زمستان ۱۴۰۰، دوره ۶، شماره ۴، پیاپی ۲۳
DOR :20.1001.1.2538161.1400.6.4.6.0

دریافت مقاله: ۱۴۰۰/۰۷/۲۷
پذیرش مقاله: ۱۴۰۰/۱۰/۲۲

نوع مقاله: پژوهشی

پیش‌بینی بیماری کرونا از روی علائم قابل مشاهده با استفاده از یادگیری ماشین

مهدی سرچاهی
دانشجوی کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار، موسسه آموزش عالی خاوران، مشهد، ایران
پست الکترونیکی: sm2249@yahoo.com

الهام مهدی‌پور*
استادیار گروه مهندسی کامپیوتر، موسسه آموزش عالی خاوران، مشهد، ایران
پست الکترونیکی: e.mahdipour@khi.ac.ir

چکیده

این روزها جامعه بشریت شاهد افزایش مرگ و میر ناشی از جهش‌های مختلف بیماری کرونا است. بیماری کرونا علائم متفاوتی در بدن هر فرد دارد؛ اما اغلب گونه‌های آن در مراحل اولیه بیماری علائمی دارند که قابل مشاهده توسط فرد نیز هستند. هدف از این پژوهش، پیش‌بینی بیماری کرونا از روی علائم اولیه بیماری هست. در این راستا جهت شناسایی و پیش‌بینی بیماری کرونا از الگوریتم‌های یادگیری ماشین همانند بیزین ساده، رگرسیون لجستیک، درخت تصمیم، تحلیل تشخیص خطی، K-نزدیک‌ترین همسایه و ماشین بردار پشتیبان استفاده شده است. برای ارزیابی کارایی الگوریتم‌های فوق از دو مجموعه داده واقعی در پایگاه داده Kaggle استفاده شده است. برای پیاده‌سازی الگوریتم‌ها از زبان برنامه‌نویسی پایتون بهره گرفته شده است. نتایج حاصل از اجرا نشان می‌دهد الگوریتم درخت تصمیم با بالاترین میزان دقت دارای بیشترین کارایی در پیش‌بینی بیماری کرونا است.

واژه‌های کلیدی: پیش‌بینی، بیماری کرونا، یادگیری ماشین، کووید-۱۹.

۱- مقدمه

پاندمی ویروس کووید-۱۹ در حال تبدیل شدن به یک بحران سلامتی در جهان است. متاسفانه با وجود گذشت دو سال از شروع این بیماری هنوز هم برخی از جهش‌های این ویروس ناشناخته هستند. بیماری کرونا از دسامبر ۲۰۱۹ در بازار غذاهای دریایی مردم ووهان، درجنوب چین ظاهر شد و به سرعت در سراسر جهان پخش شد. در آن زمان، شیوع ویروس به‌عنوان یک وضعیت اورژانسی سلامت عمومی ازسوی سازمان بهداشت جهانی اعلام شد که متاسفانه این وضعیت هنوز هم ادامه دارد.

هانگ و همکاران طی تحقیقی ثابت کردند در ووهان چین مواردی از بیماری کرونا ویروس-۲۰۱۹ یا کووید-۱۹ به شکل بیماری ذات‌الریه که به آن بیماری تنفسی حاد nCov-2019 نیز می‌گویند ظاهر شده است [۱]. آن‌ها مشخصات اپیدمیولوژیکی، بالینی، آزمایشگاهی، رادیولوژی ونتایج درمانی و بالینی این بیماران را گزارش کردند. براساس گزارش آن‌ها از ٤١ بیمار بستری در بیمارستان که دارای عفونت آزمایشگاهی nCoV-2019 بودند. بیشتر بیماران آلوده مرد بودند(۳۰ [۷۳٪] از ٤١

* نویسندهٔ مسئول

نفر). کمتر از نیمـــی از آن‌ها به بیماری‌های زمینه‌ای مبتلا بودنـــد (۱۳نفـــر [۳۲٪]) ازجمله دیابت (هشـــت نفر [۲۰٪])، فشارخون بالا (شش نفر [۱۵٪]) و بیماری‌های قلبی عروقی (شش نفر [۱۵٪]). همچنین متوسط سن افراد ۴۹ سال بوده اســـت. ۲۷ نفر (٦٦٪) از ٤١ بیمار در معرض بازارغذاهای دریایی هونان در ووهـــان قرارگرفته بودند. علائم متداول در شـــروع بیمـــاری تـــب (٤٠ [۹۸٪] از ٤١ بیمار)، ســـرفه (۳۱ نفر [۷٦٪])، و خســـتگی (۱۸ نفر [٤٤٪])، علائم شایع‌تر تولیدخلط (۱۱ نفر [۲۸٪] از ۳۹ نفر)، ســـردرد (سه نفر [۸٪] از ۳۸ نفر)، هموپتزی (دو نفر [۵٪] از ۳۹ نفر) و اسهال (یکی [۳٪] از ۳۸ نفر) بود. تنگـــی نفس در ۲۲ بیمار (۵۵٪) از ٤٠ بیمار (زمان متوسط ازشروع بیماری تا تنگی نفس ۸ روز) ایجادشده بود.

ازدمیر در مطالعات خود نشان داد که بیماری کرونا یک بیماری عفونی است که بر اثر ویروس سندروم حاد تنفسی SARS-CoV-2 ایجاد می‌شـــود [۲]. او نشـــان داد مدت زمان پنهان بودن عفونت سندرم حاد تنفسی SARS-CoV-2، ۱٤روز پس از مواجهه فرض می‌شـــود، که اغلب در حدود چهار تا پنج روز طول می‌کشد. همچنین ثابت کرد افراد در هر ســـنی ممکن است به عفونت SARS-CoV-2 مبتلا شوند، اگرچه افراد میان سال و مسن اکثریت را تشکیل می‌دهند. تحقیقات او نیز نشان داد ویژگی‌های بالینی معمول شامل تب، سرفه خشک، خستگی، گلو درد، سردرد، تهوع، استفراغ و اسهال است. او به این نتیجه رسید که هنوز هیچ ویژگی بالینی منحصر به فردی وجود ندارد که بیماری کووید-۱۹ را از سایر عفونت‌های ویروســـی مجاری هوایی فوقانی/ تحتانی متمایز کند.

وولفل و همکاران در پژوهشی نشان دادند که ویروس کرونا به شـــکل فعال در دســـتگاه تنفســـی فوقانی تکثیر می‌شود [۳]. آن‌ها نشـــان دادند ویروس عفونی به آسانی از نمونه‌هـــای گرفته‌شـــده از گلو یا ریه جدا می‌شـــود، اما ویروس باوجود غلظت‌های بالای RNA از نمونه‌های مدفوع جدا نمی‌شـــود. طبق آزمایش‌های آن‌هـــا، نمونه‌های خون و ادرار، هرگـــز وجود ویروس را مشـــخص نکردند. آن‌ها نشـــان دادند تکثیر فعال در گلو با وجود واسطه‌های RNA و ویروسی در نمونه‌های گلو، شکل می‌گیرد. آن‌ها به طور مـــداوم، جمعیت‌های ویروس متمایز را درگلو و نمونه‌های ریـــه یک بیمار ردیابی کردند و تکثیر مســـتقل آن را اثبات نمودند.

حافـــظ و همـــکاران [٤] طی مطالعـــه و تحقیق بر روی کارهای انجام شـــده ثابت کردند بیمـــاری کرونا از طریق درمان عمومی، درمان نشـــانه‌های بیماری، با اســـتفاده از داروهای ضد ویروسی، اکسیژن درمانی و توسط سیستم ایمنی بدن بهبود می‌یابد. آن‌ها نشـــان دادند که باید موارد احتمالی را در اســـرع وقت شناسایی و افراد مشکوک را از موارد تایید شـــده کووید-۱۹ جدا کرد، تا از انتقال احتمالی عفونت به ســـایر بیماران و کارکنان مراقبت‌های بهداشتی جلوگیری نمود.

طبـــق تحقیقات می‌تـــوان گفت شناســـایی ویژگی‌های جمعیت آلوده به شناســـایی عوامل خطـــر و تعیین اهداف درمانی موثرتر کمک می‌کند[۵]. نتایج بررســـی غلامپور و همکارانش نشان داد که بیماری‌های پیری و زمینه‌ای مانند دیابت و بیماری انسدادی مزمن ریوی، عواملی هستند که خطر ابتلا به کووید-۱۹را در افراد بیشتر می‌کند.

برخی از تحقیقات نیز اقدام به شناسایی بیماری کرونا با اســـتفاده از یادگیری عمیق نمودند. نارین و همکاران با استفاده از شـــبکه عصبی هم‌آمیختی[1] عمیق با دقتی بالای ۹۹ درصد موفق به کشف بیماری کرونا از روی تصاویر اشـــعه X قفسه سینه شـــدند [٦]. اوکبدان و همکاران نیز با استفاده از پردازش تصاویر قفسه سینه به کمک یادگیری عمیـــق و مدل VGG توانســـتند با دقتی بـــالای ۹۹ درصد بیماری کرونا را تشـــخیص دهند [۷]. چاندرا و همکارانش نیـــز یک مدل طبقه‌بندی سلســـله مراتبی ارائـــه نمودند که بر اســـاس رای اکثریت مدل‌هـــای طبقه‌بندی‌کننده اقدام به شناسایی بیماری کرونا می‌نماید[۸].

آریســـتا در تحقیقی کارایی الگوریتـــم درخت تصمیم

1- convolutional

و رگرسیون لجستیک را برای شناسایی بیماری کرونا مقایسه می‌کند [۹]. نتایج نشان داد درخت تصمیم به دقت ۹۸٪ و رگرسیون لجستیک به دقت ۹۷٪ برای شناسایی کرونا رسیده است. الاسلام و همکارانش نیز با استفاده از رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی و XGBoost اقدام به دسته‌بندی بیماران کرونایی نمودند [۱۰] که نتیجه تحقیق آن‌ها نیز بیانگر کارایی مطلوب درخت تصمیم و XGBoost با رسیدن به دقت ۹۸٪ است. ویلاویچنسیو و همکارانش [۱۱] در تحقیقی بر روی الگوریتم‌های یادگیری ماشین با استفاده از نرم‌افزار داده‌کاوی وکا[2] و روش اعتبارسنجی متقابل ۱۰-فولد[3] نشان دادند که روش ماشین بردار پشتیبان با دقت ۹۸/۸۱ ٪ به بالاترین دقت برای شناسایی بیماری کرونا دست یافته است. یوسف [۱۲] نیز در پژوهش خود نشان داد که شناسایی بیماری کرونا با استفاده از شبکه‌های عصبی و روش اعتبارسنجی متقابل ٤-فولد[4] با ماکزیمم دقت ۹۸/۱٦ ٪ امکان‌پذیر است. در این راستا تحقیقات زیاد دیگری همچون [۱۳]، [۱٤]، [۱٥] و [۱٦] انجام شده است که همگی نشان می‌دهند بیماری‌زایی این ویروس، دستگاه تنفس را تحت تاثیر قرار می‌دهد و علائمی مشابه سرماخوردگی ساده را ایجاد می‌کند. علائم ویروس کرونا شامل: اختلالات تنفسی، آبریزش بینی، سرفه خشک، سرگیجه، گلو درد و بدن درد است که می‌تواند با سردرد و تب نیز همراه باشد و تا چند روز طول بکشد. در افرادی با نقص دستگاه ایمنی، افراد مسن و کودکان این علائم می‌تواند شدیدتر شده و منجر به سینه پهلو و برونشیت شود.

هدف از این تحقیق شناسایی بیماری کرونا از روی علائم و ویژگی‌های قابل مشاهده توسط فرد بیمار است. این تحقیق با استفاده از الگوریتم‌های یادگیری ماشین به پیش‌بینی بیمار بودن یا نبودن افراد می‌پردازد. ساختار مقاله حاضر در ادامه به شرح زیر است: بخش دوم به مرور اجمالی الگوریتم‌های یادگیری ماشین به کار گرفته شده در این پژوهش می‌پردازد. بخش سوم مجموعه داده‌های مورد استفاده در این تحقیق را معرفی می‌کند. بخش چهارم نتایج حاصل از پیاده‌سازی الگوریتم‌های یادگیری ماشین برای شناسایی و پیش‌بینی بیماری کرونا را نشان می‌دهد. بخش پنجم نتیجه‌گیری حاصل از این پژوهش را نشان می‌دهد.

۲- الگوریتم‌های یادگیری ماشین

با توجه به توسعه روزافزون داده‌ها و اطلاعات در دنیای امروز و پیچیدگی و زمان‌بر بودن محاسبات داده‌ها، به مرور طی چندین سال گذشته شاهد پیشرفت و تحول در زمینه توسعهٔ تجزیه و تحلیل داده‌ها ومحاسبات با کارایی بالا بوده‌ایم. هوش مصنوعی و یادگیری ماشین نیز بخش مهمی از این توسعه بوده و پیشرفت‌های چشم‌گیری داشته است که حاصل آن ارائه الگوریتم‌های مختلف یادگیری ماشین می‌باشد. این الگوریتم‌ها را می‌توان به سه دسته یادگیری بانظارت، یادگیری بدون نظارت و یادگیری تقویتی تقسیم نمود [۱۷]. با توجه به تحقیقات انجام شده، در این تحقیق، ما از الگوریتم‌های طبقه‌بندی در دسته یادگیری با نظارت استفاده نموده‌ایم تا کارایی سایر الگوریتم‌های یادگیری ماشینی که در تحقیقات بررسی نشده است را بررسی نماییم. از آنجایی که امروزه اغلب پژوهشگران با کارایی این الگوریتم‌ها آشنایی دارند در ادامه مروری کوتاه بر هر یک از الگوریتم‌های به‌کار گرفته شده در این تحقیق داریم.

۲-۱ الگوریتم طبقه‌بندی بیزین ساده[5]

این روش(NB) یک الگوریتم طبقه‌بندی مبتنی بر قضیه بیز با فرض استقلال بین متغیرهای تصادفی است. به زبان ساده، طبقه‌بندی‌کننده بیزین ساده فرض می‌کند که وجود یک ویژگی خاص در یک کلاس یا برچسب مستقل ازحضور هر ویژگی دیگر است [۱۷]. بنابراین الگوریتم بیزین ساده را می‌توان یک مدل برمبنای احتمال شرطی

2- Weka
3- 10-fold cross validation
4- 4 fold cross validation
5- Naive Bayes Algorithm

در نظر گرفت. فرض کنید $X=(x_1,...,x_n)$ برداری از n ویژگی را بیان کند که به صورت متغیرهای مستقل هستند. به این ترتیب می‌توان احتمال رخداد C_k یعنی $p(C_k x_1,...,x_n)$ را به‌عنوان یکی از حالت‌های کلاس رخدادهای مختلف به ازاء kهای متفاوت، به شکل رابطه (۱)نمایش داد [۱۸].

$$p(C_k|X) = \frac{p(C_k)p(X|C_k)}{p(X)} \qquad (۱)$$

۲-۲ الگوریتم رگرسیون لجستیک[۶]

رگرسیون لجستیک(LR) یک مدل ریاضی است که برای مدل‌سازی دودویی استفاده می‌شود و برای رگرسیون منطقی چندین توسعه پیچیده دیگر وجود دارد. دراصل، رگرسیون لجستیک یک مدل رگرسیون است که پیش‌بینی می‌کند داده یا ورودی داده شده به طبقه خاصی تعلق دارد یا خیر؛ در واقع رگرسیون لجستیک احتمال خروجی را پیش‌بینی می‌کند. رگرسیون لجستیک دارای نکات مهم بسیاری از جمله سادگی پیاده‌سازی، کارایی محاسباتی، کارایی مبتنی بر آموزش، سهولت تنظیم است.برای ویژگی‌های ورودی، نیازی به مقیاس بندی ندارد [۱۹].

۲-۳ الگوریتم درخت تصمیم[۷]

درخت تصمیم (DT) الگوریتمی برای یادگیری ماشین بانظارت برای حل مسائل رگرسیونی و طبقه‌بندی با تقسیم مکرر داده‌ها بسته به یک متغیر خاص است. در درخت تصمیم، گره‌های داخلی آزمون‌هایی بر روی ویژگی‌ها یا الگوهای ورودی هستند و برگ‌ها بیانگر کلاس یا ردهٔ مربوط به آن الگو می‌باشند. در این درخت هر گره با توجه به مقادیر ویژگی مربوطه از طریق شاخه‌هایی به گره پایینی متصل می‌شود. هر الگوی ورودی با توجه به مقادیر ویژگی‌هایش مسیری از ریشه درخت تا برگ را طی می‌کند که گره برگ بیانگر کلاس آن الگو است.

درخت تصمیم برای تقریب توابع گسسته به‌کار می‌رود و نسبت به نوفه مقاوم است. از مزایای درخت تصمیم این است که می‌توان الگوی طی شده از ریشه تا برگ را به‌صورت یک قانون اگر-آنگاه نشان داد [۱۸]. هدف درخت تصمیم، ساخت مدلی است که بتواند با یادگیری قوانین تصمیم‌گیری ساده که از رویدادهای آموزش به دست آمده است، متغیرهدف داده تست راپیش‌بینی نماید [۱۹].

۲-۴ تحلیل تشخیص خطی[۸]

تحلیل تشخیص خطی(LDA) روشی آماری می‌باشد که از جمله در یادگیری ماشین و بازشناسی الگو برای پیداکردن ترکیب خطی خصوصیاتی که به بهترین صورت دو یا چند کلاس از اشیاء را از هم جدا می‌کند نیز استفاده می‌شود [۲۰]. تحلیل تشخیص خطی بسیار شبیه به تحلیل واریانس و تحلیل رگرسیون است. در هر سه روش آماری فوق متغیر وابسته به صورت یک ترکیب خطی از متغیرهای دیگر مدل‌سازی می‌شود. با این حال دو روش تحلیل واریانس و رگرسیون متغیر وابسته را از نوع فاصله‌ای در نظر می‌گیرند در حالی که تحلیل تشخیص خطی برای متغیرهای وابسته اسمی یا رتبه‌ای به کار می‌رود. از این رو تحلیل تشخیص خطی به رگرسیون لجستیک شباهت بیشتری دارد [۲۱].

همچنین می‌توان گفت روش تحلیل تشخیص خطی ارتباط نزدیکی با تحلیل واریانس و تحلیل رگرسیون دارد زیرا این روش‌ها نیز سعی دارند یک متغیر مستقل را به‌عنوان ترکیبی خطی از ویژگی‌های دیگر بیان کنند. این متغیر مستقل در LDA به شکل برچسب یک کلاس بیان می‌شود. روش LDA سعی در مدل‌سازی تفاوت بین کلاس‌های مختلف داده‌ها نیز دارد. در واقع از LDA زمانی استفاده می‌شود که مقادیر ویژگی‌ها و مشاهدات، پیوسته باشند [۲۲].

۲-۵ الگوریتم K- نزدیک‌ترین همسایه[۹]

الگوریتم K- نزدیک‌ترین همسایه(KNN) یک روش آماری بدون پارامتر است که برای طبقه‌بندی آماری و

6- Logistic Regression
7- Decision Tree
8- Linear Discriminant Analysis
9- K-Nearest Neighbors

رگرسیون استفاده می‌شود [۲۳]. در هر دوحالت K شامل نزدیک‌ترین مثال آموزشی در فضای داده‌ای می‌باشد و خروجی آن بسته به نوع مورد استفاده در طبقه‌بندی و رگرسیون متغیر است. در حالت طبقه‌بندی با توجه به مقدار مشخص شده برای K، به محاسبه فاصله نقطه‌ای که می‌خواهیم برچسب آن را مشخص کنیم با نزدیک‌ترین نقاط می‌پردازد و با توجه به تعداد رای حداکثری این نقاط همسایه، در رابطه با برچسب نقطه مورد نظر تصمیم‌گیری می‌کند. برای محاسبه این فاصله می‌توان از روش‌های مختلفی استفاده کرد که یکی از مطرح‌ترین این روش‌ها، فاصله اقلیدسی است. در حالت رگرسیون نیز میانگین مقادیر به‌دست آمده از K خروجی آن می‌باشد. از آنجا که محاسبات این الگوریتم بر اساس فاصله است نرمال‌سازی داده‌ها می‌تواند به بهبود عملکرد آن کمک کند [۱۸].

مزایای الگوریتم KNN عبارتند از [۲٤]: تکنیک ساده‌ای است که به راحتی پیاده‌سازی می‌شود.ساخت مدل ارزان است. این الگوریتم یک طرح طبقه‌بندی انعطاف‌پذیر و مناسب برای کلاس‌های چندگانه است. به‌طور معمول میزان خطای شناسایی آن حداکثر دو برابر میزان خطای بیز است.

معایب KNN نیز به شرح زیر است [۲٤]: طبقه‌بندی نمونه‌های ناشناخته نسبتا هزینه‌بر است. نیازمند محاسبه فاصله از نزدیک‌ترین همسایگان است. با افزایش اندازه مجموعه داده آموزشی، بار محاسباتی الگوریتم پیچیده‌ترمی‌شود. نوفه و ویژگی‌های نامرتبط باعث کاهش دقت دسته‌بندی مدل می‌شود.

۲-۶ ماشین بردار پشتیبان[10]

ماشین بردار پشتیبان (SVM) یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کار طبقه‌بندی‌کننده SVM طبقه‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کند خطی را انتخاب کند که حاشیه اطمینان بیشتری داشته باشد. این روش قبل از تقسیم خطی برای این‌که ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را به وسیله تابع هسته یا کرنل به فضای با ابعاد بالاترمی‌برد [۲۵]. تابع هسته مهم‌ترین پارامتر الگوریتم SVM برای تحلیل داده‌ها و تشخیص الگو است.

تابع هسته نقاط جدید را با استفاده از ابرصفحه به فضای ویژگی ترسیم می‌کند. هسته‌های مختلف مورد استفاده در این روش عبارتند از: خطی، گوسی، نمایی، چندجمله‌ای، هسته‌های ترکیبی و غیره. دقت طبقه‌بندی‌کننده SVM بستگی به انتخاب مناسب تابع هسته دارد [۲٦]. در الگوریتم SVM، هر نمونه داده به‌عنوان یک نقطه در فضای n بعدی روی نمودار پراکندگی داده‌ها ترسیم می‌شود (n تعداد ویژگی‌هایی است که یک نمونه داده دارد) و مقدار هر ویژگی مربوط به داده‌ها، یکی از مؤلفه‌های مختصات نقطه روی نمودار را مشخص می‌کند. سپس با ترسیم یک خط راست، داده‌های مختلف و متمایز از یکدیگر را دسته‌بندی می‌کند[۱۸].

۳-مجموعه داده

در این پژوهش برای اثبات کارایی الگوریتم‌های یادگیری ماشین در شناسایی بیماری کرونا از دو مجموعه داده واقعی موجود در پایگاه داده Kaggle استفاده شده است. در ادامه شرح ویژگی‌های هر یک از این مجموعه داده‌ها آورده شده است.

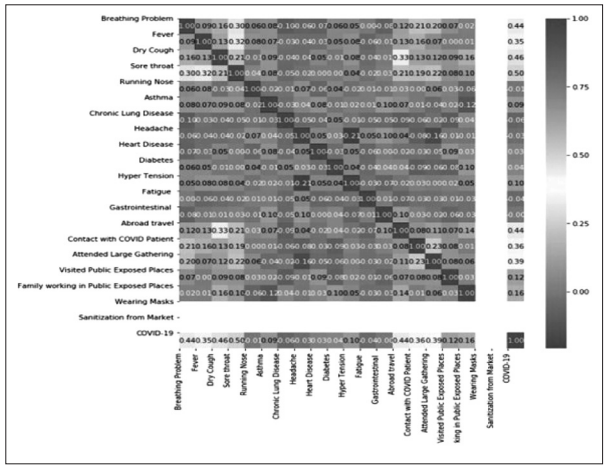
۳-۱ مجموعه داده اول

مجموعه داده اول شامل ۲۱ ویژگی یا علامت بیماری می‌باشدکه ویژگی کووید-۱۹ را به‌عنوان برچسب کلاس در نظر گرفتیم [۲۷]. این مجموعه داده، علائم بیماری را برای ٥٤۳٤ فرد ثبت نموده است که در این بین ۱۰۵۱ نفر سالم و ٤۳۸۳ نفر مبتلا به بیماری کرونا وجود دارد. با توجه به علائمی که افراد بیمار دارند و مقادیری که در ویژگی‌های این مجموعه داده ثبت شده است؛ ما با استفاده از الگوریتم‌های یادگیری ماشین مثبت یا منفی بودن

10- Support Vector Machine

جدول ۱: ویژگی‌های مجموعه داده اول [۲۷]

ردیف	ویژگی	ردیف	ویژگی	ردیف	ویژگی
۱	Breathing Problem	۸	Headache	۱۵	Contact with COVID Patient
۲	Fever	۹	Heart Disease	۱۶	Attended Large Gathering
۳	Dry Cough	۱۰	Diabetes	۱۷	Visited Public Exposed Places
۴	Sore throat	۱۱	Hyper Tension	۱۸	Family working in Public Exposed Places
۵	Running Nose	۱۲	Fatigue	۱۹	Wearing Masks
۶	Asthma	۱۳	Gastrointestinal	۲۰	Sanitization from Market
۷	Chronic Lung Disease	۱۴	Abroad travel	۲۱	COVID-19



شکل۱: نمودار همبستگی ویژگی‌های مجموعه داده اول نسبت به یکدیگر و نسبت به برچسب کلاس(COVID-19)

کرونـــا را در ویژگی کووید-۱۹ تشـــخیص می‌دهیم. تمام ویژگی‌هـــای این مجموعـــه داده در جدول ۱ آورده شـــده است؛ همان‌طورکه مشـــاهده می‌کنید این ویژگی‌ها از نوع طبقه‌بندی دودویی[11] هســـتند که با بله یـــا خیر مقداردهی شـــده‌اند. بله به معنای این است که فرد آن علامت بیماری را داشته است و خیر به معنای این است که فرد آن علامت را در خود مشاهده نکرده است. لازم به ذکر است تحقیقات [۱۲-۹] نیـــز از این مجموعه داده اســـتفاده نمودند که در بخش ۵ نتایج تجربی پژوهش حاضر با نتایج آنان مقایسه شده است.

برای این‌کـــه بتوانیم ویژگی‌های موثر در شناســـایی بیمـــاری را پیدا کنیم باید ویژگی‌های مرتبط با یکدیگر و یا مرتبط با برچسب کلاس را حذف کنیم. برای یافتن ارتباط بین ویژگی‌ها ما از نمودار همبستگی[12] استفاده نمودیم.

در شکل ۱ ملاحظه می‌کنید که همبستگی به دست آمده برای کل ویژگی‌های مجموعه داده، مقادیری بین ۰/۲۱- و ۱ دارد و ویژگی‌ها نسبت به یکدیگر دارای پیوستگی بالایی نیســـتند. همچنین همان‌گونه که مشاهده می‌شود هیچ‌کدام از ویژگی‌ها همبستگی مستقیمی با برچسب کلاس ندارند. بنابرایـــن، مـــا بدون حـــذف ویژگی اقدام به شناســـایی و پیش‌بینی بیماری در مجموعه داده اول نمودیم.

۳-۲ مجموعه داده دوم

مجموعه داده دوم شامل ۲۰ ویژگی است که ویژگی آخر آن نتیجه آزمایش کرونا است و ما این ویژگی را به‌عنوان برچسب کلاس انتخاب نمودیم [۲۸]. در این مجموعه داده نیز از روی علائم بیماری می‌توان نتیجه آزمایش کرونا را پیش‌بینی نمود که ما این کار را با استفاده از الگوریتم‌های یادگیری ماشـــین انجام دادیم. جدول ۲ تمامی ویژگی‌های این مجموعه داده را نشان می‌دهد.

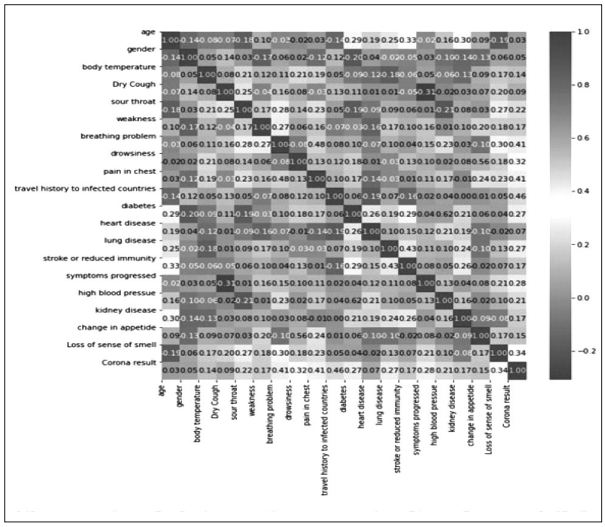
در ایـــن مجموعه داده به غیر از ویژگی ســـن و دمای بدن ســـایر ویژگی‌هـــا به‌صورت دودویی بـــا صفر و یک مقداردهی شـــده‌اند که صفر به معنی نداشتن آن علامت و یک به معنی داشـــتن آن علامت بیماری است. این مجموعه داده اطلاعات ۱۲۷ نفر را ثبت نموده است که در این بین ۳۳ نفر سالم (مقدار صفر)، ۶۹ نفر کرونای خفیف (مقدار یک) و ۲۵ نفر کرونای شدید (مقدار ۲) داشتند. بنابراین پیچیدگی مجموعه داده دوم در این است که سه کلاس داریم. شکل۲

11- Binary Categorical

12- Correlation

جدول ۲: ویژگی‌های مجموعه داده دوم [۲۸]

ردیف	ویژگی	ردیف	ویژگی	ردیف	ویژگی
۱	Age	۸	Drowsiness	۱۵	Symptoms progressed
۲	Gender	۹	Pain in chest	۱۶	High blood pressure
۳	Body temperature	۱۰	Travel history to infected countries	۱۷	Kidney disease
۴	Dry Cough	۱۱	Diabetes	۱۸	Change in appetide
۵	Sour throat	۱۲	Heart disease	۱۹	Loss of sense of smell
۶	Weakness	۱۳	Lung disease	۲۰	Corona result
۷	Breathing problem	۱۴	Stroke or reduced immunit		



شکل۲: همبستگی ویژگی‌های مجموعه داده دوم با یکدیگر و با برچسب کلاس

همبستگی بین ویژگی‌های مجموعه داده دوم را با یکدیگر و با برچسب کلاس یعنی ویژگی نتیجه کرونا[۱۳] نشان می‌دهد.

همان‌طور که در شکل۲ مشاهده می‌کنید ویژگی‌های این مجموعه داده نیز ارتباط زیادی با یکدیگر و با برچسب کلاس ندارند زیرا مقادیر پیوستگی کوچک و قابل صرفنظر کردن است لذا می‌توان از همه آن‌ها برای پیش‌بینی بیماری کرونا استفاده نمود.

۴-پیاده‌سازی و نتایج تجربی

همان‌طور که ذکر شد این پژوهش از آن جهت حائز اهمیت است که امروزه شناسایی این بیماری در مراحل اولیه می‌تواند منجر به پیشگیری از سرایت به دیگران گردد و کنترل بیماری را ساده‌تر کند. در بخش دوم نیز ذکر شد که در این پژوهش از الگوریتم‌های یادگیری ماشین بیزین ساده، رگرسیون لجستیک، درخت تصمیم، تحلیل تشخیص خطی، K- نزدیک‌ترین همسایه و ماشین بردار پشتیبان برای شناسایی و پیش‌بینی بیماری کرونا استفاده شده است. جدول ۳ نتایج حاصل از پیاده‌سازی الگوریتم‌های یادگیری ماشین فوق را بر روی دو مجموعه داده مورد آزمایش نشان می‌دهد.

برای پیاده‌سازی الگوریتم‌های یادگیری ماشین فوق، ما از زبان برنامه‌نویسی پایتون و محیط ژوپیتر نوتبوک از بُن‌سازهٔ آناکوندا بهره گرفتیم. از ۷۰ درصد هر مجموعه داده برای آموزش الگوریتم‌ها و از ۳۰ درصد باقی‌مانده برای آزمایش الگوریتم‌ها استفاده نمودیم. الگوریتم‌ها بر روی سیستمی باپردازشگر Intel(R) Core(TM) i5-2410M CPU 2.30 GHz، ٤ گیگابایت حافظه اصلی[۱۴]، ۲ گیگابایت حافظه گرافیکی[۱۵] اجرا شده‌اند.

برای ارزیابی کارایی الگوریتم‌های یادگیری ماشین و مقایسه آن‌ها با یکدیگر ما از معیارهای ارزیابی دقت[۱۶]، صحت[۱۷]، پوشش[۱۸]، امتیاز-F[۱۹] استفاده نمودیم. همچنین برای این‌که نشان دهیم هر الگوریتم چقدر در پیش‌بینی بیماری کرونا موفق بوده است ماتریس درهم‌ریختگی[۲۰] هر الگوریتم و زمان اجرای آن را نیز نمایش دادیم. برای آشنایی با معیارهای ارزیابی و نحوه محاسبه ماتریس

13- Corona result

14- RAM
15- GPU
16- Accuracy
17- Precision
18- Recall
19- F-score
20- Confusion Matrix

جدول۳: مقایسه ارزیابی الگوریتم‌های یادگیری ماشین

الگوریتم	مجموعه داده	نوع ارزیابی	دقت	صحت	پوشش	امتیاز-F	ماتریس‌درهم‌ریختگی	زمان اجرا (میلی ثانیه)
بیزین ساده (NB)	اول	آموزش	۰/۷۵	۰/۷۲	۰/۸۴	۰/۷۱	[۷۳۳ ۰] [۹۳۰ ۲۱۴۰]	۲۶
		آزمایش	۰/۷۵	۰/۷۲	۰/۸۴	۰/۷۱	[۳۱۸ ۰] [۳۹۳ ۹۲۰]	۲۶
	دوم	آموزش	۱	۱	۱	۱	[۲۱ ۰ ۰] [۰ ۵۰ ۰] [۰ ۰ ۱۷]	۱۷
		آزمایش	۰/۸۹	۰/۹۴	۰/۸۸	۰/۹۰	[۸ ۴ ۰] [۰ ۱۹ ۰] [۰ ۰ ۸]	۲۲
رگرسیون لجستیک (LR)	اول	آموزش	۰/۹۷	۰/۹۶	۰/۹۴	۰/۹۵	[۶۵۶ ۷۷] [۲۸ ۳۰۴۲]	۵۲
		آزمایش	۰/۹۶	۰/۹۵	۰/۹۲	۰/۹۴	[۲۷۵ ۴۳] [۱۴ ۱۲۹۹]	۴۰
	دوم	آموزش	۰/۸۷	۰/۸۹	۰/۸۳	۰/۸۵	[۱۸ ۳ ۰] [۲ ۴۷ ۱] [۰ ۵ ۱۲]	۳۲
		آزمایش	۰/۶۶	۰/۶۷	۰/۶۳	۰/۶۴	[۸ ۴ ۰] [۱ ۱۴ ۴] [۰ ۴ ۴]	۳۸
درخت تصمیم (DT)	اول	آموزش	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۷	[۷۱۸ ۱۵] [۴۶ ۳۰۲۴]	۳۲
		آزمایش	۰/۹۷	۰/۹۵	۰/۹۷	۰/۹۶	[۳۰۶ ۱۲] [۲۶ ۱۲۸۷]	۲۳
	دوم	آموزش	۱	۱	۱	۱	[۲۱ ۰ ۰] [۰ ۵۰ ۰] [۰ ۰ ۱۷]	۱۴
		آزمایش	۱	۱	۱	۱	[۱۲ ۰ ۰] [۰ ۱۹ ۰] [۰ ۰ ۸]	۱۵
تحلیل تشخیص خطی (LDA)	اول	آموزش	۰/۹۵	۰/۹۴	۰/۹۱	۰/۹۲	[۶۲۱ ۱۱۲] [۴۹ ۳۰۲۱]	۶۰
		آزمایش	۰/۹۴	۰/۹۳	۰/۸۹	۰/۹۱	[۲۶۰ ۵۸] [۲۷ ۱۲۸۶]	۴۶
	دوم	آموزش	۰/۸۵	۰/۸۶	۰/۸۰	۰/۸۲	[۱۹ ۲ ۰] [۳ ۴۶ ۱] [۰ ۷ ۱۰]	۲۹
		آزمایش	۰/۵۶	۰/۵۵	۰/۵۰	۰/۵۱	[۷ ۵ ۰] [۴ ۱۳ ۲] [۰ ۶ ۲]	۳۶
K-نزدیک‌ترین همسایه (KNN)	اول	آموزش	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۷	[۷۱۷ ۱۶] [۴۹ ۳۰۲۱]	۵۴۹
		آزمایش	۰/۹۷	۰/۹۵	۰/۹۶	۰/۹۶	[۳۰۴ ۱۴] [۲۳ ۱۲۹۰]	۳۰۰
	دوم	آموزش	۰/۸۴	۰/۸۳	۰/۸۴	۰/۸۳	[۱۸ ۳ ۰] [۶ ۴۲ ۲] [۰ ۳ ۱۴]	۳۶
		آزمایش	۰/۶۹	۰/۶۹	۰/۶۳	۰/۶۵	[۹ ۳ ۰] [۲ ۱۵ ۲] [۰ ۵ ۳]	۳۴

ماشین بردار پشتیبان (SVM)	اول	آموزش	۰/۹۷	۰/۹۷	۰/۹۴	۰/۹۶	[۶۵۸ ۷۵] [۱۳ ۳۰۵۷]	۲۶۸
		آزمایش	۰/۹۶	۰/۹۶	۰/۹۲	۰/۹۴	[۲۷۲ ۴۵] [۸ ۱۳۰۵]	۲۶۰
	دوم	آموزش	۰/۷۹	۰/۹۱	۰/۶۵	۰/۶۶	[۰ ۳ ۱۸] [۰ ۵۰ ۰] [۲ ۱۵ ۰]	۲۸
		آزمایش	۰/۶۹	۰/۸۷	۰/۵۶	۰/۵۷	[۰ ۵ ۷] [۰ ۱۹ ۰] [۱ ۷ ۰]	۲۸

درهم‌ریختگی می‌توانید به مرجع [۱۴] مراجعه نمایید.

همان‌طور که در جدول ۳ مشاهده می‌کنید الگوریتم‌های یادگیری ماشین با دقت مناسبی قادر به پیش‌بینی بیماری کرونا از روی علائم بیماری هستند. اگر بخواهیم الگوریتم‌های یادگیری ماشین جدول ۳ را از نظر کارایی با یکدیگر مقایسه کنیم بدین نتیجه می‌رسیم که برای مجموعه داده اول بیشترین دقت شناسایی بر روی داده‌های آزمایشی مربوط به الگوریتم درخت تصمیم با دقت ۰/۹۸و K-نزدیک‌ترین همسایه با میزان دقت ۰/۹۷ است.

برای مجموعه داده دوم نیز بیشترین دقت شناسایی بر روی داده‌های آزمایشی مربوط به الگوریتم درخت تصمیم با میزان دقت ۱ یا صد در صد است. پس از آن الگوریتم بیزین ساده با دقت ۰/۸۹ بر روی داده‌های آزمایشی مجموعه داده دوم بهترین کارایی را داشته است. همچنین اگر بخواهیم کارایی الگوریتم‌ها را از نظر دقت و زمان اجرای مناسب در نظر بگیریم، الگوریتم درخت تصمیم دارای بهترین کارایی است زیرا در زمان اجرای کمتری قادر به آموزش و پیش‌بینی داده‌ها است.

۵- بحث

با در نظر گرفتن دو مجموعه داده مورد بررسی می‌توان گفت علائم مشترک بیماری کرونا در اغلب بیماران عبارتند از: مشکل تنفسی، تب، گلو درد، سرفه خشک، خستگی. همچنین اغلب بیماران دارای بیماری‌های قبلی همانند دیابت، بیماری قلبی، بیماری ریوی، فشارخون بوده‌اند و در اغلب افراد مبتلا به کرونا، سفر خارج از کشور نیز مشاهده شده است. همان‌طور که نتایج حاصل از پیاده‌سازی نشان می‌دهد الگوریتم‌های یادگیری ماشین قادر به شناسایی بیماری کرونا در مراحل اولیه هستند.

برای درک بهتر و مقایسه کارایی الگوریتم‌ها با یکدیگر، میانگین کارایی الگوریتم‌های یادگیری ماشین به‌کار گرفته شده در این پژوهش با بیشترین دقت به‌دست آمده از سایر تحقیقاتی که از مجموعه داده اول این تحقیق و ویژگی‌های آن جهت شناسایی بیماری کرونا استفاده نموده‌اند، از نظر معیارهای ارزیابی در جدول ٤ آورده شده است. همچنین برای این‌که مقایسه در شرایط مشابه صورت گیرد تنها تحقیقاتی انتخاب شدند که داده‌ها را به دو دسته آموزش و آزمایش تقسیم‌بندی نمودند. نتایج عینا از تحقیقات گرفته شده است و اگر معیار مورد نظر در تحقیقی محاسبه نشده باشد، در جدول٤ با خط تیره نشان داده شده است. نتایج ذکر شده، حاصل میانگین نتیجه هر معیار ارزیابی بر روی داده‌های آموزش و آزمایش است. نتایج این تحقیق با دیگر تحقیقات بر روی مجموعه داده دوم مقایسه نشده است زیرا این مجموعه داده اخیرا ثبت شده است و هنوز تحقیق منتشر شده‌ای در این راستا وجود نداشت.

همان‌طور که در جدول ٤ مشاهده می‌شود سایر تحقیقات نیز با استفاده از درخت تصمیم به دقت خوبی دست یافتند، از طرفی در این پژوهش نشان داده شد که درخت تصمیم بر روی مجموعه داده دوم نیز از دقت ۱۰۰ درصد برخوردار است. این امر بیانگر کارایی مطلوب درخت تصمیم و کسب جایگاه اول در مقایسه بین الگوریتم‌ها است. اما اگر به جدول ٤ دقت کنید برخی تحقیقات از ۸۰٪

جدول۴: مقایسه کارایی الگوریتم‌های یادگیری ماشین به‌کارگرفته شده با بهترین نتایج سایر تحقیقات

تحقیق	نام الگوریتم	درصد داده آموزش و آزمایش	دقت	صحت	پوشش	امتیاز-F
آریستا [۹]	درخت تصمیم	۸۰٪ آموزش، ۲۰٪ آزمایش	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۹
	رگرسیون لجستیک		۰/۹۷	۰/۹۶	۰/۹۵	۰/۹۶
الاسلام [۱۰]	درخت تصمیم	۷۰٪ آموزش، ۳۰٪ آزمایش	۰/۹۸	۰/۹۵	۰/۹۸	–
	رگرسیون لجستیک		۰/۹۷	۰/۹۵	۰/۹۵	–
	XGBoost		۰/۹۸	۰/۹۵	۰/۹۸	–
پژوهش حاضر	بیزین ساده	۷۰٪ آموزش، ۳۰٪ آزمایش	۰/۷۵	۰/۷۲	۰/۸۴	۰/۷۱
	رگرسیون لجستیک		۰/۹۷	۰/۹۶	۰/۹۳	۰/۹۵
	درخت تصمیم		۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۷
	تحلیل تشخیص خطی		۰/۹۵	۰/۹۴	۰/۹۰	۰/۹۲
	K-نزدیک‌ترین همسایه		۰/۹۸	۰/۹۶	۰/۹۷	۰/۹۷
	ماشین بردار پشتیبان		۰/۹۷	۰/۹۷	۰/۹۳	۰/۹۵

داده‌ها برای آموزش و ۲۰٪ برای آزمایش استفاده نمودند که این امر باعث کسب دقت بالاتری می‌شود. از آنجایی که تقسیم داده‌ها به ۷۰٪ آموزش و ۳۰٪ آزمایش مرسوم‌تر و نتایج حاصل از آن پایدارتر و قابل استنادتر است؛ در این پژوهش از نسبت ۷۰ به ۳۰ استفاده شده است.

به‌طور کلی اگر بخواهیم الگوریتم‌های یادگیری ماشین را از نظر کارایی با یکدیگر مقایسه کنیم بدین نتیجه می‌رسیم که برای مجموعه داده اول از نظر دقت شناسایی، الگوریتم DT در رتبه اول، KNN در رتبه دوم، SVM در رتبه سوم، LR در رتبه چهارم، LDA در رتبه پنجم و NB در رتبه ششم قرار دارد.

برای مجموعه داده دوم نیز از نظر دقت شناسایی، الگوریتم DT در رتبه اول، NB در رتبه دوم، KNN در رتبه سوم، SVM در رتبه چهارم، LR در رتبه پنجم و LDA در رتبه ششم قرار دارد.

با توجه به تفاوت در نوع ویژگی‌ها و توزیع داده‌ها و تعداد کلاس‌های موجود در هر مجموعه داده این تفاوت در رتبه‌بندی قابل قبول است؛ بنابراین آنچه حائز اهمیت است رسیدن به قابلیت تشخیص بالای ۹۷٪ توسط برخی از الگوریتم‌های یادگیری ماشین است که در سایر تحقیقات نیز این امر اثبات شده است.

لذا در این پژوهش برای ارزیابی کارایی الگوریتم‌های یادگیری ماشین به‌کار گرفته شده و مقایسه آن‌ها با یکدیگر، ما از آزمون آماری T-test نیز استفاده نمودیم. ما فرضیه‌های موجود جهت مقایسه طبقه‌بندها را به شکل زیر در نظر گرفتیم:

- H0: هر دو الگوریتم یادگیری ماشین A و B روی مجموعه داده کارایی یکسانی دارند.
- H1: هر دو الگوریتم یادگیری ماشین A و B روی مجموعه داده کارایی یکسانی ندارند.

با توجه به مقادیر احتمال p-value ما سطح خطای $\alpha = 0.05$ را برای رد فرضیه صفر در نظر گرفتیم. بنابراین مقدار p-value کمتر از ۰/۰۵ بیانگر رد فرضیه صفر است یعنی کارایی دو الگوریتم یادگیری ماشین یکسان نیست و مقدار p-value بالای ۰/۰۵ بیانگر اثبات فرضیه صفر و یکسان بودن کارایی دو الگوریتم یادگیری ماشین است. جدول ٤ نتایج حاصل از مقایسه الگوریتم‌های یادگیری ماشین بر اساس p(A, B) که بیانگر مقدار p-value بین دو الگوریتم یادگیری ماشین A و B است رابر روی مجموعه داده اول نشان می‌دهد.

جدول ٥ نیز نتایج حاصل از آزمون آماری T-test را بر روی مجموعه داده دوم نشان می‌دهد.

همان‌طور که در جدول‌های ٤ و ٥ مشاهده می‌کنید اغلب الگوریتم‌های یادگیری ماشین کارایی متفاوتی با

جدول۴: مقایسه کارایی الگوریتم‌ها بر اساس p-value

	DT	NB	LDA	LR	KNN	SVM
DT	-	2.08E-281	8.01E-16	7.87E-05	0.83089	0.00509
NB	-	-	9.38E-198	2.12E-242	1.98E-279	3.17E-254
LDA	-	-	-	2.23E-05	3.96E-15	7.46E-08
LR	-	-	-	-	0.00018	0.24468
KNN	-	-	-	-	-	0.00959
SVM	-	-	-	-	-	-

جدول۵: مقایسه کارایی الگوریتم‌ها بر اساس p-value بر روی مجموعه داده دوم

	DT	NB	LDA	LR	KNN	SVM
DT	-	0.044	1.81E-09	2.87E-05	3.42E-08	1.51E-05
NB	-	-	1.06E-06	0.00505	1.45E-05	0.00294
LDA	-	-	-	0.02251	0.54674	0.03578
LR	-	-	-	-	0.09191	0.85267
KNN	-	-	-	-	-	0.13315
SVM	-	-	-	-	-	-

یکدیگر دارند. در این بین الگوریتم ماشین بردار پشتیبان کارایی نسبتا یکسانی با رگرسیون لجستیک داشته است. در مجموعه داده اول کارایی DT با KNN نسبتا یکسان است و تفاوت معناداری بینشان وجود ندارد. در مجموعه داده دوم نیز کارایی LDA با KNN نسبتا یکسان بوده است و کارایی KNN نیز با SVM نسبتا یکسان بوده است. بر این اساس می‌توان گفت درخت تصمیم همانند آنچه از سایر معیارهای ارزیابی برآورد شده بود کارایی کاملا متفاوتی با دیگر الگوریتم‌های یادگیری ماشین دارد و بر اساس جدول‌های ٤ و ٥ این تفاوت معنادار است.

از آنجایی که الگوریتم درخت تصمیم در هر دو مجموعه داده توانسته است به دقت مطلوب و بالایی دست یابد، می‌توان این روش را یکی از مناسب‌ترین روش‌های یادگیری ماشین جهت شناسایی بیماری کرونا معرفی نمود. نتایج تحقیقات مرتبط [۱۲-۹] نیز دلالت بر برتری روش درخت تصمیم دارند. این امر نشان می‌دهد از میان الگوریتم‌های یادگیری ماشین، درخت تصمیم، پرچم‌دار شناسایی بیماری کرونا از روی علائم و ویژگی‌های قابل مشاهده است.

۶-نتیجه‌گیری

با توجه به پیشرفت روزافزون بیماری کرونا این پژوهش سعی در شناسایی و پیش‌بینی بیماری از روی علائم ابتدایی و قابل مشاهده توسط فرد را دارد. در این پژوهش برای پیش‌بینی بیماری از الگوریتم‌های یادگیری ماشین شامل بیزین ساده، رگرسیون لجستیک، درخت تصمیم، تحلیل تشخیص خطی، K- نزدیک‌ترین همسایه و ماشین بردار پشتیبان، بهره گرفته شد. نتایج حاصل از پیاده‌سازی الگوریتم‌ها به زبان پایتون در محیط ژوپیتر نوتبوک نشان داد که الگوریتم درخت تصمیم دارای بهترین عملکرد از نظر میزان دقت شناسایی و زمان اجرا می‌باشد. همچنین نتایج نشان می‌دهد که با توجه به دقت شناسایی بترتیب هر یک از الگوریتم‌های بیزین ساده، K- نزدیک‌ترین همسایه، ماشین بردار پشتیبان، رگرسیون لجستیک وتحلیل تشخیص خطی نیزعملکرد مطلوبی دارند. در ادامه این تحقیق، نگارندگان قصد دارند کارایی الگوریتم‌های یادگیری ماشین را بر روی مجموعه داده به‌دست آمده از بیمارستان‌ها و مراکز درمانی داخل کشور اجرا نمایند و میزان قدرت شناسایی الگوریتم‌ها را با توجه به جهش‌های گوناگون بیماری کرونا ارزیابی نمایند. در این راستا می‌توان یک وب اپلیکیشن برای شناسایی بیماری کرونا طراحی نمود تا افراد با خودآزمایی سلامتشان از شیوع بیشتر این بیماری جلوگیری نمایند.

مراجع

[1] Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang et al, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.", The lancet, Vol. 395, no. 10223, pp. 497-506, 2020.

[2] Ozdemir, Oner, "Coronavirus disease 2019 (COVID-19): diagnosis and management", Erciyes Medical Journal, Vol. 42, no. 3, pp. 242-248, 2020.

[3] Wölfel, Roman, Victor M. Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A. Müller, Daniela Niemeyer et al, "Virological assessment of hospitalized patients with COVID-2019", Nature, Vol. 581, no. 7809, pp. 465-469, 2020.

[4] Hafeez, Abdul, Shmmon Ahmad, Sameera Ali Siddqui, Mumtaz Ahmad, and Shruti Mishra, "A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention", EJMO, Vol. 4, no. 2, pp. 116-125, 2020.

[5] Gholampour, Yousef, Banafsheh Tehranineshat, Hojatolah Najafi, Mojtaba Farjam, Zahra Rahimi, and Mostafa Bijani, "A Study of Demographic Characteristics, Clinical Manifestations, Radiologic and Lab Findings of Patients Hospitalized with COVID-19 in the South of Iran", Journal of Fasa University of Medical Sciences, Vol. 10, no. 3, pp. 2456-2465, 2020.

[6] Narin, Ali, Ceren Kaya, and Ziynet Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks", Pattern Analysis and Applications, pp. 1-14, 2021.

[7] Oukebdane, Mohammed Anis, Samir Ghouali, Emad Kamil Hussein, Mohammed Seghir Guellil, Amina Elbatoul Dinar, Walid Cherifi, Abd Ellah Youcef Taib, and Boualem Merabet, "COVIDz: Deep learning for coronavirus disease detection", In Computational Intelligence Techniques for Combating COVID-19, pp. 355-378. Springer, Cham, 2021.

[8] Chandra, Tej Bahadur, Kesari Verma, Bikesh Kumar Singh, Deepak Jain, and Satyabhuwan Singh Netam, "Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble", Expert systems with applications, Vol. 165, no. 113909, 2021.

[9] Arista, Artika, "Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19", Sinkron: jurnal dan penelitian teknik informatika, Vol. 7, no. 1, pp. 59-65, 2022.

[10] Al-Islam, Ferdib,Ghosh, Mounita, "COV-Doctor: A Machine Learning Based Scheme for Early Identification of COVID-19 in Patients", In Proceedings of the International Conference on Big Data, IoT, and Machine Learning, pp. 39-50. Springer, Singapore, 2022.

[11] Villavicencio, Charlyn Nayve, Julio Jerison Escudero Macrohon, Xavier Alphonse Inbaraj, Jyh-Horng Jeng, and Jer-Guang Hsieh, "COVID-19 Prediction applying supervised machine learning algorithms with comparative analysis using WEKA", Algorithms, Vol. 14, no. 7: 201, 2021.

[12] Yusuf, Rahadian, "Comparing Different Supervised Machine Learning Accuracy on Analyzing COVID-19 Data using ANOVA Test", In 2020 6th International Conference on Interactive Digital Media (ICIDM), pp. 1-6. IEEE, 2020.

[13] Taheri, Sara, "A Review on Coronavirus Disease (COVID-19) and What is Known about it", Depiction of Health 11, no. 1, pp. 87-93, 2020.

[14] Chen, Nanshan, Min Zhou, Xuan Dong, Jieming Qu, Fengyun Gong, Yang Han, Yang Qiu et al, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study", The lancet 395, no. 10223, pp. 507-513, 2020.

[15] Guan, Wei-jie, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou, Jian-xing He, Lei Liu et al, "Clinical characteristics of coronavirus disease 2019 in China", New England journal of medicine, Vol. 382, no. 18, pp. 1708-1720, 2020.

[۱۶] **زارع محمود، استفاده از تکنیکهای یادگیری عمیق برای تشخیص موارد کووید-۱۹ با استفاده از تصاویر اشعه ایکس قفسه سینه، مجله علوم رایانشی، شماره ۲۲، پاییز ۱۴۰۰، (DORL: ۲۰,۱۰۰۱,۱,۲۵۳۸۱۶۱,۱۴۰۰,۶,۳,۴,۶).**

[17] Mahesh, Batta, "Machine Learning Algorithms-A Review", International Journal of Science and Research (IJSR). [Internet] Vol. 9, pp. 381-386, 2020.

[18] Han, Jiawei, Jian Pei, and Micheline Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[19] Ibrahim, Ibrahim, and Adnan Abdulazeez, "The role of machine learning algorithms for diagnosing diseases", Journal of Applied Science and Technology Trends, Vol. 2, no. 01, pp. 10-19, 2021.

[20] Tharwat, Alaa, Tarek Gaber, Abdelhameed Ibrahim, and Aboul Ella Hassanien, "Linear discriminant analysis: A detailed tutorial", AI communications, Vol. 30, no. 2, pp. 169-190, 2017.

[21] Xanthopoulos, Petros, Panos M. Pardalos, and Theodore B. Trafalis, "Linear discriminant analysis", In Robust data mining, pp. 27-33, Springer, New York, NY, 2013.

[22] Ioffe, Sergey, "Probabilistic linear discriminant analysis", In European Conference on Computer Vision, pp. 531-542. Springer, Berlin, Heidelberg, 2006.

[23] Peterson, Leif E, "K-nearest neighbor", Scholarpedia, Vol. 4, no. 2, p. 1883, 2009.

[24] Ray, Susmita, "A quick review of machine learning algorithms", In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), pp. 35-39, IEEE, 2019.

[25] Noble, William S, "What is a support vector machine?", Nature biotechnology, Vol. 24, no. 12, pp. 1565-1567, 2006.

[26] Delen, Dursun, "A comparative analysis of machine learning techniques for student retention management", Decision Support Systems, Vol. 49, no. 4, pp. 498-506, 2010.

[27] Symptoms and COVID Presence, Available on:https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence, 2020.

[28] Covid19 Patient Symptoms,Available on: https://www.kaggle.com/bitsofishan/covid19-patient-symptoms, 2021.