پیشنهاد یک راه‌کار فناورانه موثر جهت تشخیص زودهنگام بیماری کووید-۱۹: مطالعه مبتنی بر یادگیری ماشین داده محور

رئوف نوپور[۱] مصطفی شنبه‌زاده[۲] هادی کاظمی‌آرپناهی[۳*]

۱. کارشناسی ارشد، فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران. ORCID: 0000-0003-3370-2375
۲. گروه فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی ایلام، ایلام، ایران.
۳. گروه فناوری اطلاعات سلامت، دانشگاه علوم پزشکی آبادان،آبادان، ایران.

مجله اطلاع‌رسانی پزشکی نوین؛ دوره هفتم؛ شماره اول؛ بهار ۱۴۰۰؛ صفحات: ۶۸-۷۸.

چکیده

هدف: *تشخیص صحیح، دقیق و به موقع بیماری کووید-۱۹ با استفاده از فناوری‌های هوش مصنوعی و یادگیری ماشین نقش مهمی در بهبود شاخص‌های بیماری، استفاده بهینه از منابع محدود بیمارستانی و کاهش بار کاری کارکنان خط مقدم پاندمی خواهد داشت. بنابراین هدف پژوهش حاضر ارزیابی کارایی الگوریتم‌های منتخب داده‌کاوی در تشخیص بیماری کووید-۱۹ خواهد بود.*

روش‌ها: *پژوهش حاضر یک مطالعه گذشته‌نگر و توصیفی کاربردی است. در این مطالعه از داده‌های بیماران بستری شده با تشخیص قطعی کووید-۱۹ در بازه زمانی ۲۷ اسفند ۱۳۹۸ لغایت ۲۰ آذر ۱۳۹۹ که در پایگاه داده پرونده الکترونیک سلامت بیماری کووید-۱۹ بیمارستان آیت‌الله طالقانی شهرستان آبادان ثبت شده است، استفاده گردید. پس از اعمال معیارهای ورود و خروج برای شناسایی نمونه‌ها در نهایت ۴۰۰ رکورد به عنوان ورودی و تغذیه وارد نرم‌افزار داده کاوی وکا ورژن ۳.۹ شد. داده‌ها با استفاده از ملاک کای دو برای تعیین متغیرها به منظور آموزش الگوریتم‌ها، عملکرد آن‌ها براساس معیارهای مختلف ارزیابانه در ماتریس آشفتگی مورد مقایسه قرار گرفتند.*

نتایج: *براساس مقایسه عملکرد الگوریتم‌های داده‌کاوی با توجه به معیارهای ارزیابانه در ماتریس آشفتگی، الگوریتم J-48 با میزان حساسیت، دقت، و ضریب همبستگی ماتیوس به ترتیب ۰/۸۵، ۰/۸۵، ۰/۶۸ عملکرد بهتری نسبت به سایر الگوریتم‌های داده‌کاوی برای تشخیص بیماری کووید-۱۹ داشت. ۳ متغیر وجود ضایعات ریوی، تب و سابقه تماس با افراد مظنون به کرونا با در نظر گرفتن شاخص جینی ایندکس برای تعیین نقطه تقسیم، به ترتیب با میزان جینی ایندکس ۰/۲۱۷، ۰/۲۰۵ و ۰/۱۸۸ به عنوان مهم ترین فاکتورهای موثر در تشخیص کرونا در نظر گرفته شدند.*

نتیجه‌گیری: *استفاده از روش‌های داده کاوی منتخب و به طور خاص الگوریتم J-48 قابلیت بالایی در تشخیص به موقع و اثربخش بیماری کووید-۱۹ در قالب سیستم‌های پشتیبان تصمیم یار بالینی خواهد داشت.*

کلیدواژه‌ها: *کووید-۱۹، هوش مصنوعی، یادگیری ماشین، داده‌کاوی.*

نوع مقاله: پژوهشی

دریافت مقاله: ۱۳۹۹/۱۲/۲۳ اصلاح نهایی: ۱۴۰۰/۲/۲۱ پذیرش مقاله: ۱۴۰۰/۳/۲۰

ارجاع: نوپور رئوف، شنبه‌زاده مصطفی، کاظمی‌آرپناهی هادی. پیشنهاد یک راه‌کارفناورانه موثر جهت تشخیص زودهنگام بیماری کووید-۱۹: مطالعه مبتنی بر یادگیری ماشین داده محور. مجله اطلاع‌رسانی پزشکی نوین. ۱۴۰۰؛ (۱)۷: ۶۸-۷۸.

مقدمه:

بیماری کووید-۱۹ یک بیماری عفونی مشترک بین انسان و حیوان، مسری و منتشره می‌باشد که عامل بیماری‌زایی آن ویروسی به نام سندرم تنفسی حاد شدید کرونا ویروس ۲ (SARS-CoV-2) می‌باشد که باعث آسیب‌های جدی به دستگاه تنفسی، ایجاد ذات‌الریه و در مواردی منجر به مرگ انسان می‌گردد [۱]. این بیماری ابتدا در دسامبر سال ۲۰۱۹ در شهر ووهان، استان هوبیی و کشور چین گزارش شد و به تدریج به سراسر کشور و جهان انتشار پیدا کرد [۲-۴]. از مهم‌ترین علائم این بیماری می‌توان به مواردی همچون تب، سرفه خشک، تنگی نفس، سردرد، درد شکمی، علائم گوارشی، استفراغ، احساس سفتی در سینه و وجود

نویسنده مسئول:
هادی کاظمی‌آرپناهی
گروه فناوری اطلاعات سلامت، دانشگاه علوم پزشکی آبادان،آبادان، ایران.
تلفن: ۹۱۳۸۲۰۰۰۲۷ ۹۸+ پست الکترونیکی: h.kazemi@abadanums.ac.ir ORCID: 0000-0002-8882-5764

ضایعات ریوی و نارسایی‌های تنفسی اشاره کرد [۵،۶]. این بیماری ویروسی از طریق ترشحات دهان و بینی به هنگام سرفه یا عطسه و یا هنگام صحبت کردن از افراد آلوده به سالم منتقل می‌شود [۷،۸]. دوره کمون این بیماری در بیشتر افراد تا ۱۴ روز می‌باشد اما در مواردی تا ۲۴ روز نیز گزارش شده است [۶]. تعداد موارد ابتلا و فوتی در سراسر جهان در حال افزایش است، این در حالی است که ایران جزء ۱۰ کشور اول از لحاظ میزان شیوع این بیماری بوده است [۹]. بر طبق اطلاعات منتشر شده از سازمان بهداشت جهانی، تا ۱۶ اوت سال ۲۰۲۰ تعداد موارد مبتلابه بیماری کرونا به ۲۲۴۹۲۳۱۲ نفر رسیده است که تعداد مرگ و میر ناشی از این بیماری ۷۸۸۵۰۳ نفر اعلام شده است، در ایران نیز در این دوره تعداد ۳۵۲۵۵۸ مورد مبتلا و ۲۰۲۶۴ مرگ ناشی از این بیماری رخ داده است [۱۰]. با توجه به سرعت انتشار بالای این بیماری و افزایش تعداد موارد مبتلا به آن و مرگ و میرها در سطح جهانی و نیز آثار سوء بر وضعیت‌های اقتصادی و اجتماعی جوامع، بنابراین استفاده از روش‌های فناورانه نوین و پیشرفته جهت تشخیص سریع این بیماری در مراحل اولیه، جداسازی و قرنطینه موارد مبتلا و اعمال نظارت‌های دقیق و مؤثر در پیشگیری از این بیماری بسیار مهم تلقی می‌شود [۱۱]. همچنین، به دلیل ماهیت مبهم و حساس این بیماری و فقدان روش‌های درمانی اثبات شده از جمله داروی مؤثر برای ایمن‌سازی و درمان، کمبود منابع آزمایشگاهی با هدف ارزیابی و در دسترس نبودن روش‌هایی جهت کشف سریع این بیماری، بنابراین استفاده از روش‌های نوآورانه و غیرتهاجمی برای شناسایی آن در مراحل اولیه بسیار مفید و تعیین‌کننده خواهد بود [۱۲]. از مهم‌ترین فناوری‌های نوین و پیشرفته امروزی که نقش به سزایی در کنترل همه‌گیری کووید-۱۹ دارد، می‌توان به روش‌های هوش مصنوعی (یادگیری ماشینی و یادگیری عمیق)، رایانش ابری، اینترنت اشیاء، داده‌کاوی داده‌های بزرگ و فناوری‌های رباتیک و سیستم‌های هوشمند اشاره کرد [۱۳].

فرآیند داده‌کاوی شامل کشف الگوهای مهم از میان حجم انبوهی از داده‌ها می‌باشد که ترکیبی از یادگیری ماشین، علم آمار و سیستم‌های پایگاه داده می‌باشد [۱۴]. با توجه به حجم عظیم داده‌های موجود در حیطه پزشکی، این فرآیند تاکنون نقش بسیار مؤثری در مدیریت بیماری‌های گوناگون از قبیل پیش‌آگهی، تشخیص و درمان داشته است [۱۵]. تاکنون پژوهش‌های متعددی در زمینه‌ی کاربرد روش‌های داده‌کاوی در مدیریت و کنترل بیماری کووید-۱۹ با هدف نظارت و پیش‌بینی روندهای بیماری و ارزیابی‌های همه‌گیرشناسی، ارائه مدل‌های تشخیصی، تعیین بهترین مدل درمانی و مراقبتی و پیش‌گویی احتمال فوت انجام پذیرفته است. در مطالعه‌ی مروری که توسط Albahri و همکارانش در مورد نقش داده‌کاوی و هوش مصنوعی در زمینه‌ی کشف و تشخیص بیماری کووید-۱۹ انجام گرفت، در مورد معیارهای کمی و کیفی پژوهش‌های انجام شده در این حوزه بحث شد. در نهایت، نتایج نشان دادند که استفاده از این فناوری به منظور ارائه مدل‌های تشخیصی و شناسایی بهینه‌ترین و مؤثرترین الگوریتم داده‌کاوی می‌تواند در شناسایی به موقع، مؤثر و اقتصادی بیماری کمک قابل توجهی کند [۱۶].

با توجه به شیوع بالای بیماری در کشورمان و وجود برخی محدودیت‌ها و کمبودهای منابع، از این رو، هدف از انجام مطالعه حاضر ایجاد یک مدل مؤثر و کارآمد تشخیصی براساس مقایسه عملکرد الگوریتم‌های داده‌کاوی برای بیماری کووید-۱۹ و براساس معیارهای تشخیصی مختلف می‌باشد تا بتوان با تشخیص به موقع این بیماری از میزان شیوع و به تبع آن میزان مرگ ناشی از این بیماری در سطح جامعه کاست و همچنین بتوان به پزشکان به عنوان یک راه‌کار مناسب جهت افزایش دقت تشخیصی آنان کمک نمود.

مواد و روش‌ها:

مطالعه حاضر یک پژوهش گذشته‌نگر است که به صورت توصیفی-کاربردی در سال ۱۳۹۹ در بیمارستان مرکزی تشخیص و درمان بیماری کووید-۱۹ تابعه دانشگاه علوم پزشکی آبادان انجام شده است. هدف مطالعه حاضر تشخیص بیماری کووید-۱۹ با استفاده از تکنیک‌های گوناگون داده‌کاوی و ارائه بهترین مدل تشخیصی بود. در ادامه جزئیات روش انجام مطالعه در قالب هفت بخش شامل ۱- جامعه پژوهش، ۲- انتخاب نمونه، ۳- توصیف مجموعه داده، ۴- معیارهای ورود و خروج، ۵- پیش پردازش، ۶- طراحی مدل‌ها و ۷- ارزیابی مدل‌ها ارائه شده است.

جامعه پژوهش شامل حدود ۲۷۸۰ مراجعه‌کننده مشکوک به ابتلاء به بیماری کووید-۱۹ بود که در بازه زمانی بین ۲۷ اسفند ۱۳۹۸ لغایت ۲۰ آذر ۱۳۹۹ جهت انجام بررسی‌های تشخیصی و دریافت خدمات مراقبتی به بخش‌های سرپایی و درمانگاهی بیمارستان طالقانی شهرستان آبادان مراجعه کردند. از مجموع ۲۷۸۰ مراجعه‌کننده، ۱۴۳۵ بیمار تحت آزمایش Rt-PCR قرار گرفتند و آزمایش ۶۸۶ مورد از آن‌ها مثبت اعلام شد. از این تعداد، ۳۰۲ بیمار به دلیل وجود علائم شدیدتر بیماری در بیمارستان بستری شدند. به علاوه از میان مجموع ۷۴۹ بیمار با نتیجه منفی آزمایش RT-PCR، ۱۶۹ بیمار به دلیل شدت علائم شبه آنفلوانزا در بیمارستان بستری شدند. داده‌های بیماران بستری شده با تشخیص

قطعی بیماری کووید-۱۹ و علائم شبه کووید-۱۹ به عنوان ورودی به مطالعه انتخاب و از سیستم پرونده الکترونیک سلامت استخراج شدند. این داده‌ها در قالب کلاس‌های داده‌ای پایه / عمومی، علائم و نشانه‌ها، بیماری‌های زمینه‌ای، نتایج آزمایشگاهی و پیامد درمان (کد صفر زنده و یک فوتی) در قالب ۴۰ متغیر اولیه بالینی ثبت شدند. اطلاعات تعداد ۴۳۵ فرد مبتلا به بیماری کووید-۱۹ و فاقد بیماری که نتیجه بررسی تشخیصی در آن‌ها مثبت یا منفی گزارش شده بود، در پرونده الکترونیک پزشکی بیمار به همراه ۴۰ معیار تشخیصی در آن مرکز ذخیره گردید؛ معیارهای تشخیصی مورد استفاده در پژوهش شامل یافته‌های جمعیت شناختی، داده‌های عمومی و پایه، بالینی و مراقبتی، سوابق پزشکی و شخصی و داده‌های همه‌گیرشناسی بودند. متغیر خروجی پژوهش همان نتیجه تشخیصی بیماری کووید-۱۹ بود که در آن عدد صفر به افرادی نسبت داده می‌شد که نتیجه آزمون تشخیصی بیماری کووید-۱۹ در آن‌ها منفی بود و به عنوان افراد فاقد بیماری طبقه‌بندی می‌شدند و عدد یک متعلق به افرادی بود که نتیجه آزمون تشخیصی در آن‌ها مثبت بود و به عنوان افراد مبتلابه بیماری کرونا در نظر گرفته شدند. استخراج داده از پایگاه داده پرونده الکترونیک سلامت به واسطه مراجعه حضوری پژوهشگران و ثبت در قالب فایل برنامه کاربردی Excel انجام پذیرفت. معیارهای ورود به مطالعه شامل بستری بودن فارغ از نوع بخش (عمومی یا ویژه) در بیمارستان طالقانی شهرستان آبادان در بازه زمانی بین ۲۷ اسفند ۱۳۹۸ لغایت ۲۰ آذر ۱۳۹۹ به دلیل مشکلات ناشی از کووید-۱۹ یا شک ابتلاء به این بیماری بود. به علاوه سن بالاتر از ۱۸ سال و نیز عدم فوت بیماران از دیگر معیارهای ورود به مطالعه بود. متعاقباً معیار خروج شامل موارد ثبتی در خارج از بازه زمانی مدنظر، سن پایین‌تر از ۱۸ سال، فوت بیماران و وضعیت نامشخص بیماران، وجود فیلدهای اطلاعات ناقص، موارد منفی بیماری کووید ۱۹ و موارد مثبتی که برای ادامه درمان به منزل ارجاع داده شدند، بود. از طرفی با توجه به ماهیت گذشته‌نگر بودن مجموعه داده، برخی از رکوردها دارای فیلدهای خالی بودند. به منظور برطرف شدن این مشکل رکوردهایی که تعداد فیلد ناقص آن‌ها بیش از ۷۰ درصد بود، از مطالعه حذف شد. فیلدهای خالی سایر رکوردها از طریق جایگزینی با میانگین پر شد. به علاوه تلاش شد تا فیلدهای دارای خطا، اعداد غیرطبیعی، ناهماهنگ و غیریکپارچه به واسطه بررسی محققین شناسایی و از طریق تماس با پزشک معالج و جمع‌کننده داده برطرف شود.

سپس قبل از انجام فرآیند داده‌کاوی با توجه به تعداد فیلدهای تشخیصی موجود در پژوهش و هم‌چنین به منظور بالا بردن کارایی نتایج حاصل از الگوریتم‌های داده‌کاوی مختلف، برخی از معیارهای تشخیصی در پژوهش که از اهمیت کمتری برخوردار بودند، از پژوهش حذف گردیدند تا فرآیند داده‌کاوی با حداکثر سرعت و با استفاده از مهم‌ترین معیارهای تشخیصی در پژوهش انجام گردد و نتایج حاصل از معیارهای مختلف ارزیابی عملکرد مطلوب‌تری را نشان دهد؛ بنابراین، از روش مجذور کای پیرسون با توجه به وجود متغیر کیفی دوحالتی و چندحالتی در پژوهش، استفاده گردید. اساس این روش آماری در واقع بیانگر مقایسه بین مقدار واقعی به دست آمده و مقدار منتظره می‌باشد که مشخص می‌کند که تفاوت بین مقادیر واقعی و منتظره از نظر آماری معنادار می‌باشد و یا این‌که تفاوت بین آن‌ها ناچیز و نتایج کاملاً تصادفی می‌باشد. برای تعیین معنادار بودن روابط بین هر یک از متغیرهای مستقل پژوهش (معیارهای تشخیصی بیماری کووید-۱۹) با متغیر وابسته (مثبت یا منفی بودن نتیجه تشخیصی)، از ضریب همبستگی کای دو پیرسون در سطح P-Value استفاده گردید و معیارهای تشخیصی که میزان همبستگی آن‌ها با نتیجه تشخیصی در سطح P-Value<۰/۰۵ معنادار بود، به عنوان مهم‌ترین معیارهای تشخیصی در ابتلا به بیماری کرونا در نظر گرفته شدند. فرمول محاسبه ضریب کای دو به همراه نحوه محاسبه فراوانی منتظره به ترتیب در روابط ۱ و ۲ نشان داده شده است.

$$X^2 = \sum_{i=1}^{n} (f_{e(i)} - f_{o(i)})^2 / f_{e(i)} \quad : \text{رابطه۱}$$

$$f_{e(i)} = \frac{n_i * n_j}{n} \quad \text{رابطه۲:}$$

در رابطه ۱، Fo فراوانی مشاهده شده و Fe فراوانی مورد انتظار می‌باشد که از رابطه ۲ به دست می‌آید. همچنین در این پژوهش، روش انتخاب مهم‌ترین معیار تشخیصی براساس رابطه ۱ (ملاک کای‌دو) صورت گرفته است.

پس از تعیین مهم‌ترین معیارهای تشخیصی براساس روش پیرسون و ملاک مجذور کای دو، داده‌کاوی با استفاده از نرم‌افزار داده‌کاوی وکا ورژن ۳/۹ انجام گردید. چهار الگوریتم داده‌کاوی معروف Multi-layer Perceptron (MLP)، درخت تصمیم J-48، Bayesian Net و Logistic regression (LR) که در اکثر پژوهش‌ها از آن‌ها استفاده گردیده و کارایی نسبتاً بالایی را نسبت به سایر الگوریتم‌ها از خود نشان داده‌اند، با در نظر گرفتن ۱۰درصد اعتبارسنجی متقابل جهت ارزیابی

آن‌ها استفاده گردید. در نهایت به منظور ارزیابی کارایی و کیفیت طبقه‌بندی الگوریتم‌های داده‌کاوی از ماتریس آشفتگی (جدول ۱) و معیارهای مختلف ارزیابی حاصل از آن از قبیل Precision، Sensitivity و Matthews Correlation Coefficient استفاده شد.

جدول ۱- ماتریس آشفتگی

	نمونه‌های طبقه‌بندی‌شده توسط مدل	
نمونه‌های واقعی	+	-
+	True Positive (TP)	False Positive (FP)
-	False Negative (FN)	True Negative (TN)

در این پژوهش معیارهای مثبت صحیح و منفی صحیح نمایانگر تعداد موارد بیمار و سالم است که به درستی توسط مدل طبقه‌بندی شده بودند، مثبت کاذب تعداد افراد سالم می‌باشد که به اشتباه توسط مدل بیمار تشخیص داده شده بودند و منفی کاذب تعداد افراد بیماری بوده است که توسط مدل به عنوان سالم در نظر گرفته شده بودند. در نهایت، مناسب‌ترین مدل تشخیصی ابتلا به بیماری کووید-۱۹ براساس مقایسه عملکرد الگوریتم‌ها و انتخاب الگوریتم داده‌کاوی مبتنی بر بالاترین عملکرد صورت گرفت. لازم به ذکر است در این پژوهش ملاحظات اخلاقی شامل حفظ محرمانگی و رازداری داده‌ها رعایت شده است.

یافته‌ها:

پس از اعمال معیارهای ورود و خروج به مطالعه، از مجموع ۴۷۱ رکورد بیمار بستری، در مجموع اطلاعات ۴۳۵ بیمار که دارای تشخیص مثبت و منفی ناشی از انجام آزمایش RT-PCR بودند، وارد مطالعه شدند. پس از انجام پیش پردازش و حذف تعداد ۳۵ رکورد دارای فیلد ناقص (بیش از ۷۰ درصد) در نهایت ۴۰۰ رکورد به عنوان ورودی و تغذیه وارد نرم‌افزارهای داده‌کاوی شد. از این تعداد ۲۵۰ مورد بیمار قطعی کووید-۱۹ و ۱۵۰ مورد بیمار بستری شده دارای علایم مشابه کووید-۱۹ که جواب آزمایش آن‌ها منفی بوده است. پس از وزن‌دهی به متغیرهای پژوهش با استفاده از مجذور کای، مهم‌ترین معیارهای تشخیصی بیماری کووید-۱۹ در سطح P-Value>۰/۰۵ به دست آمدند. براساس یافته‌های حاصل، متغیر وجود ضایعات ریوی در فرد با ضریب کای دو ۱۷۹/۲۱ به عنوان مهم‌ترین معیار تشخیصی در نظر گرفته شد (جدول ۲).

جدول ۲- مهم‌ترین معیارهای تشخیصی کووید-۱۹ براساس مجذور کای پیرسون در سطح P-value <۰/۰۵

نام متغیر	نوع متغیر	ویژگی متغیر	میزان کای دو	P-Value
وجود ضایعه ریوی	دوحالتی	ندارد دارد	۱۷۹/۲۱	۰/۰۰۱>
تب	دوحالتی	ندارد دارد	۱۱۳/۲۶	۰/۰۰۱>
سابقه تماس با افراد مشکوک به کرونا	دوحالتی	ندارد دارد	۱۱۱/۲۶	۰/۰۰۱>
میزان اشباع اکسیژن خون	چندحالتی	٪۹۵ ٪۸۵-٪۹۵ ٪۸۵>	۱۰۲/٤	۰/۰۰۱
آبریزش بینی	دوحالتی	ندارد دارد	۹٦/٤	۰/۰۰۱>
تنگی نفس	دوحالتی	ندارد دارد	۹۰/۱	۰/۰۰۱>
علائم گوارشی	دوحالتی	ندارد دارد	۸۱/۷	۰/۰۰۱
تهوع و استفراغ	دوحالتی	ندارد دارد	۷۵/۵	۰/۰۰۱>

سابقه مسافرت به مناطق پرخطر	دوحالتی	ندارد دارد	۶۳/۳	۰/۰۰۱
سابقه استفاده از داروهای تضعیف‌کننده سیستم ایمنی	دوحالتی	ندارد دارد	۵۸/۲	۰/۰۰۱>
سابقه نارسایی تنفسی	دوحالتی	ندارد دارد	٤٦/٦	۰/۰۰۱>
سابقه عفونت دستگاه تنفسی	دوحالتی	ندارد دارد	٤۰/۲	۰/۰۰۱
سرفه	دوحالتی	ندارد دارد	۳٤/۲	۰/۰۰۱>
نوع محل جغرافیایی ازلحاظ خطر کرونا	چندحالتی	خطر کم خطر متوسط خطر بالا	۳۲/۱	۰/۰۵>
گلودرد	دوحالتی	ندارد دارد	۳۲/۱	۰/۰۰۱
سردرد	دوحالتی	ندارد دارد	۳۰/٤	۰/۰۰۱>
ضعف و ناتوانی	دوحالتی	ندارد دارد	۳۰/۱	۰/۰۰۱>
خطرات شغلی	چندحالتی	بی‌خطر کم‌خطر خطر متوسط پرخطر	۲۹/۸	۰/۰۵>
سطح هوشیاری	چندحالتی	فاقد نسبی کامل	۲٤/٤	۰/۰۰۱
احساس لرزش	دوحالتی	ندارد دارد	۲۳/۱	۰/۰۵>
ضعیف شدن حواس پنج‌گانه	دوحالتی	ندارد دارد	۲۱/۸	۰/۰۵>

نتایج حاصل از طبقه‌بندی الگوریتم‌های داده‌کاوی از افراد مبتلا به بیماری و افراد فاقد بیماری براساس ماتریس آشفتگی در جدول ۳ نشان داده شد. براساس اطلاعات حاصل از جدول ۳ الگوریتم داده‌کاوی LR با طبقه‌بندی ۲۲۰ مورد صحیح از افراد مبتلا به بیماری کووید-۱۹ از میان ۲۵۰ نمونه و ۱۲۰ نمونه صحیح از افراد فاقد بیماری از میان ۱۵۰ فرد فاقد بیماری بهترین طبقه‌بندی را نسبت به سایر الگوریتم‌های داده‌کاوی داشته است.

جدول ۳- نتایج حاصل از طبقه‌بندی افراد مبتلابه بیماری کرونا و فاقد بیماری

نام الگوریتم	میزان TP	میزان FP	میزان FN	میزان TN
LR	۲۲۰	۳۰	۳۰	۱۲۰
J-48	۲۱۲	۳۸	۳٤	۱۱٦
MLP	۲۱٦	۳٤	۳۸	۱۱۲
Bayesian Net	۲۰۹	٤۱	۲۹	۱۲۱

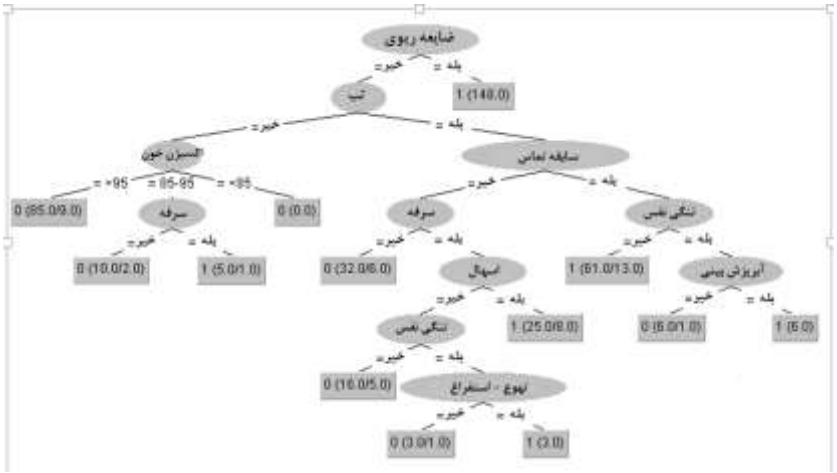
نتایج حاصل از Precision، Sensitivity و Matthews Correlation Coefficient الگوریتم‌های مختلف داده‌کاوی در جدول ۴ ارائه گردیده است. نتایج حاصل از مقایسه عملکرد الگوریتم‌ها براساس معیارهای منتخب ارزیابانه نشان داد که الگوریتم داده‌کاوی J-48 با میزان Sensitivity، Precision و Matthews Correlation Coefficient به

ترتیب ۰/۸۵، ۰/۸۵ و ۰/۶۸ عملکرد بهتری نسبت به سایر الگوریتم‌های داده‌کاوی داشته است.

جدول ۴- مقایسه عملکرد الگوریتم‌های داده‌کاوی برحسب معیارهای ارزیابانه

نام الگوریتم	میزان Sensitivity	میزان Precision	میزان MCC
LR	۰/۸۵	۰/۸۵	۰/٦۸
J-48	۰/۸۲	۰/۸۲۱	۰/٦۱۸
MLP	۰/۸۲۳	۰/۸۲۳	۰/٦۲۳
Bayes Net	۰/۸۲۵	۰/۸۲۹	۰/٦۳٤

در شکل ۱ درخت J-48 هرس شده با ضریب اطمینان ۱۵ درصد جهت تشخیص بیماری کووید-۱۹ ترسیم شده است. براساس درخت ترسیم شده، متغیر وجود ضایعات ریوی در فرد مبتلا به بیماری کووید-۱۹ با داشتن آنتروپی نسبی بالاتر نسبت به سایر متغیرها در گره ریشه (راس درخت) قرار گرفته است، این متغیر توانسته با توجه به اهمیت بالای خود کل درخت تصمیم را براساس طبقه‌بندی افراد از نظر داشتن ضایعات ریوی و فقدان آن به دو زیر درخت تقسیم کند به نحوی که افراد دارای ضایعات ریوی (۱۴۸ نمونه) به عنوان افراد با تشخیص مثبت بیماری کووید (زیر درخت سمت راست) و باقیمانده نمونه‌های پژوهش (فاقد ضایعات ریوی) در صورت وجود یافته‌های بالینی دیگر بر حسب شرایط در دو گروه مبتلا و غیر مبتلا به بیماری کرونا در نظر گرفته شوند (زیر درخت سمت چپ).



شکل ۱: درخت تصمیم هرس شده J-48 تشخیص کووید-۱۹

بحث و نتیجه‌گیری:

انقلاب فناوری‌های بهداشت دیجیتالی در ابتدای دهه ۲۰۲۰ میلادی با همه‌گیری بیماری کووید-۱۹ به واسطه استفاده گسترده از فناوری‌های هوش مصنوعی، رایانش ابری، اینترنت اشیاء و داده‌کاوی داده‌های بزرگ رقم خورد [۱۷،۱۸]. امروزه استفاده از فناوری‌های محاسباتی به‌روز و پیشرفته در قالب سیستم‌های یادگیری ماشین، یادگیری عمیق و داده‌کاوی داده‌های بزرگ، قابلیت‌های تحلیلی عمیق، اثربخش و غیرتهاجمی را به منظور حمایت از تصمیم‌گیری پزشکان و سیاست‌گذاران بهداشتی نسبت به روش‌های سنتی آماری، بررسی‌های بالینی و ارزیابی‌های آزمایشگاهی فراهم کرده است [۱۹].

با گسترش شیوع بیماری کووید-۱۹ و درگیری شدید بسیاری از جوامع به این بیماری، سیستم بهداشت و درمان بسیاری از کشورها ظرفیت پاسخ‌گویی به نیازهای روزافزون بیماران برای تشخیص، درمان و مراقبت از بیماری را نخواهند داشت. در چنین شرایطی طراحی و پیاده‌سازی سیستم‌های تصمیم‌یار بالین مجهز به الگوریتم‌های پیش‌بینی کننده سفارشی و اثربخش به منظور شناسایی موارد مبتلا و افتراق آن از افراد سالم و نیز افراد مبتلا به سایر بیماری‌های تنفسی مشابه حیاتی خواهد بود [۲۰،۲۱].

فناوری‌های تصمیم‌یار بالین با توجه به ماهیت حساس، مبهم و چند بعدی این بیماری، نبود درمان مشخص، طول دوره کمون و بهبودی بالا،

ارزیابی‌های پاراکلینیکی نسبتاً دشوار و تشخیص افتراقی با سایر بیماری‌های تنفسی مشابه نقش مهمی در بهبود کیفیت تشخیص و ارائه مراقبت‌های سفارشی و بیمار محور دارد. به علاوه با افزایش تعداد موارد آلوده، شناسایی سریع و اثربخش بیماران، اولویت‌دهی صحیح منابع بهداشتی، غربالگری و اعمال برنامه‌های قرنطینه‌سازی در کاهش بار بیماری کمک‌کننده است [۲۲]. بنابراین هدف اصلی مطالعه حاضر ارزیابی مقایسه‌ای عملکرد برخی از مدل‌های داده‌کاوی مانند MLP، درخت تصمیم J-48، Bayes Net و LR در تشخیص بیماری کووید-۱۹ و تمایز بین موارد آلوده از افراد سالم برای بهبود کیفیت و اثربخشی تصمیمات بالینی بود.

در زمینه‌ی کاربرد داده‌کاوی (یادگیری ماشین و یادگیری عمیق) کارهای تحقیقاتی متعددی انجام شده که در برخی از آن‌ها مقایسه الگوریتم‌های مختلف به منظور پیشنهاد کارآمدترین آن‌ها به منظور ارائه یک مدل تشخیصی بهینه در دستور کار بوده است. نتایج مطالعه Alakus و همکارانش بهم نظور طراحی مدل تشخیصی بیماری کووید-۱۹ مبتنی بر یادگیری عمیق نشان داد میزان صحت، درجه F1 (Score)، دقت، فراخوانی و AUC سیستم به منظور تشخیص موارد بیماری به ترتیب ۸۸/۶۶٪، ۹۱/۸۹٪، ۸۶/۷۵٪، ۹۹/۴۲٪ و ۶۲/۵۰٪ بود [۲۳].

در مطالعه Moftakhar قابلیت شبکه عصبی مصنوعی (ANN) و مدل خود همبسته میانگین متحرک (ARIMA) در تشخیص بیماری کووید-۱۹ به منظور کمک به تصمیم‌گیری پزشکان و سیاست‌گذاران بهداشتی مقایسه و نتایج حاکی از صحت و دقت بالاتر ARIMA در تشخیص موارد بود. در پژوهش Narin و همکاران (۲۰۲۰) از سه نوع شبکه عصبی مصنوعی ResNet50، inception V3 و Inception-ResNetV2 برای تشخیص بیماری کرونا انجام گرفت، نتایج حاصل از پژوهش نشان داد که شبکه عصبی مصنوعی ResNet50 با میزان صحت ۹۸ درصد کارایی بالاتری نسبت به سایر الگوریتم‌های داده‌کاوی داشته است [۲۴]. در پژوهش Elaziz و همکاران از روش یادگیری ماشین برای تحلیل عکس قفسه سینه در افراد استفاده گردید و الگوریتم قادر بود تا افراد مبتلا به کووید-۱۹ و افراد سالم را براساس تحلیل عکس قفسه سینه تشخیص دهد و نتایج حاصل از ارزیابی الگوریتم میزان صحت ۹۶ درصد و ۹۸ درصد را در دو مجموعه داده از تصاویر را نشان داد [۲۵]. در پژوهشی که توسط Brunce و همکاران انجام گرفت از روش یادگیری عمیق قابل تفسیر برای کشف سریع بیماری‌های ریوی و کرونا استفاده شد، با انجام تحلیل آماری بر روی ۶۵۲۳ عکس قفسه سینه متعلق به مؤسسات مختلف، نتایج حاصل از پژوهش میزان صحت ۹۷ درصد را در طی ۲/۵ ثانیه در تشخیص بیماری نشان داد [۲۶].

در مطالعه Rodriguez از طریق مقایسه دو روش ریاضی (Gompertz و Logistic) و یک روش محاسباتی (ANN) نتایج حاکی از عملکرد بهتر مدل‌های محاسباتی بود [۲۷]. در پژوهش حاضر براساس مقایسه عملکرد الگوریتم‌های داده‌کاوی براساس معیارهای مختلف ارزیابی از قبیل Precision، Sensitivity و MCC حاصل از الگوریتم‌های مختلف، الگوریتم داده‌کاوی J-48 به ترتیب با میزان‌های ۸۵ درصد، ۸۵ درصد و درصد عملکرد بالاتری تشخیصی نسبت به سایر الگوریتم‌ها داشته است. در این پژوهش برخلاف مطالعات گذشته که برای تشخیص بیماری از مدل‌های شبکه عصبی بر روی داده‌های تصویری استفاده شده است، از داده‌های بالینی در قالب درخت تصمیم به منظور شناسایی موارد مبتلاء از سالم استفاده گردید. مشخصاً معیارهای ارزیابانه مطالعات گذشته نشان‌دهنده دقت و صحت بالاتری تشخیصی مدل‌های طراحی شده است، با این وجود استفاده از درخت تصمیم و ساختارهای سلسله مراتبی در تشخیص بیماری نیز می‌تواند در کمک به متخصصین مراقبتی راهگشا باشد.

در ساختار درخت طراحی شده دو مورد از قوانین استخراج شده در تشخیص بیماری کووید-۱۹ براساس قانون اگر-آنگاه از طریق پیمایش مسیر درخت از گره ریشه تا برگ آن اشاره شده است. در قانون پیشنهادی اول، اگر فرد فاقد ضایعات ریوی باشد، دارای تب باشد، سابقه تماس با افراد مشکوک به بیماری کووید-۱۹ داشته، دچار تنگی نفس و آبریزش بینی باشد و درخت تصمیم J-48 آن را به عنوان افراد مبتلا به بیماری طبقه‌بندی می‌کند. در این درخت تصمیم هشت نمونه از پژوهش در گره برگ آبریزش بینی واقع در انتهای سمت راست در زیر درخت چپ قرار دارد. درخت تصمیم J-48 توانسته است ۶ نمونه را با این الگو به عنوان تشخیص مثبت بیماری کووید-۱۹ طبقه‌بندی کند. در قانون دوم، اگر فرد فاقد ضایعات ریوی باشد، دارای تب، سابقه تماس با افراد مشکوک به بیماری کووید-۱۹ داشته، دچار تنگی نفس و فاقد آبریزش بینی باشد: درخت تصمیم J-48 آن را به عنوان افراد با تشخیص منفی به بیماری کووید-۱۹ طبقه‌بندی می‌کند. در این ساختار، شش نمونه از پژوهش را با الگوی فوق دسته‌بندی کرد که البته یک نمونه آن با وجود الگوی موجود در قانون ۲ به عنوان تشخیص مثبت در نظر گرفته و به اشتباه طبقه‌بندی شده بود.

مهم‌ترین محدودیت‌های پژوهش حاضر، محدود بودن تعداد نمونه‌ها در پایگاه داده انتخابی، تک مرکزی بودن پایگاه داده و نیز وجود برخی رکوردهای اطلاعاتی غیر یکپارچه، ناقص، دارای خطا و موارد غیرطبیعی در فیلدهای اطلاعاتی بود. از طریق اعمال الگوریتم‌های یادگیری ماشین در پایگاه‌های داده بزرگ برگرفته از چند مرکز و نیز توجه به کمیت و کیفیت مستندسازی می‌توان قابلیت‌های الگوریتم‌ها را در تشخیص صحیح موارد بالا برد. به علاوه در پژوهش حاضر از داده‌های آزمایشگاهی به عنوان ورودی سیستم‌ها استفاده نشد که توجه به این معیار عملکرد تشخیصی و پیش‌بینی سیستم‌ها را بهبود خواهد بخشید.

در نهایت می‌توان نتیجه گرفت بیماری کووید-۱۹ از عفونت‌های ویروسی بسیار مسری می‌باشد که به سرعت از فردی به فرد دیگر منتقل می‌شود. تشخیص این بیماری در مراحل اولیه بیماری بسیار مهم تلقی شده و می‌تواند از میزان شیوع این بیماری در سطح جامعه تا حد بسیاری جلوگیری نماید. استفاده از مدل‌های تشخیصی می‌تواند به پزشکان در تشخیص زودهنگام بیماری کووید-۱۹ و به تبع آن قرنطینه افراد با تشخیص مثبت بیماری کووید-۱۹ و پیشگیری آن مفید باشد. در پژوهش حاضر از الگوریتم‌های داده‌کاوی جهت ایجاد مدل تشخیصی مؤثر بیماری کووید-۱۹ استفاده گردید. نتایج حاصل از مقایسه عملکرد چهار الگوریتم معروف داده‌کاوی با استفاده از معیارهای مختلف ارزیابی نشان داد که الگوریتم درخت تصمیم J-18 می‌تواند به عنوان مدل تشخیصی مناسب و با عملکرد بالا در کمک به پزشکان در تشخیص به موقع و دقیق این بیماری مؤثر واقع گردد. پیشنهاد می‌گردد که در پژوهش‌های آتی جهت ایجاد مدل تشخیصی مؤثر بیماری کووید-۱۹ از تعداد نمونه‌های بیشتری استفاده گردد تا کارایی و دقت عملکردهای الگوریتم‌ها بهبود یابد، همچنین جامعه پژوهش به صورت چند مرکز درآید تا نمونه‌های متنوع‌تری برحسب مناطق مختلف در ایجاد مدل تشخیص در نظر گرفته شود، پیشنهاد می‌گردد از طریق مشورت با چند متخصص بالینی مهم‌ترین قوانین حاصل از درخت تصمیم استخراج شود تا بتوان از آن در طراحی سیستم‌های تصمیم‌یار مبتنی بر پایگاه قوانین استفاده نمود و نهایتاً از طریق جستجوی مقالات و کتب علمی جدید عوامل تشخیصی ناشناخته و مهمی که در ایجاد مدل‌های تشخیصی جهت تشخیص زودهنگام بیماری توسط پزشکان موثرند را در نظر گرفت و در پژوهش استفاده نمود.

تشکر و قدردانی:

از مسئولین محترم معاونت تحقیقات دانشگاه علوم پزشکی آبادان و همچنین بیمارستان طالقانی آبادان که در انجام پژوهش تیم پژوهش را یاری‌رسان بودند، کمال تشکر و قدردانی را داریم.

تأییدیه اخلاقی:

این مقاله برگرفته از یک طرح پژوهشی مصوب جلسه شورای پژوهشی مورخ ۹۹/۰۸/۲۱ و جلسه کمیته اخلاق دانشگاه علوم پزشکی آبادان با کد اخلاق IR.ABADANUMS.REC.1399.222 است.

تعارض منافع:

کار پژوهشی حاضر فاقد تضاد منافع است.

سهم نویسندگان:

رئوف نوپور (نویسنده اول) داده کاوی ۳۵ درصد؛ مصطفی شنبه‌زاده (نویسنده دوم) گزارش‌دهی نتایج ۳۰ درصد؛ هادی کاظمی‌آرپناهی (نویسنده سوم و مسئول) طراحی مطالعه ۳۵ درصد.

حمایت مالی:

این مقاله از طرف هیچ نهاد یا موسسه‌ای حمایت مالی نشده است و منابع مالی از طرف نویسندگان تأمین شده است.

Reference

1. Tang D, Comish P, Kang R. The hallmarks of COVID-19 disease. PLoS Pathog. 2020; 16(5):e1008536. DOI: 10.1371/journal.ppat.1008536
2. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. 2020; 24:91-8. DOI: 10.1016/j.jare.2020.03.005

3. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, et al. Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology. 2020; 296(2):e15-25. DOI: 10.1148/radiol.2020200490
4. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: A mathematical modelling study. Lancet Infect Dis. 2020; 20(5):553-8. DOI: 10.1016/S1473-3099(20)30144-4
5. Wang Z, Yang B, Li Q, Wen L, Zhang R. Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China. Clin Infect Dis. 2020; 71(15):769-77. DOI: 10.1093/cid/ciaa272
6. Xu G, Yang Y, Du Y, Peng F, Hu P, Wang R, et al. Clinical pathway for early diagnosis of COVID-19: Updates from experience to evidence-based practice. Clinic Rev Allerg Immunol. 2020; 59:89-100. DOI: 10.1007/s12016-020-08792-8
7. Stadnytskyi V, Bax CE, Bax A, Anfinrud P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. PNAS. 2020; 117(22):11875-7. DOI: 10.1073/pnas.2006874117
8. Celesti A, Ruggeri A, Fazio M, Galletta A, Villari M, Romano A. Blockchain-based healthcare workflow for tele-medical laboratory in federated hospital IoT clouds. Sensors. 2020; 20(9):2590. DOI: 10.3390/s20092590
9. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SR. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill. 2020; 6(2):e18828. DOI: 10.2196/18828
10. James P, Das R, Jalosinska A, Smith L. Smart cities and a data-driven response to COVID-19. Dialogues Hum Geogr. 2020; 10(2):255-9. DOI:10.1177/2043820620934211
11. Peck KR. Early diagnosis and rapid isolation: Response to COVID-19 outbreak in Korea. Clin Microbiol Infect. 2020; 26:805-7. DOI: 10.1016/j.cmi.2020.04.025
12. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. A new COVID-19 patients detection strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowl Based Syst. 2020; 205:106270. DOI: 10.1016/j.knosys.2020.106270
13. Chamola V, Hassija V, Gupta V, Guizani M. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access. 2020; 8:90225-65. DOI: 10.1109/ACCESS.2020.2992341
14. Han J, Kamber M, Pei J. Data mining concepts and techniques. 3 rd ed. USA: Elsevier; 2011. P 1-28.
15. Iavindrasana J, Cohen G, Depeursinge A, Müller H, Meyer R, Geissbuhler A. Clinical data mining: A review. Yearb Med Inform. 2009; 18(01):121-33. PMID: 19855885
16. Albahri AS, Hamid RA, Alwan JK, Al-Qays ZT, Zaidan AA, Zaidan BB, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): A systematic review. J Med Syst. 2020; 44(7):1-11. DOI: 10.1007/s10916-020-01582-x
17. Talebpour M, Hadadi A, Oraii A, Ashraf H. Rationale and design of a registry in a referral and educational medical center in Tehran, Iran: Sina Hospital COVID-19 registry (SHCo-19R). Front Emerg Med. 2020; 4(2s):e53. DOI: 10.22114/ajem.v0i0.361
18. Javaid M, Haleem A, Vaishya R, Bahl S, Suman R, Vaish A. Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. Diabetes Metab Syndr. 2020; 14(4):419-22. DOI: 10.1016/j.dsx.2020.04.032
19. Ye J. The role of health technology and informatics in a global public health emergency: Practices and implications from the COVID-19 pandemic. JMIR Med Inform. 2020; 8(7):e19866. DOI: 10.2196/19866.
20. Chaudhari SN, Mene SP, Bora RM, Somavanshi KN. Role of internet of things (IOT) in pandemic COVID-19 condition. Int J Eng Res Appl. 2020; 10(6):57-61. DOI: 10.9790/9622-1006035761
21. Bayram M, Springer S, Garvey CK, Özdemir V. COVID-19 digital health innovation policy: A portal to alternative futures in the making. OMICS. 2020; 24(8):460-9. DOI: 10.1089/omi.2020.0089
22. Hassen HB, Ayari N, Hamdi B. A home hospitalization system based on the Internet of Things, Fog computing and cloud computing. Inform Med Unlocked. 2020; 20:100368. DOI: 10.1016/j.imu.2020.100368

23. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. Chaos Solitons Fractals. 2020; 140:110120. DOI: 10.1016/j.chaos.2020.110120

24. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. Pattern Anal Applic. 2021; 10849. DOI: 10.1007/s10044-021-00984-y

25. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. PLOS ONE. 2020; 15(6):e0235187. DOI: 10.1371/journal.pone.0235187

26. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Comput Methods Programs Biomed. 2020; 196:105608. DOI: 10.1016/j.cmpb.2020.105608

27. Torrealba-Rodriguez O, Conde-Gutiérrez RA, Hernández-Javier AL. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. Chaos Solitons Fractals. 2020; 138:109946. DOI: 10.1016/j.chaos.2020.109946

Proposing an effective technological solution for the early diagnosis of COVID-19: a data-driven machine learning study

Raoof Nopour[1] Mostafa Shanbehzadeh[2] Hadi Kazemi Arpanahi[3]*

1. MSc, Health Information Technology, Faculty of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran. ORCID: 0000-0003-3370-2375
2. Department of Health Information Technology, Faculty of Allied Medical Sciences, Ilam University of Medical Sciences, Ilam, Iran.
3. Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran.

(Received 19 Mar, 2021 Accepted 10 Jun, 2021)

Original Article

Abstract

Aim: Accurate and timely diagnosis of COVID-19 using artificial intelligence and machine learning technologies will play an important role in improving the disease indicators, optimal utilization of limited hospital resources and reducing the burden on pandemic healthcare providers. Therefore, this study aimed to evaluate the efficiency of selected data mining algorithms based on their performance for COVID-19 diagnosis.

Methods: The present study was a retrospective applied-descriptive study that was conducted in 2020. In this study, the data of patients admitted with a definitive diagnosis of Covid-19 from March 17, 2020 to December 10, 2020 were extracted from the Electronic Medical Record (EMR) database in Ayatollah Taleghani Hospital in Abadan. After applying the inclusion and exclusion criteria to identify the samples, 400 records were entered into the data mining software. The data were compared using chi-square criterion to determine the variables of teach algorithms and their performance based on different evaluation criteria in the turbulence matrix.

Results: Comparing the performance from data mining algorithms based on different evaluation criteria in the turbulence matrix revealed that the J-48 algorithm with the sensitivity, precision, and Matthews Correlation Coefficient (MCC) of 0.85, 0.85 and 0.68 respectively had better performance than the other data mining algorithms for the disease diagnosis. The 3 variables of lung lesion existence, fever, and history of contact with suspected COVID-19 patients, by considering Gini Index to determine the point of division, with Gini index of 0.217, 0.205 and 0.188 respectively were considered as the most important diagnostic indicators of COVID-19.

Conclusion: Using selected data mining methods, particularly J-48 algorithm will greatly aid the timely and effective diagnosis of COVID-19 in the form of clinical decision support systems.

Keywords: COVID-19, Machine Learning, Artificial Intelligence, Data Mining.

How to cite this article: Nopour R, Shanbehzadeh M, Kazemi Arpanahi H. Proposing an effective technological solution for early diagnosis of COVID-19: a data driven machine learning study. Journal of Modern Medical Information Sciences. 2021; 7(1):68-78.

Correspondence:
Hadi Kazemi-Arpanahi
Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran.
Tel: + 989138200027 Email: h.kazemi@abadanums.ac.ir ORCID: 0000-0002-8882-5764