

تذکرات مهم

- ۱) این پروپوزال تنها در صورتی در گروه و شورای پژوهشی قابل طرح است که قبل از طرح، در سایت ثبت پروپوزال واحد که آدرس آن در سایت واحد موجود می باشد به تایید استاد/اساتید راهنما و مشاور رسیده و به ترتیب به گروه تخصصی و شورای پژوهشی دانشکده ارسال شده باشد.
- ۲) تاریخ تصویب پروپوزال در شورای گروه تخصصی با تاریخ تصویب در شورای پژوهشی دانشکده نباید بیش از یکماه باشد.
- ۳) تاریخ تصویب پروپوزال در شورای پژوهشی دانشکده و ارسال پروپوزال به معاونت پژوهش و فناوری واحد نباید بیش از یکماه نباشد.
- ۴) دانشجو ملزم است از طریق سیستم ثبت پروپوزال تمامی مراحل را تا زمان تایید معاون پژوهش و فناوری پیگیری نماید. در صورت عدم پیگیری مسئولیت رد پروپوزال در هر یک از مراحل ارزیابی به عهده دانشجوست و تاخیر در زمان تصویب و دفاع متوجه واحد نخواهد بود.
- ۵) لازم است دانشجو پس از تصویب نهایی، ابلاغیه تصویب پروپوزال را از پژوهش دانشکده تحویل گیرد.
- ۶) با توجه به اینکه پرداخت کمک هزینه های رساله/ پایان نامه ها مربوط به فعالیتهای آزمایشگاهی و کارگاهی منوط به تکمیل جدول بند ۶- استفاده از امکانات آزمایشگاهی واحد و تجهیزات و مواد مورد نیاز - می باشد، لازم است این بخش تکمیل گردد.



دانشگاه آزاد اسلامی

واحد علوم و تحقیقات (تهران)

Science and Research Branch, Islamic Azad University

فرم پیشنهاد تحقیق پایان نامه‌ی کارشناسی ارشد

عنوان تحقیق به فارسی:

ارایه‌ی یک روش بهبود یافته برای پیش‌بینی ضرورت بستری‌شدن بیماران کووید ۱۹
در بخش مراقبت‌های ویژه با استفاده از تکنیک‌های ترکیبی داده‌کاوی

نام دانشجو: مهنام

نام خانوادگی دانشجو: پدرام

دانشکده: مکانیک، برق و کامپیوتر

رشته تحصیلی: ارشد مهندسی کامپیوتر

گروه تخصصی: مهندسی نرم‌افزار

نیمسال ورود به مقطع جاری: دوم ۹۵

گرایش: نرم‌افزار

نیمسال شروع به تحصیل: دوم ۹۵

نام و نام خانوادگی استاد (اساتید) راهنما:

نام و نام خانوادگی استاد (اساتید) مشاور:

خانم دکتر مریم رستگارپور

۱-

۱-

توجه: لطفاً این فرم با مساعدت و هدایت استاد راهنما تکمیل شود.

۱- اطلاعات مربوط به دانشجو:

نام:.....مهنام.....نام خانوادگی:.....پدرام.....شماره دانشجویی:.....۹۵۰۵۱۵۳۸۲.....
مقطع:.....ارشد.....رشته تحصیلی:.....کامپیوتر.....گروه تخصصی:.....نرم افزار.....
گرایش:.....نرم افزار.....نام دانشکده: مکانیک، برق و کامپیوتر..... سال ورود به مقطع جاری:.....۱۳۹۵.....
نیمسال ورودی:.....دوم.....

آدرس پستی در تهران، بزرگراه شیخ فضل الله - شهرک فرهنگیان جدید - بلوک ۱۰ - واحد ۷
تلفن ثابت محل سکونت: ۸۸۲۵۳۶۴۲... تلفن همراه: ۰۹۳۵۲۵۳۶۳۵۹... پست الکترونیک: mahnamp@yahoo.com

آدرس پستی در شهرستان:.....
تلفن ثابت محل سکونت:.....تلفن محل کار:.....دورنگار:.....

۲- اطلاعات مربوط به استاد راهنما:

تذکرات:

- دانشجویان دوره کارشناسی ارشد می توانند حداکثر دو استاد راهنما و یک استاد مشاور انتخاب نمایند.
- در صورتی که اساتید راهنما و مشاور **مدعو** می باشند، لازم است سوابق تحصیلی، آموزشی و پژوهشی کامل ایشان (رزومه کامل) شامل فهرست پایان نامه های کارشناسی ارشد و رساله های دکتری دفاع شده و یا در حال انجام که اساتید مدعو، راهنمایی و یا مشاوره آنرا بر عهده داشته اند، به **همراه مدارک مربوطه** و همچنین آخرین حکم کارگزینی (حکم هیأت علمی) ضمیمه گردد.
- اساتید راهنما و مشاور موظف هستند قبل از پذیرش پروپوزال، به سقف ظرفیت پذیرش خود توجه نموده و در صورت تکمیل بودن ظرفیت پذیرش، از ارسال آن به دانشکده و حوزه پژوهشی و یا در نوبت قرارداد و ایجاد وقفه در کار دانشجویان جداً پرهیز نمایند.

اطلاعات مربوط به استاد راهنمای اول:

نام و نام خانوادگی : مریم رستگارپور..... آخرین مدرک تحصیلی ^{دانشگاهی} _{حوزوی} : دکتری.....
تخصص اصلی: ...هوش مصنوعی..تخصص جنبی : ...پردازش تصاویر و داده کاوی... رتبه دانشگاهی
(مرتبه علمی) :استادیار...

سنوات تدریس ^{کارشناسی ارشد} _{دکتری} : ۱۰ سال..... شغل و سمت فعلی : هیات علمی تمام وقت.....

آدرس محل کار: دانشگاه آزاد اسلامی واحد ساوه

تلفن منزل : تلفن همراه: ۰۹۱۲۷۵۵۷۵۸۸.. محل کار: دورنگار:
پست الکترونیک (Email) : rastgarpour@gmail.com

نحوه همکاری با واحد علوم و تحقیقات :

☒ مدعو

☐ نیمه وقت

☐ تمام وقت

اطلاعات مربوط به استاد راهنمای دوم:

نام و نام خانوادگی:..... آخرین مدرک تحصیلی ^{دانشگاهی} _{حوزوی} :
عضو هیأت علمی دانشگاه

تخصص اصلی:..... رتبه دانشگاهی (مرتبه علمی):..... تلفن همراه:
تلفن منزل یا محل کار:..... نام و نام خانوادگی به زبان انگلیسی:
نحوه همکاری با واحد علوم و تحقیقات:
☐ تمام وقت ☐ نیمه وقت ☐ مدعو

اطلاعات مربوط به استاد مشاور:

نام و نام خانوادگی:..... آخرین مدرک تحصیلی ^{دانشگاهی} _{حوزوی} :
عضو هیأت علمی دانشگاه

تخصص اصلی:..... رتبه دانشگاهی (مرتبه علمی):..... تلفن همراه:
تلفن منزل یا محل کار:..... نام و نام خانوادگی به زبان انگلیسی:
نحوه همکاری با واحد علوم و تحقیقات:
☐ تمام وقت ☐ نیمه وقت ☐ مدعو

الف- عنوان تحقیق

- عنوان به زبان فارسی:

پیش‌بینی ضرورت بستری شدن بیماران کووید ۱۹ در بخش مراقبت‌های ویژه با استفاده از روش‌های یادگیری ماشین

- عنوان به زبان انگلیسی (آلمانی، فرانسه، عربی):

تذکره: صرفاً دانشجویان رشته‌های زبان آلمانی، فرانسه و عربی مجازند عنوان پایان‌نامه خود را به زبان مربوطه در این بخش درج نمایند و برای بقیه دانشجویان، عنوان بایستی به زبان انگلیسی ذکر شود.

Developing an improved ICU admission predictor for Covid-19 patients with ensemble learning

ب- تعداد واحد پایان‌نامه: ۶

ج- بیان مسأله اساسی تحقیق به طور کلی (شامل تشریح مسأله و معرفی آن، بیان جنبه‌های مجهول و مبهم، بیان متغیرهای مربوطه و منظور از تحقیق):

در سال‌های اخیر و به دنبال پیشرفت‌های چشمگیر در زمینه‌ی جمع‌آوری و ذخیره‌ی داده‌گان حیاتی، توسعه‌ی مدل‌های یادگیرنده‌ی بسیار دقیق و بهینه و همچنین طراحی پردازنده‌های دیجیتال قوی، الگوریتم‌های هوش مصنوعی و روش‌های یادگیری ماشین با اقبال گسترده‌ای در کاربردهای تشخیصی پزشکی مواجه شده‌اند. در حالت ایده‌آل، این الگوریتم‌ها که با استفاده از یک مجموعه‌ی داده‌گان بسیار-بُعدی و با تکیه بر تشخیص‌های قبلی پزشک تعلیم می‌یابند، یک سیستم تصمیم‌گیری کامپیوتری را در اختیار پزشک قرار می‌دهند که با دقت و سرعت بالایی قادر به کلاس‌بندی داده‌گان ورودی و تشخیص بیماری است. اما، در کاربردهای دنیای واقعی، طراحی و توسعه‌ی چنین سیستمی بسیار چالش برانگیز بوده و با محدودیت‌هایی از قبیل ناکافی بودن تعداد نمونه‌های تعلیم، دشواری در دسترسی به داده‌گان با کیفیت، جامع و بی‌طرف، و همچنین تعداد بالای متغیرهای ورودی مواجه است. حتی در صورت دستیابی به مجموعه‌ی داده‌گان مناسب، تعلیم یک مدل تصمیم‌گیرنده‌ی دقیق و سریع، نیازمند امکانات پیشرفته‌ی سخت‌افزاری و نرم‌افزاری بوده و در بسیاری از موارد تضمینی نیست که مدل، قدرت تعمیم بالایی در مواجهه با داده‌گان جدید داشته باشد. بنابراین، در بسیاری از تحقیقات منتشر شده در زمینه‌ی کاربرد روش‌های هوش مصنوعی در طبقه‌بندی داده‌گان پزشکی، هدف اصلی معطوف به بهبود کیفیت داده‌گان، بهبود روش‌های استخراج ویژگی‌های مناسب، و نیز افزایش دقت و حساسیت مدل‌های یادگیرنده بوده است.

در تحقیقاتی که با موضوع تشخیص کامپیوتری بیماری کووید ۱۹ و یا تشخیص ضرورت بستری شدن بیماران کووید ۱۹ در بخش مراقبت‌های ویژه صورت می‌گیرد، علاوه بر محدودیت‌های مذکور، این مشکل نیز وجود دارد که هنوز مشخص نیست ثبت کدام یک از علایم و سیگنال‌های حیاتی (به عنوان ویژگی‌های ورودی مدل تصمیم‌گیرنده) نقش مهم‌تری در دقت و حساسیت پیش‌بینی دارند. از آن‌جا که ثبت این داده‌گان، هزینه‌های مالی و زمانی بالایی را به بیمار، آزمایشگاه و سیستم درمانی تحمیل می‌کند، رده‌بندی و کاهش بُعد ویژگی‌های ورودی نیز دارای اهمیت کاربردی بالایی در این تحقیقات است. به عبارت دیگر، با هدف کاهش بُعد مساله، کاهش هزینه‌های محاسباتی و بهبود عملکرد مدل، ابتدا متغیرهای نامرتبط^۱ (زاید) و غیرضروری^۲ با

^۱ irrelevant

روش های خوشه بندی ویژگی های ورودی، شناسایی و حذف می شوند و سپس، مدل یادگیرنده با استفاده از ویژگی های باقی مانده تعلیم می یابد. به علاوه، شایان ذکر است که در اغلب تحقیقاتی که تاکنون در ارتباط با کاربردهای هوش مصنوعی در تشخیص بیماری کووید ۱۹ و یا شدت آن صورت گرفته، از تصاویر CT ریه برای تعلیم و آزمایش مدل استفاده می شود. اگرچه ممکن است ثبت این تصاویر دقت بالایی مدل یادگیرنده را در پی داشته باشد، مراحل ثبت تصویر، نه تنها از نظر زمانی و اقتصادی هزینه هایی را به بیمار تحمیل می کند، بلکه او را در موقعیت ناخوشایندتری نسبت به مراحل جمع آوری نمونه های آزمایش خون قرار می دهد.

با توجه به این مقدمه ی کوتاه، در تحقیق حاضر، از دادگان پزشکی غیرتصویری در تشخیص ضرورت بستری شدن بیماران کووید ۱۹ در بخش مراقبت های ویژه استفاده می شود و سه هدف زیر دنبال می گردد:

۱- پاکسازی دادگان و پر کردن جای خالی دادگان از دست رفته و یا ثبت نشده با روش های مناسب درون یابی.

۲- رده بندی و کاهش بُعد ویژگی های ورودی برای تعیین پیش بینی کننده های بهتر

۳- تعلیم یک مدل یادگیری ماشین که با دقت و حساسیت بالاتری نسبت به مدل های موجود قادر به طبقه بندی دادگان ورودی است، مقاوم به نویز بوده و قابلیت تعمیم بالایی دارد.

دقت اغلب مدل های طبقه بندی کننده وابسته به تعادل تعداد نمونه های هر کلاس در مجموعه ی دادگان است. در بیماری هایی مانند کووید ۱۹ که اولاً مجموعه دادگان زیادی در مورد آنها جمع آوری نشده است و ثانیاً جهش های مکرر ویروس عامل بیماری، موجب پیدایش سویه های متنوع با علائم متفاوت شده است، تعداد نمونه های بیمار به طور قطع، کمتر از تعداد نمونه های جمع آوری شده از گروه کنترل (افراد سالم) خواهد بود که این موضوع، موجب سوگیری مدل به سمت تشخیص نمونه ها به عنوان سالم (یا احتمال کم بستری در بخش مراقبت های ویژه) شده. به علاوه، در مجموعه ی دادگانی که به صورت عمومی منتشر شده است، علائم حیاتی بیماران در برخی بازه های زمانی ثبت نشده اند و این جاهای خالی باید با پیش فرض هایی در مورد ثبات وضعیت بیمار و یا به کمک روش های درون یابی مناسب پر شوند. لذا، برای مواجهه با این مشکل، اولاً علاوه بر معیار «دقت»، معیارهای «حساسیت» تشخیص نمونه های بیمار با احتمال بالای نیاز به بستری در بخش مراقبت های ویژه و FI-score نیز به عنوان معیار ارزیابی عملکرد مدل طبقه بندی کننده مورد استفاده قرار می گیرند. ثانیاً، با استفاده از روش های برون یابی، تعدادی نمونه ی مصنوعی^۳ مرتبط با مشاهدات دارای احتمال بستری بالا تولید می گردد و عملکرد مدل در دو حالت «فقط استفاده از داده های اصلی» و «استفاده از ترکیب داده های اصلی و مصنوعی» مقایسه می شود.

د- اهمیت و ضرورت انجام تحقیق (شامل اختلاف نظرها و خلاءهای تحقیقاتی موجود، میزان نیاز به موضوع، فواید احتمالی نظری و عملی آن و همچنین مواد، روش و یا فرآیند تحقیقی احتمالاً جدیدی که در این تحقیق مورد استفاده قرار می گیرد:

با توجه به اطلاعاتی که تاکنون از نحوه ی انتشار ویروس SARS-CoV-2 جمع آوری شده است، امیدواری درباره ریشه کنی جهانی این ویروس در کوتاه مدت و یا دستیابی به ایمنی جمعی (در مقیاس محلی) با تردید روبروست. دو سال پس از آغاز همه گیری، پس از بیش از ۲۶۰ میلیون مورد بیماری، بیش از ۵ میلیون مرگ و اثرات منفی اقتصادی، اجتماعی و روانی، در تاریخ ۴ آذرماه ۱۴۰۰ سازمان بهداشت جهانی اولین مورد از سویه اومیکرون^۴ را تایید کرد. اطلاعات و تحلیل اطلاعات درباره این

² redundant

³ synthetic

⁴ omicron

سویه‌ی جدید در حال تکمیل است، با این حال به نظر می‌رسد که سویه‌ی جدید سرعت انتشار بیشتری از سویه‌های قبلی دارد و راحت‌تر از سد دفاعی بدن عبور می‌کند [۱].

هر چند برخی شرکت‌های داروسازی به تازگی تولید نسخه‌ی به‌روز شده واکسن کووید را برای مقابله با سویه‌ی اومیکرون آغاز کرده‌اند [۲]، شیوع سریع و گسترده‌ی ویروس که در بسیاری موارد نیز بدون علامت‌های معمول رخ می‌دهد، ریشه‌کنی آن را دشوار نموده و تحقق ایده‌ی ایمنی جمعی تنها با فرض واکسیناسیون تمام جمعیت جهان ممکن خواهد بود [۳]. بنابراین، جان بسیاری از مردم، به ویژه سالمندان، همچنان در خطر بوده و یافتن راه‌های مؤثر برای تشخیص زودهنگام این بیماری در افراد، از اولویت‌های تحقیقاتی سال‌های اخیر به شمار می‌رود. از منظر هزینه‌های مراقبتی و برنامه‌ریزی‌های کلان بیمارستانی، پیش‌بینی نیاز افراد به بستری شدن در بخش مراقبت‌های ویژه‌ی بیمارستان نیز حائز اهمیت بوده و توجه محققان را به خود معطوف نموده است. لذا، در این تحقیق پیشنهاد می‌شود که برای تشخیص سریع این که کدامیک از مبتلایان ممکن است دچار علائم مراحل حاد این بیماری و نیازمند به مراقبت‌های ویژه شوند روش‌های مبتنی بر یادگیری ماشین به کار گرفته شوند.

ه- مرور ادبیات و سوابق مربوطه (بیان مختصر پیشینه تحقیقات انجام شده در داخل و خارج کشور پیرامون موضوع تحقیق و نتایج آنها و مرور ادبیات و چارچوب نظری تحقیق):

تحقیقاتی که تاکنون در زمینه‌ی تشخیص بیماری کووید ۱۹ و یا پیش‌بینی احتمال بستری بیمار در بخش مراقبت‌های ویژه انجام شده اند نشان می‌دهند با دسترسی به نتایج معاینات بالینی افراد مبتلا و به کارگیری الگوریتم‌های یادگیری ماشین می‌توان با دقت مناسبی، آینده‌ی وضعیت بیمار از نظر بهبودی یا تشدید علائم بیماری (به ویژه ناهنجاری‌های شدید ریوی) را پیش‌بینی نمود. روش‌های ارائه شده در اغلب این پژوهش‌ها مشتمل بر دو بخش استخراج/انتخاب ویژگی‌ها و تعلیم مدل یادگیرنده هستند. مرحله‌ی انتخاب ویژگی که با هدف کاهش بُعد مسئله صورت می‌گیرد، منجر به کاهش پیچیدگی‌های محاسباتی (از نظر مدت زمان تعلیم، تعداد پارامترهای مدل و ذخیره‌ی مدل تعلیم یافته) شده و امکان تحلیل و شناخت بهتر متغیرهای مؤثر را در اختیار قرار می‌دهد. در اغلب مطالعاتی که بر روی دادگان جمع‌آوری شده از مبتلایان به کووید ۱۹ انجام شده، از یکی از روش‌های فیلتر یا wrapper استفاده شده است. در روش فیلتر که اغلب به صورت مهندسی ویژگی‌های موجود پیاده‌سازی می‌شود، مشخصات آماری ویژگی‌ها و یا همبستگی آن‌ها با خروجی مطلوب مدل مبنای حذف یا حفظ متغیرهای موجود قرار می‌گیرند. واریانس متغیرها، آنتروپی، ضریب Gini، بهره‌ی اطلاعات^۵، میزان همبستگی^۶ متغیرها و شاخص Chi-Square از جمله شاخص‌هایی هستند که برای انتخاب ویژگی‌های مناسب در تحلیل دادگان پزشکی مربوط به بیماری کووید ۱۹ مورد استفاده قرار گرفته‌اند [۴-۶].

در یکی دیگر از این پژوهش‌ها، روش مهندسی ویژگی‌ها که نیازمند دانش مقدماتی نسبت به مجموعه‌ی دادگان است، با روشی موسوم به wrapper جایگزین شده که در آن، مدل طبقه‌بندی‌کننده روی زیرمجموعه‌های کوچکی از مجموعه‌ی دادگان اصلی تعلیم یافته و عملکرد آن ارزیابی می‌شود. سپس، زیر مجموعه‌ای که بهترین عملکرد را به دنبال داشته است به عنوان بردار ویژگی برای تعلیم مدل اصلی مورد استفاده قرار می‌گیرد [۷]. بر روی دادگان غیر پزشکی و در هنگام استفاده از مدل‌هایی مانند درخت تصمیم و Naive Bayes، نشان داده شده که این روش در مقایسه با روش فیلتر برای انتخاب ویژگی‌ها، منجر به تعلیم مدل یادگیرنده (در مسئله‌ی طبقه‌بندی دادگان ورودی) با دقتی بالاتر می‌شود [۸]. در یکی از پژوهش‌های مربوط به تشخیص کامپیوتری کووید ۱۹، از این روش برای انتخاب ویژگی‌های مناسب برای تعلیم یک مدل طبقه‌بندی‌کننده‌ی Extreme Gradient

^۵ information gain

^۶ correlation

Boosting (XGBoost) استفاده شده است که با تعلیم تکراری مدل مشابه و با استفاده از بردارهای ویژگی وزن دار، تعداد ویژگی‌ها ابتدا از ۱۶۵ به ۵۰ و سپس به ۲۰ ویژگی کاهش یافته است [۷].

در ادبیات تحقیق یادگیری ماشین و کاربردهای آن در حل مسائل مهندسی، علاوه بر روش‌های فیلتر و wrapper از الگوریتم ژنتیک، روش‌های جستجوی مستقیم و معکوس ترتیبی، جستجوی مستقیم و معکوس ترتیبی تعمیم یافته [۹] و ترکیب الگوریتم‌های خوشه بندی و رده بندی ویژگی‌ها [۱۰] نیز برای انتخاب بهترین ویژگی‌ها استفاده می‌شود که در تحقیقات مربوط به تشخیص بیماری و یا ارزیابی شدت کووید ۱۹ به کار گرفته نشده‌اند. به علاوه، روش‌های استخراج ویژگی مانند Matrix Factorization، یادگیری منیفلد^۷ الگوریتم‌های مبتنی بر Autoencoder نیز هنوز راهی به این تحقیقات نیافته‌اند و بررسی تأثیر آنها بر عملکرد مدل، به صورت بالقوه، موضوع مناسبی برای تحقیق است.

در بخش تعلیم مدل یادگیرنده در تحقیقاتی که به منظور تعیین احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه انجام گرفته اند، الگوریتم‌های مبتنی بر درخت تصمیم، رگرسیون لجستیک و Random Forest با استفاده از ویژگی‌های بالینی و نتایج آزمایش‌های خون بیماران تعلیم داده شده‌اند [۵-۷]. در یکی از این پژوهش‌ها و با استفاده از داده‌های جمع‌آوری شده در دو بیمارستان در چین گزارش شده است که با استخراج ویژگی‌هایی مانند میزان آنزیم آلانین آمینوترانسفراز در کبد، درصد درد ماهیچه و میزان هموگلوبین خون می‌توان با دقتی حدود ۷۰٪، وقوع موارد حاد را پیش‌بینی نمود [۴]. در تحقیق دیگری که با هدف پیش‌بینی وقوع حالت‌های حاد در افراد مبتلا به کووید ۱۹ با استفاده از الگوریتم XGBoost انجام شده است، ویژگی‌هایی مانند سن، وجود آسیب‌ها و بیماری‌های کلیوی، افزایش LDH، تند نفسی و هیپرگلیسمی به عنوان ویژگی‌های اصلی پیش‌بینی کننده معرفی شده‌اند [۱۱]. نتایجی که در تحقیق [۱۲] منتشر شده نیز نشان می‌دهد که با تعلیم یک مدل XGBoost و این بار با استفاده از یک مجموعه داده‌ها مربوط به ۳۷۵ بیمار (۲۰۱ بهبود یافته) در بیمارستان دانشگاه تونگجی در ووهان، احتمال مرگ ناشی از این بیماری با دقت ۹۳٪ پیش‌بینی شده است. در این پژوهش نیز ویژگی‌هایی مانند LDH، لنفوسیت، و پروتئین واکنشی سی به عنوان ویژگی‌های مهم برای پیش‌بینی معرفی شده‌اند. در تحقیقات مشابهی، همین الگوریتم با استفاده از ویژگی‌های دیگری مانند نرخ تنفس، ضربان قلب، شاخص توده بدنی، میزان نیتروژن و کراتینین خون، بر روی داده‌ها جمع‌آوری شده در بیمارستان‌های مختلف تعلیم داده شده و پیش‌بینی احتمال مرگ با دقت‌های بالاتر از ۹۰٪ گزارش شده است [۱۵-۱۳].

علاوه بر مدل‌های مبتنی بر XGBoost، از مدل ماشین بردار پشتیبان و رگرسیون لجستیک نیز برای پیش‌بینی حالت‌های حاد و نیز احتمال مرگ مبتلایان استفاده شده است [۱۹-۱۶]. در این پژوهش‌ها ویژگی‌هایی از بیومارکرهای سرمی (به عنوان مثال، کلسیم، اسید لاکتیک و آلبومین، گلوکاتایون، لنفوسیت‌های T بالغ و پروتئین تام) برای تعلیم و یا تخمین پارامترهای مدل استفاده شده است.

در تشخیص بیماری کووید ۱۹ با استفاده از داده‌ها مربوط به نتایج آزمایش خون ۲۷۹ مراجعه کننده (۱۷۷ مبتلا)، در مقایسه با الگوریتم‌های مختلفی مانند درخت تصمیم، k-نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، naive Bayes و رگرسیون لجستیک، مدل Random Forest با دقت ۸۲٪، حساسیتی برابر با ۹۲٪ و تشخیص ۶۵٪ بهترین عملکرد را نشان داده است [۲۰]. نتایج این

⁷ manifold learning

تحقیق همچنین نشان می‌دهد که این روش نسبت به نقص دادگان و یا عدم تعادل در تعداد نمونه‌های مشاهده شده در هر کلاس نیز حساسیت کمتری دارد.

و- جنبه جدید بودن و نوآوری در تحقیق:

- برای مدیریت جای خالی دادگان از دست رفته و یا ثبت نشده^۸ از الگوریتم KNNImputer و برای حذف داده‌های پرت از الگوریتم (iForest) isolation forest استفاده می‌شود.
- از آن‌جا که دادگان در مواردی با نرخ‌های نمونه‌برداری متفاوت جمع‌آوری شده‌اند، یکسان‌سازی طول بردارهای ویژگی از طریق نمونه‌برداری و یا نمونه‌افزایی ضروری به نظر می‌رسد. بدین منظور در این پژوهش برای اولین بار از روش synthetic minority over-sampling technique (SMOTE) برای مجموعه دادگان بیماران کووید استفاده خواهد شد.
- برای استخراج ویژگی‌های مرتبط، برخلاف استفاده از روش XGBoost در پژوهش‌های گذشته، از روش‌های جایگزین مانند الگوریتم ژنتیک، روش‌های جستجوی ترتیبی یا ترکیب الگوریتم‌های خوشه‌بندی و رده‌بندی ویژگی‌ها برای انتخاب بهترین ویژگی‌ها استفاده می‌شود.
- در این تحقیق، پیش‌بینی احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه از طریق تعلیم یک مدل ترکیبی داده‌کاوی^۹ دنبال خواهد شد. مدل پیشنهادی ابتدا از چهار نوع طبقه‌بندی‌کننده شناخته شده‌ی random forest، extra trees، و رگرسیون لجستیک و ماشین بردار پشتیبان که هر کدام معماری و مشخصات یادگیری مخصوص به خود را دارند ساخته می‌شود. در مرحله بعدی و برای بالا بردن کارایی، روش‌های یادگیری ترکیبی (بگینگ و/یا بوستینگ) وارد عمل شده و یک مدل پیش‌بینی‌کننده تعلیم می‌یابد.

ز- اهداف مشخص تحقیق (شامل اهداف آرمانی، کلی، اهداف ویژه و کاربردی):

- بهبود روش‌های پیش‌پردازش دادگان به منظور پر کردن جاهای خالی در مجموعه دادگان و همچنین متعادل کردن تعداد نمونه‌های دو کلاس با استفاده از روش‌های نمونه‌افزایی.
- تعیین متغیرهای تأثیرگذار (از میان نتایج آزمایش خون، علائم حیاتی و میزان گازهای موجود در خون شریانی) در تشخیص، به منظور کاهش هزینه‌های ثبت آزمایشگاهی دادگان.
- بهبود دقت تشخیص احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه از طریق تعلیم یک مدل ترکیبی داده‌کاوی بر روی بردار ویژگی‌های رده‌بندی شده.

ح- در صورت داشتن هدف کاربردی، نام بهره‌وران (سازمان‌ها، صنایع و یا گروه ذینفعان) ذکر شود (به عبارت دیگر محل اجرای مطالعه موردی):

- آزمایشگاه‌های ثبت دادگان پزشکی
- مراکز بهداشتی و بیمارستانی

ط- سؤالات تحقیق:

- اعمال روش‌های پیش‌پردازش دادگان و متعادل نمودن تعداد نمونه‌های دو کلاس با استفاده از روش‌های نمونه‌افزایی چه تأثیری بر دقت عملکرد الگوریتم طبقه‌بندی دارد؟

^۸ missing data

^۹ ensemble learning

- از میان متغیرهای به دست آمده از ثبت نتایج آزمایش خون، علائم حیاتی و میزان اکسیژن موجود در خون شریانی، کدام بردار ویژگی بیشترین قابلیت پیش‌بینی‌کنندگی را دارد؟
- تعلیم یک مدل ترکیبی داده‌کاوی بر روی بردار ویژگی‌های رده‌بندی شده با چه دقت، حساسیت و تشخیصی قادر به پیش‌بینی احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه در مبتلایان به کووید ۱۹ است؟

ی- فرضیه‌های تحقیق:

- با دسترسی به معاینات بالینی و آزمایش خون بیماران و به کارگیری الگوریتم‌های یادگیری ماشین می‌توان با دقت بالایی، آینده‌ی وضعیت بیمار مبتلا به کووید ۱۹ از نظر بهبودی یا تشدید علایم بیماری (به ویژه ناهنجاری‌های شدید ریوی) را پیش‌بینی نمود.
- مدل ترکیبی داده‌کاوی با دقت بالاتری بر روی بردار ویژگی‌های رده‌بندی شده تعلیم می‌یابد.

ک- تعریف واژه‌ها و اصطلاحات فنی و تخصصی (به صورت مفهومی و عملیاتی):

- معاینات بالینی؛ دانش درک و شناخت علائم پزشکی و نشانه‌های بیماری که با حس‌های پنجگانه معمولی و بدون نیاز به تجهیزات ویژه پزشکی یا به عبارتی بر بالین بیمار نیز، قابل شناخت و درک بوده و در نخستین مراجعه بیمار و به صورت سریایی توسط پزشک یا پرستار تشخیص و مدرک می‌شوند.
- الگوریتم Random Forest: یکی از روش‌های طبقه‌بندی و رگرسیون است که از ترکیبی از درختان تصمیم‌گیری ساخته می‌شود. همه درختان تصمیم‌گیری تحت نوع خاصی از تصادفی‌سازی در طول فرآیند یادگیری رشد می‌کنند و برای یک طبقه‌بندی، هر درخت می‌تواند تصمیم بگیرد و کلاسی که بیشترین رای را داشته باشد، طبقه‌بندی نهایی را تعیین می‌کند.
- روش‌های ترکیبی داده‌کاوی: روش‌هایی برای تجمع مجموعه محدودی از الگوریتم‌های یادگیری مختلف به منظور حصول نتایج بهتر نسبت به هر یک از الگوریتم‌های یادگیری موجود در مجموعه.
- حساسیت؛ یکی از شاخص‌های دقت آزمایش. حساسیت، حاصل تقسیم موارد مثبت واقعی به حاصل جمع موارد مثبت واقعی و موارد منفی کاذب است.
- تشخیص؛ یکی از شاخص‌های دقت آزمایش. تشخیص، حاصل تقسیم موارد منفی واقعی به حاصل جمع موارد منفی واقعی و مثبت کاذب است.
- امتیاز اف ۱؛ یکی از شاخص‌های دقت آزمایش. نوعی میانگین پارامترهای دقت و بازیابی در یک پیش‌بینی است.

۵- روش تحقیق:

الف- شرح کامل روش تحقیق بر حسب هدف، نوع داده‌ها و نحوه اجراء (شامل مواد، تجهیزات و استانداردهای مورد استفاده در قالب مراحل اجرایی تحقیق به تفکیک):

تذکر: درخصوص تفکیک مراحل اجرایی تحقیق و توضیح آن، از به کار بردن عناوین کلی نظیر، «گردآوری اطلاعات اولیه»، «تهیه نمونه‌های آزمون»، «انجام آزمایش‌ها» و غیره خودداری شده و لازم است در هر مورد توضیحات کامل در

¹⁰ clinical examinations

¹¹ sensitivity

¹² specificity

¹³ F1 score

رابطه با منابع و مراکز تهیه داده‌ها و ملزومات، نوع فعالیت، مواد، روش‌ها، استانداردها، تجهیزات و مشخصات هر یک ارائه گردد.

- **پیش پردازش داده‌گان:** این مرحله با هدف پاکسازی داده‌ها، پر کردن جای خالی داده‌های ثبت نشده و نمونه‌افزایی به منظور متعادل نمودن تعداد نمونه‌های دو کلاس انجام می‌شود. برای پر کردن جاهای خالی به پیش‌فرض‌های پزشکی مانند احتمال ثبات وضعیت بیمار در مدت زمان بین دو نمونه‌گیری استناد می‌شود، از الگوریتم KNNImputer استفاده می‌شود و یا میانگین دو اندازه‌گیری قبلی و بعدی محاسبه می‌گردد. برای حذف داده‌های پرت از الگوریتم isolation (iForest) استفاده خواهد شد. از آن‌جا که داده‌ها در مواردی با نرخ‌های نمونه‌برداری متفاوت جمع‌آوری شده‌اند، یکسان‌سازی طول بردارهای ویژگی از طریق نمونه‌برداری و یا نمونه‌افزایی نیز ضروری به نظر می‌رسد. به علاوه در صورت عدم تعادل تعداد مشاهدات دو کلاس، از روش synthetic minority over-sampling technique (SMOTE) برای نمونه‌افزایی در بانک داده بیماران کووید استفاده خواهد شد.
- **انتخاب/استخراج ویژگی:** با توجه به اینکه یکی از اهداف این تحقیق، یافتن مهمترین متغیرهای پیش‌بینی‌کننده است، الگوریتم‌های متنوع انتخاب/استخراج ویژگی بر روی داده‌گان اعمال شده و اثر آن‌ها بر دقت، حساسیت و تشخیص طبقه‌بندی‌کننده بررسی و گزارش می‌شود. به طور مشخص، روش‌های جستجوی مستقیم و معکوس ترتیبی و تعمیم یافته‌ی آن‌ها، ترکیب الگوریتم‌های خوشه‌بندی و رده‌بندی ویژگی‌ها، و نیز روش‌های مبتنی بر Matrix Factorization مانند Sparse PC, principal components analysis (PCA), Non-negative matrix factorization (NMF) و ICA) برای انتخاب بهترین ویژگی‌ها و یا تابعی از آن‌ها استفاده می‌شود.
- **تعلیم مدل طبقه‌بندی‌کننده:** پیش‌بینی احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه از طریق تعلیم یک مدل ترکیبی داده‌کاوی دنبال خواهد شد. مدل پیشنهادی ابتدا از چهار نوع طبقه‌بندی‌کننده شناخته شده‌ی extra trees، random forest، رگرسیون لجستیک و ماشین بردار پشتیبان که هر کدام معماری و مشخصات یادگیری مخصوص به خود را دارند ساخته می‌شود. در مرحله بعدی و برای بالا بردن کارایی، روش‌های یادگیری ترکیبی (بگینگ و/یا بوستینگ) وارد عمل شده و یک متا مدل پیش‌بینی‌کننده تعلیم می‌یابد. عملکرد این مدل با محاسبه‌ی شاخص‌های دقت، حساسیت، تشخیص و امتیاز اف ۱ ارزیابی می‌گردد. مدل‌های پایه^۴ برای مقایسه یک مدل Random Forest و یک مدل XGBoost خواهند بود که با استفاده از تمامی ویژگی‌ها تعلیم می‌یابند.
- **ارزیابی مدل:** به منظور ارزیابی عملکرد مدل‌ها، شاخص‌های دقت^۵ (Acc)، حساسیت (Se)، تشخیص (P) و امتیاز اف-۱ گزارش می‌شوند. از آن‌جا که مسئله‌ی مورد نظر یک مسئله‌ی طبقه‌بندی دو کلاسی (باینری) است، با داشتن مقادیر مثبت درست^۶ (TP)، منفی درست^۷ (TN)، مثبت کاذب^۸ (FP) و منفی کاذب^۹ (FN) این شاخص‌ها به صورت زیر محاسبه می‌شوند:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Se = \frac{TP}{TP + FN}$$

^۴ baseline

^۵ accuracy

^۶ true positive

^۷ true negative

^۸ false positive

^۹ false negative

$$P = \frac{TP}{TP + FP}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN}$$

ب- متغیرهای مورد بررسی در قالب یک مدل مفهومی و شرح چگونگی بررسی و اندازه گیری متغیرها:

مجموعه داده شامل ۲۳۱ ویژگی اندازه گیری شده در گروه های زیر است که از آن ها به عنوان متغیرهای مسئله استفاده خواهد شد:

- اطلاعات جمعیت شناسی (سن، جنس، محل زندگی، ...)
 - بیماری های قبلی بیمار
 - نتایج آزمایش خون (مانند کمینه، بیشینه و میانگین هموگلوبین، آلبومین، گلوکز، کلسیم، کراتینین و PH خون و ...)
 - علایم حیاتی (دمای بدن، فشار خون، ضربان قلب، فشار دهلیزی و بطنی، ریتم تنفس و ...)
 - غلظت گازهای موجود در خون (مانند اکسیژن اشباع)
- طبقه بندی کننده بر اساس این ورودی ها بیماران را به دو گروه ملایم / شدید (نیازمند به بستری شدن در بخش مراقبت های ویژه) تقسیم می کند.
- خروجی نهایی پژوهش، شاخص های دقت، حساسیت، تشخیص و امتیاز اف ۱ خواهد بود.

ج- شرح کامل روش (میدانی، کتابخانه ای) و ابزار (مشاهده و آزمون، پرسشنامه، مصاحبه، فیش برداری و غیره) گردآوری داده ها:

متغیرهای حیاتی و نتایج آزمایش خون و غلظت اکسیژن خون برای ۳۸۴ بیمار مبتلا به کووید ۱۹ در پنجره های زمانی ۲ تا ۱۲ ساعته اندازه گیری و ثبت شده اند.

د- جامعه آماری، روش نمونه گیری و حجم نمونه (در صورت وجود و امکان):

مجموعه داده: داده های مورد استفاده از این تحقیق اطلاعات جمعیتی بیمار، سوابق بیماری های قبلی، نتایج آزمایش خون، علائم حیاتی و میزان گازهای موجود در خون شریانی که از ۳۸۴ بیمار در بیمارستان سیریولبانز در سائوپولو برزیل جمع آوری شده است و به صورت رایگان در اختیار عموم قرار گرفته و از پایگاه اینترنتی Kaggle قابل دسترسی است [۲۱].

متغیرهای حیاتی و نتایج آزمایش خون و غلظت اکسیژن خون برای ۳۸۴ بیمار مبتلا به کووید ۱۹ در پنجره های زمانی ۲ تا ۱۲ ساعته اندازه گیری و ثبت شده اند.

ه- روش ها و ابزار تجزیه و تحلیل داده ها:

الگوریتم های مورد استفاده عبارتند از: SMOTE، XGBoost، Random Forest، روش های جستجوی مستقیم و معکوس ترتیبی و تعمیم یافته ی آن ها و همچنین روش های مبتنی بر Matrix Factorization. این الگوریتم ها در صورت دسترسی در نرم افزار RapidMiner پیاده سازی و اجرا می شوند. در غیر این صورت، این مدل ها به زبان برنامه نویسی پایتون و با استفاده از کتابخانه های Pandas، Scikit-learn، Numpy و SciPy طراحی، پیاده سازی و ارزیابی می شوند.

۶- استفاده از امکانات آزمایشگاهی واحد:

- آیا برای انجام تحقیقات نیاز به استفاده از امکانات آزمایشگاهی واحد علوم و تحقیقات می باشد؟ بلی

☐ خیر ☒

در صورت نیاز به امکانات آزمایشگاهی لازم است نوع آزمایشگاه، تجهیزات، مواد و وسایل مورد نیاز در این قسمت مشخص گردد.

نوع آزمایشگاه	تجهیزات مورد نیاز	مواد و وسایل	مقدار مورد نیاز

- آیا برای انجام تحقیقات نیاز به حمایت از سایر مراکز خارج از واحد علوم و تحقیقات می باشد؟

☐ بلی ☒ خیر

در صورت نیاز نام مراکز و نحوه حمایت (مالی، امکانات و تجهیزات و ...) مشخص گردد.

امضاء استاد راهنما:

امضاء مدیر گروه تخصصی:

۷- زمان بندی انجام تحقیق:

الف- تاریخ شروع:۱۴۰۰/۰۷/۰۱ ب- مدت زمان انجام تحقیق: ۸ ماه ج- تاریخ
اتمام:۱۴۰۱/۰۳/۰۱

تذکر: لازم است کلیه فعالیت ها و مراحل اجرایی تحقیق (شامل زمان ارائه گزارشات دوره ای) و مدت زمان مورد نیاز برای هر یک، به تفکیک پیش بینی و در جدول مربوطه درج گردیده و در هنگام انجام عملی تحقیق، حتی الامکان رعایت گردد.

پیش‌بینی زمانبندی فعالیت‌ها و مراحل اجرایی تحقیق و ارائه گزارش پیشرفت کار

ردیف	شرح فعالیت	زمان کل (ماه)	زمان اجرا به ماه											
			۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۱	بررسی منابع، جمع‌آوری پیشینه نظری و ادبیات تحقیق	۱												
۲	مطالعات تکمیلی درباره نمونه‌افزایی به روش SMOTE، روش‌های استخراج ویژگی‌ها و طبقه‌بندی‌کننده‌ی Random Forest	۱												
۳	طراحی بخش پیش‌پردازش شامل نمونه‌افزایی و استخراج ویژگی‌ها	۲												
۴	طراحی طبقه‌بندی‌کننده	۲												
۵	ارزیابی نتایج، مقایسه با روش‌های پیشین و نتیجه‌گیری	۱												
۶	نگارش پایان‌نامه	۱												
۷														
۸														
۹														
۱۰														
۱۱														
۱۲														

توجه: ۱- زمان و نوع فعالیت‌های اجرایی پایان‌نامه، حتی‌الامکان باید با مندرجات جدول منطبق باشد.

۲- حداقل زمان قابل قبول برای پیش‌بینی مراحل مطالعاتی و اجرایی پایان‌نامه کارشناسی ارشد ۶ ماه و حداکثر ۱۲ ماه می‌باشد.

این فرم باید توسط دانشجو تکمیل شود

فرم سازمان مرکزی دانشگاه آزاد اسلامی

فرم الف - فرم اطلاعات پایان نامه کارشناسی ارشد

نام واحد دانشگاهی: واحد علوم و تحقیقات	
عنوان پایان نامه کارشناسی ارشد: بهبود پیش بینی ضرورت بستری شدن بیماران کووید ۱۹ در بخش مراقبت های ویژه با استفاده از روش های یادگیری ماشین	
نام و نام خانوادگی دانشجو: مهنام پدرام شماره دانشجویی: ۹۵۰۵۱۵۳۸۲	نیمسال تحصیلی: اول ۱۴۰۰ تعداد واحد پایان نامه: ۶
رشته تحصیلی: ارشد مهندسی کامپیوتر گرایش: نرم افزار کد رشته:	
<input type="checkbox"/> فنی و مهندسی <input type="checkbox"/> کشاورزی	<input type="checkbox"/> علوم انسانی <input type="checkbox"/> هنر
<input type="checkbox"/> علوم پایه	
نام و نام خانوادگی استاد راهنما ۱: خانم دکتر مریم رستگار پور رشته تحصیلی: هوش مصنوعی کد شناسایی استاد راهنما:	
مرتبه علمی: استادیار <input checked="" type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/>	
نام و نام خانوادگی استاد مشاور ۱:	
رشته تحصیلی:	
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/> مربی <input type="checkbox"/> کد شناسایی استاد راهنما:	

با توجه به اینکه پژوهش های انجام شده در حیطه زیست پزشکی به منظور رعایت و بکارگیری اصول و موازین اخلاقی در پژوهش های انسانی و حیوانی، حفظ حقوق پژوهشگران و آزمودنی ها در بررسی های انجام شده و انتشار نتایج، چاپ مقالات در تعدادی از مجلات داخلی و خارجی و همچنین اخذ حمایت های تشویقی از برخی نهادها نیاز به اخذ کد اخلاق دارند. برخی از موضوعات نیازمند به اخذ کد اخلاق عبارتند از:

۱- موضوعات مشتمل بر آزمودنی انسان

- * موضوعات مرتبط با بررسی های مستقیم یا غیر مستقیم فیزیولوژیکی بر روی انسان ها (بررسی اثرات داروهای مختلف، کارآزمایی های بالینی)، مکمل های غذایی و فعالیت های ورزشی بر روی انسان)
- * موضوعات مرتبط با بررسی اعضا و نمونه های جدا شده از انسان (پژوهش در مورد گامت و رویان، ژنتیک پزشکی، سلول های بنیادی و اخذ نمونه های خون، سلول، مایع مغزی نخاعی و اسپرم)
- * موضوعات مرتبط با بررسی های مستقیم یا غیر مستقیم روانی و ذهنی بر روی انسان ها (مداخلات آموزشی، مطالعات رفتاری، سبک زندگی، عملکردی و مدیریتی)
- * موضوعات مرتبط با در معرض خطر قرار گرفتن انسان ها و محیط زیست آنها مانند کار بر روی بیماری های عفونی، مواد سرطان زا و ترکیبات زیان آور.
- * موضوعات مرتبط با مطالعه بر روی گروه های آسیب پذیر (مانند بررسی های انجام شده روی بیماران اسکیزوفرنی، سندرم داون و ...)

۲- موضوعات مشتمل بر آزمودنی حیوان

- * این موضوعات شامل استفاده از هر نوع و هر تعداد حیوان آزمایشگاهی برای تمامی بررسی های مستقیم و غیر مستقیم، اخذ نمونه های مختلف، استفاده از مواد زیان آور برای پژوهشگر و آسیب های محیط زیستی ناشی از پژوهش می باشد.
- لذا خواهشمند است با در نظر گرفتن مطالب فوق، مجریان محترم در روند اخذ کد اخلاق قرار گیرند. لطفاً در نظر داشته باشید:
- * در صورت نیاز به اخذ کد اخلاق، بلافاصله بعد از تصویب پروپوزال و قبل از شروع کار عملی مدارک مورد نیاز به یکی از کمیته های اخلاق تابعه وزارت بهداشت، درمان و آموزش پزشکی ارجاع شود.
- * عطف به نامه شماره ۷۰۰/۵۰۱۴/د مورخ ۱۳۹۷/۱۱/۲۳ دبیرخانه کمیته ملی اخلاق در پژوهش های زیست پزشکی، کد اخلاق برای پژوهش های در حال انجام، پایان یافته و یا مقالات آماده چاپ به هیچ عنوان صادر نمی شود.
- * کد اخلاق در واحد علوم و تحقیقات، از طریق کمیته اخلاق پزشکی مستقر در دانشکده پزشکی واحد، صادر می گردد.

امضای اساتید راهنما

امضای مدیر پژوهش

توجه

امضاء استاد راهنما و مدیر گروه در این بخش، به منزله تایید علمی و حقوقی پروپوزال می باشد. لذا پس از بررسی کامل علمی، این بخش امضاء و تایید گردد.

نام و نام خانوادگی استاد راهنما:

امضاء

تاریخ

نام و نام خانوادگی مدیر گروه تخصصی:

امضاء

تاریخ

1. Karim, S. S. A., & Karim, Q. A. (2021). Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *The Lancet*, 398(10317), 2126-2128.
2. Pfizer and BioNTech provide update on Omicron variant. Pfizer. (n.d.). Retrieved December 24, 2021, from <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-provide-update-omicron-variant>
3. Collis, H. (2021, December 22). Who forecasts coronavirus pandemic will end in 2022. POLITICO. Retrieved December 24, 2021, from <https://www.politico.eu/article/who-forecasts-coronavirus-pandemic-will-end-in-2022>
4. Jiang, Xiangao, et al. "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity." *Computers, Materials & Continua* 63.1 (2020): 537-551.
5. Heldt, Frank S., et al. "Early risk assessment for COVID-19 patients from emergency department data using machine learning." *Scientific reports* 11.1 (2021): 1-13.
6. Ezz, Mohamed, Murtada K. Elbashir, and Hosameldeen Shabana. "Predicting the need for icu admission in covid-19 patients using xgboost." *Computers, Materials and Continua* (2021): 2077-2092.
7. Aznar-Gimeno, Rocío, et al. "A Clinical Decision Web to Predict ICU Admission or Death for Patients Hospitalised with COVID-19 Using Machine Learning Algorithms." *International Journal of Environmental Research and Public Health* 18.16 (2021): 8677.
8. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273-324.
9. Ferri, F. J., Pudil, P., Hatef, M., & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. In *Machine Intelligence and Pattern Recognition* (Vol. 16, pp. 403-413). North-Holland.
10. Haq, A. U., Zhang, D., Peng, H., & Rahman, S. U. (2019). Combining multiple feature-ranking techniques and clustering of variables for feature selection. *IEEE Access*, 7, 151482-151492.
11. Vaid, S. Somani, A.J. Russak, J.K. De Freitas, F.F. Chaudhry, I. Paranjpe, et al. Machine learning to predict mortality and critical events in covid-19 positive New York city patients: a cohort study *J Med Internet Res*, 49 (6) (2020), pp. 1918-1929
12. L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jin, M. Zhang, et al. A machine learning-based model for survival prediction in patients with severe covid-19 infection *MedRxiv* (2020)
13. E. Rechtman, P. Curtin, E. Navarro, S. Nirenberg, M.K. Horton Vital signs assessed in initial clinical encounters predict covid-19 mortality in an nyc hospital system *Sci Rep*, 10 (2020), pp. 1-6
14. D. Bertsimas, G. Lukin, L. Mingardi, O. Nohadani, A. Orfanoudaki, B. Stellato, H. Wiberg, S. Gonzalez-Garcia, C.L. Parra-Calderon, K. Robinson, et al. Covid-19 mortality risk assessment: an international multi-center study *PloS One*, 15 (2020), Article e0243262
15. X. Guan, B. Zhang, M. Fu, M. Li, X. Yuan, Y. Zhu, J. Peng, H. Guo, Y. Lu Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized covid-19 patients: results from a retrospective cohort study *Ann Med*, 53 (2021), pp. 257-266,
16. A.L. Booth, E. Abels, P. McCaffrey Development of a prognostic model for mortality in covid-19 infection using machine learning *Mod Pathol* (2020), pp. 1-10
17. L. Sun, F. Song, N. Shi, F. Liu, S. Li, P. Li, W. Zhang, X. Jiang, Y. Zhang, L. Sun, X. Chen, Y. Shi Combination of four clinical indicators predicts the severe/critical symptom of patients infected covid-19 *J Clin Virol*, 128 (2020), p. 104431
18. H. Yao, N. Zhang, R. Zhang, M. Duan, T. Xie, J. Pan, E. Peng, J. Huang, Y. Zhang, X. Xu, et al. Severity detection for the coronavirus disease 2019 (covid-19) patients using a machine learning model based on the blood and urine tests *Frontiers in cell and developmental biology*, 8 (2020), p. 683
19. C. Hu, Z. Liu, Y. Jiang, O. Shi, X. Zhang, K. Xu, et al. Early prediction of mortality risk among patients with severe COVID-19, using machine learning *Int J Epidemiol*, 49 (6) (2020), pp. 1918-1929
20. Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *Journal of medical systems*, 44(8), 1-12.
21. Sírío-Libanês, H. (2020, June 22). Covid-19 - clinical data to assess diagnosis. Kaggle. Retrieved December 24, 2021, from <https://www.kaggle.com/S%C3%ADrio-Libanes/covid19>