ. Prevalence of comorbidities among individuals with COVID-19: A rapid review of current literature. *Am J Infect Control*. 2021;49(2):238-46. <https://doi.org/https://doi.org/10.1016/j.ajic.2020.06.213>.
56. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *Clin Infect Dis*. 2020;71(15):833-40.
57. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool to

- early predict severe 2019-novel coronavirus pneumonia (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *MedRxiv*. 2020. <https://doi.org/10.1101/2020.03.17.20037515>.
58. Estiri H, Strasser ZH, Klann JG, Naseri P, Waghlikar KB, Murphy SN. Predicting COVID-19 mortality with electronic medical records. *NPJ Digit Med*. 2021;4(1):1-10. <https://doi.org/10.1038/s41746-021-00383-x>.
59. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care*. 2020;24(1):1-4. <https://doi.org/10.1186/s13054-020-2833-7>.
60. Hajifathalian K, Sharaiha RZ, Kumar S, Krisko T, Skaf D, Ang B, et al. Development and external validation of a prediction risk model for short-term mortality among hospitalized US COVID-19 patients: A proposal for the COVID-AID risk tool. *PloS one*. 2020;15(9):e0239536. <https://doi.org/10.1371/journal.pone.0239536>.
61. Lee LY, Cazier J-B, Starkey T, Briggs SE, Arnold R, Bisht V, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol*. 2020;21(10):1309-16. [https://doi.org/10.1016/S1473-3099\(20\)30442-3](https://doi.org/10.1016/S1473-3099(20)30442-3).
62. Albiges L, Foulon S, Bayle A, Gachot B, Pommeret F, Willekens C, et al. Determinants of the outcomes of patients with cancer infected with SARS-CoV-2: results from the Gustave Roussy cohort. *Nat Cancer*. 2020;1(10):965-75. <https://doi.org/10.1038/s43018-020-00120-5>.
63. Arab S, Ebrahimi Z, Izadi A, Madanchi H, Yarmohammadi M, Darban M, et al. Relationship between the history of kidney disease, clinical findings, hospitalization duration, and mortality in COVID-19 patients. *Faslnameh Kumish*. 1400;24(1):102-8. eng.
64. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool to early predict severe 2019-novel coronavirus pneumonia (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *MedRxiv*. 2020. <https://doi.org/10.1101/2020.03.17.20037515>.
65. Estiri H, Strasser ZH, Klann JG, Naseri P, Waghlikar KB, Murphy SN. Predicting COVID-19 mortality with electronic medical records. *NPJ Digit Med*. 2021;4(1):1-10. <https://doi.org/10.1038/s41746-021-00383-x>.
66. Podder P, Mondal MRH, editors. Machine Learning to Predict COVID-19 and ICU Requirement. 2020 11th International Conference on Electrical and Computer Engineering (ICECE); 2020. IEEE.

Research Article

Triage of patients with COVID19-: Using Ensemble learning method for risk factor analysis and death prediction

Neda Sadat ¹, Sharareh R. Niakan Kalhori ^{2,3,*}, Shahrzad Darvishi ⁴, Jamileh Kiani ¹, Farhad Abbasi ⁵, Batool Amiri ¹, Erfan Javanmardi ¹, Safiyeh Daneshi ¹

¹ Clinical Research Development Unit, Persian Gulf Martyrs Hospital, Bushehr University of Medical Sciences, Bushehr, Iran

² Department of Health Information Management, Faculty of Paramedicine, Tehran University of Medical Sciences, Tehran, Iran

³ PLRI Informatics Research Institute, Medical School Hannover, Technical University of Braunschweig, Braunschweig, Germany

⁴ Imam Khomeini Hospital, Bushehr University of Medical Sciences, Bushehr, Iran

⁵ Department of Infectious Diseases, Faculty of Medicine, Bushehr University of Medical Sciences, Bushehr, Iran

*Corresponding Author: : Sharareh R. Niakan Kalhori: Department of Health Information Management, Faculty of Paramedicine, Tehran University of Medical Sciences, Tehran, Iran. Email: sh-rniakank@tums.ac.ir

Received 17/05/2022; Accepted 18/06/2023

Abstract

Background: Early identification of high-risk patients with COVID19- using non-laboratory data at the time of admission may help the effective use of limited healthcare resources and improve clinical decision-making which reduces cost and time, and consequently the death of patients.

Objectives: This study aims to provide an intelligent approach to triaging COVID19- patients by analyzing effective risk factors and predicting the risk of death due to COVID19- using an ensemble learning method.

Methods: : This is a descriptive-applied study that was conducted in 2021. Non-laboratory data were used during the admission of 4558 confirmed patients with COVID19- referred to the Shohaday-e-Khalij-e-Fars Hospital in Bushehr registered in the medical care information system. After data preprocessing, the risk factors affecting the death were identified and ranked in importance. The ensemble learning (voting) method was used to develop the death prediction model, and the confusion matrix criteria evaluated its performance.

Results: From a total of 5433 patient records, the data of 4558 cases were included in the study, of which %45.5) 2222) were women and %64.5) 2663) were men. The average age of the patients was 47.6. Out of all the investigated factors, 17 characteristics were identified as effective in predicting the death of patients, and their ranking indicated that the first six factors for predicting the death of patients are state of blood oxygen, age, state of consciousness, fever, cough, and body pain, respectively. The proposed model estimated the risk of death with an accuracy of %85, sensitivity of %85, specificity of %0.85, and AUROC of 0.91.

Conclusions: The results of the study provide a low-cost, fast, and innovative solution for the early identification of patients with COVID19- at risk of death to triage them more effectively, which can be used by the managers of healthcare systems to manage resources by continuously updating new data and providing appropriate treatments for high-risk patients.

Keywords: COVID19-, Coronavirus Infections, Machine Learning, Risk Factors, Triage