# تذکرات مهم

## 1)­ 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.

## 2) تاریخ تصویب پروپوزال در شورای گروه تخصصی با تاریخ تصویب در شورای پژوهش دانشکده نباید بیش از یکماه باشد.

## 3)­ تاریخ تصویب پروپوزال در شورای پژوهشی دانشکده و ارسال پروپوزال به معاونت پژوهش و فناوری واحد نباید بیش از یکماه نباشد.

## 4) 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.

## 5)لازم است دانشجو پس از تصویب نهایی، ابلاغیه تصویب پروپوزال را از پژوهش دانشکده تحویل گیرد.

**6)­ با توجه به اینکه پرداخت کمک هزینه های رساله/ پایان نامه ها مربوط به فعالیتهای آزمایشگاهی و کارگاهی منوط به تکمیل جدول بند 6- استفاده از امکانات آزمایشگاهی واحد و تجهیرات و مواد مورد نیاز – می باشد، لازم است این بخش تکمیل گردد.**

دانشگاه آزاد اسلامي

**واحد علوم و تحقيقات (تهران)**

**Science and Research Branch, Islamic Azad University**

**فرم پيشنهاد تحقيق**

**پايان‏نامه‌ی كارشناسي ارشد**

**عنوان تحقيق به فارسي:**

**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 دانشکده: مکانیک، برق و کامپیوتر**

**نام خانوادگی دانشجو: پدرام گروه تخصصی: مهندسی نرم‌افزار**

**رشته تحصيلي: ارشد مهندسی کامپیوتر گرايش: نرم‌افزار**

**نيمسال ورود به مقطع جاري: دوم ۹۵ نيمسال شروع به تحصيل: دوم ۹۵**

**نام و نام خانوادگی استاد (اساتيد) راهنما: نام و نام خانوادگی استاد (اساتيد) مشاور:**

**1- خانم دکتر مریم رستگارپور 1-**

**توجه:** لطفاً اين فرم با مساعدت و هدايت استاد راهنما تكميل شود.

1. اطلاعات مربوط به دانشجو:

نام:....مهنام...................................نام‏خانوادگي:..........پدرام.............................شماره‌دانشجويي:...۹۵۰۵۱۵۳۸۲ ......

مقطع:............ارشد........................رشته‌تحصيلي:........کامپیوتر..................................گروه‌تخصصي:..نرم‌افزار..........

گرایش:.....نرم‌افزار.....نام‌دانشكده: مکانیک، برق و کامپیوتر...... سال ورود به مقطع جاري: .......۱۳۹۵.............

نيمسال ورودي:.......دوم......

آدرس پستي در تهران، بزرگراه شیخ فضل‌الله – شهرک فرهنگیان جدید – بلوک ۱۰ – واحد ۷

تلفن ثابت محل سكونت: .۸۸۲۵۳۶۴۲... تلفن همراه:..۰۹۳۵۲۵۳۶۳۵۹... پست الكترونيك: mahnamp@yahoo.com

آدرس پستي در شهرستان:............................................................................................................................................

تلفن ثابت محل سكونت: ..............................تلفن محل كار: .................................. دورنگار:.................................

1. اطلاعات مربوط به استاد راهنما:

**تذكرات:**

* + دانشجويان دوره كارشناسی ارشد می‌توانند حداکثر دو استاد راهنما و یک استاد مشاور انتخاب نمايند.
  + در صورتي كه اساتيد راهنما و مشاور **مدعو** مي باشند، لازم است سوابق تحصيلي، آموزشي و پژوهشي كامل ايشان (رزومه كامل) شامل فهرست پایان‌نامه‌های کارشناسی ارشد و رساله‏هاي دكتري دفاع شده و يا در حال انجام كه اساتيد مدعو، راهنمايي و يا مشاوره آنرا بر عهده داشته‏اند، **به همراه مدارك مربوطه** و همچنين آخرين حكم كارگزيني (حكم هيأت علمي) ضميمه گردد.
  + اساتيد راهنما و مشاور موظف هستند قبل از پذيرش پروپوزال، به سقف ظرفيت پذيرش خود توجه نموده و در صورت تكميل بودن ظرفيت پذيرش، از ارسال آن به دانشكده و حوزه پژوهشي و يا در نوبت قراردادن و ايجاد وقفه در كار دانشجويان جداً پرهيز نمايند.

**اطلاعات مربوط به استاد راهنماي اول:**

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

سنوات تدریس : .....10 سال...... شغل و سمت فعلی : .......هیات علمی تمام وقت............

آدرس محل كار: دانشگاه آزاد اسلامی واحد ساوه

تلفن منزل : ................... تلفن همراه: ..09127557588.. محل كار: ................ دورنگار: ................

پست الكترونيك *(Email)* : rastgarpour@gmail.com

نحوه همکاری با واحد علوم و تحقیقات :

تمام وقت نیمه وقت مدعو

**اطلاعات مربوط به استاد راهنماي دوم:**

دانشگاهي

نام و نام خانوادگي:........................................................آخرين مدرك تحصيلي ـــــــــــــــ :.....................................

حوزوي

عضو هيأت علمي دانشگاه ........................................

تخصص اصلي:........................... رتبه دانشگاهي (مرتبه علمي): ........................ تلفن همراه: .......................................

تلفن منزل يا محل كار:...................................... نام و نام خانوادگي به زبان انگليسي: ....................................................

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

**اطلاعات مربوط به استاد مشاور:**

دانشگاهي

نام و نام خانوادگي:........................................................آخرين مدرك تحصيلي ـــــــــــــــ :.....................................

حوزوي

عضو هيأت علمي دانشگاه ........................................

تخصص اصلي:........................... رتبه دانشگاهي (مرتبه علمي): ........................ تلفن همراه: .......................................

تلفن منزل يا محل كار:...................................... نام و نام خانوادگي به زبان انگليسي: ....................................................

نحوه همکاری با واحد علوم و تحقیقات:

تمام وقت نیمه وقت مدعو

4- اطلاعات مربوط به پايان‏نامه:

الف- عنوان تحقیق

- عنوان به زبان فارسی:

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

- عنوان به زبان انگليسي(آلماني، فرانسه، عربي):

**تذكر:** صرفاً دانشجويان رشته‏هاي زبان آلماني،‌ فرانسه و عربي مجازند عنوان پايان‏نامه خود را به زبان مربوطه در اين بخش درج نمايند و براي بقيه دانشجويان، عنوان بايستي به زبان انگليسي ذكر شود.

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

ب– تعداد واحد پايان‏نامه: 6

ج- بيان مسأله اساسي تحقيق به طور كلي (شامل تشريح مسأله و معرفي آن، بيان جنبه‏هاي مجهول و مبهم، بيان متغيرهاي مربوطه و منظور از تحقيق):

در سال‌های اخیر و به دنبال پیشرفت‌های چشمگیر در زمینه‌ی جمع‌آوری و ذخیره‌ی دادگان 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 دادگان 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 دادگان 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 دادگان با کیفیت، جامع و بی‌طرف، و همچنین تعداد بالای متغیرهای ورودی مواجه است. حتی در صورت دستیابی به مجموعه‌ی دادگان 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 دادگان جدید داشته باشد. بنابراین، در بسیاری از تحقیقات منتشر شده در زمینه‌ی کاربرد روش‌های هوش مصنوعی در طبقه‌بندی دادگان پزشکی، هدف اصلی معطوف به بهبود کیفیت دادگان، 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 بستری‌شدن 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 دادگان، 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[[1]](#footnote-2) (زاید) و غیرضروری[[2]](#footnote-3) با روش‌های خوشه‌بندی ویژگی‌های ورودی، شناسایی و حذف می‌شوند و سپس،‌ مدل یادگیرنده با استفاده از ویژگی‌های باقی‌مانده تعلیم می‌یابد. 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 می‌شود. اگرچه ممکن است ثبت این تصاویر دقت بالای مدل یادگیرنده را در پی داشته باشد، مراحل ثبت تصویر،‌ نه تنها از نظر زمانی و اقتصادی هزینه‌هایی را به بیمار تحمیل می‌کند،‌ بلکه او را در موقعیت ناخوشایندتری نسبت به مراحل جمع‌آوری نمونه‌های آزمایش خون قرار می‌دهد.

با توجه به این مقدمه‌ی کوتاه، در تحقیق حاضر، از دادگان پزشکی غیرتصویری در تشخیص ضرورت بستری‌شدن بیماران کووید ۱۹ در بخش مراقبت‌های ویژه استفاده می‌شود و سه هدف زیر دنبال می‌گردد:

۱- پاکسازی دادگان و پر کردن جای خالی دادگان از دست رفته و یا ثبت نشده با روش‌های مناسب درون‌یابی.

۲- رده‌بندی و کاهش بُعد ویژگی‌های ورودی برای تعیین پیش‌بینی کننده‌های بهتر

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

دقت اغلب مدل‌های طبقه‌بندی‌کننده وابسته به تعادل تعداد نمونه‌های هر کلاس در مجموعه‌ی دادگان است. در بیماری‌هایی مانند کووید ۱۹ که اولاً مجموعه دادگان زیادی در مورد آن‌ها جمع‌آوری نشده است و ثانیاً جهش‌های مکرر ویروس عامل بیماری، موجب پیدایش سویه‌های متنوع با علایم متفاوت شده است، تعداد نمونه‌های بیمار به طور قطع، کمتر از تعداد نمونه‌های جمع‌آوری شده از گروه کنترل (افراد سالم) خواهد بود که این موضوع،‌ موجب سوگیری مدل به سمت تشخیص نمونه‌ها به عنوان سالم (یا احتمال کم بستری در بخش مراقبت‌های ویژه) شده. به علاوه، در مجموعه‌ی دادگانی که به صورت عمومی منتشر شده است،‌ علایم حیاتی بیماران در برخی بازه‌های زمانی ثبت نشده‌اند و این جاهای خالی باید با پیش‌فرض‌هایی در مورد ثبات وضعیت بیمار و یا به کمک روش‌های درون‌یابی مناسب پر شوند. لذا، برای مواجهه با این مشکل، اولاً علاوه بر معیار «دقت»، معیارهای «حساسیت» تشخیص نمونه‌های بیمار با احتمال بالای نیاز به بستری در بخش مراقبت‌های ویژه و F1-score نیز به عنوان معیار ارزیابی عملکرد مدل طبقه‌بندی‌کننده مورد استفاده قرار می‌گیرند. ثانیاً، با استفاده از روش‌های برون‌یابی، تعدادی نمونه‌ی مصنوعی[[3]](#footnote-4) مرتبط با مشاهدات دارای احتمال بستری بالا تولید می‌گردد و عملکرد مدل در دو حالت «‌فقط استفاده از داده‌های اصلی» و «استفاده از ترکیب داده‌های اصلی و مصنوعی» مقایسه می‌شود.

د- اهمیت و ضرورت انجام تحقيق (شامل اختلاف نظرها و خلاءهاي تحقيقاتي موجود، ميزان نياز به موضوع، فوايد احتمالي نظري و عملي آن و همچنين مواد، روش و يا فرآيند تحقيقي احتمالاً جديدي كه در اين تحقيق مورد استفاده قرار مي‏گيرد:

با توجه به اطلاعاتی که تاکنون از نحوه‌ی انتشار ویروس SARS-CoV-2 جمع‌آوری شده است، امیدواری درباره ریشه‌کنی جهانی این ویروس در کوتاه‌مدت و یا دستیابی به ایمنی جمعی (در مقیاس محلی) با تردید روبروست. دو سال پس از آغاز همه‌گیری ، 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[[4]](#footnote-5) را تایید کرد. اطلاعات و تحلیل اطلاعات درباره این 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 این‌که کدامیک از مبتلایان ممکن است دچار علایم مراحل حاد این بیماری و نیازمند به مراقبت‌های ویژه شوند روش های مبتنی بر یادگیری ماشین به کار گرفته شوند.

ه- مرور ادبیات و سوابق مربوطه (بيان مختصر پیشینه تحقيقات انجام شده در داخل و خارج کشور پيرامون موضوع تحقیق و نتايج آنها و مرور ادبیات و چارچوب نظري 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 علایم 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[[5]](#footnote-6)، میزان همبستگی[[6]](#footnote-7) 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 [۶-۴].

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 طبقه‌بندی کننده‌یExtreme Gradient Boosting (XGBoost) استفاده شده است که با تعلیم تکراری مدل مشابه و با استفاده از بردارهای ویژگی وزن‌دار،‌ 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 الگوریتم های خوشه بندی و رده‌بندی 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[[7]](#footnote-8) و 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 [۷-۵]. 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، تندنفسی 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 دادگان 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 دادگان 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، گلوتاتیون ، لنفوسیت‌های T بالغ و پروتئین تام) برای تعلیم و یا تخمین پارامترهای مدل استفاده شده است.

در تشخیص بیماری کووید ۱۹ با استفاده از دادگان 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 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 جديد بودن و نوآوري در تحقيق:

* برای مدیریت جای خالی دادگان از دست رفته و یا ثبت نشده[[8]](#footnote-9) از الگوریتم KNNImputer و برای حذف داده‌های پرت از الگوریتم isolation forest (iForest) استفاده می‌شود.
* از آن‌جا که دادگان در مواردی با نرخ های 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[[9]](#footnote-10) 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 میابد.

ز- اهداف مشخص تحقيق (شامل اهداف آرماني، کلی، اهداف ويژه و كاربردي):

* 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 رده بندی شده.

ح– در صورت داشتن هدف كاربردي، نام بهره‏وران (سازمان‏ها، صنايع و يا گروه ذينفعان) ذكر 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 نمونه افزایی چه تأثیری بر دقت عملکرد الگوریتم طبقه‌بندی دارد؟
* 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- فرضيه‏هاي 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 رده بندی شده تعلیم می‌یابد.

ک- تعريف واژه‏ها و اصطلاحات فني و تخصصی (به صورت مفهومی و عملیاتی):

* معاینات بالینی[[10]](#footnote-12): 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 سرپایی 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 طبقه بندی، هر درخت می تواند تصمیم بگیرد و کلاسی که بیشترین رای را داشته باشد، طبقه بندی نهایی را تعیین می کند.
* روش‌های ترکیبی داده‌کاوی: روش‌هایی برای تجمیع مجموعه محدودی از الگوریتم‌های یادگیری مختلف به منظور حصول نتایج بهتر نسبت به هر یک از الگوریتم‌های یادگیری موجود در مجموعه.
* حساسیت[[11]](#footnote-13): یکی از شاخص‌های دقت آزمایش. حساسیت، حاصل تقسیم موارد مثبت واقعی به حاصل جمع موارد مثبت واقعی و موارد منفی کاذب است.
* تشخیص[[12]](#footnote-14): یکی از شاخص‌های دقت آزمایش. تشخیص، حاصل تقسیم موارد منفی واقعی به حاصل جمع موارد منفی واقعی و مثبت کاذب است.
* امتیاز اف۱[[13]](#footnote-15): یکی از شاخص‌های دقت آزمایش. نوعی میانگین پارامترهای دقت و بازیابی در یک پیشبینی است.

5- روش تحقیق:

الف- شرح كامل روش تحقیق بر حسب هدف، نوع داده‌ها و نحوه اجراء (شامل مواد، تجهيزات و استانداردهاي مورد استفاده در قالب مراحل اجرايي تحقيق به تفكيك):

تذكر: درخصوص تفكيك مراحل اجرايي تحقيق و توضيح آن، از به كار بردن عناوين كلي نظير، «گردآوري اطلاعات اوليه»، «تهيه نمونه‏هاي آزمون»، «انجام آزمايش‏ها» و غيره خودداري شده و لازم است در هر مورد توضيحات كامل در رابطه با منابع و مراكز تهيه داده‏ها و ملزومات، نوع فعاليت، مواد، روش‏ها، استانداردها، تجهيزات و مشخصات هر يك ارائه گردد.

* **پیش پردازش دادگان:** 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 forest (iForest) استفاده خواهد شد. از آن‌جا که داده‌ها در مواردی با نرخ های نمونه‌برداری متفاوت جمع‌آوری شده‌اند، یکسان‌سازی طول بردارهای ویژگی از طریق نمونه برداری و یا نمونه افزایی نیز ضروری به نظر میرسد. به علاوه درصورت 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 دادگان 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 principal components analysis (PCA)، Sparse PC، independent components analysis (ICA) و Non-negative matrix factorization (NMF) 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 میابد. عملکرد این مدل با محاسبه‌ی شاخص‌های دقت،‌ حساسیت، تشخیص و امتیاز اف۱ ارزیابی میگردد. مدل‌های پایه[[14]](#footnote-16) 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[[15]](#footnote-17) (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[[16]](#footnote-18) (TP)، منفی درست[[17]](#footnote-19) (TN)، مثبت کاذب[[18]](#footnote-20) (FP) و منفی کاذب[[19]](#footnote-21) (FN) این شاخص‌ها به صورت زیر محاسبه می‌شوند:

ب- متغيرهاي مورد بررسي در قالب یک مدل مفهومی و شرح چگونگی بررسی و اندازه‌گیری متغیرها:

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

* اطلاعات جمعیت شناسی (سن، جنس، محل زندگی، ...)
* بیماری‌های 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 )

طبقه‌بندی‌کننده بر اساس این ورودی‌ها بیماران را به دو گروه ملایم / شدید ( نیازمند به بستری شدن در بخش مراقبت‌های ویژه) تقسیم می‌کند.

خروجی نهایی پژوهش، شاخص‌های دقت، حساسیت، تشخیص و امتیاز اف۱ خواهد بود.

ج– شرح کامل روش (ميداني، كتابخانه‏اي) و ابزار (مشاهده و آزمون، پرسشنامه، مصاحبه، فيش‏برداري و غيره) گردآوري داده‏ها:

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 اندازه‌گیری و ثبت شده‌اند.

د– جامعه آماري، روش نمونه‏گيري و حجم نمونه (در صورت وجود و امکان):

مجموعه داده: داده‌های مورد استفاده از این تحقیق اطلاعات جمعیتی بیمار، سوابق بیماری‌های قبلی، نتایج آزمایش خون، علائم حیاتی و میزان گازهای موجود در خون شریانی که از ۳۸۴ بیمار در بیمارستان سیریولبانز در سائوپلو برزیل جمع‌آوری 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 پنجره‌های زمانی ۲ تا ۱۲ ساعته اندازه‌گیری و ثبت شده‌اند.

هـ- روش‌ها و ابزار تجزيه و تحليل داده‏ها:

الگوریتم‌های مورد استفاده عبارتند از: SMOTE، Random Forest ، XGBoost، روش‌های جستجوی مستقیم و معکوس ترتیبی و تعمیم یافته‌ی آن‌ها و همچنین روش‌های مبتنی بر Matrix Factorization. . این الگوریتم‌ها در صورت دسترسی در نرم‌افزار RapidMiner پیاده‌سازی و اجرا می‌شوند. در غیر این‌صورت، این مدل ها به زبان برنامه نویسی پایتون و با استفاده از کتابخانه‌های Pandas ،Scikit-learn ، Numpy و SciPy طراحی، پیاده‌سازی و ارزیابی می‌شوند.

6- استفاده از امكانات آزمايشگاهي واحد:

* + آيا براي انجام تحقيقات نياز به استفاده از امكانات آزمايشگاهي واحد علوم و تحقيقات مي‌باشد؟ بلي  خير

در صورت نياز به امكانات آزمايشگاهي لازم است نوع آزمايشگاه، تجهيزات، مواد و وسايل مورد نياز در اين قسمت مشخص گردد.

|  |  |  |  |
| --- | --- | --- | --- |
| **نوع آزمايشگاه** | **تجهيزات مورد نياز** | **مواد و وسايل** | **مقدار مورد نياز** |
|
|
|  |  |  |  |

* + آیا برای انجام تحقیقات نیاز به حمایت از سایر مراکز خارج از واحد علوم و تحقیقات می‌باشید؟

بلي  خير

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

**امضاء استاد راهنما: امضاء مدیر گروه تخصصی:**

7- زمان‌بندي انجام تحقيق:

الف- تاريخ شروع:...........01/07/1400...................... ب- مدت زمان انجام تحقيق: ۸ ماه ج- تاريخ اتمام:................01/03/1401.......

تذكر: لازم است كليه فعاليت‏ها و مراحل اجرايي تحقيق (شامل زمان ارائه گزارشات دوره‏اي) و مدت زمان مورد نياز براي هر يك، به تفكيك پيش‏بيني و در جدول مربوطه درج گرديده و در هنگام انجام عملي تحقيق، حتي‏الامكان رعايت گردد.

**پيش‏بيني زمانبندي فعاليت‏ها و مراحل اجرايي تحقيق و ارائه گزارش پيشرفت كار**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **رديف** | **شرح فعاليت** | **زمان كل**  **(ماه)** | **زمان اجرا به ماه** | | | | | | | | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| 1 | بررسی منابع، جمع آوری پیشینه نظری و ادبیات تحقیق | 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | مطالعات تکمیلی درباره نمونه‌افزایی به روش SMOTE، روش‌های استخراج ویژگی‌ها و طبقه‌بندی‌کننده‌ی Random Forest | ۱ |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | طراحی بخش پیش‌پردازش شامل نمونه‌افزایی و استخراج ویژگی‌ها | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | طراحی طبقه‌بندی‌کننده | 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | **ارزیابی نتایج، مقایسه با روش‌های پیشین و نتیجه‌گیری** | **۱** |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | **نگارش پایان نامه** | **۱** |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 12 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**توجه: 1- زمان و نوع فعاليت‌هاي اجرايي پايان‏نامه، حتي‏الامكان بايد با مندرجات جدول منطبق باشد.**

**2- حداقل زمان قابل قبول براي پيش‏بيني مراحل مطالعاتي و اجرايي پايان‏نامه كارشناسي ارشد 6 ماه و حداكثر 12 ماه مي‏باشد.**

اين فرم بايد توسط دانشجو تكميل شود

فرم سازمان مركزي دانشگاه آزاد اسلامي

فرم الف- فرم اطلاعات پايان‌نامه كارشناسي ارشد

|  |
| --- |
| نام واحد دانشگاهي: واحد علوم و تحقيقات |
| عنوان پايان‌نامه كارشناسي ارشد: **بهبود پیش‌بینی ضرورت بستری‌شدن بیماران کووید ۱۹**  **در بخش مراقبت‌های ویژه با استفاده از روش‌های یادگیری ماشین** |
| نام و نام‏خانوادگي دانشجو: مهنام پدرام نيمسال تحصيلي: اول 1400  شماره‏دانشجويي: 950515382 تعداد واحد پايان‌نامه: 6 |
| رشته تحصيلي: ارشد مهندسی کامپیوتر گرايش: نرم افزار كد رشته:  فني و مهندسي  علوم انساني  علوم پايه  كشاورزي  هنر |
| نام و نام‏خانوادگي استاد راهنما 1: خانم دکتر مریم رستگار پور رشته تحصيلي: هوش مصنوعی  مرتبه علمي: استاديار دانشيار  استاد  كد شناسايي استاد راهنما: |
| نام و نام‏خانوادگي استاد مشاور 1: رشته تحصيلي:  مرتبه علمي: استاديار  دانشيار  استاد  مربي  كد شناسايي 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:

**1-­ 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 ...)

2**- 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 5014/700/د مورخ 23/11/1397 دبیرخانه کمیته ملی اخلاق در پژوهش های زیست پزشکی، کداخلاق 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 می‌باشد. لذا پس از بررسی کامل علمی، این بخش امضاء و تایید گردد.  نام و نام خانوادگی استاد راهنما: نام و نام خانوادگی مدیر گروه تخصصی:  امضاء امضاء  تاریخ تاریخ |

منابع:

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. LuClinical 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. Sírio-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

1. irrelevant [↑](#footnote-ref-2)
2. redundant [↑](#footnote-ref-3)
3. synthetic [↑](#footnote-ref-4)
4. omicron [↑](#footnote-ref-5)
5. information gain [↑](#footnote-ref-6)
6. correlation [↑](#footnote-ref-7)
7. manifold learning [↑](#footnote-ref-8)
8. missing data [↑](#footnote-ref-9)
9. ensemble learning [↑](#footnote-ref-10)
10. clinical examinations [↑](#footnote-ref-12)
11. sensitivity [↑](#footnote-ref-13)
12. specificity [↑](#footnote-ref-14)
13. F1 score [↑](#footnote-ref-15)
14. baseline [↑](#footnote-ref-16)
15. accuracy [↑](#footnote-ref-17)
16. true positive [↑](#footnote-ref-18)
17. true negative [↑](#footnote-ref-19)
18. false positive [↑](#footnote-ref-20)
19. false negative [↑](#footnote-ref-21)