مجله تخصصی اپیدمیولوژی ایران؛ پاییز ۱۳۹۷، دوره ۱٤، شماره ۳: صفحات: ۲۸۲-۲۷۲.

مقاله پژوهشی

تعیین عوامل خطرزا و ارایه مدل پیش‌آگهی آمبولی ریه بیماران بستری با استفاده از شبکه‌های بیزی

فرزانه فیض‌منش [۱]، علی‌اصغر صفائی [۲]

[۱] کارشناس ارشد، گروه انفورماتیک پزشکی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

[۲] استادیار، گروه انفورماتیک پزشکی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

نویسنده رابط: علی‌اصغر صفائی، نشانی: تهران، بزرگراه جلال آل احمد، دانشگاه تربیت مدرس، دانشکده علوم پزشکی، گروه انفورماتیک پزشکی،تلفن: ۸۲۸۸۴۵۸۱ ،

پست الکترونیک: aa.safaei@modares.ac.ir

تاریخ دریافت: ۹۶/۱۰/۰۳؛ پذیرش: ۹۷/۰۳/۳۰

مقدمه و اهداف: آمبولی ریه یک رویداد بالقوه کشنده و در عین حال شایع است که در سال‌های اخیر باعث افزایش تدریجی تعداد بستری‌های ناشی از آن در بیمارستان‌ها شده است. به همین دلیل، یکی از چالش برانگیزترین بیماری‌ها نزد پزشکان به حساب می‌آید. هدف اصلی از این پژوهش، گزارش یک پروژه تحقیقاتی به منظور مقایسه الگوریتم‌های مختلف داده‌کاوی برای انتخاب دقیق‌ترین مدل برای پیش‌بینی وقوع آمبولی ریه در بیماران بستری است که به فراهم کردن دانش مورد نیاز کادر درمانی در تصمیم‌گیری بهتر کمک می‌کند.

روش کار: در این پژوهش تلاش شد تا با استفاده از روش‌های مختلف یادگیری ماشین، یک مدل پیش‌بینی طراحی شود که بهترین عملکرد در پیش‌بینی احتمال وقوع آمبولی ریه در بیماران بستری در معرض خطر را داشته باشد. از میان الگوریتم‌های داده‌کاوی، از شبکه‌های بیزی و الگوریتم‌های درخت تصمیم J48، رگرسیون لجستیک و نیز بهینه‌سازی حداقل متوالی استفاده شد. داده‌های مورد استفاده تحقیق، عوامل خطرزا و سوابق گذشته مربوط به بیماران بستری بخش ریه بیمارستان شریعتی تهران بود.

یافته‌ها: بررسی‌ها صورت گرفته نشان می‌دهد که صحت و ویژگی در تمام مدل‌های پیش‌بینی از عملکرد مطلوبی برخوردار بوده است، و مدل بیزی در پیش‌بینی وقوع آمبولی ریه، بیشترین میزان حساسیت را داشت.

نتیجه‌گیری: یافته‌ها نشان می‌دهند اگر چه تفاوت کمی در عملکرد مدل‌های پیش‌بینی وجود دارد، اما در این گونه داده‌ها برای پیش‌بینی وقوع آمبولی ریه در بیماران بستری، مدل شبکه بیزی ابزار مناسب‌تری است، که می‌تواند به‌عنوان روش حمایتی در کنار تصمیم‌های پزشکی قرار گیرد تا صحت پیش‌بینی بیماری‌ها را ارتقاء بخشد.

واژگان کلیدی: آمبولی ریه، یادگیری ماشین، مدل پیش‌بینی، شبکه‌های بیزی، درخت تصمیم، رگرسیون لجستیک، بهینه‌سازی حداقل متوالی، عوامل خطرزا

مقدمه

آمبولی ریه انسداد در شریان اصلی مربوط به ریه یا یکی از شریان‌های جانبی است که بافت ریه را خونرسانی می‌کند. اغلب آمبولی‌های ریوی از عروق لگنی یا اندام تحتانی برخاسته و زمانی که ترومبوز حاد ورید عمقی بالای زانو، بدون درمان رها شود، آمبولی ریوی بالینی در نیمی از بیماران رخ خواهد داد (۱). علائم آمبولی ریه غیر اختصاصی بوده و در سایر بیماری‌های ریه، پلور و قلب و گوارش نیز دیده می‌شود. بسته به بیماری‌های همراه، علائم و نشانه‌های بالینی ممکن است پوشانده شوند و تشخیص دشوارتر شود (۲). ترمبوآمبولی ریه همیشه با عوامل مداخله‌گرانه درمانی دستخوش تغییر بوده است. از سوی دیگر فاکتورهای متعدد ارثی و اکتسابی در افزایش خطر ایجاد ترومبوآمبولی ریه کاملاً شناخته و معرفی شده‌اند. در ۱۵ سال گذشته در ادبیات طب به طور چشمگیر بر نقش فاکتورهای ژنتیکی در تمامی انواع آمبولی‌های ریه (همراه با عوامل خطرزا و بدون عوامل خطرزا) تأکید شده است (۳). مرگ ناشی از ابتلا به آمبولی ریه ۳۱ درصد برآور شده است؛ این در حالی است که درمان حمایتی اینوتروپیک صورت گرفته و ایست قلبی تنفسی اتفاق نیافتد (۴). این بیماری بیش از ۱۵ درصد از همه مرگ‌های بیمارستانی و ۳۱-۲۱ درصد از مرگ‌هایی که در ارتباط با بارداری و زایمان است را در اروپا و امریکا به خود اختصاص داده است (۵). آمبولی ریه یک بیماری تهدیدکننده حیات است. تعدادی از ۷۵-۵۰ درصد موارد ترومبوآمبولی وریدی در بخش جراحی و ICU بیمارستان‌ها رخ می‌دهد. بیشتر بیماران بستری در بخش مراقبت‌های ویژه در معرض خطر بروز ترومبوآمبولی وریدی، به علت عدم تحرک و یا

[Downloaded from irje.tums.ac.ir on 2025-04-04]

بیماری‌های پس‌زمینه‌ای هستند. همچنین در ایران میزان بروز ترومبوآمبولی وریدی در بیماران بستری ۹ مورد در هر ۱۰۰۰۰ نفر و با ۱۱ درصد میزان مرگ‌ومیر است. میانگین سالانه بالغان مبتلا به ترومبوز وریدی عمقی در ایران ۲،۲۸۸،۲۷۲ نفر است. میانگین سالانه شیوع ترومبوز وریدی عمقی در بیماران بستری در ایران با ریسک خطر ابتلا به آن ۳۹۵/۱۶- ۱۲۹/۹۰ مورد در هر ۱۰۰۰ بیمار بستری است (۶). از این رو ترومبوآمبولی وریدی یکی از عوامل تهدیدکننده حیات محسوب می‌شود. با توجه به آمارهای ذکر شده، توجه به وقوع آمبولی ریه از اهمیت بالایی برخوردار است. وجود عوامل خطرزای متعدد در ایجاد آمبولی ریه، تجزیه و تحلیل این بیماری و یافتن مهم‌ترین عوامل مؤثر را کاری پیچیده و دشوار ساخته است. از سوی دیگر، داده‌های موجود در مورد بیماران بسیار پیچیده و بزرگ و نیز گسترده و متنوع است، بنابراین پردازش و تجزیه و تحلیل این داده‌ها با روش‌های سنتی اغلب پیچیده، زمان‌بر و هزینه‌بر خواهد بود؛ از این رو سامانه‌های هوش مصنوعی، به منظور حمایت از کارکنان مراقبت بهداشتی در پیش‌بینی و تشخیص بیماری، ذخیره و بازیابی اطلاعات در پایگاه داده، تشخیص و تفسیر تصاویر پزشکی، ارائه طرح درمان و کنترل بیماری آمده‌اند (۷). در حوزه پزشکی، یادگیری ماشین به منظور استنتاج قوانین تشخیص و پیش‌بینی بیماری‌ها و حل مشکلات آن‌ها کاربرد زیادی دارد (۸). الگوریتم‌هایی مانند درخت‌های تصمیم، ماشین‌های بردار پشتیبان و شبکه‌های بیزی در سامانه‌های اطلاعات سلامت را می‌توان نام برد که هر یک در حوزه خاصی مورد استفاده قرار می‌گیرند (۹). مطالعه‌ها نشان می‌دهد روش‌های متفاوتی در تشخیص و پیش‌بینی آمبولی ریه به‌کار گرفته شده است، که در ادامه به چند نمونه آن اشاره می‌شود.

آقارضایی و همکاران (۲۰۱۶) به پیش‌بینی سطح خطر آمبولی ریه و ترومبوز وریدی عمقی با استفاده از شبکه‌های مصنوعی پرداختند. در این پژوهش، تعداد ۳۱ عامل خطر، به منظور ارزیابی وضعیت ۲۹۴ بیمار بستری مبتلا به آمبولی ریه در سه بیمارستان آموزشی وابسته به دانشگاه علوم پزشکی کرمان مورد بررسی قرار گرفت. دو نوع شبکه مصنوعی Feed-Forward Back Propagation و Elma Back Propagation در این پژوهش مقایسه شدند. از طریق یک مدل شبکه عصبی بهینه‌سازی شده دقت و شاخص سطح خطر ۹۳/۲۳ درصد به‌دست آمد و سپس یافته‌های حاصل از راه مدل با یافته‌های به‌دست آمده از روش تشخیصی اسکن پرفیوژن بیماران مبتلا مقایسه شد. ۸۶/۶۱ درصد از بیماران که از روش تشخیصی اسکن پرفیوژن، با خطر بالا تشخیص داده شده بودند نیز به درستی توسط روش پیشنهادی در این مطالعه تشخیص داده شدند (۱۰). جعفری و همکاران (۲۰۱۵) با استفاده از شبکه‌های عصبی مصنوعی، به طراحی سامانه تصمیم‌یار پزشکی برای کمک به تشخیص آمبولی ریه پرداختند. یافته‌های پژوهش، صحت تشخیص موارد بیمار به کمک شبکه عصبی مصنوعی با الگوریتم لونبرگ مارکواردت ۹۷ درصد و میانگین مربعات خطا ۰/۰۱۲۰ را نشان داده است. بنابراین می‌توان از آن به‌عنوان یک سامانه تصمیم‌یار برای تشخیص‌های دقیق، صحیح، به‌موقع و ارزان استفاده کرد (۱۱).

D. LUCIANI و همکاران (۲۰۰۶) در پژوهش خود با عنوان «سامانه خبره کمک تشخیص آمبولی ریه مبتنی بر شبکه‌های بیزی» این سامانه روی داده‌های به‌دست آمده از یک مطالعه کوهورت آینده‌نگر ۷۵۰ نفر از بیماران بیمارستانی در ایتالیا که در آن تشخیص صحت آمبولی ریوی با استفاده از آنژیوگرافی ریه تأیید شده بود، به‌کار گرفته شد. یافته‌های ارزیابی سامانه تصمیم‌یار طراحی شده در این پژوهش (BayPAD) در مدل اولیه، دقتی برابر ۸۸/۶ درصد داشت (۱۲). Logia tang و همکاران (۲۰۱۰) در پژوهشی با عنوان «یک روش تشخیصی برای آمبولی ریه به کمک شبکه‌های عصبی مصنوعی و با استفاده از داده‌های بالینی» آن‌ها ازمدل شبکه عصبی پس انتشار خطا استفاده کردند و داده‌های مربوط به ۱۱۲ بیمار مشکوک به آمبولی ریه ویژگی‌های ددیمر، علائم بالینی، اطلاعات دموگرافیک، گاز های خون شریانی، را مورد استفاده قرار دادند «۱۳ ».

با توجه به اهمیت پیشگیری، تشخیص و درمان آمبولی ریه و لزوم انجام اقدامات درمانی مناسب و به‌موقع، این مطالعه به‌منظور مقایسه عملکرد روش‌های مختلف یادگیری ماشین در پیش‌بینی وقوع آمبولی ریه در بیماران بستری انجام شد.

روش کار

داده‌های مورد نیاز برای اجرای پژوهش، مهم‌ترین عوامل خطرزای وقوع آمبولی ریه مربوط به بیماران بستری بخش ریه در ۶ ماه پایانی سال ۱۳۹۵ بود که از بخش ریه و مدارک پزشکی بیمارستان شریعتی تهران جمع‌آوری شد. با مطالعه پژوهشی که تاکنون در این زمینه انجام شده و مشورت پزشکان متخصص ریه، ۳۵ ویژگی مرتبط با مسأله پژوهش مشخص شدند. منبع جمع‌آوری اطلاعات، شرح حال کارآموز و کارورز، گزارش پرستار و یادداشت پزشک معالج در سیر بیماری روزانه بود. تمامی اطلاعات به‌دست آمده از بیماران در مطالعه در فرم مخصوص ثبت شده و

[Downloaded from irje.tums.ac.ir on 2025-04-04]

فرم‌های اطلاعاتی به وسیله پژوهش‌گر به صورت بانک اطلاعاتی در فرمت اکسل جمع‌آوری شد تا برای تحلیل‌های آماری قابل استفاده شود. در پیش‌پردازش، برای بهبود کیفیت داده‌های واقعی، به دلیل ثبت نشدن یا عدم شناسایی برخی از ویژگی‌ها، با مشورت متخصصان، ابتدا مواردی که از حوزه این پژوهش خارج بودند، حذف شدند؛ مانند هورمون درمانی، شرایط ترومبوفیلی شناخته شده، سابقه سقط، بیماری حاد التهابی روده[۱] و پرونده‌هایی که به هر دلیل نتیجه آن‌ها مشخص نبود. در مجموع، با تأیید پزشک متخصص، در این پایگاه، داده‌ها با ۲۴ ویژگی با متغیر مختلف اندازه‌گیری شدند که به طور مستقیم یا غیر مستقیم با وقوع آمبولی ریه در بیماران بستری مرتبط بودند. متغیرهایی که در این پژوهش، مورد بررسی قرار گرفته‌اند، در مجموع ۲۴ ویژگی منتخب، شامل جنس، سن، سرطان فعال[۲]، سابقه ترومبوآمبولیک ریوی[۳]، سابقه ترومبوز وریدی عمقی[۴]، شاخص توده بدن[۵]، بی‌حرکتی طولانی مدت، سندروم آنتی فسفولیپید[۶]، لوپوس[۷]، بهجت[۸] پنومونی[۹]، بیماری مزمن انسدادی ریوی[۱۰]، سکته قلبی[۱۱]، نارسایی قلبی[۱۲]، نارسایی بطن چپ قلب[۱۳]، سکته مغزی[۱۴]، عفونت[۱۵]، سندروم نفروتیک[۱۶]، فشارخون بالا[۱۷]، جراحی اخیر، سوء مصرف داخل وریدی مواد[۱۸]، سیگار، دیابت و کاتتر وریدی مرکزی[۱۹]، به عنوان متغیر ورودی سامانه و ۱ متغیر وابسته (هدف)، به‌عنوان خروجی سامانه، که در این‌جا خصیصه آمبولی ریه[۲۰] انتخاب می‌شود و شامل دو حالت وقوع و عدم وقوع است (Presence/Absence) در مورد هر پرونده، جمع‌آوری شده است. فیلدهای پایگاه داده جمع‌آوری شده در جدول شماره ۱ مشخص شده است.

[۱] Inflammatory Bowel Disease
[۲] Active cancer
[۳] Pulmonary thromboembolism
[۴] Deep vein thrombosis
[۵] Body Mass Index
[۶] Antiphospholipid syndrome; APS
[۷] Systemic lupus erythematosus
[۸] Behçet's disease
[۹] Pneumonia
[۱۰] Chronic obstructive pulmonary disease; COPD
[۱۱] Myocardial infarction
[۱۲] congestive heart failure
[۱۳] Ischemic Heart Disease
[۱۴] Cerebro vascular
[۱۵] Sepsis
[۱۶] Nephrotic syndrome
[۱۷] Hypertension
[۱۸] IV drug abuser
[۱۹] Central venous catheter (CV Line)
[۲۰] Pulmonary embolism

بر اساس مطالعه‌های انجام شده، از میان مرسوم‌ترین الگوریتم‌های کاربردی یادگیری ماشین مربوط به پیش‌بینی یافته‌های بیماران مبتلا به سرطان، استفاده از روش‌هایی مانند شبکه عصبی و ماشین بردار پشتیبان، بسیار متداول است (۱۵-۱۴). در زیر به نمونه‌ای از پژوهش‌های انجام شده در زمینه پیش‌بینی بیماری‌ها با استفاده از الگوریتم‌های یادگیری ماشین اشاره شده است:

Lu و همکاران در زمینه پیش‌بینی بیماری عروق کرونری از شبکه بیزی، ماشین بردار پشتیبان و K نزدیک‌ترین همسایه استفاده کردند و نشان دادند که مدل شبکه بیزی در صحت و دقت نسبت به سایر روش‌های ذکر شده برتری دارد (۱۶).

Wang و همکاران، از مدل شبکه های بیزی و ماشین بردار پشتیبان و رگرسیون لجستیک در پیش‌بینی سرطان ریه استفاده نموده‌اند و نشان دادند که مدل شبکه بیزی در صحت و دقت نسبت به سایر روش‌های ذکر شده برتری دارد (۱۲).

Konstantina و همکاران در پژوهشی با استفاده از از شبکه‌های عصبی، شبکه‌های بیزی، ماشین بردار پشتیبان و درخت تصمیم در زمینه تحقیقات سرطان، به توسعه مدل پیش‌بینی سرطان پرداختند. یافته‌های این پژوهش نشان داد که دو روش شبکه‌های عصبی و ماشین بردار پشتیبان نسبت به سایر روش‌ها برتری دارند (۱۵).

بر اساس مروری که در این پژوهش بر مطالعه‌های گذشته در زمینه پیش‌بینی بیماری‌ها صورت پذیرفت، در زمینه پیش‌بینی پزشکی از سال ۲۰۰۷ تا ۲۰۱۷ میلادی صورت گرفت تا کنون ۴ دسته الگوریتم یادگیری ماشین در حوزه پیش‌بینی بیماری مورد استفاده واقع شده‌اند که در زیر دسته‌بندی‌ها ارائه شده و در ادامه شرح مختصری از آن‌ها را خواهیم داشت:

- **مدل‌سازی داده‌ها با روش رگرسیون لجستیک[۲۱]**

رگرسیون لجستیک، رایج‌ترین الگوریتم آماری است که در تحقیقات بالینی به منظور ارزیابی ارتباطات بین ویژگی‌های بیمار و متغیرهای دوحالتی به کار گرفته می‌شود. مدل رگرسیون لجستیک، تعمیم‌یافته مدل خطی آن است و مناسب استفاده در برآورد حداکثر درستنمایی هستند. این روش نسبتا ساده پارامتری، به‌طور گسترده در ممیزی بالینی[۲۲] مورد استفاده قرار می‌گیرد (۱۷).

[۲۱] Logistic regression
[۲۲] Audit Clinical

[Downloaded from irje.tums.ac.ir on 2025-04-04]

- **مدل‌سازی داده‌ها با روش ماشین بردار پشتیبان[1]**

ماشین بردار پشتیبان یکی از روش‌های یادگیری با نظارت است که هم برای دسته‌بندی و هم برای رگرسیون قابل استفاده است. ماشین بردار پشتیبان در اصل یک دسته‌بندی کننده دو کلاسی است که کلاس‌ها را توسط یک مرز خطی از هم جدا می‌کند. اخیراً از این روش یادگیری، بیشتر از سایر روش‌های یادگیری با نظارت دیگر در پیش‌بینی/ پیش‌آگهی سرطان استفاده می‌شود (۱۵).

- **مدل‌سازی داده‌ها با روش درخت تصمیم**

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی است. ساختار درخت تصمیم در یادگیری ماشین، یک مدل پیش‌بینی کننده است که حقایق مشاهده شده در مورد یک پدیده را به استنتاج‌هایی در مورد مقدار هدف آن پدیده نقش می‌کند (۱۸).

برای پیاده‌سازی مدل پیش‌بینی وقوع آمبولی ریه مبتنی بر شبکه‌های بیزی از نرم‌افزار GENIE استفاده شده است. همچنین پیاده‌سازی روش‌های مختلف یادگیری ماشین، با استفاده از نرم‌افزار Weka صورت پذیرفت، در هر یک از مدل‌ها، وقوع آمبولی ریه در بیماران بستری با توجه به عوامل خطرزای بر وقوع آن پیش‌بینی می‌شود. و در نهایت یافته‌های پیش‌بینی حاصل از هر یک در دو حالت (مجموعه داده پیش و پس از انتخاب ویژگی) با یکدیگر مقایسه می‌شوند.

به کمک انتخاب ویژگی‌ها[2]، ویژگی‌های مورد نیاز از بین سایر ویژگی‌ها انتخاب، و اهمیت و سهم هر ویژگی با توجه به متغیر هدف[3] به روش فیلتر (روش InfoGainAttributeEval) مورد بررسی قرار گرفت.

به‌منظور بررسی اثر بخشی مدل پیش‌بینی احتمال وقوع آمبولی ریه از معروف‌ترین معیارهای ارزیابی الگوریتم‌های یادگیری ماشین استفاده شد که در رابطه‌های ۱ تا ۳ این سه معیار معرفی شده‌اند.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (۱)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (۲)$$

$$Specificity = \frac{TN}{FP + TN} \quad (۳)$$

[1] Support Vector Machine
[2] Attribute Selection
[3] Pulmonary Embolism

یافته‌ها

پس از این‌که پایگاه داده جمع‌آوری شده در فرمت CSV ذخیره شد، به طراحی مدل پیش‌بینی با رویکرد شبکه بیزی و سایر روش‌های یادگیری ماشین در دو حالت (مجموعه داده پیش و پس از انتخاب ویژگی) پرداخته شد که در ادامه به صورت مجزا به توصیف هر یک پرداخته خواهد شد:

مدل پیش‌بینی آمبولی ریه با استفاده از شبکه‌های بیزی

پس از این‌که داده‌های جمع‌آوری شده در فرمت Csv در نرم‌افزار GENIE بارگذاری شد، به طراحی مدل‌های شبکه بیزی مورد نظر در پژوهش پرداخته شد. از میان روش‌های مختلف یادگیری ساختار شبکه بیزی با توجه به توصیف مسأله مورد نظر در پژوهش، از روش مبتنی بر رتبه استفاده کرده و از بین الگوریتم‌های مختلف ساخت شبکه موجود در نرم‌افزار GENIE، از الگوریتم ANB[4] استفاده شد. با اعمال الگوریتم یادگیری، مدل شبکه بیزی برای محاسبه وقوع آمبولی ریه به‌دست آمد. شکل ۱ مدل شبکه بیزی به‌دست آمده برای محاسبه احتمال وقوع آمبولی ریه را نشان می‌دهد.

سپس عوامل مؤثر در وقوع آمبولی ریه در نمونه پژوهشی تعیین شدند و به ترتیب میزان تأثیر در وقوع بیماری فهرست شدند. در این پژوهش و بر اساس داده‌هایی که از بیمارستان شریعتی تهران دریافت شد، مهم‌ترین عامل در وقوع آمبولی ریه با اهمیت نسبی ۰/۳۴۹۴ متغیر سابقه ترومبوآمبولیک ریوی است. شکل شماره ۲، اهمیت نسبی عوامل مؤثر در وقوع آمبولی ریه نسبت به متغیر هدف که در این پژوهش Pulmonary Embolism است، را مشخص می‌کند در مرحله بعد، به طراحی مدل پیش‌بینی وقوع آمبولی ریه، مبتنی بر شبکه‌های بیزی و با استفاده از انتخاب ویژگی، پرداخته شدند. انتخاب خصیصه، به روش فیلتر (روش InfoGainAttributeEval در نرم‌افزار Weka) صورت پذیرفت. اهمیت عوامل تأثیرگذار در وقوع آمبولی ریه به‌ترتیب نزولی عبارت بودند از: سابقه ترومبوآمبولیک ریوی، سابقه ترومبوز ورید عمقی، شاخص توده بدن، لوپوس، سندروم آنتی فسفولیپید، نارسایی قلبی، جراحی اخیر، پنومونی، کاتتر ورید مرکزی، جنس، دیابت، سکته قلبی، سیگار، بی‌حرکتی طولانی مدت، سکته مغزی، نارسایی بطن چپ قلب، سندروم نفروتیک، بهجت، سوء مصرف داخل وریدی مواد، بیماری مزمن انسدادی ریوی، سرطان فعال،

[4] Augmented naïve Bayes

[Downloaded from irje.tums.ac.ir on 2025-04-04]

عفونت، فشار خون بالا، سن. پس از مشخص شدن اهمیت هر یک از عوامل به ترتیب نزولی درصد احتمال هر کدام از آن‌ها در وقوع آمبولی ریه، به طراحی مدل بهینه پیش‌بینی آمبولی ریه پرداخته شد. گفتنی است برای بررسی مناسبت مدل از سناریوی زیر استفاده شد:

با توجه به مسأله پژوهش و ضرورت پیش‌بینی صحیح وقوع آمبولی ریه، برای حصول مدل بهینه پیش‌بینی آمبولی ریه، در گام نخست، از یک مجموعه داده تک عنصری که عامل با بالاترین درصد احتمال در وقوع آمبولی ریه بود، شروع کرده و سپس الگوریتم یادگیری ساختار شبکه بیزی مورد استفاده در پژوهش (ANB) روی این مجموعه داده اعمال شد و معیارهای ارزیابی حساسیت و سطح زیر منحنی[1] را برای هر مجموعه داده‌ای محاسبه شد. به همین ترتیب برای مجموعه داده‌های بعدی (شامل ۲ ویژگی، ۳ ویژگی، ۴ ویژگی و ..... ۲۴ ویژگی) نیز معیارهای ارزیابی یاد شده محاسبه شدند. مقایسه این زیر مجموعه از داده‌ها به صورت بصری در نمودار شماره ۱ نمایش داده شده است.

همان‌طور که در نمودار یاد شده مشاهده می‌شود، بر اساس این معیارهای ارزیابی، بهترین تعداد زیرمجموعه بهینه از مجموعه کل ویژگی‌های داده‌های پژوهش، در تعداد ویژگی‌های برابر ۸ است و بر اساس این تعداد در شکل ۳ مدل پیش‌بینی آمبولی ریه با استفاده ازانتخاب ویژگی، طراحی شده است.

مقایسه مدل شبکه‌های بیزی با تکنیک‌های یادگیری ماشین

در این قسمت، نتایج مدل‌های پیش‌بینی وقوع آمبولی ریه با استفاده از سایر روش‌های یادگیری ماشین شرح داده شده است: برای ارزیابی مدل‌ها از شاخص‌های دقت، حساسیت و ویژگی و سطح زیر منحنی استفاده شد.

درخت تصمیم J48

در این پژوهش از درخت تصمیم J48 نرم‌افزار Weka که مبتنی بر الگوریتم C4.5 است، استفاده شده است. این درخت تصمیم با استفاده از مجموعه داده پیش و پس از انتخاب ویژگی، مورد سنجش قرار گرفت. گفتنی است ویژگی‌های دارای شرایط پس از اعمال انتخاب ویژگی و پس از سنجش معیارهای ارزیابی حساسیت و مقدار سطح زیر منحنی بر اساس سناریوی پیشنهادی، شامل ۸ ویژگی {HPTE، HDVT، BMI>30، Antiphospholipid، SLE، CHF، Rsurgery، Pneumonia} هستند.

رگرسیون لجستیک

در این پژوهش، از رگرسیون لجستیک نرم‌افزار Weka، استفاده شده است. پس از اعمال روش فیلتر برای استخراج ویژگی روی داده‌ها و پس از سنجش معیارهای ارزیابی حساسیت و مقدار سطح زیر منحنی برای آن‌ها بر اساس سناریوی پیشنهادی، ویژگی‌های دارای شرایط، شامل سه ویژگی {HPTE، HDVT، BMI>30} هستند.

بهینه‌سازی حداقل متوالی

در این پژوهش، از بهینه‌سازی حداقل متوالی (SMO) (Sequential Minimal Optimization) که الگوریتم یادگیری ماشین بردار پشتیبان (SVM) است و اغلب سریع‌تر از آن آموزش داده می‌شود، استفاده شده است. پس از اعمال روش فیلتر برای استخراج ویژگی روی داده‌ها و پس از سنجش معیارهای ارزیابی حساسیت و مقدار سطح زیر منحنی برای آن‌ها بر اساس سناریوی پیشنهادی، ویژگی‌های دارای شرایط، شامل دو ویژگی {HPTE، HDVT} هستند.

مقایسه نتایج مدل‌های شبکه بیزی با روش‌های یادگیری ماشین

در جدول شماره ۲، نتایج ارزیابی مدل‌های پیش‌بینی آمبولی ریه، پیش از انتخاب ویژگی (با ویژگی‌های جدول شماره ۱) بدین صورت است: در جدول ۳ نیز، نتایج ارزیابی مدل‌های پیش‌بینی آمبولی ریه، پس از انتخاب ویژگی نشان داده شده است.

با توجه به نتایج ارزیابی مدل‌های پیش‌بینی (جدول شماره ۲)، و همان‌طور که انتظار می‌رفت، عملکرد برآوردها با تغییر در تعداد ویژگی‌ها به مقدار چشمگیری تغییر نمی‌کند. تنها معیار ارزیابی که پس از انتخاب ویژگی دستخوش تغییرات مطلوبی در مدل پیش‌بینی وقوع آمبولی ریه شد، معیار حساسیت بود که از اهمیت ویژه‌ای در پیش‌بینی و تشخیص‌های پزشکی برخوردار است.

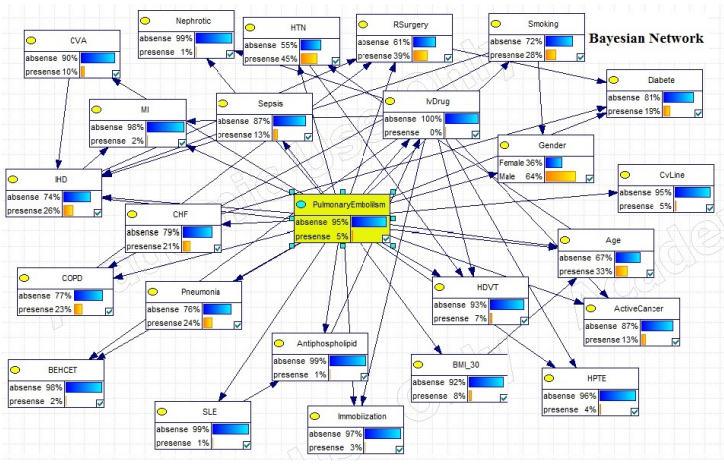
در یک بررسی سیستماتیک در سال ۲۰۱۶ میلادی نشان داده شده است که پیش‌بینی اغلب بیماری‌ها با رویکرد شبکه‌های بیزی دارای عملکرد بهتری نسبت به سایر روش‌های دسته‌بندی یادگیری ماشین است (۱۹). در نمودار شماره ۲، یافته‌های مقایسه مدل‌های شبکه بیزی با روش‌های یادگیری ماشین به صورت بصری نشان داده شده است.

[1] Area Under the Curve; AUC

[Downloaded from irje.tums.ac.ir on 2025-04-04]

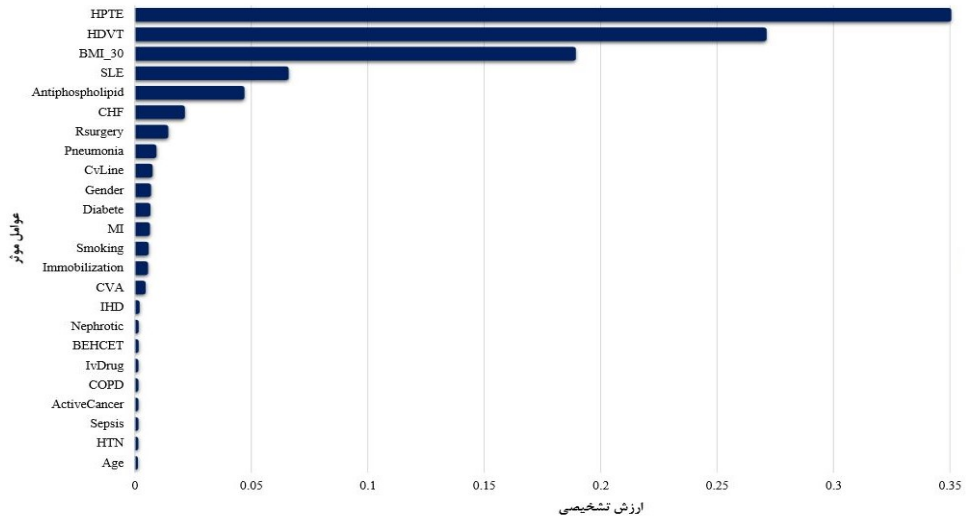
جدول شماره ۱ – متغیرهای اصلی انتخاب شده، توصیف و نوع آن‌ها برای بررسی در این مطالعه

شماره	ویژگی‌ها	نوع			نقش آن در پژوهش		واحد اندازه گیری
		کیفی	کمی گسسته	کمی پیوسته	مستقل	وابسته	
۱	جنس	*			*		زن/ مرد
۲	سن		*		*		سال
۳	سرطان فعال	*			*		دارد/ ندارد
۴	سابقه PTE	*			*		دارد/ ندارد
۵	سابقه DVT	*			*		دارد/ ندارد
۶	شاخص توده بدن (BMI>30)			*	*		دارد/ ندارد
۷	بی‌حرکتی طولانی مدت	*			*		دارد/ ندارد
۸	سندروم آنتی فسفولیپید	*			*		دارد/ ندارد
۹	لوپوس (SLE)	*			*		دارد/ ندارد
۱۰	بهجت	*			*		دارد/ ندارد
۱۱	پنومونی (ذات‌الریه)	*			*		دارد/ ندارد
۱۲	بیماری مزمن انسدادی ریوی (COPD)	*			*		دارد/ ندارد
۱۳	سکته قلبی (MI)	*			*		دارد/ ندارد
۱۴	نارسایی قلبی (CHF)	*			*		دارد/ ندارد
۱۵	نارسایی بطن چپ قلب (IHD)	*			*		دارد/ ندارد
۱۶	سکته مغزی (CVA)	*			*		دارد/ ندارد
۱۷	عفونت	*			*		دارد/ ندارد
۱۸	سندروم نفروتیک	*			*		دارد/ ندارد
۱۹	فشارخون بالا (HTN)	*			*		دارد/ ندارد
۲۰	جراحی اخیر	*			*		دارد/ ندارد
۲۱	سوء مصرف داخل وریدی مواد	*			*		دارد/ ندارد
۲۲	سیگار	*			*		دارد/ ندارد
۲۳	دیابت	*			*		دارد/ ندارد
۲۴	کاتتر وریدی	*			*		دارد/ ندارد
۲۵	آمبولی ریه	*			*	*	دارد/ ندارد

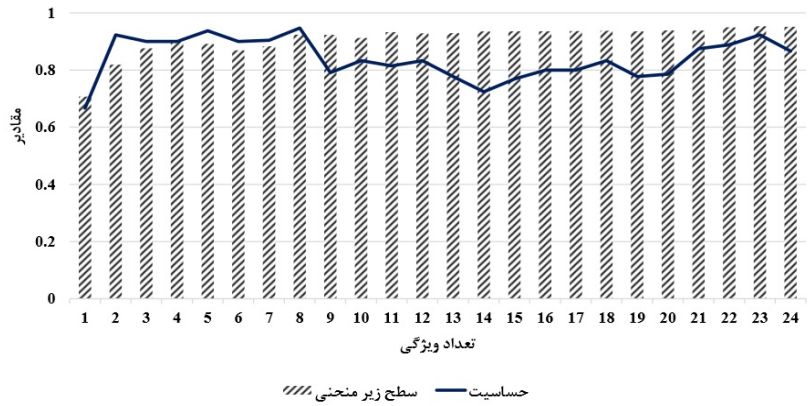


شکل شماره ۱ – مدل پیش‌بینی آمبولی ریه با شبکه‌های بیزی بدون انتخاب ویژگی

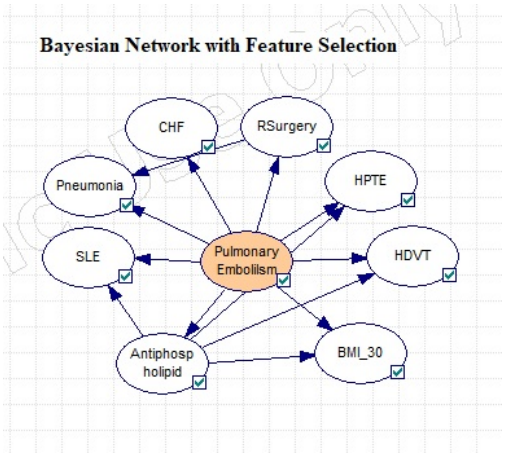
[Downloaded from irje.tums.ac.ir on 2025-04-04]



شکل شماره ۲- اهمیت عوامل تأثیرگذار در وقوع آمبولی ریه در بیماران بستری



گراف شماره ۱- انتخاب زیرمجموعه ویژگی‌ها در طراحی مدل پیش‌بینی با انتخاب ویژگی



شکل شماره ۳ – مدل پیش‌بینی آمبولی ریه با استفاده از انتخاب ویژگی

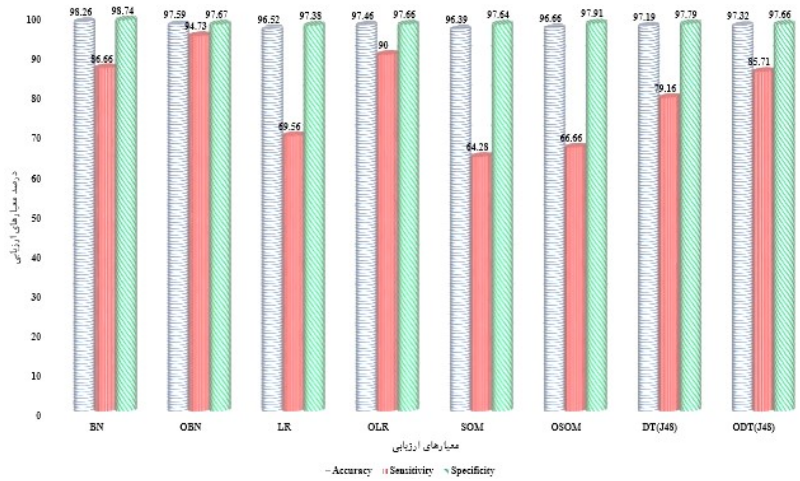
جدول شماره ۲ – نتایج ارزیابی مدل‌های پیش‌بینی آمبولی ریه، پیش از انتخاب ویژگی

[Downloaded from irje.tums.ac.ir on 2025-04-04]

سنجه ارزیابی عملکرد / روش	صحت	حساسیت	ویژگی	سطح زیر منحنی
شبکه بیزی	۰/۹۸۲۶۴۴	۰/۸۶۶۶۶۶	۰/۹۸۷۴۸۲	۰/۹۵۸۲۰۳
درخت تصمیم	۰/۹۷۱۹۶۳	۰/۷۹۱۶۶۶	۰/۹۷۷۹۳۱	۰/۷۷۹
رگرسیون لجستیک	۰/۹۷۵۲۸۷	۰/۶۹۵۶۵۲	۰/۹۷۳۸۲۹	۰/۸۶۱
بهینه‌سازی حداقل متوالی	۰/۹۶۳۹۵۲	۰/۶۴۲۸۵۷	۰/۹۷۶۴۲۱	۰/۷۵۰

جدول شماره ۳ – نتایج ارزیابی مدل‌های پیش‌بینی آمبولی ریه، پس از انتخاب ویژگی

سنجه ارزیابی عملکرد / روش	صحت	حساسیت	ویژگی	سطح زیر منحنی
شبکه بیزی	۰/۹۷۵۹۶۷	۰/۹۴۷۳۶۸	۰/۹۷۶۷۱۲	۰/۹۲۰۷۰۸
درخت تصمیم	۰/۹۷۳۲۹۸	۰/۸۵۷۱۴۲	۰/۹۷۶۶۴۸	۰/۷۴۳
رگرسیون لجستیک	۰/۹۷۴۶۳۳	۰/۹	۰/۹۷۶۸	۰/۸۶۳
بهینه‌سازی حداقل متوالی	۰/۹۶۶۶۲۲	۰/۶۶۶۶۶۶	۰/۹۷۹۱۳۷	۰/۷۷۹



گراف شماره ۲ – مقایسه نتایج بررسی روش‌های دسته‌بندی گوناگون با استفاده از معیارهای ارزیابی

بحث

امروزه پیش‌بینی[1] نقش مهم و تعیین کننده‌ای را در پزشکی داشته و روند تشخیص، درمان و کنترل بیماری‌ها را تسهیل می‌بخشد (۲۰). از آن‌جا که در پژوهش‌های علوم پزشکی مسأله سلامت انسان مطرح است، پیش‌بینی درست نتایج اهمیت بیشتری می‌یابد.

بنابراین باید از روش‌هایی استفاده شود که پیش‌بینی بر اساس آن‌ها دارای کمترین خطا و بیش‌ترین اعتماد باشد. از جمله روش‌هایی که توجه بسیاری از پژوهشگران را به خود معطوف داشته است، شبکه‌های بیزی است. شبکه بیزی یکی از تکنیک‌های داده‌کاوی است که در سال‌های اخیر به عنوان یک تکنیک قدرتمند برای بررسی در حوزه‌های پیچیده و دارای عدم قطعیت به کار می‌رود و یک روش اساسی برای طبقه‌بندی و شناخت الگو محسوب می‌شود (۲۴-۲۱). این شبکه برای مدل‌سازی تشخیص و پیش‌بینی بیماری‌های گوناگون به‌طور موفقیت‌آمیزی کاربرد داشته است. این روش برای تشخیص پنومونی[2] و سرطان سینه[3] به کار برده شده است (۲۶-۲۵).

آمبولی ریه یک بیماری تهدیدکننده حیات است. شیوع بالای این بیماری، تشخیص نسبتاً مشکل، عدم دسترسی آسان به

[1] Prediction

[2] Pneumonia
[3] Breast Cancer

[Downloaded from irje.tums.ac.ir on 2025-04-04]

روش‌های تشخیصی و هزینه بالای انجام آن‌ها که در تشخیص آمبولی ریه مورد نیاز هستند، باعث ایجاد مشکلاتی در سامانه بهداشت و درمان کشور شده است، بنابراین چنین به‌نظر می‌رسد که ارایه مدلی برای تعیین و میزان اثرگذاری عوامل خطرزای آمبولی ریه می‌تواند فرآیند غربالگری را تسهیل کند به طوری که در صورت غیر نرمال بودن وضعیت بیمار، اقدامات تشخیصی دقیق‌تری روی ایشان انجام شود و به این ترتیب شانس زنده ماندن بیمار، افزایش یابد. با توجه به اهمیت پیشگیری، تشخیص و درمان آمبولی ریه و لزوم انجام اقدامات درمانی مناسب و به‌موقع، این مطالعه را به‌منظور ارزیابی اپیدمیولوژی، ویژگی‌های اساسی و تعیین عوامل خطرزای آمبولی ریه بیماران بستری در ۶ ماهه پایانی سال ۱۳۹۵ در بخش ریه بیمارستان شریعتی تهران انجام شد.

در این مطالعه تلاش شد که ضمن توجه به منابع تجربی و تشخیصی پزشکی، روابط بین عوامل خطرزای آمبولی ریه در کنار سایر عوامل مؤثر بر وقوع آن مورد بررسی قرار گیرد. یافته‌های به‌دست آمده از این پژوهش، گویای این مطلب است که می‌توان مدعی شد، متغیرهای سابقه ترومبوآمبولیک ریوی، سابقه ترومبوز ورید عمقی، شاخص توده بدن، جراحی اخیر، بی‌حرکتی طولانی مدت، لوپوس، سندروم آنتی فسفولیپید، سکته قلبی و پنومونی، بیشترین تأثیر را بر روند پیش‌بینی وقوع آمبولی ریه دارند و این نتیجه با مطالعه‌های نظری در زمینه اثرپذیری این متغیرها مصداق دارد. در مقایسه مدل‌های شبکه بیزی با روش‌های یادگیری ماشین که در این پژوهش استفاده شد، همان‌طور که انتظار می‌رفت، عملکرد برآوردها با تغییر حجم نمونه به مقدار چشمگیری تغییر نمی‌کند. به عبارت دیگر، دقت برآوردها تحت تأثیر حجم نمونه نیست. تنها معیار ارزیابی که پس از انتخاب ویژگی دستخوش تغییرات چشمگیری شد، معیار حساسیت بود که در روش‌های بیزی، رگرسیون لجستیک و درخت تصمیم به میزان قابل توجهی افزایش یافت. همچنین، همان‌طور که در یافته‌های روش‌های رگرسیون لجستیک، بهینه‌سازی حداقل متوالی و درخت تصمیم دیده می‌شود، استناد به دقت به‌عنوان تنها سنجه ارزیابی کننده عملکرد مدل می‌تواند گمراه کننده باشد. زیرا دقت در این روش‌ها مطلوب است، اما حساسیت پایین است.

شبکه‌های بیزی یکی از بهترین روش‌های پیش‌بینی بیماری‌ها در مقایسه با نظر متخصصان و الگوریتم‌های موجود دیگر است که می‌تواند به‌عنوان روش حمایتی در کنار تصمیمات پزشکی قرار گیرد تا صحت پیش‌بینی بیماری‌ها را ارتقا دهد.

در توجیه عملکرد بهتر روش شبکه‌های بیزی در زمینه پیش‌بینی بیماری‌ها در مقایسه با سایر روش‌ها، می‌توان به نکات زیر اشاره کرد:

- ساختار شبکه‌های بیزی رابطه داخلی بین ویژگی‌ها را بیان می‌کنند.
- شبکه‌های بیزی می‌توانند برای مجموعه‌های ناقص از داده‌ها هم استفاده شوند.
- پژوهشگران می‌توانند از دانش تخصصی خود در فرایند کشف دانش با این رویکرد استفاده کنند در حالی که سایر تکنیک‌های داده کاوی به طور عمده روی داده‌های کدگذاری شده برای کشف دانش اعمال می‌شوند.
- متخصصان می‌توانند ساختار آن را به راحتی دریابند و بر این اساس، به آسانی خروجی مدل شبکه‌های بیزی را درک و تفسیر کنند.
- الگوریتم‌های شبکه بیزی بر این دانش کدگذاری شده‌اند تا کارایی و دقت خود را در فرآیند مدل‌سازی و عملکرد پیش‌بینی مدل افزایش دهند.
- شبکه‌های بیزی در تعامل متغیرهای ورودی[1]، عملکرد بهتری دارند. گرچه در برخی موارد ممکن است درخت تصمیم‌گیری یا درخت رگرسیون[2] دسته‌بندی دقیق‌تری تولید کنند. چرا که آن‌ها تنها روابط بین خروجی و متغیرهای ورودی را در نظر می‌گیرند.
- مدل شبکه‌های بیزی در مقایسه با سایر تکنیک‌های داده‌کاوی دارای پیش‌بینی دقیق‌تری در وضعیتی هستند که اطلاعات کامل در دسترس نیست.

نشان داده شد، شبکه‌های بیزی دارای کارایی بالاتری نسبت به سایر روش‌های یادگیری ماشین در زمینه پیش‌بینی وقوع آمبولی ریه در بیماران بستری است.

تشکر و قدردانی

مقاله حاضر برگرفته از پایان‌نامه کارشناسی ارشد در رشته انفورماتیک پزشکی می‌باشد که به شماره ۵۲د/۷۷۵۷ در دانشکده علوم پزشکی دانشگاه تربیت مدرس تصویب گردید. در خاتمه از مسئولان محترم بیمارستان شریعتی تهران که در گردآوری مجموعه

[1] Input variables

[2] Classification and Regression Trees(CART)

[Downloaded from irje.tums.ac.ir on 2025-04-04]

داده مربوط از مسؤولان محترم بیمارستان شریعتی تهران که در گردآوری مجموعه داده‌های مربوط همکاری نمودند، صمیمانه تشکر و قدردانی می‌شود.

منابع

1. Silverstein M, Heit J, David M, Petterson TM, O'Fallon W, Melton III J. Trends in the Incidence of Deep Vein Thrombosis and Pulmonary Embolism. Arch Intern Med, 1998; 158: 585–93.
2. Remy-Jardin M, Pistolesi M, Goodman LR, Gefter WB, Gottschalk A, Mayo JR, et al. Management of suspected acute pulmonary embolism in the era of CT angiography: a statement from the Fleischner Society. Radiology, 2007; 245: 315–29.
3. Morris TA. Natural History of Venous Thromboembolism. Crit Care Clin, 2011; 27: 869–84.
4. Arora H, Painter N, Kumar P. Acute pulmonary embolism during an endoscopic retrograde cholangiopancreatography. Ann Card Anaesth, 2014; 17: 145.
5. Concepts C, Agnelli G, Becattini C, States U. Acute pulmonary embolism. N Engl J Med, 2010; 266–74.
6. Fedullo PF, Tapson VF. The evaluation of suspected pulmonary embolism. New England Journal of Medicine, 2003; 349: 1247-56.
7. Progress M, Tapson VF. Acute Pulmonary Embolism, 2009; 1037–52.
8. Michie D, Spiegelhalter D, Taylor C. Machine learning, neural and statistical classification, 1994.
9. Li M, Zhou Z. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Trans Syst Man, 2007.
10. Agharezaei L, Agharezaei Z, Nemati A, Bahaadinbeigy K, Keynia F, Baneshi MR, Iranpour A, Agharezaei M. The Prediction of the Risk Level of Pulmonary Embolism and Deep Vein Thrombosis through Artificial Neural Network. Acta Informatica Medica, 2016.
11. Jafari M. The design of the medical decision-making system helps the diagnosis of pulmonary embolism using artificial neural networks, Master of Science (MSc), Tehran, Medical Sciences of Tehran, 2015.
12. Luciani D, Cavuto S, Antiga L, Miniati M, Monti S, Pistolesi M, et al. Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. Emerg Med J, 2007; 24: 157–64.
13. Tang L, Wang L, Pan S, Su Y, Chen Y. A neural network to pulmonary embolism aided diagnosis with a feature selection approach. InBiomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on 2010 Oct 16,. 6: 2255-60.
14. Simes R. Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer. J Chronic Dis, 1985.
15. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal. 2015; 13: 8-17.
16. Cooper JN, Wei L, Fernandez SA, Minneci PC, Deans KJ. Pre-operative prediction of surgical morbidity in children: Comparison of five statistical models. Computers in biology and medicine. 2015; 57: 54-65.
17. Hosmer, DW.; Lemeshow S. Applied logistic regression. New York: Wiley; 2000; 330-9.
18. Witten I, Frank E, Hall M, Pal C. Data Mining: Practical machine learning tools and techniques, 2016 oct 1; 97-136.
19. Langarizadeh M, Moghbeli F. Applying naive bayesian networks to disease prediction: a systematic review. Acta Inform Medica, 2016 oct; 24: 364
20. Hendriksen JMT, Geersing GJ, Moons KGM, de Groot JAH. Diagnostic and prognostic prediction models. J Thromb Haemost, 2013; 11: 129–41.
21. Heckerman D. Bayesian networks dor data mining. Data Min Knowl Discov, 1997 Mar 1; 1: 79-119.
22. Fayyad U, Piatetsky-Shapiro G, Smyth P. Advances in knowledge discovery and data mining, 1996 Mar 15; 17: 37
23. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference, 2014 Jun 28; 53-79
24. Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: The combination of knowledge and statistical data, 1995 Sep 1; 20: 197-243
25. Aronsky D, Haug P. Automatic identification of patients eligible for a pneumonia guideline. Proc AMIA Symp, 2000: 12-16.
26. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney G A, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. Radiology, 2006; 240: 666–73.

[Downloaded from irje.tums.ac.ir on 2025-04-04]

Iranian Journal of Epidemiology Atumn 2018; 14 (3): 272-282.
Original Article

Determining the Risk Factors and Presenting a Prognostic Model for Pulmonary Embolism in Hospitalized Patients using Bayesian Networks

Feizmanesh F, Safaei AA
1- Master of Science, Tarbiat Modares University, Medical Informatics, Department of Medical Informatics, Faculty of Medical Sciences, Tehran, Iran
2- Assistant Professor, Tarbiat Modares University, Computer Engineering Software, Department of Medical Informatics, Faculty of Medical Sciences, Tehran, Iran
Corresponding author: Safaei, aa.safaei@modares.ac.ir
(Received 24 December 2017**; Accepted** 20 June 2018**)**

Background and Objectives: Pulmonary embolism is a potentially fatal and prevalent event that has led to a gradual increase in the number of hospitalizations in recent years. For this reason, it is one of the most challenging diseases for physicians. The main purpose of this paper was to report a research project to compare different data mining algorithms to select the most accurate model for predicting pulmonary embolism in hospitalized patients. This model would provide the knowledge needed by the medical staff fir better decision making.

Methods: In this research, we designed a prediction model using different methods of machine learning that would best predict the probability of pulmonary embolism in patients at risk. Among data mining algorithms, Bayesian network, decisions tree (J48), logistic regression (LR), and sequential minimal optimization (SMO) were used. The data used in the study included risk factors and past history of patients admitted to the Lung Department of Shariati Hospital, Tehran, Iran.

Results: The results showed that the accuracy and specificity of all prediction models were satisfactory. The Bayesian model had the highest sensitivity in predicting pulmonary embolism.

Conclusion: Although the results showed a little difference in the performance of prediction models, the Bayesian model is a more appropriate tool to predict the occurrence of pulmonary embolism in hospitalized patients in this type of data. It can be considered a supportive approach along medical decisions to improve disease prediction.

Keywords: Pulmonary embolism, Machine learning, Prediction models, Bayesian networks, Decision Tree, Logistic Regression, Sequential minimal optimization, Risk factors

[Downloaded from irje.tums.ac.ir on 2025-04-04]