



## ارائه مدل جدید برای تشخیص سریع بیماری‌های حاد تنفسی با استفاده از الگوریتم‌های

### یادگیری ماشین

مهران نظامی<sup>(۱)</sup> عوض نقی پور<sup>(۲)\*</sup> بهنام صفیری ایرانق<sup>(۳)</sup>

(۱) گروه مهندسی کامپیوتر، مؤسسه آموزش عالی نبی اکرم (ص)، تبریز، ایران

(۲) گروه مهندسی کامپیوتر، مؤسسه آموزش عالی نبی اکرم (ص)، تبریز، ایران \*

(۳) گروه مهندسی کامپیوتر، مؤسسه آموزش عالی نبی اکرم (ص)، تبریز، ایران

(تاریخ دریافت: ۱۴۰۲/۰۹/۰۷ تاریخ پذیرش: ۱۴۰۳/۰۲/۱۷)

#### چکیده

کرونا ویروس، ویروس سارس و آنفلوانزای خوکی یک بیماری ناشی از سندروم حاد تنفسی است. این ویروس‌ها به سبب سرایت فوری در بین انسان‌ها به ابزارهای پیشرفته برای شناسایی عوامل خطرناک مرگ‌ومیر با دقت بالا نیاز دارند. روش‌های یادگیری ماشین مستقیماً به این موضوع می‌پردازند و ابزارهای ضروری برای شناخت و هدایت مداخلات بهداشت عمومی هستند. در این مقاله از یادگیری ماشین برای بررسی اهمیت جمعیت‌شناختی و بالینی استفاده شده است. ویژگی‌های مورد بررسی شامل سن، جنسیت، تب، کشورها و جزئیات بالینی مانند سرفه، تنگی نفس و... می‌باشند. چندین الگوریتم یادگیری ماشین روی داده‌های جمع‌آوری‌شده، پیاده‌سازی و اعمال گردیده که الگوریتم K- نزدیک‌ترین همسایه با بالاترین دقت (بیش از ۹۷٪) برای پیش‌بینی و انتخاب ویژگی‌هایی که به درستی وضعیت ویروس‌ها را نشان می‌دهد، عمل می‌کند.

الگوریتم‌های یادگیری ماشین، بیماری‌های همه‌گیر، پیش‌بینی

واژه‌های کلیدی: الگوریتم‌های یادگیری ماشین، بیماری‌های همه‌گیر، پیش‌بینی

\*نویسنده مسئول:

عوض نقی پور

نشانی: گروه مهندسی کامپیوتر، مؤسسه آموزش عالی نبی اکرم (ص)، تبریز، ایران

پست الکترونیکی: [naghipour@ucna.ac.ir](mailto:naghipour@ucna.ac.ir)



شیوع اخیر بیماری همه‌گیر سراسری، یک نگرانی جدی جهانی است. در غیاب درمان‌های مؤثر، راهبرد اصلی مهار، کاهش سرایت با جداسازی افراد آلوده است. با این حال، جداسازی افراد مبتلا به این بیماری‌ها بسیار نامطلوب است. برای کمک به تصمیم‌گیری سریع در مورد نیازهای درمان و قرنطینه، تعیین اینکه کدام ویژگی‌های ارائه‌شده منجر به تشخیص سریع بیماری‌های همه‌گیر باشد، مفید است. این را می‌توان با تجزیه و تحلیل ویژگی‌های بیمار، بیماری‌های همراه و علائم انجام داد. مدلی توسعه داده‌شده که از الگوریتم‌های یادگیری ماشینی نظارت‌شده برای شناسایی ویژگی‌های ارائه که تشخیص بیماری‌های همه‌گیر را با دقت بالا پیش‌بینی می‌کند، استفاده می‌کند.

مدل‌های یادگیری ماشین در بسیاری از زمینه‌ها از جمله، مراقبت‌های بهداشتی انقلابی ایجاد کرده‌اند. با کمک این مدل‌ها می‌توان روش‌های درمانی و سایر خدمات بهداشتی و پزشکی را برای افراد و بیماران توسعه داد. علم داده در پزشکی شامل مراحل جمع‌آوری، پردازش و تجزیه و تحلیل داده‌ها است. هدف از توسعه این علم پیوند دادن مراقبت‌های پزشکی و تحقیقات است. انجام این کار باعث می‌شود که بیماری‌ها را بهتر درک کرد و روش‌های درمانی را برای هر بیمار شخصی‌سازی کرد. این علم هم‌چنین امکان پیش‌بینی بهتر تأثیرات درمانی را می‌دهد. همانطور که اشاره شد این بیماری‌ها ماهیت تهاجمی دارند یعنی بیماری‌ها همه‌گیر بوده و مهم‌ترین مشکل موجود در این بیماری‌ها طبقه‌بندی افراد در معرض خطر این بیماری‌ها برای شناسایی اولیه و ویروس است.

طبق آخرین توصیه‌های سازمان بهداشت جهانی، از آزمون واکنش زنجیره‌ای پلیمرز ترانس کریپتاز معکوس<sup>۱</sup> (RT-PCR) برای تشخیص قطعی آلودگی فرد به ویروس‌ها استفاده می‌شود [۱]. این آزمون از نمونه‌گیری بزاق دهان و بینی به دست می‌آید. با این حال، کیت‌های مربوط به تشخیص ویروس‌ها مبتنی بر این آزمون محدوداند [۲]. این کیت‌ها زمان زیادی برای تشخیص نیاز دارند و نسبت به تشخیص ویروس حساسیت کمی دارند. همچنین این کیت‌ها با توجه به ماهیت تهاجمی که دارند، پزشکان و پرستاران را در معرض آلودگی به ویروس قرار می‌دهند [۳]؛ نتیجه آزمایش واکنش زنجیره‌ای پلیمرز از ترانس کریپتاز معکوس مورد استفاده برای تشخیص ویروس‌ها ممکن است منفی باشد، به خصوص در مراحل اولیه بیماری یا زمانی که بار ویروسی کم است [۴]. درنهایت، گفتنی است تشخیص مبتنی بر این آزمون مقرون به صرفه نیست. بنابراین استفاده از روش‌های نوآورانه و غیرتهاجمی برای شناسایی ویروس در مراحل اولیه بسیار مفید و تعیین‌کننده خواهد بود. مطالعات متعددی برای بررسی استفاده از یادگیری ماشین از نظر توانایی‌های تشخیصی انجام شده است. مشخص شد که در مقایسه با مجرب‌ترین پزشک که می‌تواند با دقت ۷۹/۹۷ درصد به درستی تشخیص دهد،

<sup>۱</sup> Reverse Transcription Polymerase Chain Reaction

الگوریتم‌های یادگیری ماشین می‌توانند با ۹۱/۱ درصد شناسایی کنند [۵]. در نهایت با استفاده از داده‌های تولیدشده در مراکز بهداشتی، می‌توان با تکنیک‌های یادگیری ماشین به شناسایی علائم تأثیرگذار این بیماری‌ها پی برد. با شناسایی علائم‌های تأثیرگذار این بیماری‌ها می‌توان به پیش‌گیری و درمان سریع بیماران پرداخت.

## ۱-۲- کارهای مرتبط

اخیراً شیوع سریع ویروس کرونا جدید منجر به بیماری تنفسی کروناویروس می‌شود. اولین کروناویروس‌های انسانی تصور می‌شود که از عفونت‌های مشترک بین انسان و دام ناشی می‌شوند، زیرا در جمعیت‌های مختلف خفاش هستند. عفونت‌های کروناویروس شناخته‌شده در ابتدا به‌عنوان ایجاد شرایط تنفسی بی‌ضرر انسانی در نظر گرفته می‌شد که تهدیدکننده حیات نبود. بروز اختلالات تنفسی جدی و کشنده منتسب به اعضای زیرخانواده بتا-کروناویروس در بیست سال گذشته با سندرم شدید تنفسی حاد (SARS)<sup>۱</sup> و سندرم تنفسی خاورمیانه (MERS)<sup>۲</sup> رخ داده است. عفونت‌های کروناویروس برای اولین بار در فوشان، چین در سال ۲۰۰۲ و ویروس مرس در سال ۲۰۱۲ در عربستان سعودی به وجود آمدند، که هر دو باعث ایجاد هشدار بین‌المللی و تلاش‌های مهار به دلیل گسترش سریع و نرخ بالای مرگ‌ومیر شدند [۶]. SARS و MERS به ترتیب با نرخ مرگ‌ومیر ۹/۶٪ و ۳۶٪ در میان بیماران تشخیص داده شده مرتبط بودند. این عفونت‌های کروناویروس را به‌عنوان یک تهدید مهم برای سلامت انسان با پتانسیل ایجاد عفونت‌های شدید و کشنده دستگاه تنفسی در افراد شناسایی کردند، به‌ویژه اگر عفونت فرد به‌فرد به راحتی رخ دهد [۷].

گسترش بیماری در فوریه ۲۰۲۰، نرخ مرگ‌ومیر مورد اشاره برای کروناویروس در ووهان چین، ۱/۴٪ بود. به‌عنوان مثال، آمار مرگ‌ومیر بر اساس گروه‌سنی، ایتالیا در طول مارس ۲۰۲۰، نرخ مرگ‌ومیر ۷/۲٪ را نشان داد. این ممکن است تا حدی منعکس‌کننده تفاوت‌های جمعیتی بین کشورها باشد، به‌طوری که ۲۳٪ از جمعیت ایتالیا بالای ۶۵ سال هستند. با این حال، حتی زمانی که طبق سن طبقه‌بندی شود، نرخ عفونت در ایتالیایی‌های بالای ۷۰ سال در مقایسه با چین بالاتر است [۸]. در حوزه یادگیری ماشین برای پیش‌بینی و شناسایی مرگ‌ومیر ناشی از بیماری عفونی تاکنون پژوهش‌های زیادی انجام شده است که در برخی از آن‌ها مقایسه الگوریتم‌های یادگیری به منظور شناسایی کارآمدترین الگوریتم برای ارائه یک مدل شناسایی بهینه در دستور کار بوده است. نتایج مطالعه البحری و همکاران در سال ۲۰۲۰ نشان داد که استفاده از

<sup>۱</sup> Severe Acute Respiratory Syndrome

<sup>۲</sup> Middle East Respiratory Syndrome

الگوریتم‌های یادگیری ماشین به منظور ارائه مدل‌های شناسایی و پیش‌بینی‌کننده با هدف تشخیص به‌موقع، اثربخش و اقتصادی بیماری کروناویروس در کاهش شیوع بیماری و عوارض مخرب آن بسیار موثر است [۹]. یاد او و همکاران [۱۰] پس از بررسی عملکرد چند الگوریتم یادگیری ماشین بر روی ۳۸۴۱ داده بیمار بستری مبتلاء به کروناویروس به منظور پیش‌بینی خطر احتمال مرگ‌ومیرها، الگوریتم افزایش‌گرایان شدید بهترین عملکرد را به دست آورد. در مطالعه یان [۱۱] بر روی داده‌های ۴۸۵ نفر از بیماران بستری، الگوریتم افزایش‌گرایان شدید با دقت پیش‌بینی ۹۰ درصد به عنوان کارآمدترین الگوریتم در تشخیص بدحال شدن بیماران بستری مبتلاء به کروناویروس معرفی شده است. پژوهش گائو و همکاران [۱۲]، بر روی داده‌های ۲۵۲۰ بیمار قطعی مبتلاء به کروناویروس الگوریتم شبکه‌عصبی با صحت پیش‌بینی بالای ۹۷ کارآمدتر از سایر الگوریتم‌ها بود.

ارزیابی‌های پژوهشگران [۱۳] بر روی داده‌های مربوط به ۴۰۲۹ نفر از بیماران کروناویروس نشان داد که الگوریتم پرسپترون چندلایه به نسبت الگوریتم رگرسیون لاسو در پیش‌بینی بدحال شدن بیماران و مرگ آن‌ها از قابلیت بالاتری برخوردار است. در مطالعه دوتا و همکاران [۱۴] از طریق الگوریتم‌های مختلف یادگیری ماشین بر روی پایگاه داده مشتمل بر داده‌های ۱۲۷۲ فرد مبتلاء به کروناویروس و بستری‌شده در بیمارستان، الگوریتم ماشین بردار پشتیبان صحت ۹۳ درصد به‌عنوان کارآمدترین الگوریتم برای پیش‌بینی بستری شدن در بخش مراقبت‌های ویژه و بدحال شدن بیمار معرفی شد. مشابهاً در بررسی پژوهش [۱۳] بر روی ۱۰۸۷ داده بیمار قطعی کروناویروس، الگوریتم ماشین بردار پشتیبان با دقت ۱۰۰ درصد، بهترین عملکرد را به‌نسبت دو الگوریتم دیگر در پیش‌بینی بدحال شدن و مرگ بیماران کسب کرد.

همچنین برای بررسی اهمیت جمعیت‌شناسی و مطالعات بالینی در مرگ‌ومیر کروناویروس، مطالعات بر روی بیماران بستری در بیمارستان لوندرینا، پارانا، برزیل انجام شده است که با توجه به علائم موجود در بیماران از جمله تنگی‌نفس، تب، خستگی‌بدن و غیره از ۴ مدل روش یادگیری ماشین رگرسیون لجستیک، ماشین بردار پشتیبانی، جنگل تصادفی و افزایش‌گرایان شدید جهت پیش‌بینی ویروس کرونا استفاده شده است [۱۵]. در این پژوهش نویسندگان به این نتیجه رسیده‌اند که استفاده از روش افزایش‌گرایان شدید به عملکرد عالی دست پیدا می‌کند و شبکه مربوط به افراد سالخورده و پیر بسیار متراکم‌تر از افراد جوان می‌باشد. علاوه بر این وقوع هم‌زمان بیماری‌های قلبی و دیابت ممکن است مهم‌ترین عامل برای پیش‌بینی مرگ‌ومیر کروناویروس باشد.

در مطالعات بالینی [۱۶] با استفاده از مدل‌های یادگیری ماشین از جمله افزایش‌گرایان شدید و ماشین بردار پشتیبان و غیره از گروه‌های سنی برای شناخت علائم‌های موجود کروناویروس استفاده شده است. در این مطالعه گروه‌های سنی

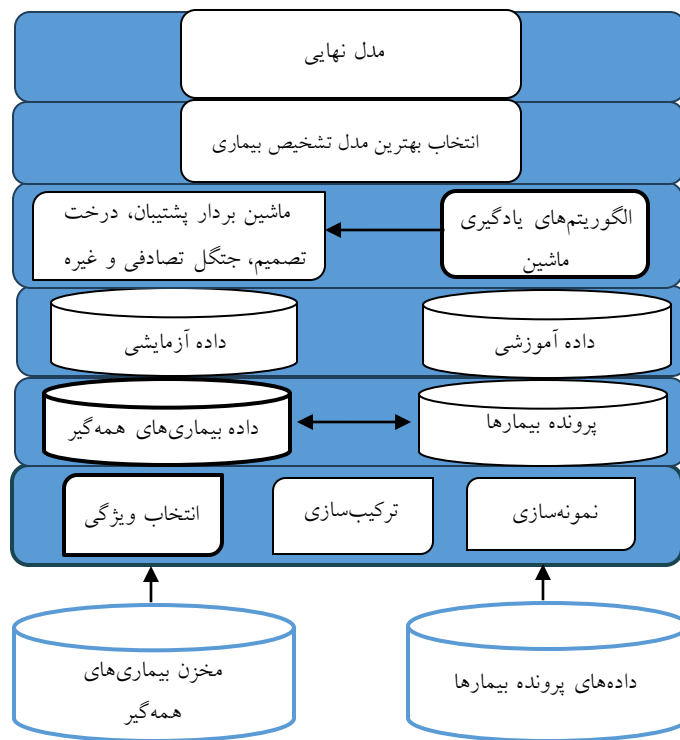
۰ تا ۲۰، ۲۱ تا ۶۰، ۶۱ تا ۹۶ و در نهایت ۰ تا ۹۶ سال مورد بررسی قرار گرفته که با توجه به مدل‌های یادگیری و میزان درست پیش‌بینی علائم برای هر یک از گروه‌سنی متفاوت می‌باشد. در [۲۲] مؤلفان به تشخیص بیماری‌های تنفسی حاد در جمعیت کودکان با استفاده از ویژگی‌های صدای سرفه و یادگیری ماشینی پرداخته‌اند. در پژوهش [۲۳] از علائم‌های بالینی که بیماران بستری مبتلا به کروناویروس داشت از الگوریتم‌های یادگیری برای طبقه‌بندی بیماران در معرض خطر مرگ‌ومیر استفاده شد. پس از شناسایی و اولویت‌بندی مهم‌ترین علائم‌های بالینی و استفاده از آن‌ها به‌عنوان ورودی به مدل‌های یادگیری ماشین، به مقایسه کارایی و عملکرد روش‌های مختلف الگوریتم‌های یادگیری ماشین در پیش‌بینی مرگ‌ومیر بیماران مبتلا به این ویروس همه‌گیر پرداخته شد. الگوریتم‌های مورد استفاده در این پژوهش شامل الگوریتم‌های رگرسیون لجستیک، ماشین‌بردار پشتیبانی، K - نزدیک‌ترین همسایه و جنگل تصادفی بود که از طریق تعداد ۱۷ علائم منتخب بالینی و آزمایشگاهی مربوط به ۱۲۲۴ داده فرد بیمار بستری‌شده مورد آموزش قرار گرفتند. معیار ارزیابی یا اعتبارسنجی الگوریتم‌های پیاده‌سازی‌شده بر اساس ماتریس به‌هم‌ریختگی بود. در نهایت پس از پیاده‌سازی مدل‌های مختلف، الگوریتم K - نزدیک‌ترین همسایه با صحت طبقه‌بندی ۹۳/۷۴ درصد عملکرد کارآمدتری را نسبت به سایر الگوریتم‌ها به‌دست آورد.

در ادامه، در بخش دوم به تشریح الگوریتم‌های به کار رفته در پژوهش پرداخته شده و مزایا و معایب هر کدام از این الگوریتم‌ها توضیح داده خواهد شد. در بخش سوم نیز معیارهای ارزیابی برای هر یک از الگوریتم‌ها مشخص شده است. در فصل چهارم به ارزیابی مدل‌ها و مقایسه نتایج مدل‌ها پرداخته شده است. در بخش پنجم نیز نتیجه‌گیری و یک جمع‌بندی کلی آورده شده است.

## ۲- مواد و روش‌ها

الگوریتم‌های یادگیری ماشین به سیستم‌ها کمک می‌کند تا به صورت خودکار یادگیری و پیشرفت داشته باشند. یک ماشین به کمک مدل‌های یادگیری ماشین می‌تواند از تجربیات و مشاهداتی که بر اساس یک مجموعه داده تجزیه و تحلیل می‌کند، آموزش ببیند. یادگیری ماشین، کاربردی از هوش مصنوعی است که برای سیستم‌ها، امکان یادگیری و پیشرفت خودکار از تجربه را بدون برنامه‌ریزی صریح فراهم می‌کند. یادگیری ماشین بر توسعه برنامه‌های رایانه‌ای است

که می‌توانند به داده‌ها دسترسی پیدا کرده و از آن برای یادگیری خود استفاده کنند [۱۷]. در شکل ۱ می‌توان مراحل یک مدل یادگیری ماشین را تا رسیدن به مدل نهایی مشاهده کرد.



شکل ۱. معماری پیشنهادی مدل تشخیص بیماری‌های همه‌گیر

## ۲-۱- لایه استخراج اطلاعات

در این پژوهش، داده‌ها از منابع مختلفی گردآوری شده است. داده‌های واقعی از داده‌های خام بیمارستانی چندین کشور که توسط وزارت بهداشت و رفاه هند با همکاری سازمان بهداشت جهانی گردآوری شده، از انبار داده سایت کاگل<sup>۱</sup> استخراج شده است. مجموعه داده مورد استفاده برای بیماری‌های همه‌گیر متشکل از ۲۷۰۰ بیمار مبتلا به بیماری‌های همه‌گیر که شامل ۲۷ ویژگی می‌باشد. این مجموعه داده شامل ویژگی‌های نظیر اطلاعات اولیه در مورد جنسیت، کشورهای مورد مطالعه، گروه‌های سنی، علائم‌ها (تب، سرفه خشک، مشکل تنفسی، درد عضلانی، بدون علامت)، سابقه بیماری (اسهال، احتقان بینی، آبریزش بینی و آیا این بیماری‌ها داشته است)، شدت بیماری (خفیف، متوسط، شدید و یا مشخص

<sup>۱</sup> <https://www.kaggle.com/code/harshaggarwal7/covid-19-symptom-analysis>

نیست) و تماس با فرد مبتلا (داشته است، نداشته است و یا مشخص نیست) که به طور مستقیم یا غیرمستقیم با بیماری های عفونی مرتبط است، می باشد. در جدول شماره ۱ اطلاعات و جامعه آماری از هر یک از ویژگی ها را می توان مشاهده کرد.

جدول ۱. اطلاعات آماری از ویژگی های مورد بررسی

درصد	تعداد	ویژگی	درصد	تعداد	ویژگی
۲۱/۳۳	۵۷۶	سن ۲۵ تا ۵۹	۱۵	۶۷/۰۷	۱۸۱۱ تب
۲۱/۱۱	۵۷۰	سن ۶۰ به بالا	۱۶	۳۳/۸۱	۹۱۳ خستگی
۴۷	۱۲۶۹	جنسیت مؤنث	۱۷	۸۱/۲۹	۲۱۹۵ سرفه خشک
۴۸	۱۲۹۶	جنسیت مذکر	۱۸	۳۳/۱۷	۹۱۲ تنگی نفس
۵	۱۳۵	جنسیت ترنس	۱۹	۹۶/۱۷	۲۶۱۳ گلو درد
۲۵/۰۳	۶۷۶	شدت خفیف بیماری	۲۰	۱/۱۸	۳۲ بدون علائم
۲۵/۲۹	۶۸۳	شدت نرمال بیماری	۲۱	۵۵/۵۹	۱۵۰۱ درد عضلانی
۲۴/۶۶	۶۶۶	شدت وخیم بیماری	۲۲	۷۱/۲۲	۱۹۲۳ گرفتگی بینی
۲۵/۰۳	۶۷۶	شدت بیماری نامشخص	۲۳	۸۹	۲۴۰۳ آبریزش بینی
۳۳	۸۹۱	سابقه تماس نامشخص	۲۴	۱۶/۲۹	۴۴۰ اسهال
۳۳/۳۳	۹۰۰	سابقه تماس دارد	۲۵	۲۷/۶۲	۷۴۶ بدون علامت
۳۳/۶۶	۹۰۹	سابقه تماس ندارد	۲۶	۱۸/۵۵	۵۰۱ سن ۰ تا ۹
-	۹	کشورها	۲۷	۲۲/۶۶	۶۱۲ سن ۱۰ تا ۱۹
				۱۶/۳۳	۴۴۱ سن ۲۰ تا ۲۴

## ۲-۲- لایه آماده سازی داده برای اجرای الگوریتم ها

در پیش پردازش داده ها به دلیل تمیز بودن داده ها (نبود داده پرت و داده نویز و یا داده از دست رفته) عمل انتخاب ویژگی<sup>۱</sup> انجام شد. برای انتخاب ویژگی داده ها از روش کای دو<sup>۲</sup> استفاده شد. دلیل استفاده از این روش این است که تمامی داده ها به صورت داده طبقه بندی شده<sup>۳</sup> می باشند.

<sup>۱</sup> Feature Selection

<sup>۲</sup> Chi 2 Test

<sup>۳</sup> Categorical Data



آزمون کای دو یا مربع کای آزمونی است که فراوانی‌های مورد انتظار را با فراوانی‌های تحقیق مقایسه می‌کند تا مشخص شود آیا تفاوت معناداری بین این دو فراوانی وجود دارد یا خیر.

$$\chi^2_c = \sum \frac{(O-E)^2}{E} \quad (1)$$

که در آن  $\chi^2$  برابر تعداد نمونه،  $c$  برابر درجات آزادی،  $O$  برابر فراوانی مشاهده شده و  $E$  برابر فراوانی مورد انتظار می‌باشد. سطح معناداری شاخص کای دو (معناداری مدل برازش شده با داده‌های موجود) برای مقادیر کم آن باید بیش از ۰/۰۵ باشد. در واقع شاخص P-Value شاخصی جهت عدم معناداری آزمون کای دو می‌باشد، که مقادیر زیر ۰/۰۵ گواه بر عدم معناداری بوده ( $H_0$ ) و در مقابل برای معناداری ( $H_1$ ) کای دو مقدار P-Value باید بالای ۰/۰۵ باشد.

جدول ۲. مقدار P-Value ویژگی‌ها در آزمون کای دو

	ویژگی	P-Value مقدار		ویژگی	P-Value مقدار
۱	تب	۰	۱۵	سن ۲۵ تا ۵۹	$72/4(e)^{-11}$
۲	خستگی	۰	۱۶	سن ۶۰ به بالا	$07/1(e)^{-12}$
۳	سرفه خشک	$14/1(e)^{-107}$	۱۷	جنسیت مؤنث	۰/۰۳
۴	تنگی نفس	۰	۱۸	جنسیت مذکر	۰/۰۰۰۵
۵	گلو درد	$92/3(e)^{-15}$	۱۹	جنسیت ترنس	۰/۰۹۵
۶	بدون علائم	۰/۰۰۰۱	۲۰	شدت خفیف بیماری	۰/۲۱
۷	درد عضلانی	۰	۲۱	شدت نرمال بیماری	۰/۳۲
۸	گرفتگی بینی	$98/2(e)^{-44}$	۲۲	شدت وخیم بیماری	۰/۵۲
۹	آبریزش بینی	$99/3(e)^{-67}$	۲۳	شدت نامشخص بیماری	۰/۶۱
۱۰	اسهال	$7/3(e)^{-139}$	۲۴	سابقه تماس نامشخص	۱
۱۱	بدون علامت	۰	۲۵	سابقه تماس دارد	۰/۸۷
۱۲	سن ۰ تا ۹	$34/1(e)^{-6}$	۲۶	سابقه تماس ندارد	۰/۸۷
۱۳	سن ۱۰ تا ۱۹	۰/۰۰۱	۲۷	کشورها	۱
۱۴	سن ۲۰ تا ۲۴	۰/۰۰۰۲			

بنابراین بعد از مشخص شدن مقدار P-Value برای هر ویژگی، ۹ ویژگی از جمله (جنسیت ترنس، شدت خفیف بیماری، شدت نرمال بیماری، شدت وخیم بیماری، شدت نامشخص بیماری، سابقه تماس نامشخص، سابقه تماس دارد، سابقه تماس ندارد و فیلد کشورها) حذف شدند.

## ۲-۳- لایه تقسیم‌بندی و تنظیم داده‌ها

چگونگی تقسیم داده به دو عامل بستگی دارد: اول تعداد ویژگی‌ها و دوم نوع مدل یادگیری. برخی از مدل‌ها برای آموزش نیاز به تعداد زیادی از داده‌ها را دارند، بنابراین تقسیم‌بندی باید به گونه‌ای باشد که در مجموعه داده آموزش بزرگی وجود داشته باشد [۱۷]. آموزش و تنظیم مدل‌هایی که ابرپارامترهای کمی دارند راحت و آسان می‌باشد و می‌توان حجم کمی از داده را برای مجموعه اعتبارسنجی در نظر گرفت. از طرف دیگر، در صورتی که مدل ابرپارامترهای زیادی داشته باشد، مجموعه اعتبارسنجی بزرگ‌تری مورد نیاز است. همچنین در صورتی که مدل یادگیری ابرپارامتر نداشته باشد یا ابرپارامترهای آن به گونه‌ای باشد که به آسانی قابل تنظیم نباشد، می‌توان از مجموعه اعتبارسنجی برای آن مدل صرف‌نظر کرد. در این پژوهش مجموعه داده آموزش برابر با ۷۰ درصد کل داده و مجموعه داده آزمایش برابر با ۳۰ درصد کل داده می‌باشد.

## ۲-۴- لایه مدل‌سازی الگوریتم‌های یادگیری ماشین

در این بخش ابتدا به شناسایی بیمارهای همه‌گیر با به‌کارگیری الگوریتم‌های یادگیری ماشین پرداخته می‌شود. از آنجایی که تشخیص بیماری همه‌گیر از حساسیت بالایی برای کادر درمان و مراکز بهداشتی برخوردار است، از الگوریتم یادگیری ماشین نظارت‌شده بر روی مجموعه داده‌های آماده‌شده استفاده شده تا بتوان با توجه به داده‌های برچسب گذاری شده که در بخش‌های قبلی توضیح داده شد، میزان صحت‌سنجی را دقیق‌تر محاسبه کرد. براساس تحقیقات صورت گرفته که در جدول شماره ۳ نشان داده شده است، الگوریتم‌های متعددی برای شناسایی بیماری‌های همه‌گیر استفاده شده است. از آنجایی که سلامت هر فرد و قطع زنجیره شیوع بیماری بسیار حائز اهمیت است در این مدل امکان افزودن سایر الگوریتم‌های یادگیری میسر می‌باشد.

جدول ۳. الگوریتم‌های یادگیری ماشین و نحوه عملکرد آن‌ها

نام الگوریتم	عملکرد
--------------	--------

ماشین بردار پیشتیبان	مبنای کاری دسته‌بندی‌کننده ماشین بردار پیشتیبان دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌شود ابرصفحه‌ای را انتخاب کرد که حاشیه اطمینان بیشتری داشته باشد.
K- نزدیک‌ترین همسایه	این الگوریتم برای مسائل طبقه‌بندی K- نزدیک‌ترین همسایه را پیدا کرده و با اکثریت آرا نزدیک‌ترین همسایگان کلاس را پیش‌بینی می‌کند.
درخت تصمیم	به سمت پایین رشد می‌کنند که از ریشه درختی که نمونه‌ها را به نحوی دسته‌بندی می‌کند و در نهایت به گره برگ می‌رسد.
افزایش گرادیان شدید	ترکیب تعدادی از مدل‌های ساده‌تر و ضعیف‌تر درخت تصمیم که به‌طور متوالی ساخته می‌شوند، امکان ایجاد پیش‌بینی‌ها را فراهم می‌کند.
جنگل تصادفی	شامل تعدادی درخت تصمیم در زیر مجموعه‌های مختلف مجموعه داده قرار دارد و برای بهبود دقت پیش‌بینی آن مجموعه داده، میانگین می‌گیرد. جنگل تصادفی به جای تکیه بر یک درخت تصمیم، پیش‌بینی را از هر درخت و براساس اکثریت آرا پیش‌بینی می‌کند و نتیجه نهایی را به عنوان خروجی در نظر می‌گیرد.

بسیاری از الگوریتم‌های یادگیری ماشین برای حل مشکلات مختلفی طراحی شده‌اند. این سوال ممکن است مطرح شود که کدام الگوریتم کارایی بالاتری نسبت به سایر الگوریتم‌ها دارد [۱۸]؟ به‌عنوان مثال در حوزه دسته‌بندی داده، باید چندین الگوریتم دسته‌بندی را برای آموزش انتخاب کرد. هر الگوریتم ویژگی‌های منحصر بفرد خودش را دارد و هیچ الگوریتمی بدون در نظر گرفتن فرضیات مسئله ارائه شده نمی‌تواند نسبت به دیگری برتری پیدا کند. بنابراین، لازم است که حداقل تعداد معدودی از الگوریتم‌های مختلف را برای آموزش مدل انتخاب و با مقایسه عملکرد آن‌ها، بهترین مدل را انتخاب کرد. خروجی الگوریتم‌های عنوان شده در جدول ۳ بر روی داده‌های جمع‌آوری شده، پیاده‌سازی و آزمون خواهد شد تا بتوان بهترین مدل تشخیص‌دهنده بیماری‌ها را جهت پیش‌بینی سریع مشخص نمود.

## ۲-۵- لایه انتخاب بهترین مدل

پس از اعمال الگوریتم‌های عنوان شده در مراحل خروجی، بهترین الگوریتم از لحاظ صحت‌سنجی نشان می‌دهد که آیا فرد به کدامیک از بیماری‌های همه‌گیر مبتلا است یا خیر؟ با توجه به نتایج به دست آمده الگوریتم K- نزدیک‌ترین همسایه با عملکرد بهتر نسبت به سایر الگوریتم‌های یادگیری توانسته است تشخیص آلودگی افراد به هر یک از بیماری‌ها را به درستی تشخیص دهد. با توجه به اینکه تأثیر هر یک از روابط زیر بر روی توصیه‌ی نهایی می‌تواند متفاوت باشد، از

روش فاصله اقلیدسی برای هر یک از آن‌ها استفاده شده است. رابطه‌ی نهایی پیشنهادی مورد استفاده برای این الگوریتم، به صورت زیر می‌باشد.

به منظور پیش‌بینی مقدار، با داشتن یک داده مشخص  $x$  می‌توان فاصله آن را از تمامی داده‌های آموزش محاسبه کرد:

$$d_i = \text{distance}(x, \text{Triani}, x) \quad (2)$$

تابع فاصله می‌تواند انواع مختلف داشته باشد. یکی از معیارهای فاصله پرکاربرد، فاصله مینکوسکی<sup>۱</sup> است. این معیار به شکل زیر محاسبه می‌شود:

$$d(a, b) = \left( \sum_{i=1}^n |a_i - b_i|^p \right)^{\frac{1}{p}} \quad (3)$$

پارامتر  $p$  می‌تواند اعداد مختلفی در بازه  $[0, +\infty)$  باشد. اگر مقدار آن برابر با ۱ باشد، رابطه به شکل زیر در می‌آید:

$$(a, b) = \sum_{i=1}^n |a_i - b_i| \quad (4)$$

به این معیار فاصله منهن<sup>۲</sup> نیز گفته می‌شود. حال اگر مقدار  $p$  برابر با ۲ در نظر گرفته شود، رابطه زیر حاصل می‌شود:

$$d(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (5)$$

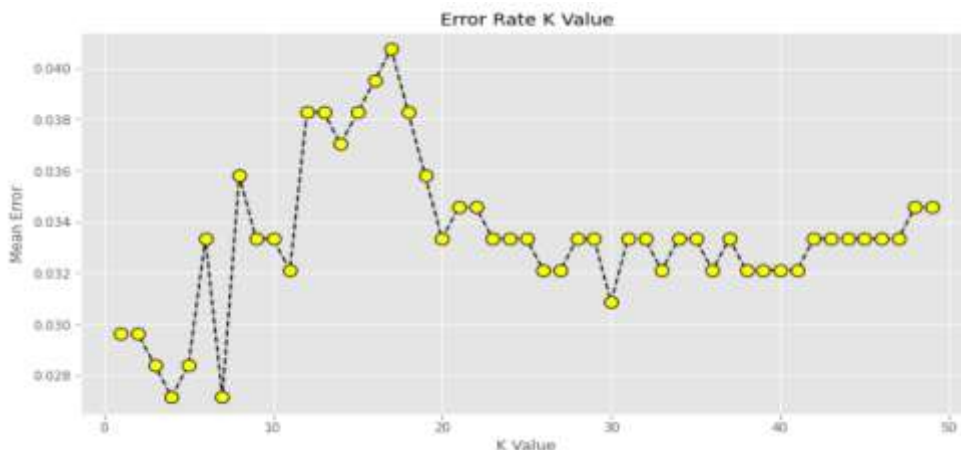
به این معیار فاصله اقلیدسی<sup>۳</sup> نیز گفته می‌شود. برای  $p$  دو مقدار ۱ و ۲ معمول‌تر است.

مقادیر به دست آمده با استفاده از محاسبه میانگین خطا برای تمامی مقادیر پیش‌بینی شده برای تعداد همسایگی الگوریتم  $K$  - نزدیک‌ترین همسایه برابر با ۴ و ۷ می‌باشند. در شکل ۲ می‌توان مقادیر به دست آمده از مقایسه میانگین خطا برای هر یک از همسایگی‌ها را مشاهده کرد. عملکرد هر دو مقادیر به دست آمده از تعداد همسایگی برابر می‌باشند.

<sup>1</sup> Minkowski Distance

<sup>2</sup> Manhattan Distance

<sup>3</sup> Euclidean Distance



شکل ۲. محاسبه نرخ خطا برای تعداد همسایگی در الگوریتم K- نزدیکترین همسایه

### ۳- مدل پیشنهادی

در این پژوهش از الگوریتم‌های ماشین بردار پشتیبان، الگوریتم افزایش گرادیان شدید، الگوریتم درخت تصمیم، الگوریتم K- نزدیکترین همسایه و الگوریتم جنگل تصادفی استفاده شده است. از آنجایی که داده‌های جمع‌آوری شده داده‌های تمیزی بودند و نیاز به عمل پاک‌سازی داده‌ها نبود، بعد از عمل انتخاب ویژگی داده‌ها در یک فایل جدید ذخیره‌سازی شدند. برای اجرای الگوریتم‌های یاد شده از زبان برنامه‌نویسی پایتون در محیط نرم‌افزاری ژوپیترا<sup>۱</sup> استفاده شده است. پس از اعمال الگوریتم‌ها بر روی مجموعه داده‌ها جهت تشخیص ۳ بیماری تنفسی به نتایج مختلفی از هر الگوریتم دست یافتیم. برای ارزیابی مدل پیشنهادی و الگوریتم‌های یادگیری ماشین به کار رفته، چهار معیار دقت، صحت، نرخ بازخوانی و مقدار F1 محاسبه شد. چهار نوع تشخیص تحت عناوین مثبت صحیح شناسایی شده<sup>۲</sup>، منفی صحیح شناسایی شده<sup>۳</sup>، مثبت کاذب شناسایی شده<sup>۴</sup> و منفی کاذب شناسایی شده<sup>۵</sup> برای ارزیابی لحاظ گردیده است.

میزان "صحت" یکی از معیارهای اندازه‌گیری عملکرد مدل‌ها، دقت طبقه‌بندی است که میزان پیش‌بینی صحیح طبقه‌بندی‌کننده را اندازه‌گیری می‌کند [۲۰، ۱۹].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (۶)$$

<sup>۱</sup> Jupyter Notebook

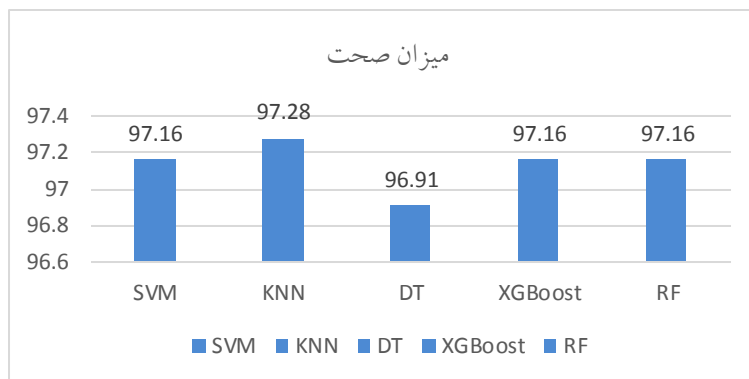
<sup>۲</sup> True Positive

<sup>۳</sup> True Negative

<sup>۴</sup> False Positive

<sup>۵</sup> False Negative

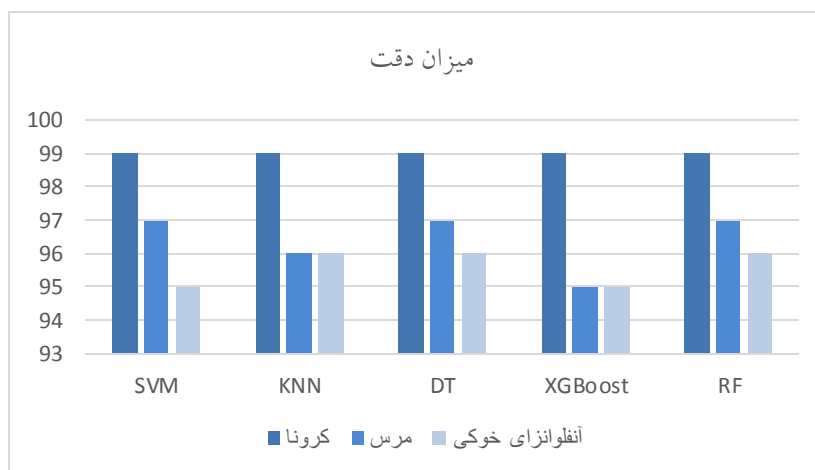
در شکل ۳ نتایج میزان صحت از اعمال الگوریتم‌های یادگیری ماشین برای شناسایی بیماری‌های همه‌گیر قابل مشاهده می‌باشد.



شکل ۳. نتایج میزان صحت از اعمال الگوریتم‌های یادگیری ماشین برای شناسایی بیماری‌های همه‌گیر

در این تحقیق با توجه به نتایج به دست آمده از اعمال پنج الگوریتم یادگیری ماشین عنوان شده، میزان صحت الگوریتم K - نزدیک‌ترین همسایه با ۹۷/۲۸ درصد به عنوان کارآمدترین الگوریتم انتخاب شد. معیار دوم "میزان دقت" است که به عنوان ارزش اخباری مثبت نیز شناخته می‌شود و نسبت نمونه‌های مرتبط در میان نمونه‌های بازیابی شده است. به عبارت دیگر، به این پرسش پاسخ می‌دهد که در واقع چه نسبتی از شناسه‌های مثبت درست بودند؟

$$\text{Precision} = \frac{TP}{TP+FP} \quad (V)$$

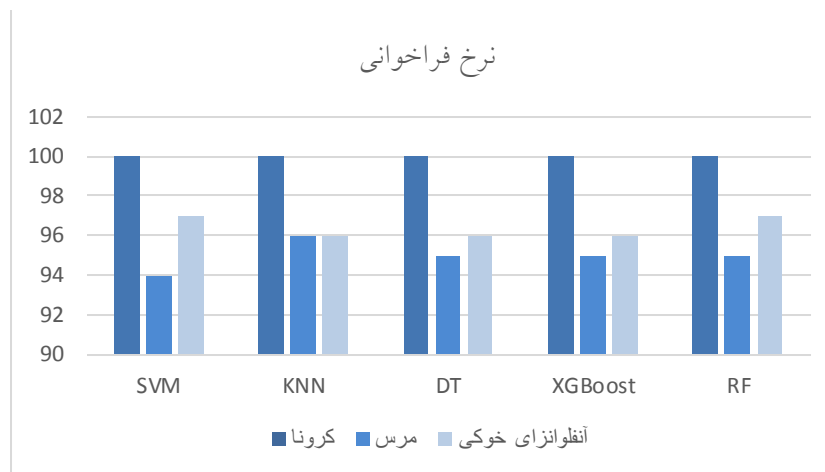


شکل ۴. نتایج میزان دقت از اعمال الگوریتم‌های یادگیری ماشین برای شناسایی بیماری‌های همه‌گیر

با توجه به پردازش بدست آمده از معیار دقت دو الگوریتم **K**- نزدیک ترین همسایه و درخت تصمیم عملکرد بهتری نسبت به سایر الگوریتم‌ها از خود نشان دادند.

معیار سوم "نرخ بازخوانی" است. این معیار به این پرسش پاسخ می‌دهد: چه نسبتی از تشخیص‌های مثبت واقعی به درستی شناسایی شده‌اند؟ به بیان ساده‌تر به این مفهوم اشاره دارد که از بین همه کلاس‌های مثبت، چقدر درست پیش‌بینی شده است و باید تا حد ممکن بالا باشد. البته به این معیار حساسیت نیز گفته می‌شود.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

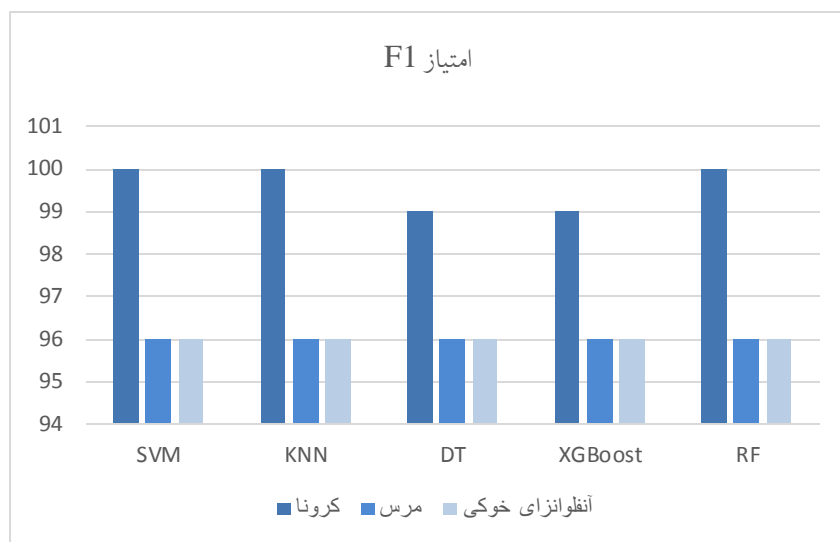


شکل ۵. نتایج نرخ فراخوانی از اعمال الگوریتم‌های یادگیری ماشین برای شناسایی بیماری‌های همه‌گیر

با توجه به پردازش بدست آمده از محاسبه نرخ بازخوانی الگوریتم جنگل تصادفی و **K**- نزدیک ترین همسایه نسبت به سایر الگوریتم‌ها باقی‌مانده عملکرد بهتری از خود نشان دادند.

معیار چهارم و آخر "امتیاز **F1**" است که میانگین وزنی دو مقدار قبلی دقت و نرخ بازخوانی است. در شکل ۶ نتایج مربوط به محاسبه این مقدار نمایش داده شده است.

$$\text{F1 Score} = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (9)$$



شکل ۶. نتایج معیار امتیاز F1 از اعمال الگوریتم‌های یادگیری ماشین برای شناسایی بیماری‌های همه‌گیر

با توجه به نتایج به دست آمده از امتیاز F1 می‌توان دریافت که سه الگوریتم ماشین بردار پشتیبان، K- نزدیک‌ترین همسایه و جنگل تصادفی نسبت به سایر رقبا عملکرد بهتری از خود نشان دادند.

#### ۴- ارزیابی و مقایسه نتایج

اهمیت ویژگی، فهرستی از ویژگی‌هاست که برای مدل مهم هستند. این فهرست به هر ویژگی یک نمره اهمیت اختصاص می‌دهد که اهمیت آن ویژگی خاص را برای پیش‌بینی نشان می‌دهد. اهمیت ویژگی؛ یکی از توابع تعبیه‌شده کتابخانه learn-Scikit است که در پیاده‌سازی بسیاری از مدل‌های یادگیری ماشینی استفاده می‌شود. برای تشخیص بهترین زیرمجموعه ممکن از ویژگی‌ها می‌توان از مقادیر اهمیت ویژگی استفاده کرد و براساس آن زیرمجموعه، مدلی قوی ساخت. در جدول شماره ۴ به بررسی این که کدامیک از ویژگی‌های مجموعه داده، به ترتیب دارای اهمیت بالا هستند نشان داده می‌شود.

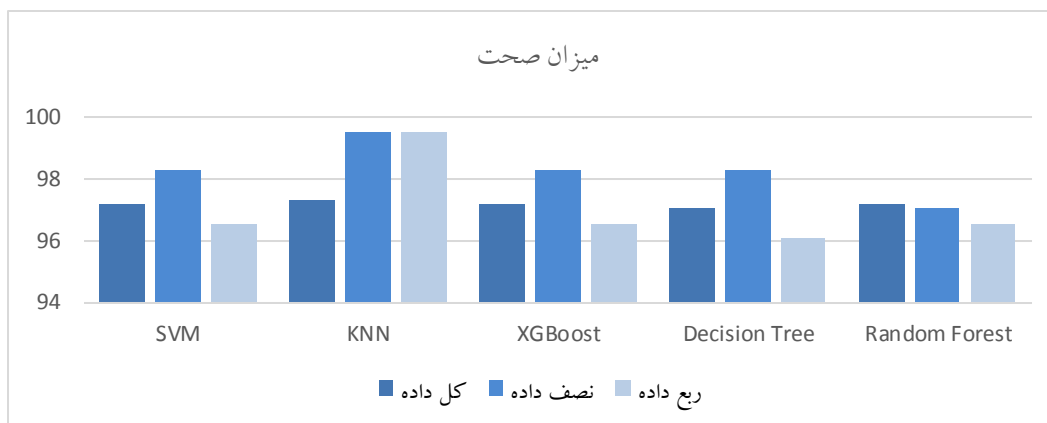
جدول ۴. اهمیت ویژگی هر یک از الگوریتم‌های یادگیری ماشین

اهمیت ویژگی	اولین علائم	دومین علائم	سومین علائم	چهارمین علائم	پنجمین علائم	ششمین علائم
ماشین بردار پشتیبان	درد عضلانی	بدون تجربه بیماری	مشکل در تنفس	احساس خستگی	تب	گروه سن ۰ تا ۹



جنسیت زن	جنسیت مرد	اسهال	گرفتگی بینی	احساس خستگی	مشکل در تنفس	- نزدیکترین همسایه K
سرفه خشک	بدون تجربه بیماری	احساس خستگی	درد عضلانی	تب	مشکل در تنفس	افزایش گرادیان شدید
سرفه خشک	احساس خستگی	تب	بدون تجربه بیماری	درد عضلانی	مشکل در تنفس	درخت تصمیم
سرفه خشک	بدون تجربه بیماری	تب	احساس خستگی	مشکل در تنفس	درد عضلانی	جنگل تصادفی

در قدم بعدی، به تأثیر تعداد داده بر هر الگوریتم می‌پردازیم. با تقسیم کردن داده به نصف و بار دیگر به یک چهارم عمل آموزش و آزمایش با معیار ارزیابی دقت را بار دیگر اجرا می‌کنیم تا نتیجه بگیریم که کدام یک از الگوریتم‌های به کار رفته در این پژوهش به تعداد نمونه‌ها وابستگی دارد. شکل ۷ مقدار عددی معیار ارزیابی صحت بر تعداد نمونه هر یک از الگوریتم‌های یادگیری را نشان می‌دهد. وقتی داده‌ها به نصف تقسیم می‌شوند به جز الگوریتم جنگل تصادفی، صحت طبقه‌بندی الگوریتم‌ها بهبود می‌یابد. همچنین داده‌ها بار دیگر به ربع تقسیم شوند (یک چهارم داده) به جز الگوریتم K- نزدیک‌ترین همسایه، دقت طبقه‌بندی کاهش می‌یابد.



شکل ۷. مقدار عددی معیار ارزیابی صحت بر تعداد نمونه هر یک از الگوریتم‌های یادگیری

## ۵- نتیجه‌گیری

نتایج حاصل از این پژوهش با بهره‌گیری از تکنیک‌های هوش مصنوعی و یادگیری ماشین نشان می‌دهد که نه تنها می‌توان بیماری‌های همه‌گیر همچون کروناویروس، ویروس مرس و آنفلوآنزای خوکی را با درصد بالایی از صحت توسط الگوریتم‌های یادگیری ماشین تشخیص داد بلکه، امکان استفاده برای سایر بیماری‌های همه‌گیر نیز وجود دارد. با توجه به

ویژگی‌های به کار رفته در این پژوهش می‌توان در مرحله اولیه شیوع و ویروس با شناسایی سریع عوامل یا همان علائم‌های آن، به جداسازی افراد آلوده و همچنین برای پزشکان و کادر بهداشتی که در معرض خطر هستند کمک کرد [۲۱]. به کمک اجرای هم‌زمان پنج الگوریتم یادگیری ماشین درخت تصمیم، ماشین بردار پشتیبان، افزایش گرادیان شدید، جنگل تصادفی و K- نزدیک‌ترین همسایه می‌توان عنوان نمود که مجموعه داده یکی از موثرترین عوامل در انتخاب الگوریتم مناسب است. نتایج آزمایش‌های صورت گرفته در مورد روش پیشنهادی بر روی داده‌های مختلف نشان می‌دهد که الگوریتم معرفی شده (K- نزدیک‌ترین همسایه) بهتر عمل کرده است. برای ارزیابی کارایی الگوریتم K- نزدیک‌ترین همسایه، این روش بر روی داده‌های متعددی اجرا و نتایج آن، با نتایج الگوریتم‌هایی مانند: درخت تصمیم، جنگل تصادفی، افزایش گرادیان شدید، ماشین بردار پشتیبان، مقایسه شد که نتایج نشان از برتری این روش در بیشتر موارد دارد. میزان موفقیت این الگوریتم در جدول‌های مختلف در بخش نتایج و ارزیابی به صورت نمودار آورده شده است که در بیشتر موارد میزان موفقیت این الگوریتم بالای ۹۷ درصد بوده است. به این صورت که به‌طور مثال داده‌ها یک بار به نصف (۱۳۵۰ نمونه) و بار دیگر به ربع (۶۷۵ نمونه) تقسیم شدند. سپس دوباره عمل آموزش و آزمایش انجام شد که تقریباً الگوریتم K- نزدیک‌ترین همسایه بهترین عملکرد را میان سایر الگوریتم‌ها داشت.

روش انتخابی در این پژوهش به‌عنوان یکی از کارآمدترین الگوریتم‌های یادگیری ماشین شناخته می‌شود. برای نشان دادن کارایی خوب الگوریتم K- نزدیک‌ترین همسایه این روش بر روی تعداد زیادی نمونه‌ها در مقیاس‌های کوچک و بزرگ مورد ارزیابی قرار گرفت و همان‌طور که ملاحظه شد نتایج خوبی در اکثر موارد از خود نشان داد. در دسته الگوریتم‌های یادگیری ماشین به حداقل رساندن خطای هر الگوریتم با کاهش نویز در داده‌ها و همچنین تنظیم ابرپارامترها باعث پدیدار شدن بهترین جواب در مدل می‌باشد. توجه به دو عامل ارائه کارآمدترین مدل پیش‌بینی‌کننده و تشخیص علائم‌های اولیه بیماری منجر به تحقق یافتن این اهداف خواهد شد. روند روش انتخابی تا حد ممکن این عامل‌های مهم را پوشش داده و در آزمایش‌های متعدد نتایج خوبی از خود نشان داده است. به‌طوری که نتایج به‌دست آمده نشان می‌دهد که کیفیت جواب‌ها یک وضعیت مطلوب دارد. به‌طور کلی می‌توان گفت:

- مطالعاتی بر روی موضوع‌ها و چالش‌های موجود در زمینه شناسایی بیماری‌های ویروسی و الگوریتم‌های یادگیری ماشین صورت گرفت.
- همچنین تحقیق و بررسی‌هایی بر روی الگوریتم‌های یادگیری، ساختار و نحوه‌ی کار آن‌ها انجام شد.
- یک الگوریتم یادگیری ارائه شد که می‌تواند به حل مسائل در زمینه شناسایی بیماری‌های ویروسی بپردازد.
- اجرای الگوریتم K- نزدیک‌ترین همسایه بر روی نمونه‌های مختلف نشان داد که الگوریتم دارای وضعیت خوبی

از لحاظ شناسایی و افزایش کیفیت جواب‌ها است.

- در بیشتر موارد الگوریتم K- نزدیک‌ترین همسایه جواب‌های بهتری نسبت به دیگر الگوریتم‌های هم‌رده از خود نشان داد.

در کارهای آتی می‌توان با افزودن سایر علائم‌های بالینی به داده‌ها میزان دقت را بالا برد. همچنین برای استفاده در مراکزهای بهداشتی و درمانی برای طبقه‌بندی اولیه بیماری‌های عفونی از سیستم پیشنهادی استفاده نمود. همچنین:

- استفاده از مدل برای سایر داده‌های استخراج شده بهداشتی توسط مراکز درمانی.
- الگوریتم پیشنهادی برای حل مسائل یادگیری با نظارت مناسب می‌باشد، لذا می‌توان با اعمال تغییراتی از آن برای حل مسائل یادگیری نیمه‌نظارت و یا بدون نظارت نیز استفاده کرد.
- همچنین با اعمال الگوریتم انتخابی به مسائل مهندسی، می‌توان آن را برای حل چنین مسائلی به کار برد. این الگوریتم مناسب مسائل دسته‌بندی می‌باشد که با اعمال تغییراتی می‌توان در مسائل رگرسیون نیز استفاده کرد.

## حمایت مالی

این پژوهش تحت حمایت مالی هیچ سازمان یا نهادی نمی‌باشد.

## مراجع

- [1] P. Dashraath et al., Coronavirus disease 2019 (COVID-19) pandemic and pregnancy, American journal of obstetrics and gynecology, Vol. 222, No. 6, pp. 521-531, 2020.
- [2] K. Syal, Guidelines on newly identified limitations of diagnostic tools for COVID-19 and consequences, Journal of medical virology, Vol. 93, No. 4, pp. 1837-1842, 2021.
- [3] C. Martin et al., Dynamics of SARS-CoV-2 RT-PCR positivity and seroprevalence among high-risk healthcare workers and hospital staff, Journal of Hospital Infection, Vol. 106, No. 1, 2020, pp. 102-106.
- [4] Çinare Oğuz, Mete Yağanoğlu, Detection of COVID-19 using deep learning techniques and classification methods: Information Processing and Management, 2022, 59, 103025, 1.
- [5] Y. Kazemi, S.A. Mirroshandel, A novel method for predicting kidney stone type using ensemble learning, Artif. Intell. Med. 84, 2018, 117–126.
- [6] Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med; 367(19), 2012, 1814-20.
- [7] Md. Martuza Ahamad, Sakifa Aktar, Md. Rashed-Al-Mahfuz, Shahadat Uddin, Pietro Lie, Haoming Xu, Matthew A. Summers, Julian M.W. Quinn, Mohammad Ali Moni, A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patient, 160, 2020,

113661, 1- 3.

- [8] L. Lan et al. Positive RT-PCR test results in patients recovered from COVID-19, *Jama*, Vol. 323, No. 15, 2020, pp. 1502-1503.
- [9] Albahri A, Hamid RA. 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: 1-11.
- [10] Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health* 2020; 2: 516-525.
- [11] Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intel.* 2020; 2: 283-288.
- [12] Gao Y, Cai GY, Fang W, Li HY, Wang SY, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Comm.* 2020; 11: 1-10.
- [13] Agieb R. Machine learning models for the prediction the necessity of resorting to ICU of covid-19 patients. *Int. J Adv. Trends Com Sci. Eng.* 2020; 6980-4.
- [14] Pijush Dutta, Shobhandeb Paul, Asok Kumar, Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19, *Electronic Devices, Circuits, and Systems for Biomedical Applications*, 2021, 521-540.
- [15] Hemanuel Passarelli-Araujo, Hisrael Passarelli-Araujo, Mariana R. Urbano, Rodrigo R. Pescim, Machine learning and comorbidity network analysis for Hospitalized patients with COVID-19 in a city in Southern Brazil: *Smart Health*, 2020, 26, 100323, 1-2.
- [16] Li, F., Li, Y.-Y., & Wang, C., Uncertain data decision tree classification algorithm. *Journal of Computer Applications*, 2009, 29, 3092–3095.
- [17] Knipe, D.M., and Howley, P.M. Orthomyxoviridae: The Viruses and Their Replication. In: Robert A. Lamb, Robert M. Krug, editor. *Fundamental of Virology*. 4<sup>th</sup> edition. Lippincott Williams & Wilkins publication; New York, 2001, PP 724-76.
- [18] M. Seera, C.P. Lim, A hybrid intelligent system for medical data classification, *Expert Syst. Appl.* 41 (5), 2014, 2239–2249.
- [19] E. Gürbüz, E. Kılıç, A new adaptive support vector machine for diagnosis of diseases, *Expert Syst.* 31 (5), 2014, 389–397.
- [20] 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.

[21] 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.

[22] Roneel V. Sharan, Hania Rahimi-Ardabil, Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review, International Journal of Medical Informatics, 176 (2023) 105093. (a)

[۲۳] شنبه‌زاده. مصطفی، ولی‌نژاد. علی، افراح. رامین، ۱۴۰۰. مقایسه کارایی الگوریتم‌های یادگیری ماشین برای ایجاد یک مدل پیش‌بینی‌کننده مرگ‌ومیر بیماران بستری در بیمارستان با کووید-۱۹، مجله کومش، ۱، ۱-۱۰.