عنوان مقاله

Realizing an Effective COVID-19 Diagnosis System Based on Machine Learning and IOT in Smart Hospital Environment

تحقق یک سیستم تشخیص موثر COVID-19 مبتنی بر یادگیری ماشین و IOT در محیط بیمارستان هوشمند

علیرضا سنگتراش-۳۹۹۱۲۳۴۱۰۵۷۰۱۸

دانشجوی دکترای معماری سیستم های کامپیوتری

تعریف مسئله و هدف اصلی مقاله:

سوال اصلی مطرح شده در مقاله چیست؟

به علت توان بالای ویروس کرونا در پخش شدن و شیوع پیدا کردن، این بیماری به سرعت در جهان در حال گسترش است. با توجه به سرعت بالای شیوع بیماری Covid-19 و افزایش روز افزون تعداد مبتلایان به این بیماری، چالش بسیار بزرگی برای مراکز درمانی و بیمارستانی بوجود آمده است.

با توجه به تعداد بالای مراجعین، شاهد ازدحام جمعیت(بیماران) و صف های طولانی در مراکز تشخیصی هستیم که این مورد هم باعث افزایش تعاملات افراد با هم می شود، هم باعث افزایش بار کاری پزشکان و تکنسین های آزمایشگاهی می شود. واضح است که مسئله ازدحام جمعیتی رابطه مستقیمی با افزایش میزان مرگ و میر افراد دارد.

بنابراین اگر بتوان روشی پیدا کرد که مسئله ازدحام جمعیت در بیمارستان ها کاهش دهد، میتوان ادعا نمود که به جامعه بشری خدمت بسیار بزرگی شده است. بنابراین سوال اصلی مطرح شده در این مقاله پیداکردن روشی برای جلوگیری از ازدحام جمعیت در بیمارستان ها از طریق هوشمندسازی بییمارستان ها می باشد که این عمل با کمک گرفتن از دو بازوی قدرتمند اینترنت اشیا و یادگیری ماشین امکان پذیر می شود.

در این مقاله یک مدل مبتنی بر ML و IoT به عنوان یک سیستم پشتیبانی از تصمیم گیری بالینی پیشنهاد شده است. که میتواند به طور موثری بار کاری را برای پزشکان و تکنسین های آزمایشگاهی کاهش دهد، مسئله ازدحام بیماران را حل کند و میزان مرگ و میر را در طی بیماری همه گیر COVID-19 کاهش دهد.

چه مشکلی باید برطرف شود؟

خدمات ارائه شده در بیمارستانها نمایانگر اهمیت مراقبتهای بهداشتی است. طیف گسترده ای از تخصص ها و خدمات به گروه های مختلف بیماران با نیازهای مختلف ارائه می شود. یکی از مشکلاتی که در بیمارستان ها تجربه می شود ازدحام جمعیت است.

این مشکل زمانی رخ می دهد که تعداد بیماران بیشتر از منابع موجود باشد، از جمله تجهیزات پزشکی، کادر پزشکی یا فضاهای مراقبت از بیمارستان است. وقتی بیمارستان ها شلوغ هستند، زمان انتظار بیماران افزایش می یابد، در حالی که کیفیت خدمات و کارایی ارائه دهندگان خدمات بهداشتی کاهش می یابد. مسئله ازدحام جمعیت یک مسئله جهانی و یک مشکل بهداشت عمومی است که افراد را فارغ از طبقه اجتماعی خود تحت تأثیر قرار می دهد. وقوع این مشکل به عوامل خاصی از جمله ورود و خروج بیماران، زمان انتظار بیماران و کارایی ارائه دهندگان خدمات بهداشتی، اقامتگاه ها و در دسترس بودن منابع نسبت داده می شود. در نتیجه ازدحام بیش از حد، کیفیت خدمات کاهش می یابد، خطر مرگ و میر افزایش می یابد و تبعیض اجتماعی در صف ها

حاکم می شود. با توجه به سرعت انتشار عفونت COVID-19، افزایش شدیدی در تمام حالات فوق الذکر مشاهده شده است. بنابراین، افزایش نرخ COVID-19 یک چالش جدید برای بخش بهداشت همراه با خدمات ارائه شده در هر روز است که انتظار می رود نرخ مرگ و میر زیاد باشد و فرایند تشخیص طولانی مدت شود.

چه ضرورتی برای مطرح شدن مسئله است؟

در طی بیماری همه گیر COVID-19، انتظار می رود تعداد زیادی از بیماران تشخیص، معالجه و کنترل شوند، بنابراین این امر بار سنگینی را بر دوش سازمان های پزشکی می گذارد. چرا که ویروس کرونا با سرعت وحشتناکی در جهان در حال گسترش می باشد و این امر کار را برای مراکز درمانی و پزشکی بسیار دشوار کرده است. به همین دلیل ضرورت یافتن روشی برای حل این چالش بزرگ به شدت احساس میشود.

استفاده از سرویس های خودکار می تواند باعث کاهش بار پزشکان، ازدحام بیش از حد و نرخ مرگ و میر شود. علاوه بر این، در بیمارستان های هوشمند، رویکردهای ML و پارادایم های اینترنت اشیا می توانند به عنوان سیستم های پشتیبانی تصمیم گیری

ضرورت اصلی مطرح شدن این مسئله سرعت بالای گسترش این بیماری و به طبع آن بالا رفتن حضور افراد و بیماران در مراکز درمانی و به وجود آمدن چالش ازدحام جمعیت می باشد.

چه روش هایی قبلا برای این کار انجام شده است؟

در ادبیات تحقیقات قابل توجهی در مورد استفاده از اینترنت اشیا (IoT) برای بهبود خدمات بهداشتی به ویژه در هنگام شیوع COVID-19 وجود دارد. هم چنین پژوهش های متعددی نیز در رابطه با استفاده از یادگیری ماشین در این زمینه وجود دارد. چرا که که بیماری همه گیر کرونا یک چالش جهانی است و همه محققان و پژوهشگران در هر حوزه ای را بر آن داشته است تا با پیدا و معرفی کردن یک روش بتوانند گامی موثر در راه برطرف کردن مشکلات ناشی از بیماری بردارند. در ادامه بعضی از این تحقیقات را با هم بررسی می نماییم.

Singh و همکاران در مورد کاربردهای متفاوت اینترنت اشیا در زمینه بهداشت و درمان صحبت کرده اند و همچنین بیان کرده اند که در حین قرنطینه، یک سیستم نظارت (مانیتورینگ) مناسب مفید است. آن ها سیستمی را معرفی کرده اند که با استفاده از آن میتوان همه بیماران در معرض خطر با استفاده از شبکه مبتنی بر اینترنت به راحتی ردیابی می شوند.

Mohammed و همکاران کاربردهای فناوری اینترنت اشیا را در زمینه پزشکی و بهداشت و درمان توضیح داده و پتانسیل آن را برجسته می کند. آنها هم چنین یک کلاه ایمنی هوشمند نصب شده با سیستم های تصویربرداری حرارتی برای شناسایی افراد آلوده در میان جمعیت ایجاد کرده است.و همچنین به سیستم تشخیص چهره مجهز شده است. در این روش اسکن جمعیت با استفاده از دوربین مادون قرمز انجام میشود و اگر درجه حرارت بالا برای شخصی تشخیص داده شود ، با استفاده از دوربین نوری چهره او را ضبط می کند. همچنین موقعیت فرد آلوده را از طریق GPS فراهم می کند.

Bradley و همکاران درباره ی مفهوم اینترنت اشیا (IoT) که شامل استفاده، پردازش و ذخیره اطلاعات در ابر است صحبت کرده اند. چنین اطلاعاتی قابل دسترسی است و می تواند توسط اشیای هوشمند متصل به ابر از طریق اینترنت به طور مستقل مورد استفاده قرار گیرد.

Dhariwal و Mehta معماری و طرح یک بیمارستان هوشمند مبتنی بر اینترنت اشیا را ارائه می دهند. آنها اعلام می کنند نمایش کاربرد عمومی IoT در سیستم های خدماتی را می توان در بیمارستان های هوشمندی مشاهده کرد که بر اساس فناوری IoT کار می کنند، ساخت چنین بیمارستان هوشمندی از بردارهای سرویس های مختلف کاربردی استفاده می کند. این نوع بیمارستان به عنوان یک نوع جدید بیمارستان در نظر گرفته می شود که عملکردهای مختلفی از جمله تشخیص، درمان، مدیریت و تصمیم گیری را با استفاده از این فناوری ها با هم ترکیب می شوند

JinJun و همکاران در رابطه با بیمارستان های هوشمند صبحت کرده اند و نیز به طور مختصر درباره ی مشکلاتی ادغام اطلاعات و هوشمندسازی در بیمارستانها مواردی را اعلام کرده اند. آن ها بیان کرده اند که بیمارستانهای یکپارچه، توصیف خاص و دقیق و پویای بیمارستانها را منعکس می کنند. هنگامی که یک بیمارستان هوشمند اجرا می شود، ممکن است نیاز به پیاده سازی سیستم برنامه ای مبتنی بر یک محیط دیجیتالی داشته باشد. با داشتن یک بیمارستان هوشمند، افراد می توانند به سرعت و با دقت اطلاعات مربوط به یک سرویس را بدست آورند. بنابراین، این سیستم هوشمند بیمارستان، اطلاعات سازی تشخیص، تصمیم گیری علمی و استاندارد سازی مدیریت را تسهیل می کند. علاوه بر این، با ترکیب سرویس نرم افزاری در سیستم بیمارستان های هوشمند، می توان اطلاعات را در بیمارستان به دست آورد و به اشتراک گذاشت. هدف از این کار افزایش اجرای تشخیص هوشمند، مدیریت هوشمند، خدمات هوشمند و درمان هوشمند است.

Shi و همکاران در رابطه با استفاده از آزمایش های اشعه ایکس و سی تی اسکن به عنوان ابزار اصلی تشخیص، که می توانند شدت COVID-19 را ارزیابی کرده صحبت کرده اند. همچنین بیان کرده اند که می توان از این ابزارها برای نظارت بر موارد اورژانسی بیماران آلوده و پیش بینی میزان پیشرفت COVID-19 استفاده کرد.

Ozturk و همکاران یک سیستم خودکار تشخیص موارد مبتلا به Covid-19 را با استفاده از یک شبکه عصبی و تصاویر ایکس ری معرفی کرده اند. آن ها بیان کرده اند که با افزایش سرعت انتقال COVID-19 در سراسر جهان، بیمارستان ها مملو از بیمارانی می شوند که شرایط سلامتی آنها رو به بهبود یا بدتر شدن است. بنابراین، آزمایش بیمار باید با بهره وری و سرعت بالایی انجام شود تا بیشترین جان ممکن را نجات دهد. فن آوری های هوشمند سرمایه های بزرگی هستند که می توانند در تشخیص و طبقه بندی موثر میزان شدت COVID-19 کمک کنند.

Yu و همکاران درباره استفاده از هوش مصنوعی در مراقبت های بهداشتی و مزایای آن پژوهشی را انجام داده اند. آن ها بیان کرده اند که در زمینه های مختلف، استفاده از هوش مصنوعی به طور گسترده ای مورد استفاده قرار گرفته است زیرا تولید نتایج دقیق تشخیص را تسهیل می کند در حالی که حجم کار سیستم های مراقبت بهداشتی را کاهش می دهد. علاوه بر این، هوش مصنوعی می تواند زمان مورد نیاز برای تصمیم گیری در طول روند تشخیص در هنگام استفاده از روش های مرسوم را کاهش دهد.

Alakus و همکاران تشخیص بیماران مبتلا به COVID-19 بر اساس مدل های ML و یافته های آزمایشگاهی بررسی کرده اند. بیان می کنند که علیرغم مزایای ارائه شده توسط تکنیک های AI از نظر تشخیص و طبقه بندیCOVID-19 ، انتخاب یک روش AI که برای تولید نتایج دقیق مناسب باشد، همچنان یک چالش اساسی است. دشواری مرتبط با انتخاب مناسب ترین روش برای تشخیص و طبقه بندی COVID-19 به دلیل وجود طیف گسترده ای از تکنیک های AI است.

Mohammed و همکاران مجموعه داده آزمایشگاهی مربوط به اطلاعات بیماران مختلفی را بررسی کرده اند. با مدلهای مختلف یادگیری عمیق این مجموعه داده آزمایش شده است و بهترین دقت بدست آمده ۹۲,۳٪ توسط مدل ترکیبی CNN-LSTM است. حتی با این دقت تشخیصی چشمگیر، هنوز مدلهای دقیق ML مورد نیاز است و هنوز هم می توان دقت تشخیص را بهبود بخشید.

روش پیشنهادی ارائه شده چیست؟

در این مقاله مدلی مبتنی بر یادگیری ماشین (ML) و اینترنت اشیا (IoT) برای تشخیص بیماران مبتلا به COVID-19 در بیمارستان های هوشمند پیشنهاد شده است. و نیز در رابطه با نقش مدلهای ML و فناوریهای مربوط به اینترنت اشیا در محیط بیمارستان هوشمند تأکید شده است.

در این مدل پیشنهادی میزان دقت تشخیص (طبقه بندی) بر اساس یافته های آزمایشگاهی از طریق مدل های ML سبک بهبود داده شده است. در این روش سه مدل ML، یعنی Naive Bayes (NB)، Random Forest (RF) و Support Vector Machine (SVM)، بر اساس مجموعه داده های آزمایشگاهی آموزش دیده و تست شدند. سه سناریوی اصلی روش شناسی

تشخیص COVID-19، مانند تشخیص بر اساس مجموعه داده های اصلی و نرمال و آنچه بر اساس انتخاب ویژگی است، ارائه شده است.

مدل پیشنهادی مبتنی بر ML و IoT می باشد که با بهره گیری از ویژگی های خاص این دو فناوری می تواند به عنوان یک سیستم پشتیبانی از تصمیم گیری بالینی مورد استفاده قرار گیرد. علاوه بر این، نتایج می تواند بار کاری را برای پزشکان کاهش دهد، مسئله ازدحام بیماران را حل کند و میزان مرگ و میر را در طی بیماری همه گیر COVID-19 کاهش دهد.

در این مقاله ادبیات موجود در مورد Covid-19 مورد بررسی قرار گرفته است.هم چنین در رابطه با تکنیک های یادگیری ماشین و اینترنت اشیا و نیز استفاده از آن ها برای رسیدن به یک مدل تشخیصی موثر در بیمارستان های هوشمند بحث شده است. و در نهایت مدل پیشنهادی را معرفی کرده اند و آن را براساس مجموعه داده های متفاوت و روش های یادگیری ماشین متفاوت تست نموده اند.

در ادامه بیشتر با این مدل پیشنهادی آشنا میشویم. و در رابطه با ساز وکار آن اطلاعات جامع تری را بیان میکنیم.

توضیح راه حل پیشنهادی مقاله برای حل مسئله:

روش پیاده سازی شده برای حل مسئله مقاله به چه صورت است؟ بصورت گام به گام توضیح داده شود.

در این مقاله مدلی مبتنی بر یادگیری ماشین (ML) و اینترنت اشیا (IoT) برای تشخیص بیماران مبتلا به COVID-19 در بیمارستان های هوشمند ارائه شده است. برای اولین بار است که از سه الگوریتم سبک ML برای پیش بینی COVID-19 بر اساس داده های آزمایشگاهی استفاده می کند. نقش ML و IoT در محیط بیمارستان های هوشمند برجسته می شود تا مشکل ازدحام در طی بیماری همه گیر COVID-19 حل شود.در این مدل، روش های RF ، Bernoulli (NB) و SVM را در نتایج تشخیص COVID-19بر اساس مجموعه داده های اصلی و نرمال شده، و آنهایی که براساس یافته های آزمایشگاهی انتخاب شده با استفاده از تکنیک انتخاب ویژگی brute-force اعتبارسنجی کرده است. و در نهایت دقت مدل تشخیص COVID-19 بهبود یافته است.

روش پیاده سازی در این مقاله به دو فاز کلی تقسیم می شود. فاز اول نقش ML و IOT را در محیط بیمارستان های هوشمند نشان می دهد. فاز دوم مراحل مدل تشخیص COVID-19 را در یک مطالعه موردی هدف توضیح می دهد. هر کدام از این فازها نیز به یک سری زیرمرحله دیگر تقسیم می شوند که گام به گام در ادامه بیان می کنیم.

فاز اول: نقش ML و IOT در محیط بیمارستانهای هوشمند

بیمارستان های هوشمند، که نوع جدیدی از بیمارستان ها هستند، مبتنی بر فناوری ML و IoT هستند که با خط سیر از سیستم های خدمات دهنده ی نرم افزاری مختلف ساخته شده اند. بیمارستان های هوشمند جدید تلقی می شوند زیرا در آن وظایف مختلفی از جمله تشخیص، درمان، مدیریت و تصمیم گیری ادغام شده است. سیر تکاملی ML و IoT در بیمارستان های هوشمند در طی بیماری همه گیر COVID-19 را می توان در سه مرحله توضیح داد.

مرحله اول: در این مرحله تولید و جمع آوری داده ها انجام می شود. بیماران مشکوک با استفاده از دستگاه های مختلف پزشکی مانندPCR ، سی تی اسکن و عکسبرداری اشعه ایکس معاینه می شوند. سپس، داده های معاینه برای تجزیه و تحلیل بیشتر تولید می شوند. برای عکسبرداری با اشعه ایکس یا سی تی اسکن، یک تکنسین آزمایشگاهی این مراحل را از راه دور از طریق اتاق کنترل از طریق پخش مستقیم تصاویر ویدئویی انجام می دهد. به این ترتیب، زمان مورد نیاز برای بررسی و تولید داده های پایه موارد

COVID-19 کاهش می یابد. علاوه بر این، با این روش، تماس و ارتباط کمتری لازم است. به عبارت دیگر، تکنسین های آزمایشگاه برای انجام معاینات نیاز به تردد زیادی ندارند.

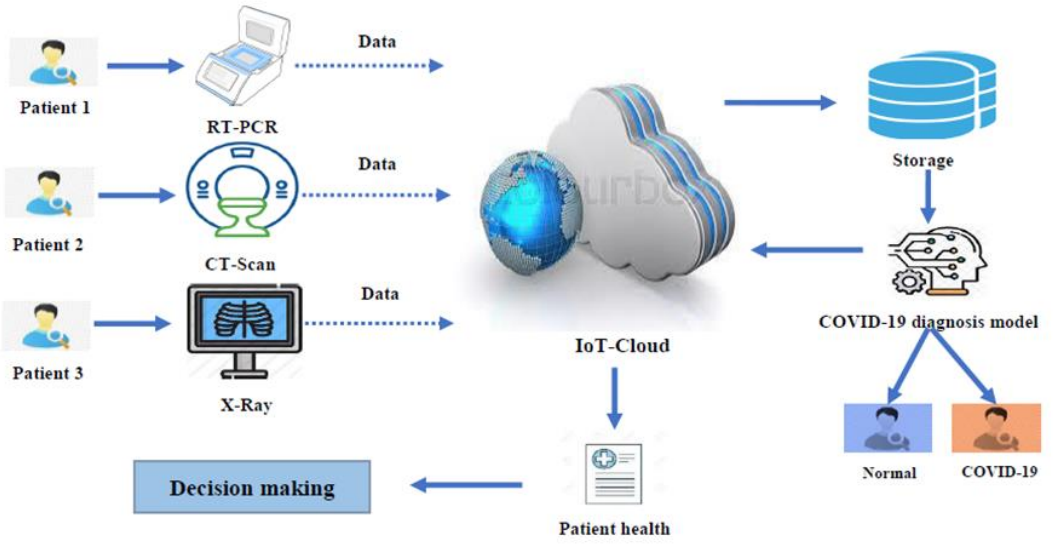
مرحله دوم: بیماران زیادی هر روز در طول بیماری همه گیر COVID-19 بررسی می شوند، بنابراین هر دقیقه حجم عظیمی از داده های بیماران تولید می شود. مشکل ازدحام جمعیت زمانی اتفاق می افتد که تعداد بیمارانی که منتظر بررسی می شوند از تعداد منابع موجود (تجهیزات پزشکی، پزشکان یا ناحیه های مراقبت بیمارستانی) بیشتر شوند. علاوه بر این، تاریخچه داده ها برای هر بیمار COVID-19 برای تجزیه و تحلیل بیشتر ارزشمند است. به دلیل محدودیت ظرفیت منابع در بیمارستان های هوشمند، یک راه حل عملی لازم است.

با این حال، داده های جمع آوری شده از مرحله اول از طریق اینترنت به ابر اینترنت اشیا منتقل می شود، جایی که سرویس ها، تحویل فوری و درخواستی زیرساخت های محاسباتی، پایگاه های داده، فضای ذخیره سازی و نرم افزارهای مورد نیاز برای پردازش و تجزیه و تحلیل داده های بیمار COVID-19 تولید شده از صدها مورد را تسهیل می کند.

در این مرحله ، داده های بیمار برای تجزیه و تحلیل بیشتر ذخیره می شود. سپس، هنگامی که درخواست تشخیص بیماری COVID-19 درخواست می شود، داده ها از ابر ذخیره ساز به یک مدل تشخیص ML منتقل می شوند. با این حال، همانطور که قبلا ذکر شد، بیماران زیادی باید هر روز در طی بیماری همه گیر COVID-19 تشخیص داده شوند. علاوه بر این، ازدحام بیش از حد بیماران، همراه با کمبود پزشک برای تشخیص بیمار، یک چالش بزرگ را ایجاد می کند، بنابراین، یک راه حل عملی لازم است. در این مرحله روش پیشنهادی ما، یک مدل ML به عنوان یک سیستم پشتیبان تصمیم گیری برای تشخیص بیماران مبتلا به COVID-19استفاده شده است. به این ترتیب می توان از مسئله ازدحام جمعیت و حجم کار پزشکان کاسته شود. معمولاً یافته های آزمایشگاهی منتقل شده برای ابر IoT در مدل های ML تغذیه می شود تا در صورت مثبت بودن یا منفی بودن بیمار از نظر COVID-19، تشخیص داده شود. اجرای مدل تشخیص در یک ایستگاه کاری انجام می شود، جایی که مدل ML می تواند به صورت آفلاین آموزش ببیند. از طریق نرم افزار یکپارچه در سیستم، فرآیند آموزش می تواند به طور خودکار با داده های جدید انجام شود تا به روزرسانی ابر و بهبود قابلیت تشخیص را تسهیل کند.

مرحله سوم: پس از بدست آوردن نتایج تشخیص توسط مدل ML، گزارش های بهداشتی بیمار تولید می شود و سپس به یک واحد تصمیم گیری ارسال می شود که در آن اقدامات بعدی که می تواند قرنطینه بیمار یا انتقال بیمار به اورژانس و غیره باشد .

نقش IOT و ML در بیمارستان های هوشمند در طی بیماری همه گیر COVID-19 و نیز ارتباط مراحل فوق با هم را در تصویر زیر مشاهده می کنید.



فاز دوم: ارائه مدل تشخیصCOVID-19

این بخش مدل پیشنهادی تشخیص براساس ML را تشریح می کند، که از سه مرحله اصلی تشکیل شده است.

مرحله اول: سه زیرمرحله در این مرحله انجام می شود.

نخست، مجموعه داده های پایه برای آزمایش از یک مجموعه داده آزمایشگاهی از بیماران مبتلا به COVID-19 در بیمارستان Albert Einstein در سائو پائولو ، برزیل استخراج شده است. نمونه های بیمار برای شناسایی افراد آلوده به COVID-19 در ابتدای سال ۲۰۲۰ جمع آوری شد. مجموعه داده های آزمایشگاهی شامل اطلاعات ۶۰۰ بیمار با ۱۸ یافته آزمایشگاهی است. در این مجموعه داده ، ۵۲۰ نفر هیچ یافته ای نداشتند و ۸۰ نفر بیمار COVID-19 بودند.

دوم، اطلاعات مربوطه ی استخراج شده از داده های خام به دلیل تأثیر مستقیم آن بر عملکرد طبقه بندی، برای طبقه بندی COVID-19 ضروری است. ورودی اصلی برای طبقه بندی موارد COVID-19 مقادیر هر یافته آزمایشگاهی برای مجموعه داده ذکر شده است. در این مرحله، ۱۸ یافته آزمایشگاهی (یعنی گلبول های قرمز خون، هموگلوبین، پلاکت، هماتوکریت، آسپارتات ترانس آمیناز، لنفوسیت ها، مونوسیت ها، سدیم، اوره، بازوفیل ها، کراتینین، سرم گلوکز، آلانین ترانس آمیناز، لکوسیت ها، پتاسیم، ائوزینوفیل ها، پروتئین واکنش پذیر C و نوتروفیل ها) در فرآیندهای استخراج ویژگی گنجانده شده اند.

سوم، فرایند تشخیص داده های بیمار COVID-19 با یک روش طبقه بندی همراه بود. مرحله تشخیص بر اساس مجموعه داده تعریف شده، ویژگی های استخراج شده و اجرای مدل های مختلف ML انجام شد. از پایتون برای تمام کارهای طبقه بندی در طول دوره تحقیق استفاده می شد. خروجی نهایی تشخیص COVID-19 در این زیرمرحله تولید شده است.

مرحله دوم:

در این فاز، تقریباً همان فرآیند ها اعمال شده، به جز اینکه فرایند نرمال سازی داده ها نیز در این زیرمرحله انجام شد. فرآیندهای استخراج و تشخیص ویژگی بر اساس مجموعه داده های نرمال انجام شده است. با این حال، چندین الگوریتم ML تلاش کردند تا روند داده ها را با مقایسه ویژگی های نقاط داده شناسایی کنند. با این وجود، به دلیل تنوع زیاد در مقیاس ویژگی ها، مسئله ای پدیدار شد. نرمال سازی به طور خاص برای ایجاد مقیاس یکسان در هر نقطه داده است زیرا هر ویژگی از اهمیت یکسانی برخوردار است. یکی از روشهای نرمال سازی داده ها که معمولاً مورد استفاده قرار می گیرد، نرمال سازی حداقل-حداکثر است که شامل تبدیل حداقل مقدار هر ویژگی به ۰، حداکثر مقدار به ۱ و هر مقدار دیگر به اعشاری بین ۰ تا ۱ است.

مجموعه داده به دلیل وجود برخی مقادیر منفی و برخی مقادیر مثبت، اختلاف زیادی در مقادیر یک ستون یا بین مقادیر موجود در همه ستون ها داشت. بنابراین، برای حل مسئله تفاوت بین مقادیر، نرمال سازی لازم بود. فرمول نرمال سازی حداقل-حداکثر به شرح زیر است:

$$v' = \frac{v-min}{\max - min} \quad (1)$$

مرحله سوم:

بر اساس مجموعه داده های نرمال شده، سه فرآیند اصلی انجام شده که فقط یک فرایند متفاوت از مراحل قبلی است، یعنی انتخاب ویژگی اعمال شده است. انتخاب ویژگی های آزمایشگاهی بر اساس روش انتخاب ویژگی brute-force انجام شده است، که به طور گسترده تمام ترکیبات داوطلب از ویژگی های ورودی را ارزیابی می کند و سپس مناسب ترین زیرمجموعه را پیدا می کند. ظاهراً، این جستجوی گسترده هزینه محاسبات بالایی را به همراه دارد و خطر نسبی overfitting را در پی دارد. (Overfit شدن به معنای این است که الگوریتم فقط داده‌هایی که در مجموعه آموزشی (train set) یاد گرفته است را می‌تواند به درستی پیش‌بینی کند ولی اگر داده‌ای کمی از مجموعه‌ی آموزشی فاصله داشته باشد، الگوریتمی که Overfit شده باشد، نمی‌تواند به درستی پاسخی برای این داده‌های جدید پیدا کند و آن‌ها را با اشتباه زیادی طبقه‌بندی می کند) بنابراین، روشهای حریصانه باید به عنوان جایگزینی برای انتخاب روش brute-force در نظر گرفته شوند. (مانند determination رو به جلو)

رویکرد ناآگاهانه Brute-force در این روش تمامی حالت های مختلف از ویژگی ها امتحان و بهترین حالت انتخاب می شود. این روش که شامل شمارش نظام مند تمام نامزدهای ممکن برای حل و چک کردن اینکه آیا هر یک از نامزدها قادر به ارضا شرط مسئله هستند یا نه می باشد. در ادامه شبه کد این الگوریتم را مشاهده می فرمایید.

#	شبه کد برای روش انتخاب ویژگی
01	شروع کردن انتخاب ویژگی brute-force
02	تنظیم و تقسیم مجموعه داده به ۸۰ درصد از داده ها برای مجموعه آموزش و ۲۰ درصد برای مجموعه آزمون.
03	فرض کنید I دارای اندازه های مختلف ویژگی(i=(0,1,....,n باشد
04	الف: فرض کنید انتخاب ویژگی I = بهترین اندازه مجموعه I ، با محاسبه به عنوان کمترین میزان
05	تقسیم داده ها در فرآیند آموزش.
06	ب: فرض کنید ویژگی آزمون = انتخاب ویژگی مجموعهi ؛ انتخاب بهترین مجموعه ویژگی به عنوان انتخاب ویژگی i ،که نمره مجموعه آزمون را به حداقل برساند

در نهایت هم پس از انجام مراحل فوق برای یافتن مدل بهینه برای تشخیص بیماران مبتلا بهCOVID-19 ، عملکرد کلیه مدل های تشخیص ML بر اساس یک آزمایش ارزیابی در هر یک از سه فاز اندازه گیری شد. آزمایش ارزیابی بر اساس اندازه گیری های زیر انجام شده است:

$$Accuracy= ((TP+TN))/ ((TP+TN+FP+FN)) \times 100 \quad (2)$$

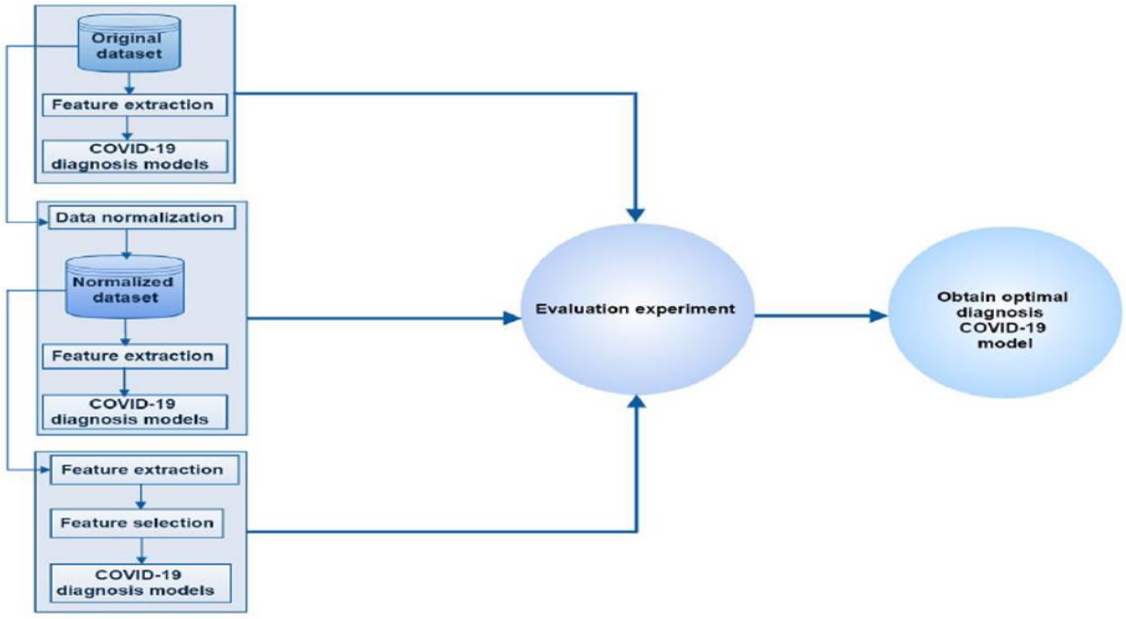
$$Precision=TP/ (TP+FP) \quad (3)$$

$$Recall=TP/ (TP+FN) \quad (4)$$

$$F1= (Precision*Recall)/ (Precision+Recall) \quad (5)$$

که در آن TP مخفف مثبت واقعی، TN نشان دهنده منفی واقعی، FP نشان دهنده مثبت کاذب و FN مخفف موارد منفی کاذب است. بر اساس نتایج آزمایش ارزیابی برای مدلهای مختلف ML، مدل بهینه مطابق با بالاترین مقادیر اندازه گیری در سه فاز انتخاب شده است.

در تصویر زیر طرح کلی و مراحل روش شناختی برای مدل پیشنهادی تشخیص COVID-19را مشاهده می فرمایید. که توضیحات مربوط به آن در بالا گفته شده است.



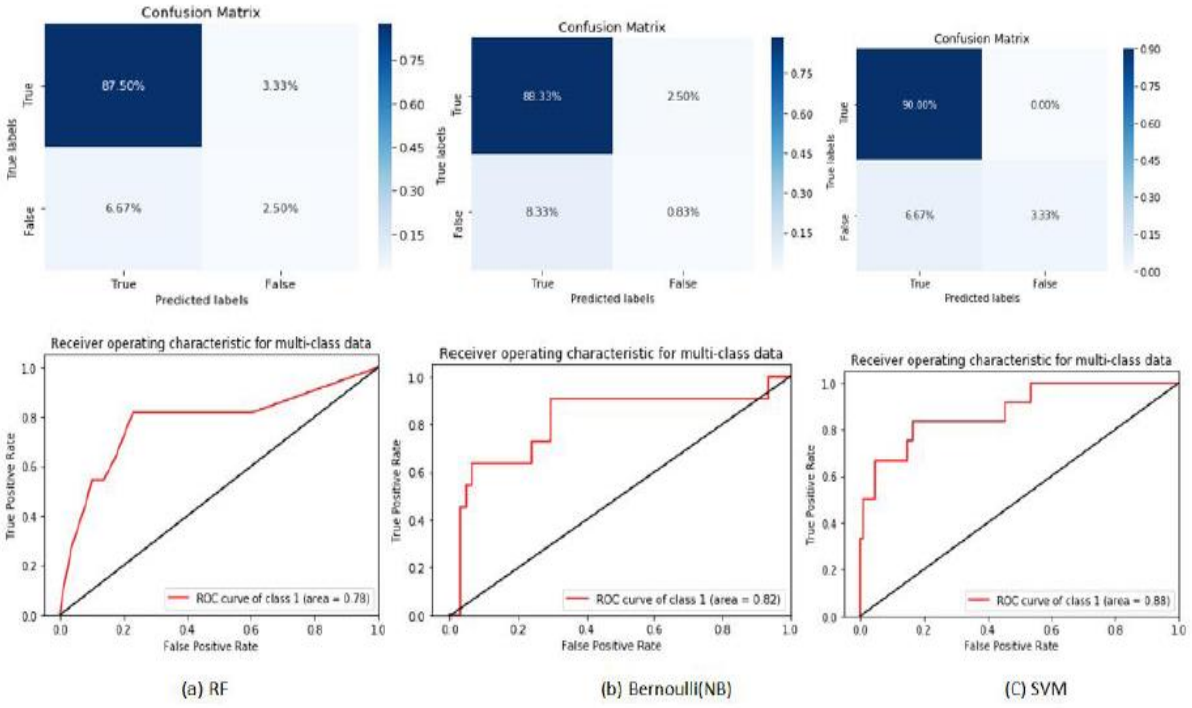
در گام آخر نیز نتایج مدل پیشنهادی در چهار زیر بخش مورد تجزیه و تحلیل قرار گرفته است: نتایج تشخیص COVID-19 بر اساس (A) مجموعه داده های اصلی و (B) نرمال شده، (C) آنهایی که براساس یافته های آزمایشگاهی انتخاب شده و (D) آنهایی که براساس مطالعات معیارند.

A. نتایج تشخیص COVID-19 بر اساس مجموعه داده اصلی

سه روش ML (به عنوان مثال ، RF، SVM و Bernoulli NB) در مجموعه داده اصلی استفاده شد. این روشها از نظر فراخوانی، صحت، دقت، امتیاز F1 و AUC ارزیابی شدند. نتایج ارزیابی روشها در جدول زیر نشان داده شده است.

روش/معیار	F1	صحت	فراخوانی	دقت	AUC
RF	89	88	90	90	0.78
Bernoulli NB	87	85	89	89.16	0.82
SVM	92	94	93	93.33	0.88

در این مطالعه، مشاهده کردیم که بهترین روش های پیش بینی با معیار AUC= 0.88 برای طبقه بندی SVM به عنوان بهترین مدل پیش بینی COVID-19 در عملکرد پیش بینی اندازه گیری می شود. پیش بینی COVID-19 براساس یافته های آزمایشگاهی کار دشواری است و فرایند چالش برانگیزی را پشت سر می گذارد زیرا جمع آوری نمونه به مدت زمان طولانی و یک فرآیند پیچیده نیاز داردطبقه بندی SVM به نتایج خوبی دست می یابد زیرا به طور گسترده ای در پیش بینی بیماری و توالی داده ها مورد استفاده قرار می گیرد، به ویژه هنگامی که داده ها شامل یک سری های زمانی باشند. نتایج ارزیابی براساس ماتریس در هم ریختگی و منحنی ROC برای سه مدل بر اساس مجموعه داده اصلی در شکل زیر نشان داده شده است.



(a) RF (b) Bernoulli(NB) (C) SVM

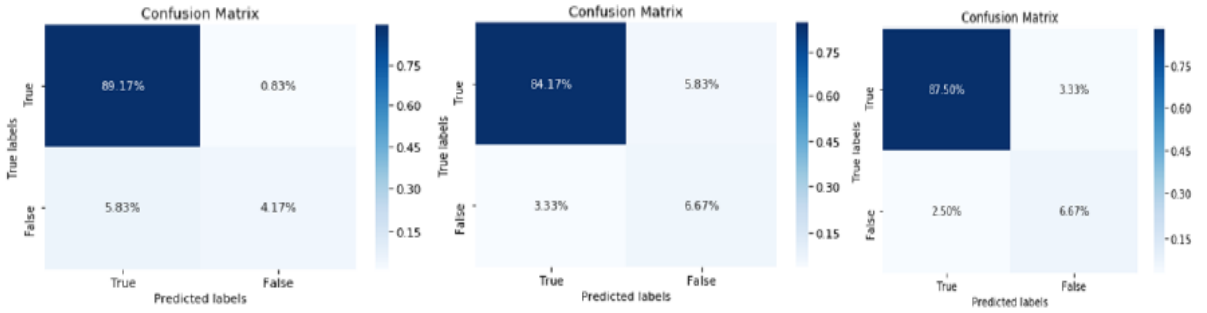
B. نتایج تشخیص COVID-19 بر اساس مجموعه داده های نرمال شده

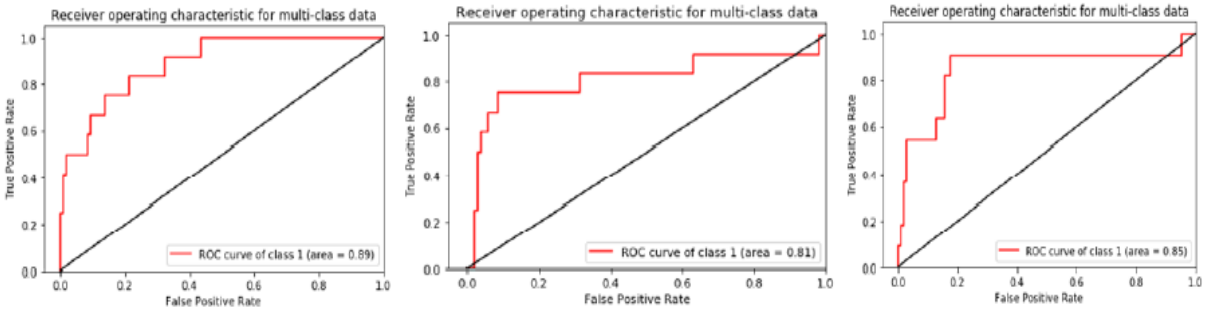
برای دستیابی به تصمیم گیری خوب و تشخیص دقیق، از روش min-max برای نرمال سازی و فیلتر مجموعه داده استفاده کردیم. از تکنیک های ML متعددی برای کشف الگوهای درون داده استفاده شده است. ما آزمایش یکسانی را با استفاده از سه

روش ML(RF، SVM و Bernoulli NB) بر روی مجموعه داده های نرمال شده تکرار کردیم و این روش های ML را از نظر فراخوانی، صحت، دقت، امتیاز F1 و AUC ارزیابی کردیم. نتایج ارزیابی در جدول زیر نشان داده شده است.

روش/معیار	F1	صحت	فراخوانی	دقت	AUC
RF	۹۲	۹۳	۹۳	۹۳,۳۳	0.89
Bernoulli NB	91	92	91	90.83	0.81
SVM	92	94	94	94.16	0.85

با توجه به نتایج حاصله مشاهده می شود که بهترین روش های پیش بینی با معیار AUC= 0.89 برای طبقه بندی RF می باشد.البته طبقه بندی SVM در سایر پارامترها نسبت به RF بهتر عمل کرده است. نتایج ارزیابی براساس ماتریس در هم ریختگی و منحنی ROC برای سه مدل بر اساس مجموعه داده اصلی در شکل زیر نشان داده شده است.





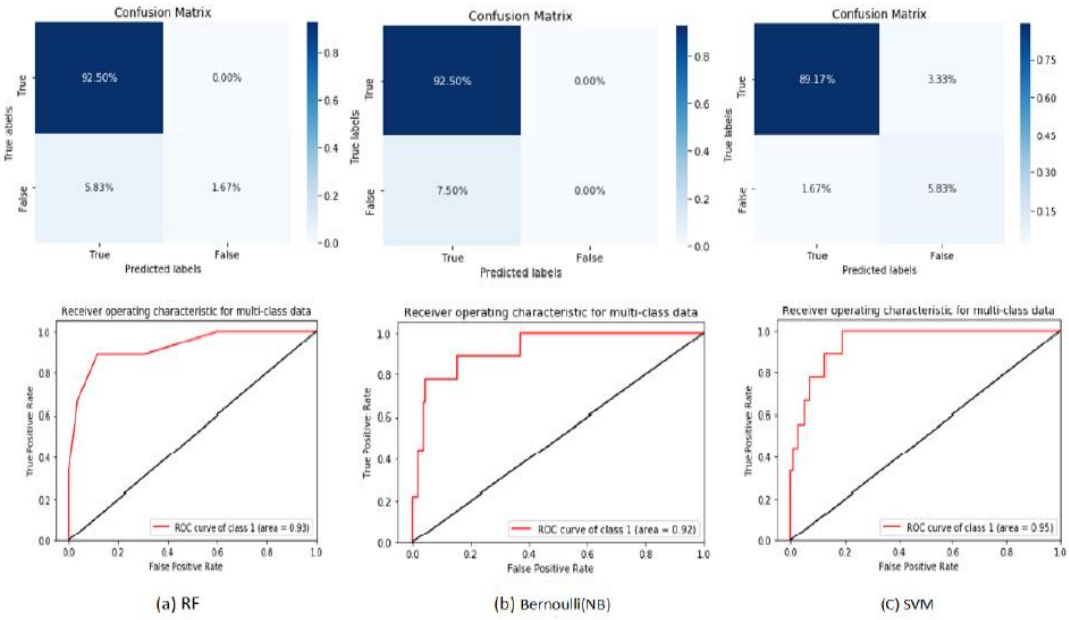
(a) RF (b) Bernoulli(NB) (C) SVM

C. نتایج تشخیص COVID-19 بر اساس یافته های آزمایشگاهی انتخاب شده

ما آزمایش سوم را روی مجموعه داده های نرمال شده تکرار کردیم و از تکنیک انتخاب ویژگی brute-force با سه روش ML استفاده کردیم. برای بهبود دقت و شناسایی بهترین مدل، ما از تکنیک brute-force به عنوان فرآیند انتخاب ویژگی برای کشف مهمترین ویژگیهای آزمایشگاهی استفاده کردیم. انتخاب ویژگی برای روشهای ML مهم است زیرا ویژگیهای ناچیز بر پیش بینی روشهای ML تأثیر می گذارد. انتخاب ویژگی باعث افزایش دقت پیش بینی و کاهش زمان اجرای روش های ML می شود. هدف تکنیک انتخاب ویژگی brute-force آن است که تمام ترکیبات قابل تصور از ویژگی های ورودی را کاملاً تشخیص دهد و پس از آن، عالی ترین زیرمجموعه کشف می شود. نتایج ارزیابی روشهای ML بر روی مجموعه داده های نرمال شده، همراه با انتخاب ویژگی brute-force ، در جدول زیر نشان داده شده است.

روش/معیار	F1	صحت	فراخوانی	دقت	AUC
RF	۹۳	۹۵	۹۴	۹۴,۱۶	0.93
Bernoulli NB	89	86	93	92.5	0.89
SVM	94	95	95	95	0.95

هنگامی که ما تفاوت بین دقت روشهای ML را مقایسه کردیم، اهمیت انتخاب ویژگی در یافته های آزمایشگاهی COVID-19 نشان داده شده است. از ۱۸ ویژگی بالینی، ما ۱۵ ویژگی برتر را انتخاب کردیم و سه ویژگی (یعنی مونوسیت ها، سدیم و ترانس آمیناز آلانین) را حذف کردیم. که تقریباً با ویژگیهای انتخاب شده براساس دیدگاه پزشکی سازگار است. با این حال، بهترین نتایج شناسایی یافته های آزمایشگاهی برای طبقه بندی SVM بدست آمد. نتایج ارزیابی براساس ماتریس در هم ریختگی و منحنی ROC برای سه مدل بر اساس مجموعه داده اصلی در شکل زیر نشان داده شده است.



(a) RF (b) Bernoulli(NB) (C) SVM

D.نتایج تشخیص COVID-19 در برابر مطالعات معیار

محک زنی(Benchmarking) مرحله اساسی است که باید در تحقیقات پردازش و تشخیص پزشکی برای تعیین کارایی و قابلیت اطمینان رویکردهای توسعه یافته استفاده شود. این روش معمولاً با استفاده از یک مجموعه داده استاندارد یا رویکردهایی انجام می شود که برای دامنه یا برنامه مشابه استفاده شده است. علاوه بر این، محک زدن با استفاده از بهترین روش ها برای یافته های آزمایشگاهی COVID-19 بر اساس روش های ML و روش های انتخاب ویژگی موجود در ادبیات حاصل می شود. جدول زیر نتایج عملکرد روش پیشنهادی در مقابل روش هایی برای یافته های آزمایشگاهی COVID-19 مطالعات معیار استفاده شده در کار حاضر را ارائه می دهد.

پژوهش	روش	دقت	F1	صحت	فراخوانی	AUC
Batista و همکاران	SVM,RF	-	0.72	-	-	0.87
Schwab و همکاران	XGB	-	-	-	-	0.66
Alakus و همکاران	CNNLSTM	92.30	0.93	-	-	0.90
کار ما	SVM	95	94	95	95	95

همانطور که در جدول نشان داده شده است، مطالعات مختلفی روی یک مجموعه داده با روشهای مختلف مرسوم ML و یادگیری عمیق مورد مقایسه و استفاده قرار گرفته است. مدل SVM ما نسبت به سایر مطالعات نتایج بهتری در دقت و سایر مقادیر به دست آورد.

روش ارائه شده به چه صورت پیاده سازی شده؟(نرم افزاری یا بصورت اثبات ریاضی دقیقا توضیح داده شود)

روش ارائه شده در این مقاله که یک مدلی مبتنی بر یادگیری ماشین (ML) و اینترنت اشیا (IoT) برای تشخیص بیماران مبتلا به COVID-19 در بیمارستان های هوشمند به صورت گام به گام ارائه شده است. روش پیاده سازی در این مقاله به دو فاز کلی تقسیم می شود. فاز اول نقش ML و IOT را در محیط بیمارستان های هوشمند نشان می دهد. فاز دوم مراحل مدل تشخیص COVID-19 را در یک مطالعه موردی هدف توضیح می دهد. که هر کدام از این فازها نیز به زیر مراحلی تقسیم شده اند که هر کدام به تفضیل توضیح داده شده است. در نهایت هم نتایج مدل پیشنهادی در چهار زیر بخش مورد تجزیه و تحلیل قرار گرفته است: نتایج تشخیص COVID-19 بر اساس (A) مجموعه داده های اصلی و (B) نرمال شده، (C) آنهایی که براساس یافته های آزمایشگاهی انتخاب شده و (D) آنهایی که براساس مطالعات معیارند. و در هر بخش براساس نتایج حاصله بهترین مدل معرفی شده است.

گام به گام تا طرح مدل پیشنهاد شده جلو رفته و در نهایت هم این مدل براساس مجموعه داده های متفاوت مورد ارزیابی قرار گرفته است.

نحوه مقایسه ایده مطرح شده با دیگر ایده های مطرح شده در مقاله

قبل از اینکه مدل پیشنهادی بیان شود به صورت مروری سایر کارهای تحقیقاتی و انجام شده در این زمینه بیان شده است و راجع به هرکدام توضیحاتی ارائه شده است. با توجه به اینکه یک سری مطالعات روی این موضوع صورت گرفته است و یک دیتا بیس در این زمینه جمع آوری شده است. سایر موضوعات مرتبط به دقت بررسی و معرفی شده اند و در باره هر کدام از آن ها یک سری توضیحاتی داده شده است.

بعد از این موارد روش خود مقاله براساس مطالعات انجام گرفته بیان شده است. طرح خود را معرفی کرده و توضیحات پیرامون آن را بیان نموده است.تا به این ترتیب در ذهن خواننده به صورت خودکار یک مقایسه ایجاد شود.

هم چنین در نهایت نتایج مدل پیشنهادی در ۴ زیر بخش براساس مجموعه داده های متفاوت با هم مقایسه شده اند و در هر کدام با توجه به معیارهای دقت،فراخوانی،صحت و امتیاز F1 این مدل ها با هم مقایسه شده اند تا بهترین آن ها در هر زیر بخش معرفی گردد. در انتها نیز با مطالعات معیار یک مقایسه کلی صورت گرفته است.

نقاط قوت و ضعف مقاله

نقاط قوت مقاله شامل موارد زیر میباشد:

این مطالعه برای اولین بار است که مدل ML و IoT را بر اساس یافته های آزمایشگاهی در محیط بیمارستان های هوشمند ارائه می دهد و از این رو یک نوآوری به حساب می آید. هم چنین رویکردهای مختلف ML با هدف بهبود دقت تشخیص در موارد COVID-19 مورد استفاده قرار گرفته است.

مطالب به صورت منسجم بیان شده است تا پله به پله به چارچوب کلی که هدف این مقاله بوده است نزدیک شوند. روی یک موضوع داغ حال حاضر جهان این مطالعه صورت گرفته است و موضوع جدیدی می باشد که میتواند به سایر محققان یک دیدگاه خوبی ارائه کند.

مشخص است که سایر تحقیقات انجام شده در این زمینه که در بخشی از مقاله آورده شده اند به دقت بررسی و تجریه و تحلیل شده اند. تا به یک نتیجه مطلوب برسند.

نقاط ضعف آن شامل موارد زیر است:

جمله بندی ها و توضیحات داده شده در هر بخش از لحاظ ادبیاتی زیاد جالب توجه نیستند. مطالبی به گونه ای بیان شده است که حتما باید پیش زمینه ای در این رابطه داشت وگرنه فهم بعضی از مطالب دشوار میشود. در بعضی جاها که نیاز به یک سری توضیحات تکمیلی بوده است، از بیان آن ها خودداری شده است.

برخی اصطلاحات به کار رفته نیاز به بیان تعاریفی بوده اند که بیان نشده اند. جزئیات مربوط به اینکه هر کدام از موارد انجام شده و بررسی ها و تحلیل ها در چه فضایی انجام شده به صورت کامل و جامع آورده نشده است.

جمع بندی و پیشنهادات برای کارهای آتی:

جمع بندی نهایی بدین شرح است:

نتایج این مطالعه تأیید می کند که (۱) بر اساس نتایج تشخیص مجموعه داده های آزمایشگاهی اصلی COVID-19، SVM در تمام معیارهای ارزیابی از مدل های ML بهتر عمل می کند، (۲) بر اساس نتایج تشخیص مجموعه داده های نرمال شده، SVM در مقایسه با سایر مدل های ML در تمام معیارها، به استثنای AUC که بالاترین مقدار توسط مدل RF ، ۸۹ درصد بدست آمده است،بهتر عمل میکند (۳) بر اساس نتایج رویکرد انتخاب ویژگی، فقط ۱۵ مورد از ۱۸ ویژگی برای تشخیص موارد COVID-19 موثر است، بنابراین بار محاسباتی و زمان تشخیص برای مدل های ML کاهش می یابد. علاوه بر این، SVM در تمام معیارهای ارزیابی از سایر مدل ها بهتر عمل می کند، به جز در صحت که SVM مقدار مساوی (۹۵٪) با RF مشترک دارد. (۴) سرانجام، در مقایسه با سایر مطالعاتی که مجموعه داده های آزمایشگاهی یکسانی را اتخاذ کرده اند، SVM پیشنهادی از نظر دقت تشخیص (تا ۲,۷٪) پیشرفت چشمگیری نشان می دهد.

نتایج این مطالعه فقط در سه مدل ML محدود است. با این حال، پیشنهاد می شود نویسندگانی که قصد دارند در این حوزه فعالیتی داشته باشند، مجموعه داده های آزمایشگاهی، مدل های ML و روش های مختلف انتخاب ویژگی را بررسی کنند.

شبیه سازی

در این مقاله مدل پیشنهادی به صورت گام به گام و براساس دو فاز کلی بیان شده و اطلاعات تکمیلی در رابطه با هر فاز جداگانه بیان شده است .در این مقاله کد نویسی یا شبیه سازی خاصی بیان نشده و توضیحاتی در این رابطه داده نشده است.

بلکه صرفا یک سری توضیحات کلی از روش ها بیان شده است و صرفا در جایی از مقاله بیان شده است که از پایتون استفاده کرده اند. ولی هیچ کد یا برنامه نویسی در بدنه ی مقاله آورده نشده است.