پیش‌بینی مرگ‌ومیر بیماران در بخش مراقبت‌های ویژه بیمارستان امام حسین(ع) تهران با تکنیک‌های داده‌کاوی

پرنیان عسگری[1] علیرضا آتشی[2] مرضیه معراجی*[3] میر محمد میری[4]

۱. دانشجوی کارشناسی ارشد، فناوری اطلاعات سلامت، گروه مدارک پزشکی و فناوری اطلاعات سلامت، دانشکده علوم پیراپزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.ORCID: 0000-0003-4426-0851

۲. گروه سلامت الکترونیک، دانشکده مجازی، دانشگاه علوم پزشکی تهران، تهران، ایران.

۳. گروه مدارک پزشکی و فناوری اطلاعات سلامت، دانشکده علوم پیراپزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

٤. گروه بیهوشی، دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

مجله اطلاع‌رسانی پزشکی نوین؛ دوره پنجم؛ شماره دوم؛ پاییز و زمستان ۱۳۹۸؛ صفحات۶۷-۵۹.

چکیده

هدف: *حجم عظیمی از داده‌ها در بخش مراقبت‌های ویژه تولید می‌شود، به نظر می‌رسد داده‌کاوی راهکار مناسبی برای استفاده‌ی بهینه از منابع باشد. شناسایی و تحلیل عوامل پرخطر مرتبط با مرگ‌ومیر، منجر به برنامه‌ریزی کاراتر و دقیق‌تر جهت بستری و انجام مداخلات خواهد شد. این پژوهش باهدف استفاده از تکنیک‌های داده‌کاوی جهت پیش‌بینی مرگ‌ومیر در بخش مراقبت ویژه صورت گرفته است.*

روش‌ها: *این پژوهش به روش مقطعی بر روی اطلاعات ۸۳۸ بیمار بستری در بخش مراقبت‌های ویژه عمومی بین سال‌های ۹۱ تا۹۷ در بیمارستان امام حسین(ع) تهران انجام گردید. الگوریتم‌هایی ماشین بردار پشتیبانی، K نزدیک‌ترین همسایه، درخت تصمیم، جنگل تصادفی و رگرسیون لجستیک جهت داده‌کاوی استفاده گردید. مراحل انجام داده‌کاوی طبق مدل کریسپ در پنج مرحله صورت گرفت. ارزیابی مدل بر اساس صحت، دقت، ویژگی، حساسیت و سطح زیر منحنی راک گزارش گردید.*

نتایج: *در ابتدا پس از بررسی متون،۲۷ فاکتور تأثیرگذار مشخص و درنهایت ۲۶ فاکتور برای انجام تکنیک‌ها مورداستفاده قرار گرفت. از میان الگوریتم‌های منتخب که در مطالعه استفاده گردید، الگوریتم رگرسیون لجستیک بر اساس سطح زیر منحنی راک (۷۶/۰)، صحت (۷۵/۶۲)، دقت (۶۸/۳۹)، حساسیت (۳۸/۶۵) ویژگی (۹۴/۵۳) عملکرد بهتری در پیش‌بینی مرگ‌ومیر نسبت به سایر الگوریتم‌های مطالعه داشت. در ضمن متغیرهای گلوکز و زمان نسبی ترومبوپلاستین بیشترین تأثیر را بر مرگ‌ومیر بر اساس مدل رگرسیون لجستیک داشت.*

نتیجه‌گیری: *تجزیه‌وتحلیل داده‌های موجود در بیماران بخش مراقبت‌های ویژه می‌تواند ابزاری مناسب و کاربردی برای پیش‌بینی مرگ‌ومیر و عوامل مرتبط با آن باشد اما با توجه به کیفیت داده‌ها نتایج متفاوت می‌باشد بااین‌حال فرآیندها و روش‌های ذکرشده در این مطالعه بیان می‌کند که قوانین استخراج‌شده از رگرسیون لجستیک می‌تواند به‌عنوان الگویی برای پیش‌بینی وضعیت مرگ‌ومیر در بخش مراقبت‌های ویژه مورداستفاده قرار گیرد.*

کلیدواژه‌ها: *مرگ‌ومیر، داده‌کاوی، بخش مراقبت ویژه، مدل پیش‌بینی.*

نوع مقاله: پژوهشی

دریافت مقاله:۱۳۹۸/۱۰/۶ اصلاح نهایی: ۹۸/۱۲/۱۶ پذیرش مقاله: ۹۸/۱۲/۲۰

ارجاع: عسگری پرنیان، آتشی علیرضا، معراجی مرضیه، میرمحمد میری. پیش‌بینی مرگ‌ومیر بیماران در بخش مراقبت‌های ویژه بیمارستان امام حسین تهران با تکنیک‌های داده‌کاوی. مجله اطلاع‌رسانی پزشکی نوین.۱۳۹۸؛ (۲)۵: ۶۷-۵۹.

مقدمه:

بخش مراقبت‌های ویژه یکی از بخش‌های بیمارستانی است که دارای کادر پزشکی بسیار مجرب و تجهیزات ویژه‌ای برای درمان بیماران می‌باشد که نیاز به نظارت شدید و پشتیبانی مداوم دارد. در حال حاضر بخش‌های مراقب ویژه با دستگاه‌های نظارتی پیشرفته و اختصاص میزان

نویسنده مسئول:

مرضیه معراجی

گروه فناوری اطلاعات سلامت، دانشکده علوم پیراپزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران.

تلفن: ۹۱۵۳۱۲۱۵۷٤۹۸+ پست الکترونیکی:Merajim1@mums.ac.ir ORCID:0000-0002-3137-0683

بالایی از منابع پزشکی در حال فعالیت می‌باشند. بااین‌وجود در این بخش‌ها هنوز هم نرخ بالایی از مرگ‌ومیر بیماران گزارش می‌شود [۱]. با توجه به شرایط فعلی که به‌طور مداوم هزینه‌های بهداشت و درمان رو به افزایش است، بخش مراقبت ویژه بالغ به حدود ۱۳ درصد از هزینه‌های بیمارستان را به خود اختصاص داده است و بااین‌وجود، کمبود منابع در این بخش‌ها دیده می‌شود [۲].

مراقبت و نظارت مستمر از بیماران بستری در بخش مراقبت ویژه تولید ثروتی از داده‌های بالینی در جهت ارائه فرصت‌هایی برای ایجاد مدل‌های پیش‌بینی نموده است. اهمیت این مدل‌ها با پشتیبانی از تصمیم‌گیری‌های بالینی و پیش‌بینی مرگ‌ومیر در ارائه مؤثر مراقبت در بخش مراقبت ویژه، امری شناخته‌شده است. در دهه‌های گذشته، داده‌های جمع‌آوری‌شده در واحدهای مراقبت‌های ویژه افزایش پیداکرده است و در مطالعات آنالیز داده‌ها و داده‌کاوی مورد استفاده قرارگرفته است [۳].

مقدار زیادی از داده‌ها جهت بهبود مراقبت از بیماران بخش مراقبت ویژه وجود دارد که هنوز بلااستفاده باقی‌مانده و یا نیاز به تحلیل بیشتر دارند. علاوه بر این، با توجه به عدم دسترسی و کمبود متخصصان انسانی به دلایل مختلف، مشغله پزشکان یا تازه‌کار بودن آن‌ها جزئیات مهم نادیده گرفته‌شده است. این در حالی است که ابزار کشف خودکار بر اساس مدل‌های مختلف پیش‌بینی می‌تواند داده‌های خام را تجزیه‌وتحلیل و با استخراج اطلاعات برای تصمیم‌گیرندگان امکان تصمیم‌گیری بهتر را فراهم نماید [۵-۳].

در مدل‌های پیش‌بینی مرگ‌ومیر در بخش مراقبت ویژه همواره درصدد پاسخ به این سؤال مهم هستیم که چگونه می‌توان خطر مرگ‌ومیر و عوامل مرتبط با آن را برای بیماران بخش مراقبت‌های ویژه شناسایی نموده و بر اساس داده‌های موجود میزان مرگ‌ومیر در بخش مراقبت‌های ویژه به حداقل رساند [۴].

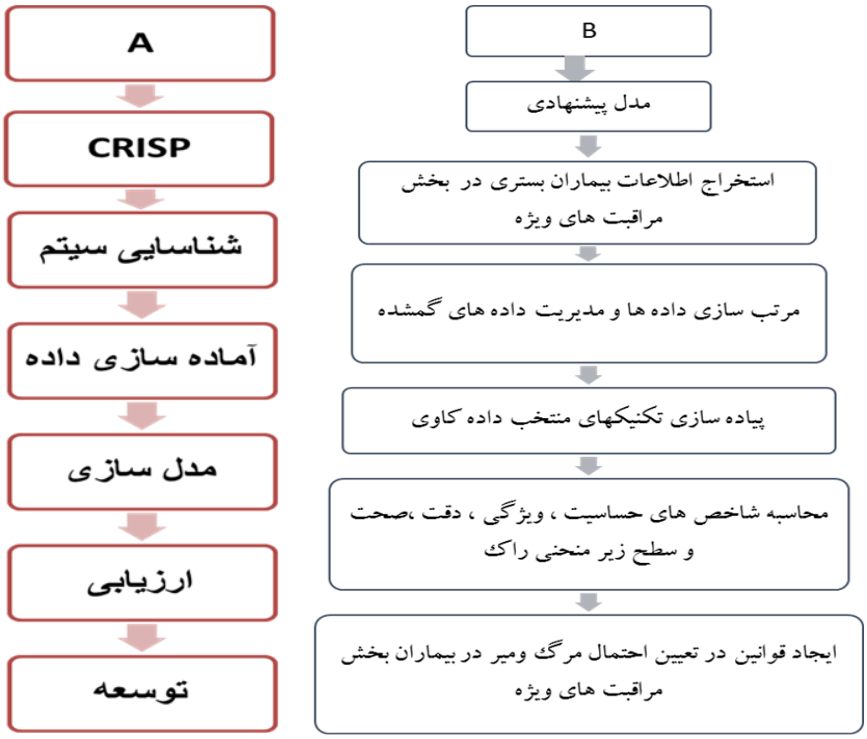
بر اساس مطالعه Vairavanو همکاران در سال ۲۰۱۷ روش‌های آماری استانداردی چون رگرسیون لجستیک از سوی کارشناسان مراقبت‌های ویژه برای پیش‌بینی خطر مرگ‌ومیر یا رویدادهای نامطلوب برای بیماران دچار ناراحتی‌ها یا آسیب‌های شدید در این بخش به‌خوبی مورد استقبال قرار گرفته است [۶].

در حال حاضر پیش‌بینی نوین با استفاده از الگوریتم‌های یادگیری ماشین مثل درخت تصمیم منتهی به ظهور تعدادی از مدل‌های پیش‌بینی در محیط‌های مختلف مراقبت بالینی شده است [۷]. نتایج تحقیق مریم مصلحی و همکاران در سال ۱۳۹۶ در بخش مراقبت‌های ویژه بیمارستان الزهرا (س) اصفهان نشان داد که بر اساس سطح زیر منحنی ROC روش نمره دهی SAPS IIمعتبر و مؤثر می‌باشد [۸]. اطلاعات حاصل از داده‌کاوی می‌تواند به‌عنوان یک منبع جهت، شناسایی روش‌های بهتر و مؤثرتر درمانی به‌ویژه در پیش‌بینی مرگ‌ومیر بیماران مورداستفاده قرار گیرد [۹].

از داده‌های بالینی موجود در بخش‌های مراقبت ویژه در جهت پیش‌بینی مرگ‌ومیر به‌طور مؤثر می‌توان استفاده نمود [۱۰]. شناسایی بیماران در معرض خطر می‌تواند، منجر به تصمیم‌گیری‌های حیاتی مانند قطع درمان، استفاده از تجهیزات لازم، بررسی خطرات درمانی، پیگیری منابع موردنیاز بخش مراقبت ویژه و کاهش طول مدت اقامت در بخش مراقبت ویژه گردد [۱۱]. این پژوهش باهدف استفاده از تکنیک‌های داده‌کاوی جهت ساخت مدل پیش‌بینی مرگ‌ومیر در بخش مراقبت‌های ویژه صورت گرفته است.

مواد و روش‌ها:

داده‌های این مطالعه از ۸۳۸ بیمار بستری در بخش مراقبت‌های ویژه بیمارستان مرکز پزشکی آموزشی، پژوهشی و درمانی امام حسین (ع) تهران بوده است. طبق نمودار ۱ پژوهش بر اساس مدل کریسپ در طی ۵ مرحله انجام‌گرفته است.



نمودار ۱. مراحل انجام پژوهش طبق مدل کریسپ

مرحله اول: شناسایی داده

در این مرحله، از پرونده بیماران بستری در بخش مراقبت‌های ویژه در محدوده زمانی ۹۱ الی ۹۷ اطلاعات بیماران استخراج گردید. معیارهای ورود به مطالعه؛ بیمارانی که در این مدت در بخش مراقبت‌های ویژه عمومی بستری شدند، جراحی قلب نداشتند، داروی روان‌درمانی مصرف نکرده و سن بالای ۱۶ سال بود. فاکتورهای تأثیرگذار طبق مطالعه آتشی و همکاران برای بخش‌های مراقبت‌های ویژه ایران شناسایی گردید [۱۲]. پارامترهای موردمطالعه در خصوص بیماران شامل؛ سن، جنس، GCS، HCT، بیلی‌روبین، پتاسیم، گلوکز، سدیم، WBC، کراتینین، فشارخون دیاستولیک، نرخ تنفس، فشارخون سیستولیک، درجه حرارت، ضربان قلب، PH، فاکتورهای اکسیژناسیون (PaCO2, FiO2, PaO2HCO3)، اوره، پلاکت، زمان نسبی ترومبوپلاستین، فشارخون متوسط شریانی، INR، آلبومین و وضعیت بیمار (فوت‌شده یا زنده) بود.

مرحله دوم: پیش‌پردازش داده‌ها

پس از جمع‌آوری داده‌ها، مرتب‌سازی و مدیریت داده‌ها با توجه به نوع و حالت داده‌ها صورت گرفت. در صورت مفقود بودن داده‌های یک متغیر، مقداری مصنوعی برای داده‌های مفقودشده که در این مطالعه برحسب میانگین در نظر گرفته شد. در ضمن پس از شناسایی داده‌های پرت تصمیم‌گیری در مورد آن‌ها صورت گرفت.

مرحله سوم: مرحله پیاده‌سازی تکنیک‌های داده‌کاوی منتخب

بر اساس مروری بر مطالعات و الگوریتم‌های پیشنهادی در مطالعات قبلی و با توجه به کیفیت داده‌ها انتخاب الگوریتم‌ها صورت گرفت. سعی شد مدل‌های انتخابی، از دودسته تکنیک‌های یادگیری ماشین و آمار کلاسیک باشند.

در این مطالعه از رگرسیون لجستیک و همچنین مدل‌های یادگیری ماشین؛ الگوریتم K نزدیک‌ترین همسایه، درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبانی استفاده گردید.

مرحله چهارم: ارزیابی و اعتبارسنجی عملکرد الگوریتم‌ها

برای ارزیابی عملکرد الگوریتم‌های پیش‌بینی کننده، معیارهای صحت (Accuracy)، دقت (precision)، حساسیت (Sensitivity) و ویژگی (Specificity) و سطح زیر منحنی راک (AUC: Area Under Curve) استفاده گردید. این معیارها با استفاده از اجزای ماتریس درهم‌آمیختگی طبق جدول ۱ تعریف و محاسبه‌شده است.

منظور از معیار صحت نسبت تعداد افراد زنده و فوت‌شده است که زنده یا فوت شدن آن‌ها به‌درستی تشخیص داده‌شده است. منظور از دقت یعنی

تعداد افرادی که فوت‌شده‌اند و مدل به‌درستی آن‌ها تشخیص داده است. حساسیت به معنی نسبتی از افرادی فوت‌شده‌اند که مدل به‌درستی آن‌ها را فوتی تشخیص داده است. بنابراین هرچه این مقدار بزرگ‌تر باشد، نشان‌دهنده دقت در تشخیص افراد فوت‌شده است. ویژگی به معنی نسبتی از افراد که زنده هستند که مدل به‌درستی آن‌ها را زنده تشخیص داده است [۱۳]. سطح زیر منحنی ROC اغلب به‌عنوان شاخصی برای تعین میزان قدرت یک مدل مورداستفاده قرار می‌گیرد. همچنین در حوزه پزشکی نیز برای ارزیابی دقت تستهای تشخیصی از سطح زیر منحنی ROC استفاده می‌شود [۱۴].

جدول ۱. ماتریس درهم‌آمیختگی

ماتریس‌درهم‌آمیختگی			
	خروجی	مقدار پیش‌بینی	
		زنده ماندن (-)	مرگ (+)
مقدار واقعی	مرگ(+)	منفی کاذب*	مثبت حقیقی*
	زنده ماندن(-)	منفی حقیقی*	مثبت کاذب*

* مثبت حقیقی (True positive): تعداد افرادی که مرده‌اند هستند و مدل به‌درستی مرده تشخیص داده است.
*مثبت کاذب (False positive): تعداد افرادی که زنده هستند اما مدل به‌اشتباه مرده تشخیص داده است.
*منفی حقیقی (True negative): تعداد افرادی که زنده هستند و مدل به‌درستی آن‌ها را زنده تشخیص داده است.
*منفی کاذب (False negative): تعداد افرادی که مرده‌اند هستند اما مدل آن‌ها را به‌اشتباه زنده تشخیص داده است [۱۵].

درنهایت برای هریک از الگوریتم‌ها منتخب به‌صورت مجزا، عملکرد آن گزارش گردیده است و الگوریتمی که نتایج بهتری نسبت به سایرین داشته است به‌عنوان بهترین مدل پیش‌بینی در عوامل مرتبط با مرگ‌ومیر در بیماران بخش مراقبت‌های ویژه گزارش شد. جهت آنالیز و مدل‌سازی از نرم‌افزار SPSS نسخه ۲۴، نرم‌افزار رپیدماینر نسخه ۷/۱۰۰۱ برای تحلیل و محاسبه منحنی‌ها، معیارها و رسم ماتریس هم‌بستگی استفاده‌شده است.

مرحله پنجم: توسعه

ارائه قوانین ایجادشده در تعیین احتمال مرگ‌ومیر در بیماران بخش مراقبت‌های ویژه و عوامل مرتبط با مرگ‌ومیر بیماران در بخش مراقبت ویژه می‌باشد.

یافته‌ها:

یافته‌ها پژوهش بر اساس مراحل پژوهش ارائه می‌گردد:

شناسایی و پیش‌پردازش داده‌ها

از ۸۳۸ پرونده بیماران بستری در بخش مراقبت‌های ویژه در محدوده زمانی ۹۱ الی ۹۷ اطلاعات بیماران استخراج گردید. میانگین سن بیماران در این مطالعه ۲۱/۱۵±۵۵/۴۵ سال، از این تعداد ۳۵/۲ درصد فوت و ۶۴/۸ درصد زنده بودند. ازنظر جنسیت، ۵۷/۴ درصد مرد و ۴۲/۶ درصد زن بودند؛ که طبق جدول ۲ از مردان و زنان به ترتیب ۲۱/۹ و ۱۳/۳ درصد فوت و ۳۵/۵ و ۲۹/۳ درصد زنده مانده‌اند.

تعداد ۲۶ متغیر موردبررسی قرار گرفت که شامل سن، جنس، GCS، HCT، بیلی‌روبین، پتاسیم، گلوکز، سدیم، WBC، کراتینین، فشارخون دیاستولیک، نرخ تنفس، فشارخون سیستولیک، درجه حرارت، ضربان قلب،PH، فاکتورهای اکسیژناسیون ($PaCO_2$,$PaO_2$$HCO_3$)، اوره، پلاکت، INR، زمان نسبی ترومبوپلاستین، فشارخون متوسط شریانی، آلبومین و وضعیت بیمار بود. جهت داده‌های مفقودشده یک مقدار مصنوعی برحسب میانگین به‌جز متغیر وضعیت بیمار و جنسیت استفاده گردید. متغیر FiO۲ به دلیل تعداد بیش‌ازحد داده‌های مفقودشده حذف گردید. متغیر دما در محدوده طبیعی ۳۶ تا ۳۸/۴ تعریف گردید که بازه محدودی برای متغیر دما در نظر گرفته شد.

متغیرها برحسب درصد فراوانی در جدول ۳ تنظیم‌شده است.

جدول ۲: درصد فراوانی متغیر پیش‌بینی برحسب جنسیت

جنسیت	زنده	فوت‌شده	جمع کل
مرد	٪۳۵/۵	٪۲۱/۹	۵۷/۴
زن	٪۲۹/۳	٪۱۳/۳	۴۲/۶

جدول ۳: درصد فراوانی متغیرهای موردمطالعه

متغیر	درصد فراوانی	متغیر	درصد فراوانی
PH	۷۵/۰۹	فشارخون دیاستولیک	۷۸/۳
نرخ تنفس	۳۶/۰۲	فشارخون سیستولیک	۸۷/۴
درجه حرارت	۹۸/۴	زمان نسبی ترومبوپلاستین	۵۴/۰۳
گلبول‌های سفید خون	۹۷/۸	پلاکت	۵۴/۰۹
اوره	۳۹	INR	۴۵/۰۱
سن	۹۹/۹	بیلی‌روبین	۴۴/۹
فشارخون متوسط شریانی	۷۰/۵	کراتین خون	۹۸/۴
سدیم	۹۸/۳	FiO_2	۱۷/۶
فشار نسبی دی‌اکسید شریانی	۷۶/۹	GCS	۹۷/۱
فشار نسبی اکسیژن شریانی	۷۶/۹	گلوکز	۹۷/۵
پتاسیم	۹۷/۹	HCO3	۹۸/۳
آلبومین	۶۸/۵	هماتوکریت	۹۸/۶
جنسیت	۹۱/۰۲	ضربان قلب	۹۸/۴

ارزیابی و اعتبارسنجی عملکرد الگوریتم‌ها

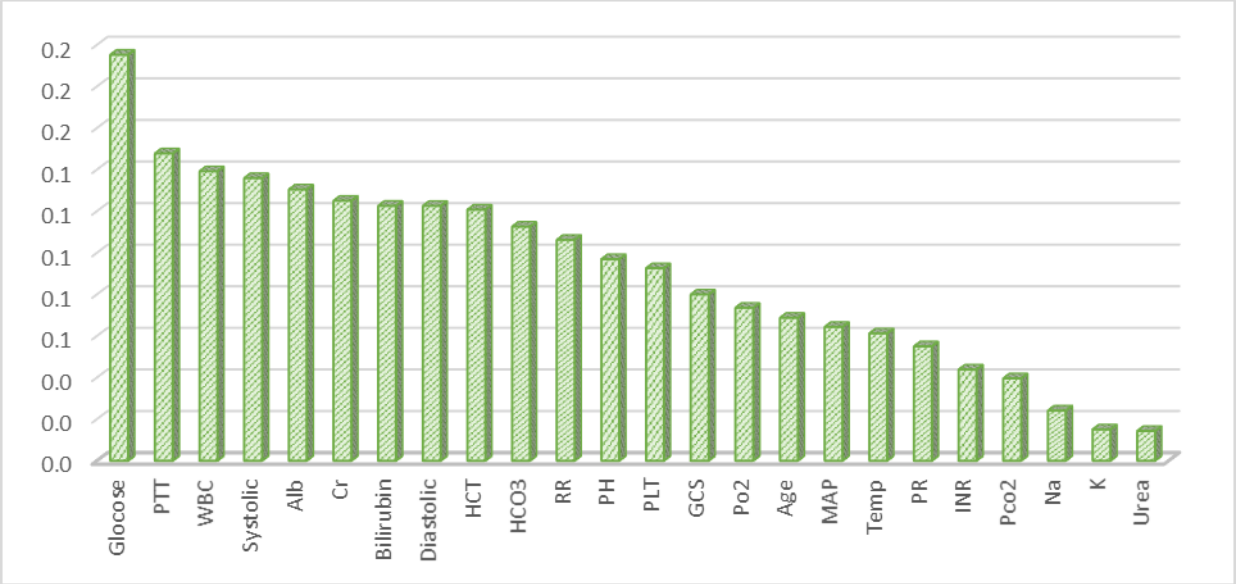
در این مرحله عملکرد الگوریتم‌های درخت تصمیم، K نزدیک‌ترین همسایه، جنگل تصادفی،SVM و رگرسیون لجستیک مورد ارزیابی قرار گرفت. سپس صحت، دقت، ویژگی، حساسیت و سطح زیر منحنی راک آن‌ها گزارش شد. مقایسه صحت، دقت، ویژگی، حساسیت و سطح زیر منحنی راک هر یک از الگوریتم‌ها در جدول ۴ نشان داده‌شده است.

دقت، صحت، سطح زیر منحنی راک، حساسیت رگرسیون لجستیک نسبت به سایر الگوریتم‌ها بالاتر بوده است. لذا رگرسیون لجستیک با صحت ۷۵/۶۲ درصد، دقت ۶۸/۳۹ درصد، سطح زیر منحنی راک ۰/۷۶ و حساسیت ۳۸/۶۵ عملکرد بهتری برای ایجاد مدل پیش‌بینی مرگ‌ومیر دارد.

جدول ۴: بررسی عملکرد الگوریتم‌های منتخب

الگوریتم / عملکرد	K نزدیک‌ترین همسایه	جنگل تصادفی	رگرسیون لجستیک	ماشین بردار پشتیبانی	درخت تصمیم
صحت	۶۹/۰۷	۷۰/۶۶	۷۵/۶۲	۷۴/۲	۷۰/۷۵
دقت	۴۷/۹۲	۵۷/۶۷	۶۸/۳۹	۶۷/۶۹	۶۳/۶۴
ویژگی	۸۷/۲۱	۹۷/۳۴	۹۴/۵۳	۹۷/۸۱	۹۴/۳۵
حساسیت	۲۶/۸۵	۸/۵۷	۳۸/۶۵	۲۷/۶۳	۶/۲۱
سطح زیر منحنی راک	۰/۶۲۲	۰/۷۱۴	۰/۷۶	۰/۷۴	۰/۵

در نمودار ۲ وزن مؤثرترین متغیرها در پیش‌بینی بر اساس رگرسیون لجستیک مشخص‌شده است. نمودار نشان می‌دهد متغیرهای گلوکز، زمان ترومبوپلاستین نسبی و تعداد گلبول سفید مؤثرین متغیرها در پیش‌بینی بر اساس رگرسیون لجستیک بودند.



نمودار ۲. بررسی مؤثرترین وزن‌ها در پیش‌بینی مرگ‌ومیر بر اساس الگوریتم رگرسیون لجستیک

بحث و نتیجه‌گیری:

اجرای یک سیستم، مبتنی بر داده‌ها می‌تواند از داده‌ها برای یافتن روابط عوامل تشخیصی، پیش‌آگهی و درمانی با استفاده از تکنیک‌های داده‌کاوی به‌خوبی در فرایند پیش‌بینی کمک کند [۱]. توسعه ابزار پیش‌بینی دقیق و علمی در حیطه مراقبت‌ها، کار ساده‌ای نیست، به این دلیل که برخی از داده‌ها به‌صورت عینی اندازه‌گیری نمی‌شوند و یا برخی از داده‌ها نیاز به‌روزرسانی دائمی در محدوده زمانی معین دارند [۲۰-۱۶]. نتایج مطالعه Terrin و همکاران در سال ۲۰۰۳ پیش‌بینی مرگ‌ومیر در بخش مراقبت ویژه با استفاده از رگرسیون لجستیک و شبکه عصبی و درخت تصمیم سطح زیر منحنی راک به ترتیب ۰/۷۶۹ و ۰/۷۲۴ و ۰/۶۶۷ گزارش گردید که بهترین عملکرد مربوط به رگرسیون لجستیک بود [۲۱]. در مطالعه Hsieh و همکاران در سال ۲۰۱۸ مقایسه درصد پیش‌بینی مرگ‌ومیر در ۳۴۱ بیمار بخش مراقبت ویژه، برتری رگرسیون لجستیک ۰/۸۵۳ نسبت به شبکه عصبی و SVM را گزارش نمود [۲۲]. مطالعه Choi و همکاران در سال ۲۰۲۰ بر روی اطلاعات ۸۹۳۷ بیمار بستری بخش مراقبت برتری رگرسیون لجستیک ۰/۸۲۷ نسبت به پرسپرون چندگانه ۰/۸۱۶ را بر اساس منحنی راک گزارش نموده است [۲۳].Vairavan و همکاران ۲۰۱۲ طی مطالعه‌ای نشان دادند که؛ روش رگرسیون لجستیک توسط متخصصان مراقبت‌های ویژه به‌منظور پیش‌بینی خطر مرگ‌ومیر برای بیماران مبتلا به بیماری‌های شدید یا جراحات بخش مراقبت ویژه، مورد استفاده قرار گرفته است [۶]. استفاده از رگرسیون لجستیک به دلیل سادگی در فهم و نمایش نتایج، می‌تواند روش مناسبی برای پیش‌بینی مرگ‌ومیر باشد اما باید توجه داشت، نتایج چنین مطالعاتی بسته به داده‌های مورداستفاده از هر بیمارستان ممکن است با بیمارستان دیگر متفاوت باشد [۱۱].

در مطالعه Howitt و همکاران نشان داد که مرگ‌ومیر بیمارانی که پس از عمل جراحی قلب در بخش مراقبت ویژه بستری‌شده‌اند تحت تأثیر عوامل و عوارض بعد از عمل قرار می‌گیرد. لذا همین مسئله بر احتمال خطر مرگ‌ومیر تأثیر می‌گذارد و باعث تأثیر در عملکرد مدل خواهد شد [۲۴]. در مطالعه LiOQue و همکاران بیان شده است مدل‌های پیش‌بینی مرگ‌ومیر ممکن است ازلحاظ آماری دقیق و کارا باشند، اما باید در عمل به‌صورت بالینی بررسی گردند، زیرا آن‌ها تنها افزایش خطر را نشان می‌دهند و یا عارضه‌ای مهم رخ می‌دهد. لذا پژوهش‌های آینده باید بر روی توسعه مدل‌ها برای اهداف خاص متمرکز شوند. به‌عنوان‌مثال، مدل‌هایی که مدل‌سازی بر اساس رویدادهای ترکیبی و داده‌های فیزیولوژیکی پیش‌بینی را گزارش می‌نمایند [۲۵]. بر اساس نتایج مطالعه Bhattacharya و همکاران در سال ۲۰۱۷ ارزیابی عملکرد مدل‌ها هنوز در حال بحث است و مطالعات بسیار کمی در موردبررسی داده‌های

ازدست‌رفته و ناقص، نوع و کیفیت داده‌ها، خطاهای انسانی هنگام ثبت صورت گرفته است [۱۱]. ازلحاظ کمیت نیز بسیاری از داده‌های موجود در پرونده کامپیوتری مثل دلیل اصلی مراجعه، تشخیص اولیه و نهایی، بیماری‌های همراه تکمیل نمی‌شود که این امر نیز باعث عدم کارآیی استفاده از داده‌ها در حوزه‌های پژوهشی می‌شود [۲۶]. برای به دست آوردن یک مدل پیش‌بینی درست، داده‌های بخش مراقبت ویژه بیمارستان‌ها باید موردبررسی قرار گیرد و مدل‌های حاصل بر اساس پیش‌بینی‌های آن‌ها ارزیابی شود [۲۷]. نتایج این مطالعات می‌تواند به‌عنوان الگویی برای پیش‌بینی وضعیت مرگ‌ومیر در بخش مراقبت‌های ویژه مورداستفاده قرار گیرد

تشکر و قدردانی:

بدین‌وسیله از متخصصین بخش مراقبت‌های ویژه بیمارستان امام حسین(ع) کمال تشکر و قدردانی رادارم.

تأییدیه اخلاقی:

این مقاله حاصل طرح تحقیقاتی مصوب دانشگاه علوم پزشکی مشهد با کد اخلاق IRMUMSREC.1398.028 و شماره طرح ۹۷۱۷۰۳ می‌باشد.

تعارض منافع:

در این مقاله تعارض منافع وجود ندارد.

سهم نویسندگان:

مرضیه معراجی (نویسنده مسئول) نگارنده مقدمه/ روش/ بحث و نتیجه‌گیری (۴۰٪)؛ پرنیان عسگری (نویسنده اول) انجام تکنیک‌ها / نگارنده بحث و نتیجه‌گیری (۳۵٪)؛ علیرضا آتشی (نویسنده دوم) ویرایشگر/ نتیجه‌گیری (۲۰٪)؛ میر محمد میری (نویسنده چهارم) ویرایشگر (۵٪)

حمایت مالی:

این مقاله از طرف معاونت تحقیقات و پژوهش دانشگاه علوم پزشکی مشهد حمایت شده است.

Reference

1. Power GS, Harrison DA. Why try to predict ICU outcomes? Curr Opin Crit Care. 2014; 20(5):544-9. Doi:10.1097/MCC.0000000000000136
2. King MA, Niven AS, Beninati W, Fang R, Einav S, Rubinson L, et al. Evacuation of the ICU: care of the critically ill and injured during pandemics and disasters: CHEST consensus statement. Chest. 2014; 146(4):e44S-e60S. Doi: 10.1378/chest.14-0735.
3. Ramon J, Fierens D, Güiza F, Meyfroidt G, Blockeel H, Bruynooghe M, et al. Mining data from intensive care patients. Advanced Engineering Informatics journal. 2007; 21(3):243-56. Doi:10.1016/j.aei.2006.12.002.
4. Silva Á, Cortez P, Santos MF, Gomes L, Neves JJAiim. Mortality assessment in intensive care units via adverse events using artificial neuralnetworks. Artif Intell Med. 2006; 36(3):223-4. Doi:10.1016/j.artmed.2005.07.006
5. Kim S, Kim W, Park RWJHir. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. J Healthc Inform Res. 2011; 17(4):232-43. Doi:10.4258/hir.2011.17.4.232
6. Vairavan S, Eshelman L, Haider S, Flower A, Seiver A. Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model. Comput Cardiol. 2012;393-396.
7. Mackenzie A. The production of prediction: What does machine learning want? Eur J Cult Stud. 2015; 18(4-5):429-45. Doi:10.1177/1367549415577384
8. Moslehi M, Alimohammadi N, Yazdannik A, Abasi S. Power of SAPSII tool in predicting mortality in ICU patients. JAP. 2017; 8(1):78-85. [In Persian]
9. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014; 7:2(3). Doi: 10.1186/2047-2501-2-3.
10. Norrie J. Mortality prediction in ICU: a methodological advance. Lancet Respir Med. 2015; 3(1):5-6. Doi:10.1016/S2213-2600(14)70268-1
11. Bhattacharya S, Rajan V, Shrivastava H. ICU

mortality prediction: A classification algorithm for imbalanced datasets. Thirty-First AAAI Conference on ArtificiIntelligence; 2017 Feb 12; Xerox Research Centre India, S Bangalore India.
12. Atashi A, Ahmadian L, Rahmatinezhad Z, Miri M, Nazeri N, Eslami SJBH, et al. Development of a national core dataset for the Iranian ICU patients outcome prediction: A comprehensive approach. BMJ Health Care Inform. 2018; 25(2):71-6. Doi: 10.14236/jhi.v25i2.953
13. Ramsay MA, Usman M, Lagow E, Mendoza M, Untalan E, De Vol EJA, et al. The accuracy, precision and reliability of measuring ventilatory rate and detecting ventilatory pause by rainbow acoustic monitoring and capnometry. ANE. 2013; 117(1):69-75. Doi: 10.1213/ANE.0b013e318290c798
14. Krzanowski WJ, Hand DJ. ROC curves for continuous data. London: Chapman and Hall/CRC; 2009.
15. Evans JR, Fisher R. Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. Appl Cogn Psychol. 2011; 25(3):501-8. Doi:10.1002/acp.1722.
16. Rosenberg A. Recent innovations in intensive care unit risk-prediction models. Current Opinion Critical Care journal. 2002; 8(4):321-30. Doi:10.1097/00075198-200208000-00009.
17. Knaus WA. APACHE 1978-2001: The development of a quality assurance system based on prognosis: Milestones and personal reflections. Arch Surg. 2002; 137(1):37-41. Doi: 10.1001/archsurg.137.1.37
18. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. Chest. 1991; 100(6):1619-36. Doi:10.1378/chest.100.6.1619
19. Le JG, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. Crit Care Med. 1984; 12(11):975-7. Doi:10.1097/00003246-198411000-00012
20. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based onaEuropean/North American multicenter study. JAMA. 1993; 270(24):2957-63. Doi: 10.1001/jama.270.24.2957.
21. Terrin N, Schmid CH, Griffith JL, D'Agostino RB, Selker HP. External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol. 2003; 56(8):721-9. Doi:10.1016/s0895-4356(03)00120-3
22. Hsieh MH, Hsieh MJ, Chen CM, Hsieh CC, Chao CM, Lai CC. Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units. Scientific Reports. Sci Rep. 2018; 8(1):17116. Doi:10.1038/s41598-018-35582-2
23. Choi, Y, Boo, Y. Comparing Logistic Regression Models with Alternative Machine Learning Methods to Predict the Risk of Drug Intoxication Mortality. Int. J. Environ. Res. Public Health. Int J Environ Res Public Health. 2020; 17(3):897. Doi:10.3390/ijerph17030897
24. Howitt SH, Caiado C, McCollum C, Goldstein M, Malagon I, Venkateswaran R, et al. Validation of three postoperative risk prediction models for intensive care unit mortalityafter cardiac surgery. Thorac Cardiovasc Surg. 2018; 66(08):651-60. Doi:10.1055/s-0037-1608897.
25. Liu Q, Xue FS, Yang GZ, Liu YY. Surgeon C. developing a risk prediction model for intensive care unit mortality after Cardiac Surgery. Thorac Cardiovasc Surg. 2018; 66(8):e1-e2. Doi: 10.1055/s-0038-1639336.
26. Shafagh O, Nateghi S, Haji MR, Goharani R, Zangi M, Ahmadi HA. Prediction of the length of stay of patients in the neuro-critical care unit using data mining technique. Anesthesiology and critical care in Iran. 2018; 2(1):22-33.[In Persian]
27. Salamati P, Razavi SM, Shokraneh F, Torabi SM, Laal M, Hadjati G, et al. Mortality and injuries among Iranians in Iraq-Iran war: a systematic review Arch Iran Med. 2013; 16(9):542-50. Doi: 013169/AIM.0012.

The Comparison of Selected Data-mining techniques in ICU Mortality Risk Prediction in Imam Hossein hospital

ParnianAsgari[1] Alireza Atashi[2] Marziyhe Meraji[3*] Mir MohammadMiri[4]

1. MSc Student, Health Information Technology, Health Information Technology, Faculty of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran.ORCID: 0000-0003-4426-0851
2. Department of Electronic Health, Virtual College,Tehran University of Medical Sciences, Tehran, Iran.
3. Department of Health Information Technology, Mashhad University of Medical Sciences, Mashhad, Iran.
4. Department of Anesthesiology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

(Received 27Dec, 2019 Accepted 10Mar, 2020)

Original Article

Abstract

Aim: Intensive Care Unit (ICU) is a ward that is critical to improving the health status of critical conditions. Data mining seems to be a good way to optimize the use of resources. Identifying and analyzing the risk factors associated with mortality will lead to more efficient and accurate planning of hospitalization and interventions. In this study, the prediction of mortality of patients in the intensive care unit of Imam Hossein Hospital in Tehran with data mining techniques is discussed.

Methods: Based on patient records and hospital information system, 838 patients admitted to the General intensive care unit between 2013 and 2019 in Imam Hossein Hospital in Tehran, the data is needed to collect this research. Algorithms used to classify patients include support vector machines, k nearest neighbor, decision tree, logistic regression and random forest that was reported based on the precision, accuracy, sensitivity, specificity, and roc under the curve.

Results: The results of this study showed, identified 26 factors affecting specific data and pre-processing of data. Among five of the algorithms used in the study, logistic regression algorithm based on the level of roc curve (0.76), accuracy percentage (75.62),precision (68.39),sensitivity (38.65) and specificity (94.53) had better performance in predicting mortality compared to other techniques of study. The variables of Glucose and Partial Thromboplastin time were the most significant effects on mortality based on the logistic regression model.

Conclusion: Data analysis in intensive care unit patients can be an appropriate and practical tool for predicting mortality and its related factors, but according to the quality of data, results are different. And the results extracted from logistic regression can be used as a model to predict the status of mortality in the intensive care unit.

Key Words: Mortality, Data Mining, Intensive Care Unit, Predictive Model.

Citation: Asgari P, Atashi A, Meraji M, Miri MM. The Comparison of Selected Data-mining techniques in ICU Mortality Risk Prediction in Imam Hossein hospital. J Mod Med Info Sci. 2020; 5(2):59-67.

Correspondence:
Marziyhe Meraji
Department of Health Information Technology, Mashhad University of Medical Sciences, Mashhad, Iran.
Tel: +989153121574 Email: Merajim1@mums.ac.ir
ORCID .0000-0002-3137-0683