بررسی اختلال پاراکلینیک موثر در مرگ‌ومیر بیماران ترومایی با استفاده از شیوه‌های داده کاوی

*مریم حسن‌زاده(MSc)[1,2]- دکتر اکبر فرهودی‌نژاد(PhD)[1]- دکتر شاهرخ یوسف‌زاده(MD)[3]

*نویسنده مسئول: گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی مهندسی، دانشگاه پیام نور، تهران، ایران- مدیریت آمار و فناوری اطلاعات، دانشگاه علوم پزشکی گیلان، رشت، ایران

پست الکترونیک: hassanzadeh@gums.ac.ir

تاریخ دریافت مقاله:۹۳/۰۸/۲۸ تاریخ ارسال۹۳/۱۲/۰۵ تاریخ پذیرش: ۹٤/۰۱/۰۷

چکیده

مقدمه: تروما شایع‌ترین سبب مرگ‌ومیر در جهان است که بیشتر بر اثر پیشامدهای جاده‌ای رخ می‌دهد و شناسایی به هنگام بیماران با آسیب‌دیدگی حاد، باعث اتخاذ درست اقدام پزشکی و در نتیجه نجات جان آنان و پرهیز از صرف هزینه‌های هنگفت درمانی خواهد شد.

هدف: یافتن بهترین الگوریتم داده کاوی برای شناسایی اختلال پاراکلینیکی موثر در مرگ‌ومیر بیماران ترومایی

مواد و روش‌ها: این پژوهش بر ۱۰۷۳ بیمار ترومایی و ۵۲ ویژگی ثبت شده در سیستم بیمارستانی مرکز آموزشی درمانی پورسینای رشت انجام شد. برای یافتن عوامل موثر و الگوی ارتباطی بین متغیرها، تکنیک‌های داده کاوی دسته‌بندی و روش‌هایی مانند درخت تصمیم، k نزدیک‌ترین همسایه و شبکه عصبی بر داده‌ها پیاده‌سازی و دقت پیشگویی آنان به روش 10-fold ارزیابی و مقایسه شد.

نتایج: از ۱۰۷۳ بیمار ترومایی، ۱۸۵ نفر(۱۷/۲٪) زن و ۸۸۸ نفر(۸۲/۸٪) مرد بودند. ۲۳۷ نفر (۲۲/۱٪) فوت شدند که بیشترین آنان (۳۰٪ معادل ۷۱ نفر) کمتر از یک هفته بستری و ۵۶ نفر (۲۳/۶٪) تا یک روز بستری بودند که ارتباط معنی‌داری بین مدت بستری و مرگ بیماران وجود داشت(p=۰/۰۰۰۱). از الگوریتم‌های اجرا شده داده کاوی، درخت تصمیم و k نزدیک‌ترین همسایه بالاترین دقت، به‌ترتیب ۹۱٪ و ۸۹٪ در دسته‌بندی و پیشگویی پیامد بیماران(مرگ یا بهبود) را داشت. به همین سبب به روش *Best First* (بهترین اولین) در درخت تصمیم از ۵۲ ویژگی بررسی شده، ۱۸ ویژگی از عوامل موثر در مرگ بیماران تشخیص داده شد.

نتیجه‌گیری: با توجه به دقت بالای الگوریتم‌های داده کاوی مانند درخت تصمیم، عوامل موثر در مرگ‌ومیر بیماران ترومایی پیشگویی و بدین وسیله با شناسایی افراد در معرض خطر، جان آنان نجات داده شد.

کلید واژه‌ها: داده کاوی/ ضربه / میزان مرگ و میر

مجله دانشگاه علوم پزشکی گیلان، دوره بیست و چهارم شماره ۹۵، صفحات: ٦۲-۵۲

مقدمه

تروما یکی از مهم‌ترین علل مرگ‌ومیر و ناتوانی در ٤ دهه‌ی اول زندگی (۳-۱) و شایع‌ترین سبب مرگ‌ومیر در سن ۵ تا ٤٤ سالگی است. برآورد شده که تا سال ۲۰۲۰ به سومین علت مرگ‌ومیر در سراسر دنیا برسد. از نظر سازمان بهداشت جهانی تروما به ویژه حوادث جاده‌ای از مهم‌ترین علل مرگ‌ومیر، صرف هزینه‌های هنگفت بیمارستانی و آسیب‌های اقتصادی به جامعه است(٤و۵).

در ایران تروما دومین علت مرگ‌ومیر پس از بیماری‌های قلبی- عروقی در تمام سنین است. بیماران ترومایی که در اثر حوادث غیرعمدی بستری شده و فوت می‌کنند به ترتیب ٤۲۵ و ٤۲نفر به ازای هر ۱۰۰۰۰۰ نفر است. تصادف موتورسیکلت و وسایل نقلیه موتوری باعث بستری و مرگ ۲۳۷- ۳۸ نفر به ازای هر ۱۰۰۰۰۰ نفر در ایران می‌شود(٦). با توجه به اینکه تروما نقش مهمی در مرگ‌ومیر پویاترین گروه جامعه دارد، بهبود مراقبت‌های بیماران ترومایی و جلوگیری از مرگ‌های پیشگیری شونده از هدف‌های بسیاری از جوامع است (۳و۷). تروما وضعیتی حساس به زمان است که نیازمند ارزیابی و رسیدگی درست در ساعت‌های نخست است (۸). اگر مراکز تروما خدمات طلائی در این‌گونه افراد ارائه دهند ممکن است زندگی تعدادی از آنان نسبت به بیمارستان‌هایی که مراکز

۱. گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی مهندسی، دانشگاه پیام نور، تهران، ایران

۲. مدیریت آمار و فناوری اطلاعات، دانشگاه علوم پزشکی گیلان، رشت، ایران

۳. گروه جراحی مغز و اعصاب، دانشکده پزشکی، مرکز تحقیقات ترومای جاده ای، دانشگاه علوم پزشکی گیلان، رشت، ایران

تروما ندارند، نجات یابد. بزرگ‌ترین چالشی که برای یک مرکز تروما وجود دارد یافتن تفاوت بین بیمارانی است که آسیب شدید دارند که می‌تواند منجر به مرگ بیانجامد و بدین سبب نیازمند انتقال سریع به مراکز تروما و شروع سریع درمان و مراقبت ویژه نسبت به بیمارانی است که مصدومیت و خطر کمتری دارند(۹).

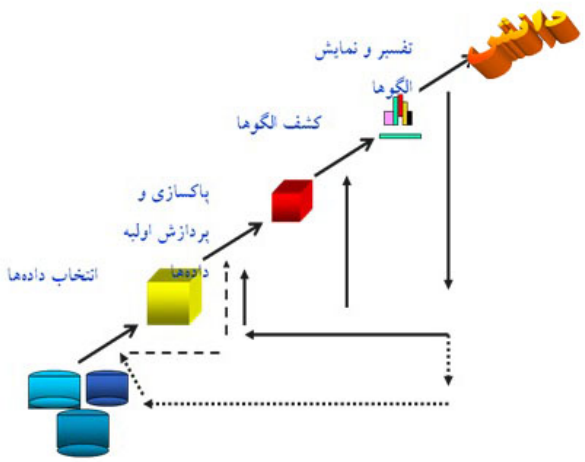
برای ارزیابی شدت تروما، تریاژ و ارزیابی نتایج کار در مراکز تروما، وجود مقیاس کمّی اندازه‌گیری شونده بایسته به نظر می‌رسد(۱و۲). روش‌های سنجش کمّی تروما از مهم‌ترین روش‌ها در ارزیابی و مقایسه‌ی استانداردهای درمان در تروماست(۱۰). از مهم‌ترین ویژگی‌های یک سیستم سنجش کمّی تروما، داشتن دقت، حساسیت و اختصاصی بودن، چشمگیر است(۱).

پژوهش‌ها نشان می‌دهد که روش‌های داده کاوی می‌تواند در شناسایی عوامل میزان شدت وخامت حال بیماران سودمند بوده و با ارائه مدل‌های پیش‌گویانه، عوامل موثر در مرگ حادثه دیدگان را پیشگویی و بدین‌ترتیب کمک شایانی به پزشکان، در اتخاذ تصمیم‌های مهم مانند جراحی نماید (۹و۱۱). هدف از روش‌های پیشگویی داده کاوی در پزشکی بالینی ساخت الگویی پیش‌گویانه است که به پزشکان کمک می‌کند تا روش‌های پیشگیری، تشخیص و برنامه‌های درمانی خود را بهبود بخشند (۱۲).

داده‌کاوی عبارت است از برگرفتن یا استخراج دانش از مجموعه‌ای بسیار حجیم از داده‌ها (۱۳). داده کاوی، جستجوی خودکار منابع داده‌ای بسیار بزرگ برای یافتن الگوها و رفتارهایی است که فراتر از تحلیل‌های ساده هستند. داده کاوی از الگوریتم‌های ریاضی پیچیده برای رده‌بندی داده‌ها و ارزیابی احتمال و رویدادها در آینده استفاده می‌کند. داده کاوی با عنوان "کشف دانش در داده‌ها" یا KDD نیز شناخته می‌شود. هدف داده کاوی در یک گستره معین علمی عبارت است از ایجاد درک از حجم زیادی از داده‌هایی که در بیشتر موارد به صورت هدایت نشده گردآوری شده‌اند(۱٤).

داده کاوی در چند گام صورت می‌گیرد که در نمودار ۱ نشان داده شده‌است: ۱- انتخاب داده‌ها: انتخاب داده‌های مرتبط از دیگر داده‌ها ۲- پیش پردازش داده‌ها شامل: الف- پاک‌سازی داده‌ها: حذف داده‌های ناکامل و نویز ب- یکپارچه‌سازی داده‌ها: ترکیب منابع چندگانه داده‌ای پ- تبدیل داده‌ها: سازگاری با فرمت مناسب برای داده کاوی ت- کاهش داده‌ها: بکارگیری روش‌های خلاصه‌سازی ۳- استفاده از تکنیک‌های داده کاوی: ساخت الگوها ٤- ارزیابی الگوها: سنجش درستی الگوها بر پایه یک سری سنجه‌های گیرایی ٥- بازنمایی دانش (۱۳و۱٥)

در هر کدام از این مراحل ممکن است چندین بار به عقب برگشته و کار دوباره تکرار شود.

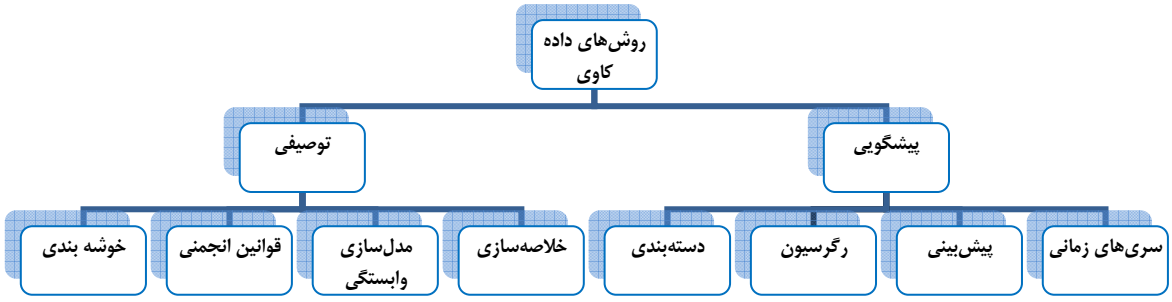


نمودار ۱. مراحل انجام داده کاوی

تکنیک‌های گوناگون داده‌کاوی را می‌توان برپایه گونه عملیاتی که انجام می‌دهند به دو دسته «پیش‌بینی‌کننده» و «توصیف‌کننده» بخش کرد. تکنیک‌های پیش‌بینی‌کننده با ساخت مدلی برای پایگاه داده، وظیفه پیش‌بینی موارد ناشناخته

را برعهده دارند. درحالی‌که تکنیک‌های تشریح‌کننده، الگوهایی فهمیدنی از داده‌ها را برای انسان کشف می‌کنند. (۱٤)

از تکنیک‌هایی که برای داده کاوی استفاده می‌شود می‌توان به روش‌های زیر اشاره کرد که در شکل ۲ نیز نشان داده شده‌است: دسته‌بندی (تکنیک پیش‌گویانه)، رگرسیون (تکنیک پیش‌گویانه)، خوشه‌بندی (تکنیک توصیفی)، کشف قواعد وابستگی (تکنیک توصیفی)، تحلیل دنباله.



شکل ۲. تکنیک‌ها و روش‌های داده‌کاوی

دسته‌بندی و پیش‌بینی دو نوع عملیات برای تحلیل داده‌ها و استخراج مدل برای توصیف دسته‌های مهم داده‌ها، فهم و پیش‌بینی رفتار آینده آنهاست. مدل‌های دسته‌بندی در تحلیل داده‌های گسسته و طبقه‌ای بکار رفته و مدل‌های پیش‌بینی یا رگرسیون بیشتر بر داده‌های پیوسته بکار می‌رود.(۱٤)

دسته‌بندی، هر جزء از داده‌ها را برپایه اختلاف بین داده‌ها به مجموعه‌های از پیش تعریف شده دسته‌ها تصویر می‌کند. باید بگوییم دسته‌بندی یادگیری با بازبینی است که دسته‌ها از پیش مشخصند. دسته‌بندی داده‌ها، فرایندی دو مرحله‌ای است. نخستین گام ساخت مدل و دومین مرحله استفاده از مدل و پیش‌بینی از راه داده‌های پیشین است.

روش‌های زیادی برای دسته‌بندی وجود دارد که از آن می‌توان به موارد: بیز ساده و شبکه‌های بیزی، شبکه‌های عصبی، نزدیک‌ترین همسایگی، درخت تصمیم و رگرسیون اشاره کرد. از الگوریتم‌های بکار رفته در داده کاوی، ۱۰ الگوریتم از برترین‌های پرکاربرد برگزیده کنفرانس بین‌المللی داده کاوی ICDM اعلام شده که دربرگیرنده: درخت C٤/٥، سنجش میانگین‌ها K-Means، شبکه عصبی SVM، Apriori، EM، PageRank، AdaBoost، KNN، Naive Bayes، درخت CART است.(۱٦)

همچنین، با پژوهش نشان داده شد که الگوریتم‌های پرکاربرد در داده‌های پزشکی شامل: درخت تصمیم C٤/٥، درخت رگرسیون (CART)، رگرسیون لوجستیک، بیز ساده، شبکه عصبی و پرسپترون چند لایه است.(۱۷)

با بکارگیری تکنیک‌های داده کاوی در داده‌های پزشکی می‌توان به شناسایی و معالجه بیماران دیابتی(۱۸)، درمان تومورهای مغزی(۱۲)، پیشگویی حمله‌های قلبی(۱۹)، تعیین عوامل موثر در بدخیمی‌ها و بهترین نوع درمان آن(۲۰) پرداخت. با کشف بهترین تصمیم برای درمان به روش خوشه‌بندی در مشخصه‌ها و علائم بیماری و تصمیم پرستار و پزشک می‌توان از بسیاری هزینه‌های ناشی از اقدام نادرست مانند جراحی پیشگیری کرد.0(۲۱و۲۲) پیش‌بینی مدت بستری بیماران باعث مدیریت بهتر منابع بیمارستانی، ارائه خدمات بیشتر و افزایش خوشنودی بیماران خواهد شد. (۲۳)

در این پژوهش تلاش شد با اجرای الگوریتم‌های داده کاوی بر طیف گسترده‌ای از بیماران پذیرفته شده در تنها بیمارستان تخصصی مغز و اعصاب استان گیلان که در سیستم بیمارستانی به ثبت رسیده‌اند، به مطالعه علل مرگ‌ومیر و پیشگویی عوامل موثر در آن بپردازیم.

مواد و روش‌ها

این تحقیق بر داده‌های مراجعان ثبت شده به اورژانس بیمارستان پورسینای رشت انجام شد که تنها مرکز آموزشی- درمانی دارای بخش تخصصی مغز و اعصاب استان گیلان بوده و به دلیل مراجعه همه بیماران حادثه دیده استان به این مرکز، بانک کاملی از بیماران ترومایی استان دارد. این سیستم

از سال ۱۳۸۵ در این بیمارستان نصب و اجرا شده است و از سال ۱۳۹۰ نیز بانک اطلاعاتی حوادث به آن افزوده شد که بدین‌ترتیب داده‌های مربوط به ٤٦٨٨٠ بیمار از سال ۱۳۹۰ تا ۱۳۹۲ استخراج و بررسی شد.

در سیستم نرم‌افزار بیمارستانی بانک‌های انبوهی وجود دارد که با توجه به ویژگی‌های مورد نیاز در این پژوهش از ۳ بانک اطلاعاتی: ۱- اطلاعات اولیه و ویژگی‌های بیمار ۲- آزمایش‌های انجام شده ۳- بانک مدارک پزشکی (ثبت نشانه‌های نخست و تشخیص نهایی بیماری) و برای استخراج داده‌ها از بانک‌های اطلاعاتی و آماده‌سازی و پاک‌سازی آنها زبان برنامه‌نویسی SQL بکار رفت. در علم داده‌کاوی مهم‌ترین و طولانی‌ترین بخش کار، پاک‌سازی و آماده سازی داده‌ها است. شناسایی و حذف داده‌های نویزی، تغییر و تبدیل لازم در نوع و مقدار داده‌ها بیشترین زمان را به خود اختصاص می‌دهد.

با توجه به این‌که تا هنگام انجام این تحقیق در این بیمارستان، علائم اولیه و بالینی بیمار ثبت نمی‌شد، با استفاده از بانک مدارک پزشکی که پس از ترخیص بیمار به ثبت تشخیص نهایی و علائم اولیه بیمار می‌پردازد، علائم اولیه و نوع بیماری استخراج شد که با توجه به اینکه ثبت این مشخصات هم‌زمان با ترخیص بیمار صورت نمی‌گیرد متأسفانه بسیاری از رکوردهای بیماران به این سبب حذف شد. در پایان از ترکیب بانک اطلاعات بیماران با بانک اطلاعات مدارک پزشکی، از ٤٦٨٨٠ بیمار مراجعه‌کننده به بیمارستان پورسینای رشت از سال ۱۳۹۰ تا ۱۳۹۲، ۲۳۲۷۲ بیمار ترومایی با کد استاندارد ICD10 تشخیص داده شد. همچنین، از ترکیب این بانک به دست آمده (ترومایی) با بانک آزمایش بیماران، با توجه به این‌که همه آزمایش‌ها بر تمام بیماران انجام نمی‌شود، تعداد متغیری از رکوردها (بین ۱۰۸٤ تا ١٦٦١٤) به دست آمد که با توجه به اهمیت آزمایش‌هایی مانند PT و ABG که تعیین کننده وخامت حال این نوع بیماران است، پس از پاک‌سازی لازم و حذف داده‌های نویزی و ناکامل، این تعداد در نهایت به ۱۰۷۳ رکورد بیمار رسید که همه آزمایش‌های مهم نظیر هماتولوژی (هموگلوبین، هماتوکریت، MCH –MCHC- MC)، پلاکت خون، RBC، WBC، سدیم، پتاسیم، PT، PTTو ABG را در بر می‌گرفت، کاهش و بدین‌ترتیب به ۵۲ ویژگی رسید و در کل متغیرها در بردارنده موارد زیر شد:

جنس، تأهل، نوع پذیرش در بیمارستان، منطقه حادثه، مکان حادثه، نوع حادثه، سن، مدت بستری، ضربه به سر، ضربه به گردن، ضربه به گلو، ضربه به شکم، ضربه به شانه، ضربه به آرنج، ضربه به دست، ضربه به مفصل ران، ضربه به زانو، ضربه به پا، ضربه به چندین قسمت از بدن، ضربه به منطقه نامشخص، تأثیر ضربه، سوختگی، سرمازدگی، مسمومیت دارویی، مسمویت غیردارویی، سایر علل خارجی و ناشناخته، عوارض پس از جراحی یا مراقبت‌های پزشکی، عوارض پس از آسیب، مسمومیت و دیگر عوامل خارجی، آزمایشات: هموگلوبین، هماتوکریتmcv ,mchc ,U mch سدیم، پتاسیم PT-control, PT-activity, PT-inr, Pt-patient ABG-، ABG-p-o2 ،ABG-be-ecf ،ABG-bb ،PTT hc03، ABG-ph, ABG-p-co2، RBC ، پلاکت خون ، WBCو وضعیت هنگام ترخیص بیمار(پیامد درمان)

برای شناسایی داده‌های نویزی، نارسا و مخدوش از آمار توصیفی توسط نرم‌افزارspss-21 استفاده و پس از اعمال پیش پردازش داده‌ها برای حذف نویز و درست کردن داده‌های نارسا توسط زبان SQL، تجزیه و تحلیل آنان با آزمون‌های کای۲ و تی (مقایسه میانگین‌ها) با نرم‌افزار spss-21 انجام شد. در نهایت برای یافتن عوامل موثر و الگوی ارتباطی بین متغیرها روش‌های دسته‌بندی داده کاوی نظیر الگوریتم درخت‌های تصمیم و C4.5، شبکه عصبی و K نزدیک‌ترین همسایه و روش خوشه‌بندی دورترین- اولین (Farthest First) با ابزار متن باز weka و زبان برنامه نویسی java بر داده‌ها پیاده‌سازی و مدل‌های مربوطه استخراج شد.

نتایج

از ۱۰۷۳ بیمار ترومایی، ۸۸۸ نفر(۸۲/۸٪) مرد بودند. ۲۳۷ نفر(۲۲/۱٪) فوت شدند. ۲۰٪ زنان بیمار و ۲۲/۵٪ مردان بیمار فوت شدند که ارتباط معنی‌داری با انجام آزمون کای‌دو بین جنس و پیامد بیماران (مرگ و بهبود) دیده نشد. از افراد بهبود یافته، ٦٩٥ نفر(٦٤/٨٪) با بهبود نسبی، ۵۹ نفر(۵/۵٪) با بهبود کامل، ۷۳ نفر(٦/٨٪) با میل شخصی و ۳ نفر(۰/۳٪) با پیگیری

بعدی از بیمارستان مرخص شده بودند، ٦ نفر(٠/٦٪) به بیمارستان دیگر منتقل شده بودند. با توجه به اهمیت مردن یا زنده بودن افراد در این تحقیق و نداشتن تناسب تعداد گروه‌ها، آنها را به دو دسته مرگ یا بهبود تقسیم کرده و آزمون‌های ارتباط سنجی در این دو گروه انجام شد.

۱۸ نفر(۱/۷٪) کمتر از یک روز بستری بودند که ۹ نفر از آنان فوت شدند. ۸۲ نفر(۷/٦٪) تا یک روز بستری بودند که ٥٦ نفر(٦٨/٣٪) فوت شدند و ٦٦ نفر(٦/٢٪) تا ٤٨ ساعت بستری بودند که ۳۱ نفر(٤٧٪) فوت شدند. همچنین، ۲۹/۵٪ افراد تا یک هفته، ۲۸/۹٪ تا دو هفته و ۲٦/۱٪ بیش از ۲ هفته بستری بودند. بین مدت بستری و پیامد (مرگ و بهبود)، با آزمون کای‌دو اختلاف معنی‌دار وجود داشت(p=۰/۰۰۰۱).

با آزمون کای‌دو ویژگی‌های وضعیت تأهل، ضربه به سر، شکم، آرنج، دست، مفصل ران، زانو، پا، چندین بخش از بدن نیز ارتباط معنی‌داری با وضعیت مرگ یا بهبود بیمار داشت (p=۰/۰۰۰۱) و مشخص شد که ویژگی‌هایی نظیر جنس، نوع پذیرش در بیمارستان، منطقه حادثه، مکان حادثه، نوع حادثه، ضربه به گردن، گلو و شانه ارتباط معنی‌داری با مرگ یا بهبود بیماران نداشت. در جدول ۱ فراوانی نسبی برخی از این ویژگی‌ها نشان داده شده‌است.

جدول ۱. فراوانی نسبی مناطق آسیب دیده بدن و پیامد خروجی بیمار

محل ضربه به بدن	وضعیت ضربه	پیامد خروجی (درصد)		مجموع	P
		بهبود تعداد(درصد)	مرگ تعداد(درصد)	تعداد(درصد)	
سر	بلی	۵۱۲ (۶۱/۲)	۱۹۳ (۸۱/۴)	۷۰۵ (۶۶/۷)	۰/۰۰۰۱
	خیر	۳۲۴ (۳۸/۸)	۴۴ (۱۸/۶)	۳۶۸ (۳۴/۳)	
گردن	بلی	۵۵ (۶/۶)	۲۲۳(۵/۹)	۶۹ (۶/۳)	۰/۷
	خیر	۷۸۱ (۹۳/۴)	۲۲۳ (۹۴/۱)	۱۰۰۴ (۹۳/۶)	
گلو	بلی	۱۵۷ (۱۸/۸)	۵۱ (۲۱/۵)	۲۰۸ (۱۹/۴)	۰/۳
	خیر	۶۷۶ (۸۱/۲)	۱۸۶ (۷۸/۵)	۸۶۵ (۸۰/۶)	
شکم	بلی	۱۳۲ (۱۵/۸)	۱۷ (۷/۲)	۱۴۹ (۱۳/۹)	۰/۰۰۱
	خیر	۷۰۴ (۸۴/۲)	۲۰۰ (۹۲/۸)	۹۲۴ (۸۶/۱)	
شانه	بلی	۶۸ (۸/۱)	۱۳ (۵/۵)	۸۱ (۷/۵)	۰/۱
	خیر	۷۶۸ (۹۱/۹)	۲۲۴ (۹۴/۵)	۹۹۲ (۹۲/۵)	
آرنج	بلی	۷۹ (۹/۴)	۱۱ (۴/۶)	۹۰ (۸/۴)	۰/۰۱
	خیر	۷۵۷ (۹۰/۶)	۲۲۶ (۹۵/۴)	۹۸۳ (۹۱/۶)	
دست	بلی	۴۵ (۵/۴)	۴ (۱/۷)	۴۹ (۴/۶)	۰/۰۱
	خیر	۷۹۱ (۹۴/۶)	۲۲۳ (۹۸/۳)	۱۰۲۴ (۹۵/۴)	
ران	بلی	۱۶۹ (۲۰/۲)	۲۳ (۹/۷)	۱۹۲ (۱۷/۹)	۰/۰۰۰۱
	خیر	۶۶۷ (۷۹/۸)	۲۱۴ (۹۰/۳)	۸۸۱ (۸۲/۱)	
زانو	بلی	۱۹۷ (۲۳/۶)	۲۴ (۱۰/۱)	۲۲۱ (۲۰/۶)	۰/۰۰۰۱
	خیر	۶۳۹ (۷۶/۴)	۲۱۳ (۸۹/۸)	۸۵۲ (۷۹/۴)	
پا	بلی	۳۸ (۴/۵)	۳ (۱/۳)	۴۱ (۳/۸)	۰/۰۲
	خیر	۷۹۸ (۹۵/۵)	۲۳۴ (۹۸/۷)	۱۰۳۲ (۹۶/۲)	
چند قسمت از بدن	بلی	۱۱۰ (۱۳/۲)	۸۴ (۳۵/۴)	۱۹۴ (۸۱/۱)	۰/۰۰۰۱
	خیر	۷۲۶ (۸۶/۸)	۱۵۳ (۶۴/۶)	۸۷۹ (۸۱/۹)	

با استفاده از آزمون مقایسه میانگین‌ها و آزمون تی نشان داده شد که متغیرهای: سن، مدت بستری، آزمایش‌ها: سدیم، هموگلوبین، هماتوکریت ،PT، ABG-bb ،ABG-be ،ABG-co2 ،ABG-hc03 ،RBC، پلاکت خون، WBC با وضعیت خروجی بیماران ارتباط معنی‌دار دارند. نتایج آزمایش نامبرده در جدول ۲ نشان داده شده‌است.

جدول ۲. نتایج آزمون مقایسه میانگین و آزمون تی بر روی متغیرهای عددی

نام متغیر	وضعیت پیامد	تعداد	میانگین	P
سن (سال)	مرگ	۲۳۷	۴۶/۳	۰/۰۰۰۱
	بهبود	۸۳۶	۳۸/۶	
مدت بستری(روز)	مرگ	۲۳۷	۷/۲	۰/۰۰۰۱
	بهبود	۸۳۶	۱۳/۱	
هموگلوبین (mg/dl)	مرگ	۲۳۷	۱۲	۰/۰۰۰۱
	بهبود	۸۳۶	۱۲/۶	
هماتوکریت (درصد)	مرگ	۲۳۷	۳۶/۶	۰/۰۰۱
	بهبود	۸۳۶	۳۸/۲	
Mch (pg)	مرگ	۲۳۷	۲۸/۷	۰/۹۹۶
	بهبود	۸۳۶	۲۸/۷	
Mchc(درصد)	مرگ	۲۳۷	۳۲/۷	۰/۰۶۷
	بهبود	۸۳۶	۳۳/۳	
Mcv (fLit)	مرگ	۲۳۷	۸۷/۶	۰/۹۱۲
	بهبود	۸۳۶	۸۷/۵	
سدیم (MEq/Lit)	مرگ	۲۳۷	۱۳۹/۲	۰/۰۳
	بهبود	۸۳۶	۱۳۸/۶	
پتاسیم (MEq/Lit)	مرگ	۲۳۷	۳/۸	۰/۲۵۵
	بهبود	۸۳۶	۳/۸	
pt_control (ثانیه)	مرگ	۲۳۷	۱۲/۹	۰/۲۳
	بهبود	۸۳۶	۱۳/۱	
pt_activity (درصد)	مرگ	۲۳۷	۸۰/۷	۰/۰۰۰۱
	بهبود	۸۳۶	۹۱/۳	
pt_inr	مرگ	۲۳۷	۱/۴	۰/۰۰۰۱
	بهبود	۸۳۶	۱/۱	
pt_patient (ثانیه)	مرگ	۲۳۷	۱۵/۷	۰/۰۰۰۱
	بهبود	۸۳۶	۱۳/۷	
ptt_patient (ثانیه)	مرگ	۲۳۷	۴۳/۶	۰/۰۰۰۱
	بهبود	۸۳۶	۳۵/۱	
ABG_bb	مرگ	۲۳۷	۴۱/۸	۰/۰۰۰۱
	بهبود	۸۳۶	۴۴/۹	
ABG_be_ecf	مرگ	۲۳۷	-۶/۱	۰/۰۰۰۱
	بهبود	۸۳۶	-۳/۲	
ABG_o2 (درصد)	مرگ	۲۳۷	۷۸/۹	۰/۰۰۹
	بهبود	۸۳۶	۸۳/۱	
ABG_p_o2 (Hg mm)	مرگ	۲۳۷	۹۱/۴	۰/۰۷۶
	بهبود	۸۳۶	۸۳/۵	
ABG_p_co2 (Hg mm)	مرگ	۲۳۷	۴۱/۱	۰/۰۰۴
	بهبود	۸۳۶	۳۸/۵	
ABG_hco3 (MEq/Lit)	مرگ	۲۳۷	۱۹/۱	۰/۰۰۰۱
	بهبود	۸۳۶	۲۱/۲	
ABG_pH	مرگ	۲۳۷	۹/۵	۰/۰۸۱
	بهبود	۸۳۶	۷/۴	
rbc_hm1 (count/ML)	مرگ	۲۳۷	۴/۲	۰/۰۰۰۱
	بهبود	۸۳۶	۴/۴	
پلاکت‌خون (count/ML)	مرگ	۲۳۷	۲۰۱۳۸۳/۹	۰/۰۰۰۱
	بهبود	۸۳۶	۲۱۹۵۹۰/۹	
Wbc (count/ML)	مرگ	۲۳۷	۱۹۸۸۲/۷	۰/۰۰۵
	بهبود	۸۳۶	۱۵۵۳۷/۵	

بهترین روش‌ها و الگوریتم‌های داده کاوی که در تحلیل داده‌های پزشکی بهترین کارایی را دارند عبارتند از: روش دسته بندی و الگوریتم‌های k نزدیک‌ترین همسایه، بیز ساده، شبکه‌های عصبی، درخت تصمیم و درخت رگرسیون (۱۷) که تمام این تکنیک‌ها بر داده‌های تحقیق اجرا و دقت درستی ارائه مدل آنان از روش اعتبارسنجی ۱۰ تایی ارزیابی شد که مقایسه آنان در جدول ۳ نشان داده شده‌است.

یکی از روش‌های دوست داشتنی ارزیابی، مدل اعتبار سنجی متقابل k تایی k-fold cross validation است. در این روش، داده‌ها به صورت تصادفی به k زیر مجموعه‌ی مجزا تقسیم می‌شوند. آموزش و تست k بار انجام می‌شود به این صورت که هر بار یکی از زیر مجموعه‌ها برای تست مدل نگه‌داشته شده و بقیه برای آموزش مدل استفاده می‌شوند. این فرایند k بار تکرار می‌شود به‌طوری که هر یک از زیر مجموعه‌ها دقیقاً یک بار برای تست مدل به کار برده می‌شوند. در نهایت نتیجه‌ی این k تکرار برای دست‌یابی به یک برآورد نهایی میانگین‌گیری می‌شوند.

به طور کلی تعداد ۱۰ در ارزیابی اعتبارسنجی متقابل یعنی 10-fold cross validation برای برآورد دقت پیشنهاد می‌شود که در این مطالعه لحاظ شده‌است.

جدول ۳. دقت درستی الگوریتم‌های داده کاوی اجرا شده بر روی داده‌ها

نام الگوریتم	دقت درستی (درصد)
درخت تصمیم	۹۱/۲۷
نزدیک‌ترین همسایه	۸۹/۴۶
درخت c4.5	۸۸/۰۴
درخت رگرسیون	۸۳/۹۶
شبکه عصبی	۷۸/۰۲
بیز ساده	۸۹/۲۳

با توجه به جدول ۳، بهترین الگوریتم مربوط به درخت تصمیم و نزدیک‌ترین همسایه (IBK) به‌ترتیب با دقت ۹۱٪ و

۸۹/۵٪ است.

درخت تصمیم به سبب توان تفسیر بسیار بالا بسیار محبوب است .درخت تصمیم یک سری پرسش است، که هر سوال یک ویژگی را بررسی می‌کند و بر پایه مقادیر آن شاخه میزند. برگ‌های درخت همان برچسب مورد نظر هستند. الگوریتم با یک گره ریشه که شامل کل داده‌هاست آغاز می‌شود و هر بار با تست یکی از ویژگی‌ها در هر گره، مجموعه را به مجموعه‌های کوچک‌تر می‌شکند. آنقدر این کار ادامه می‌یابد که تمام اعضای یک زیر مجموعه متعلق به یک گروه باشد، یا این تعداد از آستانه‌ی ویژه‌ای کمتر باشد. در نهایت درخت برای پیش‌گیری از بیش برازش overfitting داده‌ها هرس می‌شود. نحوه‌ی جستجو در درخت‌ها به صورت حریصانه است.

در استخراج ویژگی‌های موثر در مرگ بیماران از الگوریتم درخت تصمیم به شیوه بهترین- اولین (BestFirst) استفاده شد که از ۵۲ ویژگی بررسی شده، ۱۸ ویژگی که ارتباط بیشتری با مرگ داشتند پیشنهاد شد که به ترتیب اهمیت شامل موارد زیر می‌باشند:

سن، مدت بستری، ضربه به سر، زانو، چندین نقطه از بدن و آزمایش‌ها شامل: هماتولوژی (MCHC هماتوکریت)، سدیم، فاکتورهای انعقاد خون (pt-inr , pt-patient, ptt_patient)، گازهای خونی(ABG_bb, ABG_p_o2, ABG_p_co2, ABG_hco3, ABG_pH) و پلاکت خون.

در شناسایی عوامل موثر از روش خوشه‌بندی توسط الگوریتم دورترین- اولین نیز استفاده شد. الگوریتم دورترین- اولین یکی از روش‌های خوشه‌بندی است و خوشه‌های تقریبی بسیاری را به سرعت می‌یابد، بدین‌صورت که در شروع یک راه حل ایجاد می‌کند. این الگوریتم با یک نقطه تصادفی آغاز می‌شود و سپس آنقدر ادامه می‌یابد تا دورترین نقطه نسبت به اولین نقطه‌ای که انتخاب شده‌است یافت شود و آن را به مجموعه راه حل‌ها می‌افزاید و مکان هر یک از مراکز خوشه‌ها به نوبه خود در دورترین نقطه از مرکز خوشه موجود قرار دارند. این نقطه باید درون فضای داده موجود باشد. این الگوریتم تا حد زیادی خوشه بندی را تسریع می‌کند، احتمال انتخاب دوباره هر نقطه را تا حد زیادی کاهش می‌بخشد و سازگاری و تطبیق را افزایش می‌دهد. می‌توان گفت این الگوریتم نوعی الگوریتم k-means است و برای رفع برخی از اشکال‌های آن در میان گذاشته شده‌است.

با این الگوریتم بر روی داده‌ها، ویژگی‌ها در دو خوشه قرار گرفتند که خوشه اول ۹۴٪ کل داده‌ها را در بر گرفته و نشان‌دهنده درستی خوشه‌بندی است. نتیجه این خوشه‌بندی در جدول ٤ آورده شده‌است.

جدول ٤. خوشه‌بندی توسط الگوریتم دورترین-اولین (Farthest First)

نام ویژگی	خوشه ۱ (۹۴٪ داده‌ها)
جنسیت	مرد
وضعیت تأهل	متأهل
نوع پذیرش	تصادفی
منطقه وقوع حادثه	خارج از شهر و روستا
محل حادثه	بزرگراه
نوع حادثه	تصادف خودرو سوار
سن	۵۰
مدت بستری	۲۶
ضربه به سر	بلی
ضربه به گردن	خیر
ضربه به گلو	خیر
ضربه به شکم	خیر
ضربه به شانه	خیر
ضربه به آرنج	خیر
ضربه به دست	خیر
ضربه به مفصل ران	خیر
ضربه به زانو	خیر
ضربه به پا	خیر
ضربه به نقاط مختلف	خیر
ضربه به منطقه نا مشخص	خیر
تاثیراز ضربات	خیر
سوختگی	خیر
سرمازدگی	خیر
مسمومیت دارویی	خیر
مسمومیت غیردارویی	خیر
سایر علل خارجی و ناشناخته	خیر
عوارض زودرس تروما	خیر
هموگلوبین	۱۳/۹
هماتوکریت	۴۰/۴
mch	۲۸/۶
mchc	۳۴/۴۱
mcv	۸۳/۱۳

نام ویژگی	خوشه ۱ (۹۴٪ داده‌ها)
سدیم	۱۳۲
potassium	۳
pt_patient	۱۳
ptt_patient	۳۰
ABG_bb	۴۱/۶
rbc_hm1	۴/۸۶
پلاکت خون	۲۹۶۰۰۰
wbc	۹۶۰۰
وضعیت خروجی بیمار	فوت

بحث و نتیجه‌گیری

در این پژوهش ۱۰۷۳ بیمار ترومایی که از سال ۱۳۹۰ تا ۱۳۹۲ در بیمارستان پورسینای رشت مراجعه و آزمایش‌های مهمی مانند فاکتورهای انعقادخون داشتند بررسی شدند. الگوریتم پرکاربرد در پژوهش‌های مشابه بر این داده‌ها آزمایش شد که از آن میان درخت تصمیم با ۹۱٪ بالاترین ریزبینی پیشگویی را داشت که سپس به روش بهترین- اولین (Best First) از ۵۲ ویژگی بیماران مورد بررسی، ۱۸ ویژگی از مؤثرترین عوامل مرگ‌ومیر آنان تعیین شد که به ترتیب اهمیت، دربردارنده موارد زیر بود: سن، مدت بستری، ضربه به سر، زانو، چند قسمت بدن و آزمایش شامل: هماتولوژی (MCHC, هماتوکریت)، سدیم، فاکتورهای انعقادخون (pt-inr, pt-patient, ptt_patient)، گازهای خونی(ABG_bb, ABG_p_o2, ABG_p_co2, ABG_hco3, ABG_pH) و پلاکت خون.

دمسار و همکاران در پژوهش خود نشان دادند که دسته کوچکی از ویژگی‌ها می‌تواند دربردارنده اطلاعات کافی برای ساخت الگوی بخردانه دقیق پیش‌آگهی باشد. تحقیق آنان تنها بر ٦٨ بیمار و ۱۰ ویژگی صورت گرفت و با استفاده از روش‌های یادگیری ماشین موفق به شناسایی ۲ فاکتور: مقدار pH به هنگام بستری در ICU و فاکتورهای انعقادخون که از مهم‌ترین فاکتورها برای پیشگویی هستند، نائل شدند. این یافته‌ها برای پاتوفیزیولوژی مناسب بوده و ۲ مشکل از ۳ مشکل بزرگ بیماران ترومایی شدید، اسیدهای متابولیک، هیپوترمی(کاهش دما) و اختلال انعقاد خون را نشان می‌دهد. (۱۱) تحقیق ما که بر ۵۲ ویژگی صورت گرفت نیز نشان می‌دهد که افزون بر فاکتورهای انعقاد خون (pt-inr, pt-patient, ptt_patient) و گازهای خونی(ABG_bb, ABG_p_o2, ABG_p_co2, ABG_hco3, ABG_pH)، عواملی همچون سن، مدت بستری، ضربه به سر، زانو، چند نقطه از بدن و آزمایش شامل: هماتولوژی (MCHC, هماتوکریت)، سدیم و پلاکت خون نیز در مرگ بیماران موثرند و این ویژگی‌ها با الگوریتم درخت تصمیم به شیوه بهترین- اولین (Best First) تعیین گردید.

سات و همکاران برای پیشگویی مرگ ناشی از ضربه به سر، کارائی شش روش درخت تصمیم :CART, CHAID, E-CHAID, QUES, RFRC, BTCR را بر روی ١٦٠٣ بیمار ضربه به سر، از نظر حساس و ویژه بودن، پیشگویی مثبت یا منفی و میزان دقت ارزیابی شد که روش BTRC (درخت دسته‌بندی تصمیم و رگرسیون تقویت شده) با میزان دقت ۹۳٪ اختلاف معنی‌دارتری نسبت به سایر روش‌ها داشت که می‌تواند ابزار سودمندی برای تصمیم پزشکی برای پیشگویی مرگ ناشی از ضربه به سر باشد(٢٤).0 جهت پیشگویی و تعیین عوامل موثر در مرگ این افراد، پرکاربردترین الگوریتم داده کاوی در علوم پزشکی بر کلیه بیماران ترومایی (با یا بدون ضربه به سر) اجرا شد که از آن میان الگوریتم‌های درخت تصمیم و k نزدیک‌ترین همسایه (Ibk) به ترتیب بالاترین دقت دسته‌بندی ۹۱٪ و۸۹٪ نسبت به سایر الگوریتم‌ها را داشتند.

شیتز و همکاران داده‌های صحنه تصادف و ایجاد پرسشنامه شاخص رتبه‌بندی شدت تروما را اعمال کرد و با روش‌های درخت رگرسیون دو درخت طبقه‌بندی یکی بر پایه آسیب‌های شدید و دیگری آسیب‌های متوسط ایجاد کردند.(۹) در صورتی‌که داده‌های ما از اطلاعات خام بانک اطلاعاتی مقیم در بیمارستان که در حال ثبت کلیه مراجعه‌کنندگان به بیمارستان است بدست آمد و موفق به ارائه الگویی با دقت بالا برای شناسایی عوامل موثر در مرگ از راه درخت تصمیم و خوشه‌بندی شد.

۳۷۲ رکورد پزشکی با آسیب‌های له‌شدگی از زمین لرزه با تحلیل گذشته‌نگر انجام شده که مطالعه گروهی گذشته‌نگر (cohort) است. ۲۱ نشانه خطر با تحلیل رگرسیون لوجستیک

برای ۳ نتیجه مرتبط با علائم له‌شدگی ارزیابی شد. ۲ نوع از مدل (ارزیابی اول در زمین و ارزیابی دوم در بیمارستان) با تحلیل رگرسیون لوجستیک توسعه یافت. مدل مرتب‌سازی اول دربردارنده ضربان قلب، دیرکرد نجات و رنگ ادرار غیرطبیعی و دارای AUC با مقدار ۰/۷۳ بود و مدل دوم شامل WBC، نوار قلبی (تاکی‌کاردی)، رنگ ادرار غیرطبیعی و هیپرکالمی بود که AUC ۰/۷٦ را نشان داد.

تاکنون مطالعات چندانی بر بیماران ترومایی با این حجم اطلاعات و بر داده‌های راستین برای شناسایی عوامل موثر یا پیشگویی وضعیت مرگ یا بهبود آنان از راه تکنیک‌های داده کاوی دست کم در ایران صورت نگرفته است و بیشتر مطالعات مربوط به پیامدهای تروما نظیر اختلال روانی پس از آسیب مغزی(۲۵-۲۷) یا شیوه توان‌بخشی در بیماران آسیب دیده مغزی یا بررسی عوامل ایجادکننده تروما نظیر تصادف صورت گرفته است(۲۵) که همه آنها نیز لزوماً از روش‌های داده کاوی استفاده نکرده‌اند. بیشتر مطالعات به روش‌های داده‌کاوی بر سایر بیماری‌ها نظیر انواع بدخیمی‌ها و تومورهای مغزی (۱۲و۲۰) یا دیابت(۱۸) و غیره بوده‌اند.

پیشنهادها: با توجه به تکنیک‌های پیشرفته داده کاوی مانند ترکیب الگوریتم‌ها یا به کمک از ابزارهای کاهش ویژگی‌ها، می‌توان مدلی جهت پیشگویی مرگ بیماران ترومایی ارائه کرد. همچنین، پس از کاهش ویژگی‌ها و هرس قواعد زائد درخت تصمیم، به استخراج قواعد دسته‌بندی شدت آسیب‌دیدگی پرداخت که موجب یاری بیشتر جراحان در شناخت بیماران حاد خواهد شد. همچنین، با تکمیل اطلاعات اولیه بیماران مانند نشانه‌های بالینی و پاسخ آزمایش تصویربرداری و سونوگرافی در سیستم بیمارستانی می‌توان به نتایج بهتری دست یافت و پیشگویی‌های دقیق‌تری انجام داد.

تشکر و قدردانی: از کارشناسان فناوری اطلاعات مرکز (آقایان اتابکی و محمودی) و مدیر تولید شرکت میکرو افزار قشم (آقای رضوان غازیانی) شرکت تولیدکننده نرم‌افزار بیمارستانی بیمارستان پورسینای رشت که در گردآوری و استخراج داده‌ها کمال همکاری را با نویسندگان داشته‌اند و همچنین از مسئولان و کارکنان مرکز تحقیقات ترومای جاده‌ای گیلان، برای راهنمایی ارزنده‌شان کمال تشکر را داریم.

نویسندگان اعلام می‌دارند که هیچ‌گونه تضاد منافعی ندارند.

منابع

1. Ranjbar F, Malekpouri R. Assessment methods for quantitation of trauma deaths in patients presenting to the emergency department in Poursina hospital at Rasht in 2011-2012 .Thesis of MD Guilan university of medical sciences.2012: 18. [Text in Persian]
2. Sharma B.R. Triage in Trauma-Care System: A Forensic View. Journal of Clinical Forensic Medicine 2005; 12:64-73
3. Moini M, Rezaishiraz H, Zarineh A, Rasouli R.M. Evaluation of Quality of Trauma Care in a Local Hospital Using a Customization of ASCOT. European of Trauma and Emergency Surgery 2009; 35:56-60.
4. Brunicardia Ch, Andersen D, Billiar T. Schwart'z principles of Surgery 2010; 7:136.
5. Peden M, Scurfieia R, Sleet D, Mohan D,Hyder AA, Jara wan E , etal. Geneva; World Health Organization, 2004.
6. Hemmati H, Yousefzadeh Sh, Dehnadimoghadam A, Mohammadi-Melksari H, Ahmadi-Dafchahi M and Shabani S. Trauma in Guilan (North of Iran): An Epidemiologic Study. Acta Medica Iranica 2009; 47(5).
7. Laupland K B Svenson, W. L, Grant V, Ball G. Ch, Mercado M, Kirkpatrik A. Long Term Mortality Outcome of Victims of Major Trauma 2010; 41: 69-72
8. Yutaka K, Abe T, Kohshi K, Tokuda Y, Cook F. Revised Trauma Scoring System to Predict in-Hospital Mortality in the Emergency Department: Glascoe Coma Scale,Age, and Systolic Blood Pressure Score 2011; 15
9. Scheetz L, Zhang J, Kolassa J. Classification tree modeling to identify severe and moderate vehicular injuries in young and middle-aged adults. Artificial Intelligence in Medicine 2009; 45, 1-10.
10. Chaiyut T, Wongkonkitsin N, Sae Seow O, Rangsrikajee D, Jenwitheesuk K, Phugkhem A, Bhudisawadi V. Outcome of In-Patient Trauma Cases: Accident and Emergency Unit. Khon Kaen University. J Med Assoc Thai 2005; 88(1): 150-154.
11. Demšar J, Blaž Z , Aoki N, Wall M, Granchi T, Beck Robert. Feature mining and predictive model construction from severe trauma patient's data. International Journal of Medical Informatics 2001; 63: 41–50
12. Santos RS, Malheiros SMF, Cavalheiro S, Parente de Oliveira JM. A Data mining system for providing analytical information on brain tumors to public health decision makers. The Journal of Computer Methods and Programs in Biomedicine 2013; 109: 269–282.

13. Shahrabi J, Zolghadr Shojaei A. Advanced data mining, concept and algorithms. Tehran; Jahad Daneshgahi, Amirkabir , 2010[Text in Persian]
14. Ghazanfari, M. the book of Data mining and knowledge. 1392; 3-5 [Text in Persian]
15. Balac N. the book of Introduction To Data Mining. 2006; 5-10
16. Salalri M, Adibnia F. 10 algorithms of the best data mining techniques. 13'th student conference of electronic engineering of IRAN.Tehran; Tarbiat Modares University, 2011. [Text in Persian]
17. Hassanzadeh M, Farhoudinejad A. Classification of data mining algorithms to build predictive models to assess the severity of injury in trauma patients. ways of soft calculate in electronic and computer engineering region conference. Safashahr Azad University, Shiraz, Iran: 2015. [Text in Persian]
18. Antonelli D, Baralis Ea, Giulia B, Cerquitelli T, Chiusano S, Mahoto N. Analysis of diabetic patients through their examination history. The Jouranl of Expert Systems with Applications 2013; 40:4672–4678.
19. Dehghani T, Afshari S, Khalilzaseh M. Clustering algorithm K-mean genetic based for finding heart attack. 5'th IRAN data mining conference. Tehran;Amirkabir University, 2012. [Text in Persian]
20. Asadi N, Sadrodini M. Use of data mining for knowledge effective factors in cancers and finding best treatment way in Namazi hospital cancwr database. 16'th international conference of IRAN comuter groups. Tehran; Sharif University, 2011. [Text in Persian]
21. Lin WT, WU YC, Zheng JS, Chen MY. Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. Expert Systems with Applications 2011; 38:11078–11084.
22. Pardalos PM, Tomaino V, Xanthopoulos P. Optimization and data mining in medicine. 2009;17: 215-236
23. Oliaei A, Salmasi N, An efficient model to predict duration hospitalization of digestive system patients. 5'th datamining conference of IRAN. Tehran; Abirkabir Uiversity, 2012. [Text in Persian]
24. Sut N, Osman S. Comparison of regression tree data mining methods for prediction of mortality in head injury. The Journal of Expert Systems with Applications 2011; 38: 15534–15539.
25. Marcano-Cedeño A, Chausa P, García A, Cáceres C, Tormos J.M., Gómez E.J. Data mining applied to the cognitive rehabilitation of patients with acquired brain injury. The J of Expert Systems with Applications 2013; 40:1054–1060.
26. Marcano-Cede A, Chausa P, García A, Cáceres C, Tormos J.M., Gómez E.J. Artificial metaplasticity prediction model for cognitive rehabilitationoutcome in acquired brain injury patients. The J of Artificial Intelligence in Medicine 2013; 58: 91 99
27. Rezaei S, Yousefzadeh Sh, Moosavi H, Kazemnejad E, Khodadadi N. Designing a predictive model to determine the risk factors of mental disorders following traumatic brain injury. Feyz 2013; 17(1): 40 53. [Text in Persian]

Using Data Mining Techniques to Extract Clinical Disorders Affecting Mortality in Trauma Patients

Hassanzadeh M (MSc)[1,2]-Frhoudinejad A(PhD)[1]-Yousefzadeh Sh(MD)[3]

[*]**Corresponding Address:** Department of Computer and Information Technology, Faculty of Engineering, University of Payame Noor, Tehran, Iran- Management of Statistic and Information Technology, Guilan University of Medical Sciences, Rasht, Iran

Email: hassanzadeh@gums.ac.ir

Received: 31 May 2014 Revise : 27 Aug/2014 Accepted : 12 Oct/2014

Abstract

Introduction: Trauma is one of the most common causes of death in the world, which often occurs as a result of road accidents. Prompt identification of patients with acute injury, leads to take the appropriate medical actions and thus, save lives and also avoid enormous cost of treatment.

Objective: Finding the best data mining algorithms to identify clinical disorders resulting in death in trauma patients

Materials and Methods: 1,073 trauma patients hospitalized in Poursina Hospital in Rasht with their 52 recorded clinical conditions (features) have been analyzed in this research. In order to automatically identify emergency cases, a number of classification algorithms have been modified for the task, such as decision tree, K-nearest neighbor, and neural network methods. These algorithms have been trained over a wide range of features and their performance has been investigated using 10-fold cross validation.

Results: Totally, 82.8% (888) of the surveyed patients were male and17.2% (185) were female. 22.1% died, most of them (30%) in the first week after their hospitalization and 23.6% on the first day. No significant relationship has been found between the duration of hospitalization and mortality. Among the classification algorithms, decision tree and k-nearest were able to recognize death cases with higher precision, (i.e. 91% and 89%, respectively). In order to find effective factors on training a better Decision Tree classifier, the Best First algorithm was used which then selected and could identify 18 effective features (of 52 initial features).

Conclusion: Given the high accuracy of some data mining algorithms, like Decision Tree algorithm, we are able to differentiate severe trauma cases which may lead to death from those with mild injuries. Hence, their application to predict mortality in trauma patients and identify those at life risk can be investigated in real environment.

Conflict of interest: non declared

Key words: Data Mining\ Mortality\ Trauma

Journal of Guilan University of Medical Sciences, No: 95, Pages: 52-62

Please cite this article as: Hassanzadeh M, Frhoudinejad A, Yousefzadeh Sh. Using Data Mining Techniques to Extract Clinical Disorders Affecting Mortality in Trauma Patients. J of Guilan University of Med Sci 2015; 24 (95) :52-62. [Text in Persian]

1. Department of Computer and Information Technology, Faculty of Engineering, University of Payame Noor, Tehran, Iran
2. Management of Statistic and Information Technology, Guilan University of Medical Sciences, Rasht, Iran
3. Department of Neorosurgery, Faculty of Medical, Guilan Road Trauma Center, Guilan University of Medical Sciences, Rasht, Iran