Predicting the risk of myocardial infarction by Decision Tree Method

Safdari R[1], Ghazi saeedi M[2], Gharooni M[3], Nasiri M[4] Arji G[5]

Abstract

Purpose: Cardiovascular diseases are among the most common diseases in all societies. Using data mining techniques to generate predictive models to identify those at risk for reducing the effects of the disease is very helpful. The main purpose of this study was to predict the risk of myocardial infarction by Decision Tree based on the observed risk factors.

Methods: The present work was an analytical study conducted on a database containing 350 records. Data were obtained from patients admitted to Shahid Rajaei specialized cardiovascular hospital, Iran, in 2011. Data were collected using a four-sectioned data collection form. Data analysis was performed using SPSS statistical software version 12 by CRISP methodology. In the modeling section decision tree and Neural Network were used.

Results: The results of the data mining showed that the variables of high blood pressure, hyperlipidemia and tobacco smoking were the most critical risk factors of myocardial infarction. The accuracy of the decision tree model on the data was shown to be as 93/4%.

Conclusion: The best created model was decision tree C5.0. According to the created rules, it can be predicted which patient with new specified features may affected by myocardial infarction.

Keywords: Decision trees, Neural network, Myocardial infarction, Data mining

دریافت مقاله:۹۲/۰۹/۳۰ تایید مقاله:۹۳/۳/۲۰

مقایسه عملکرد درخت تصمیم گیری و شبکه عصبی در پیشگویی ابتلا به آنفارکتوس قلبی

رضا صفدری[۱]، مرجان قاضی سعیدی[۲]، منوچهر قارونی[۳]، مهدی نصیری[۴]، گلی ارجی[۵]

هدف: بیماری های قلبی عروقی از شایع ترین بیماری ها در تمامی جوامع می باشد. استفاده از تکنیک های داده‌کاوی برای ایجاد مدل‌های پیش‌گویی کننده، جهت شناسایی افراد در معرض خطر برای کاهش عوارض ناشی از بیماری بسیار کمک کننده است. هدف اصلی این مطالعه پیشگویی احتمال ابتلا افراد به آنفارکتوس قلبی با استفاده از درخت تصمیم بر اساس ریسک فاکتورهای موثر بر ابتلاست.

روش بررسی: این پژوهش از نوع تحلیلی بوده و پایگاه داده آن شامل ۳۵۰ رکورد می باشد. داده های مورد نیاز این تحقیق در سال ۱۳۹۰با استفاده از جدول مورگان از بین پرونده بیماران مراجعه کننده به بیمارستان شهید رجایی تهران بدست آمده است ابزار جمع آوری داده‌ها چک لیستی چهار قسمتی بوده است.تجزیه و تحلیل به کمک نرم افزار SPSS Clementine 12 با بکارگیری متدولوژی CRISP[۱] انجام شده است. در بخش مدل‌سازی از درخت تصمیم و شبکه عصبی استفاده شده است.

یافته ها: با توجه به مدل‌های استفاده شده مشخص شد که به ترتیب متغیرهای فشارخون بالا، چربی خون بالا و مصرف سیگار، بیشترین تاثیر را در ابتلا به آنفارکتوس قلبی دارا بودند. به کمک درخت تصمیم ایجاد شده، قوانینی استخراج شده است که می تواند به عنوان الگویی در جهت پیشگویی احتمال ابتلا افراد به آنفارکتوس قلبی استفاده شود. صحت مدل ایجاد شده با استفاده از درخت تصمیم ۹۳/۴ درصد بوده است.

[1]Cross Industry Standard Process For Data Mining

Archive of SID
www.SID.ir

نتیجه گیری: بهترین مدل ایجاد شده درخت تصمیم C5 بود. با بکارگیری قوانین ایجاد شده برای یک نمونه جدید با ویژگیهای مشخص می توان تعیین کرد که احتمال ابتلا به آنفارکتوس قلبی چقدر است.

واژگان کلیدی: درخت تصمیم، شبکه عصبی، آنفارکتوس قلبی، داده کاوی

نویسنده مسئول: گلی ارجی، Goliarji@ymail.com

آدرس: تهران،دانشگاه علوم پزشکی تهران،دانشکده پیراپزشکی

۱- دانشیار، عضو هیئت علمی گروه آموزشی مدیریت اطلاعات سلامت، دانشکده پیراپزشکی،دانشگاه علوم پزشکی تهران، ایران

۲- استادیار، عضو هیئت علمی گروه آموزشی مدیریت اطلاعات سلامت، دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران، ایران

۳- استاد بیماریهای قلب و عروق, دانشگاه علوم پزشکی تهران، ایران

۴- دانشجوی دکترای هوش مصنوعی، دانشگاه علم و صنعت ایران،تهران،ایران

۵- دانشجوی دکترای مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، ایران

مقدمه

بیماری ایسکمیک قلب (Ischemich heart diseases) بیش از هر بیماری دیگر در کشورهای توسعه یافته موجب مرگ و ناتوانی شده و هزینه های اقتصادی تحمیل می نماید (۱). با توجه به افزایش سریع بیماریهای ایسکمیک قلبی در سراسر جهان احتمالا این بیماری تا سال ۲۰۲۰ به شایعترین علت مرگ تبدیل خواهد شد (۲). در ایران، سکته قلبی اولین علت مرگ افراد بالاتر از ۳۵ سال می باشد (۳). در تعیین عوامل موثر در ابتلا به بیماری های قلبی عروقی بررسی های زیادی انجام گرفته و عوامل متعدد در آنها ذکر شده است (۶-۴).

محیط مراقبت سلامت غنی از اطلاعات و ضعیف از دانش است (۷). حجم داده های پزشکی روز به روز در حال افزایش است و پزشکان معمولا اطلاعات ارزشمندی را در خصوص بیماری‌ها و ارتباط آنها با یکدیگر و عوامل ایجاد کننده بیماری‌ها بدست می آورند (۸). اما این مجموعه داده‌های خام به خودی خود ارزشی ندارند، برای معنی بخشیدن به این داده‌ها باید آنها را تحلیل و تبدیل به اطلاعات یا بهتر از آنها دانش کرد (۹). با توجه به شیوع بیماریهای قلبی – عروقی در سراسر جهان، استفاده از روش های جدید در تحقیقات زیست پزشکی بسیار مورد توجه قرار گرفته است. داده‌کاوی می‌تواند ارتباطات و وابستگی‌های بدیعی را کشف کند که برای پزشکان بسیار مفید است.

تکنیکهای داده‌کاوی به طور کلی به دو دسته توصیف کننده و پیش‌گویی‌کننده تقسیم می شوند. وظایف توصیفی خواص عمومی داده‌ها را مشخص می‌کند و هدف آن پیدا کردن الگوهای قابل تفسیر توسط افراد برای داده‌هاست. وظایف پیش‌گویانه، پیش‌بینی رفتار آینده آنهاست و منظور از آن بکارگیری چند متغیر در پایگاه داده برای پیش‌گویی مقادیر آینده متغیرهاست (۱۲). داده کاوی گونه ای از تکنیک‌ها برای شناسایی اطلاعات و یا دانش تصمیم‌گیری از میان داده‌ها می باشد، به نحوی که با استخراج آنها در حوزه های تصمیم‌گیری، پیش‌بینی، پیش‌گویی و تخمین بتوان از آنها استفاده کرد. داده‌کاوی پزشکی دارای پتانسیل زیادی برای کشف الگوهای پنهان موجود در داده‌ها داراست که این الگوها می تواند برای تشخیص‌های بالینی مورد استفاده قرار گیرد (۹). امروزه استفاده از روش‌های متنوع داده‌کاوی و استخراج دانش برای شناسایی الگوها و ارتباطات میان متغیرهای مختلف در تولید مدل‌های پیش‌بینی‌کننده در علوم پزشکی بسیار مورد توجه قرار گرفته است (۱۰). کاربرد روش‌های داده‌کاوی در حوزه‌های مختلف پزشکی مانند تشخیص، پیش‌گویی و حتی درمان به اثبات رسیده است (۱۱).

یکی از عملکردهای پیش‌گویانه در داده‌کاوی، دسته‌بندی است. دسته بندی فرایند یافتن مدلی است که با تشخیص دسته‌ها و یا مفاهیم داده می‌تواند دسته ناشناخته اشیا دیگر را پیش‌گویی کند (۱۲). یکی از روش‌های رایج دسته‌بندی درخت تصمیم است. درخت تصمیم‌گیری روشی است که به شما در یک انتخاب خوب کمک می کند. به خصوص تصمیم‌گیری‌هایی که در بردارنده هزینه بالا و خطرات زیادی است. درختان تصمیم یک روش گرافیکی برای مقایسه رقابت جایگزین و اختصاص ارزش به آنها از طریق ترکیب عدم قطعیت ها، هزینه‌ها و بازپرداخت ها به وسیله ارزش‌های خاص عددی است (۱۳).

Archive of SID

www.SID.ir

درخت تصمیم معمولا از چندین گره[1] تشکیل می شود که با نام گره‌های ورودی و خروجی شناخته می شوند. قوانین ایجاد شده در درخت تصمیم به صورت "اگر" و "آنگاه بیان می شوند. از میان الگوریتم های مورد استفاده در ساخت درخت تصمیم، مهم‌ترین آنها الگوریتم C5[2] است که توسعه یافته ID3 [3] می باشد.

شبکه‌های عصبی مصنوعی با پردازش داده های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می کنند، به همین دلیل به این سیستم‌ها هوشمند می گویند. زیرا بر اساس محاسبات داده‌های عددی یا مثال‌ها، قوانین کلی را فرا می گیرند. در واقع شبکه‌های عصبی تکنیک هایی هستند که می توانند برای مدل‌های با ارتباطات پیچیده مورد استفاده قرار گیرند. مدل‌های شبکه عصبی مصنوعی در طبقه بندی مسائلی نظیر، پیش‌بینی مشکلات قلبی در بیماران، تشخیص فشار خون و... استفاده شده است (۱۴). پیش‌گویی خطر ابتلا به بیماری‌های قلبی به کمک سیستم‌های حمایت از تصمیم نقش مهمی در پیشگیری از بیماری ایفا می کند (۱۵). مدل‌های پیش‌گویی خطر دارای پتانسیل بالایی در جهت کمک به فرایند تصمیم‌گیری در مورد شرایط بیمار را دارا هستند. این مدل‌ها چند متغیره بوده که از برخی فاکتورهای مهم جهت پیش‌گویی و تشخیص و درمان مناسب برای افراد استفاده می‌کنند و بر اساس آنها مداخلات بهداشتی یا تغییر سبک زندگی نسبت به کسانی که در معرض خطر تشخیص داده شده اند می تواند انجام شود (۱۶).

در مطالعه‌ای با عنوان مدلی برای پیش‌گویی خطر ابتلا به بیماری‌های قلبی عروقی، مقایسه بین درخت CART[4]، رگرسیون لجستیک[5] و پرسپترون چند لایه[6] که توسط Isabell و همکارانش انجام شده است، مدل درخت تصمیم CART و شبکه عصبی به عنوان بهترین مدل در بین مدل‌های اجرا شده بیان شده است (۱۷). Christine و همکارانش در مطالعه خود نشان داده اند که درختان تصمیم نسبت به مدل رگرسیون لجستیک، قابلیت بیشتری در جهت پیش‌گویی امکان ابتلا افراد به بیماریهای قلبی دارا هستند (۱۸). در مطالعه انجام شده توسط Abdolghani جهت پیش‌گویی امکان زنده ماندن بیماران مبتلا به سرطان سینه الگوریتم C4.5 با میزان دقت ۸۶/۷ درصد به عنوان بهترین مدل پیش‌گویی کننده معرفی شده است (۱۹). در مطالعه انجام شده توسط Starc بیان شده است که مدل‌های درخت تصمیم از جمله (ID3, C4.5, CHAID and CART[7]) برای انجام تحلیل‌های اکتشافی در پایگاه‌های داده بزرگ نسبت به مدل‌های دیگر موفق‌تر عمل می‌کنند (۲۰). Ordonze و همکارانش از الگوریتم درخت تصمیم C4.5 و الگوریتم قوانین همبستگی[8] با استفاده از ۲۵ ریسک فاکتور جهت پیش‌گویی بیماری قلبی استفاده کردند و به این نتیجه رسیدند که قوانین همبستگی عموما قواعد پیش بینی ساده‌تری نسبت به درختان تصمیم ایجاد می‌کند (۲۱). در مطالعه انجام شده توسط Jianxin مدل شبکه عصبی با دقت ۸۸/۶ درصد به عنوان بهترین مدل و درخت تصمیم با ۸۸/۴ درصد به عنوان بدترین مدل پیش‌گویی کننده معرفی شده است (۲۲).

در این پژوهش ما با جمع‌آوری داده‌های مربوط به ریسک فاکتورهای موثر در ابتلا افراد به بیماری آنفارکتوس قلبی از پرونده بیماران و به کارگیری الگوریتم‌های داده کاوی به مقایسه دقت مدل‌های مختلف پرداخته و نهایتا به ارائه مدلی جهت پیش گویی احتمال ابتلای افراد به این بیماری پرداختیم، تا بتوان با استفاده از مدل ایجاد شده و براساس متغیرهای فردی و نتایج آزمایشات درباره احتمال ابتلا فرد به آنفارکتوس قلبی پیشگویی کرد. این امر از آن جهت دارای اهمیت است که می توان از آسیب‌های ناشی از روش‌های تهاجمی تشخیص بیماری‌های عروق کرونر مانند آنژیوگرافی تا حدودی پیشگیری نمود و همچنین میزان عوارض تجویز داروهای مورد استفاده برای بیماری عروق کرونر را در بیمارانی که به آن نیاز ندارند، کاهش داد.

مواد و روش ها

روشهای متعددی برای اجرای پروژه های داده کاوی وجود

[1] node
[2] Algorithm for implementing Decision Tree
[3] Iterative Dichotomiser 3
[4] classification and regression trees
[5] Logistic Regression
[6] Multi-Layer Perceptron
[7] Chi-Squared Automatic Interaction Detector
[8] Association rule

Archive of SID
www.SID.ir

دارد که یکی از روش‌های قدرتمند در این زمینه متدولوژی کریسپ می باشد (۲۳). این مقاله نیز بر اساس این متدولوژی تنظیم شده است (شکل۱). در ادامه به بررسی هر یک از این مراحل در جهت رسیدن به مدلی برای پیشگویی احتمال ابتلا به آنفارکتوس قلبی می پردازیم.

الف)شناخت سیستم

به کارگیری موفق داده کاوی مستلزم شناخت حوزه ای است که قرار است داده کاوی در آن به کار برده شود و علاوه بر آن شناخت کافی از روشها و ابزارهای داده کاوی نیز لازم است. به طور کلی تیم داده‌کاوی بایستی دانش کافی در حوزه‌ای که قرار است بررسی شود داشته باشند. در گام اول پژوهشگر با مشورت پزشک متخصص قلب و عروق و نیز با مطالعه بر روی بیماری آنفارکتوس قلبی وتعیین فاکتورهای موثر در ابتلا و همچنین روش های تشخیصی و درمانی و روش‌های پیشگیری از ابتلا به بیماری، سعی در شناخت کافی حوزه مورد بررسی داشت.

ب)مرحله آماده سازی داده ها

در این گام به جمع آوری داده‌ها از پرونده‌های بیماران پرداخته شد. داده‌های استفاده شده در این مطالعه مربوط به پرونده بیماران مراجعه کننده به بیمارستان شهید رجایی تهران در سال۱۳۹۰ بود. تعداد بیماران مراجعه کننده به این بیمارستان در سال مذکور۱۷۵۰ نفر بود. حجم نمونه با جدول مورگان۵۰۰ بدست آمد که نهایتا تعداد ۱۵۰ رکورد از آنها به علت عدم تکمیل متغیرهای مورد بررسی ،از پایگاه داده حذف گردید. در این گام بایستی داده هایی که در حال حاضر در دسترس هستند و داده هایی که برای ساخت مدل نیاز بود، تعیین شدند. برای شروع پژوهش از طریق مطالعات کتابخانه ای و مشورت با پزشک متخصص فاکتورهایی که بیشترین تاثیر را در ابتلای افراد به بیماری دارا بودند تعیین شد و در قالب چک لیست چهار قسمتی که شامل داده های دموگرافیک ، متغیرهای خونی، قندی و چربی بود و روایی آن با استفاده از نظر اساتید و متخصصین حوزه قلب و عروق مورد تایید قرار گرفت و پایایی آن بر اساس ضریب آلفای کرونباخ ۰/۸۴ بدست آمد.

متغیرهای تعیین شده، برای ایجاد مدل به دو دسته متغیر هدف و متغیرهای پیشگو دسته بندی شدند.متغیر هدف ابتلا یا عدم ابتلا به بیماری و سایر متغیرها به عنوان متغیر پیشگو مورد استفاده قرار گرفتند. در این پژوهش نیز برای مدیریت داده های از دست رفته متغیرهای زیر چون به تعداد بسیار کمی در پرونده‌های مورد مطالعه ثبت شده بودند برای اینکه در فرایند ساخت مدل خللی ایجاد نکنند از مجموعه متغیرها حذف شدند که این متغیرها شامل هورمونهای تیروئیدی (T3,T4,TSH) و متغیر HBA1C[1] و فاکتورCRP[2] بود. برای سایر متغیرهایی که مقادیر گمشده‌شان نسبت به مقادیر موجود ناچیز بود با استفاده از امکانات نرم‌افزار SPSS مقدار پیش‌فرض مربوط به آن جایگزین شدند،که این مقدار با توجه به نوع متغیر انتخاب گردید. فاکتورهایی مانند قد و وزن که به تنهایی ارزشی ندارند بلکه شاخص توده بدنی (Body Mass Index) آنها تاثیرگذار است، این شاخص به کمک آنها و با توجه به رابطه زیر بدست آمد:

(۱) شاخص توده بدنی = وزن (کیلوگرم)/قد(سانتی متر)^2

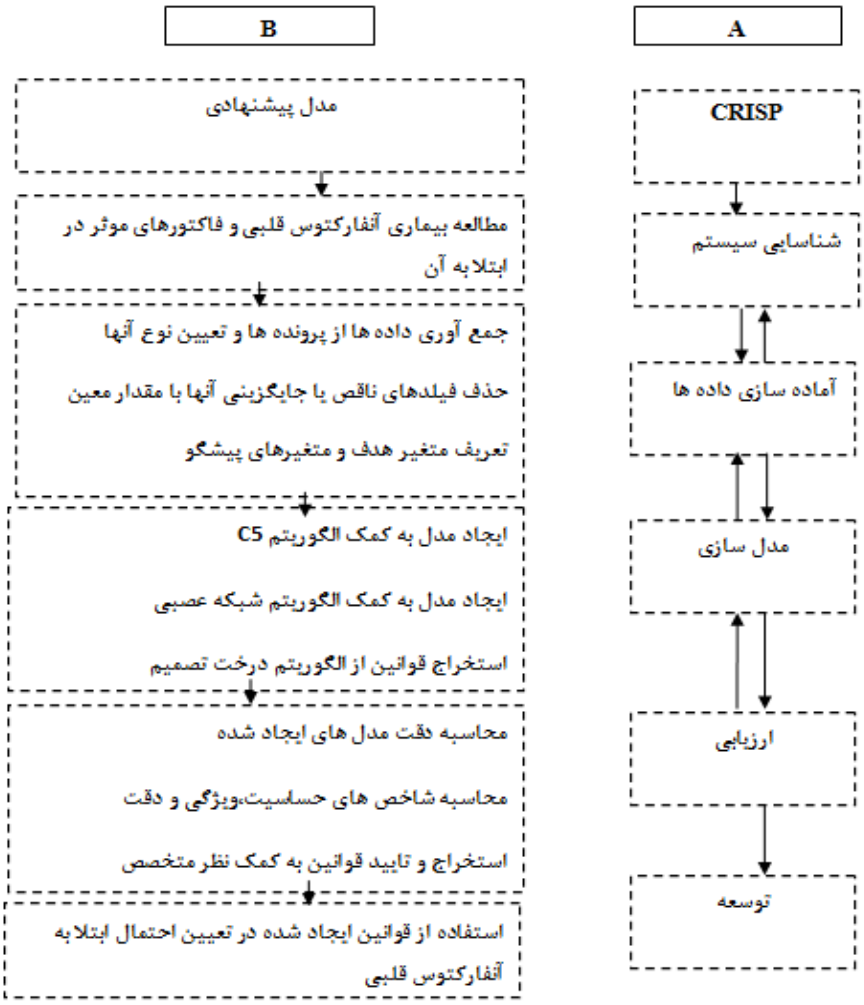
این مطالعه بر روی دو گروه بیمار ۲۰۰ نفر(۵۷/۱۴درصد) و سالم ۱۵۰ نفر (۴۲/۸۶درصد) انجام شد. افراد سالم در این مطالعه افرادی بودند که پس از بررسی‌های اولیه وجود بیماری قلبی در آنها رد شده بود و تشخیص مربوط به عدم ابتلا در پرونده این بیماران درج شده بود.

یافته ها

میانگین سنی بیماران تحت بررسی ۱۰/۸۸ ± ۵۹/۰۶ سال بود. در متغیرهای سن، فشارخون سیستولیک، قند خون ناشتا، قند خون کل، BUN و کراتینین بین دو گروه سالم و بیمار اختلاف معنی‌داری وجود داشت ($p > ۰/۰۵$)، بین سایر متغیرها اختلاف معنی داری وجود نداشت ($p>۰/۰۵$) (جدول ۱). تعداد بیماران مرد نسبت به بیماران زن بیشتر بود. سابقه خانوادگی ابتلا به بیماری‌های عروق کرونر، سابقه مصرف سیگار، سابقه اعتیاد، سابقه دیابت و چربی خون در گروه بیماران بیشتر از افراد سا لم بود. اما سابقه فشارخون بالا در گروه سالم بیشتر از افراد بیمار بود (جدول ۲).

[1]Glycosylated Hemoglobin, Type A1C
[2] Creactive protein

Archive of SID
www.SID.ir



شکل۱: گام های متدولوژی کریسپ(A) و چارچوب استفاده شده در این مطالعه(B)

جدول ۱: میانگین و انحراف معیار متغیرهای کمی مورد بررسی در افراد مورد مطالعه

نام متغیر	میانگین (انحراف معیار) افراد سالم	میانگین (انحراف معیار) بیماران	p-value
فشار خون سیستولیک (Systolic blood pressure)	۱۲۶/۶۱(۱۶/۰۹)	۱۳۳/۴۴(۲۵)	۰/۰۰۲
شاخص توده بدنی(Body Mass Index)	۲۷/۴۳(۴/۴۵)	۲۶/۸۶(۴/۱۳)	۰/۲۲
قند خون ناشتا(GFBS)	۱۱۹(۵۰/۸۴)	۱۳۸/۴۶(۷۰/۳۶)	۰/۰۱
کلسترول(Cholesterol)	۱۶۸/۸۹(۴۵/۹۸)	۱۷۳/۲۷(۱۲۷/۶۹)	۰/۷
تری گلیسرید(TG)	۱۵۴/۲۳(۸۰/۵۴)	۱۳۸/۸۸(۷۴/۹۹)	۰/۰۹
لیپوپروتئین کم چگال(LDL)	۱۰۱/۸۳(۳۹/۴۵)	۹۷/۹۲(۳۴/۷۲)	۰/۳
لیپوپروتئین پرچگال(HDL)	۴۰/۳۰(۱۰/۱۴)	۳۸/۵۴(۱۲/۲۵)	۰/۱
اوره(BUN)	۱۸/۶۱(۹/۰۸)	۲۲/۸۸(۱۷/۶۱)	۰/۰۰۴
کراتینین(Creatinine)	۱/۱(۰/۴۶)	۱/۲۶(۰/۷۹)	۰/۰۲

Archive
www.SID.ir

جدول ۲: توزیع فراوانی متغیرهای کیفی در افراد مورد مطالعه

نام متغیر	افراد سالم (درصد)فراوانی	بیماران (درصد) فراوانی
جنسیت		
مرد	۶۸(۴۵/۳)	۱۴۲(۷۱)
زن	۸۲(۵۴/۷)	۵۸(۲۹)
سابقه خانوادگی بیماری عروق کرونر		
مثبت	۱۹(۱۲/۷)	۴۰(۲۰)
منفی	۱۳۱(۸۷/۳۶)	۱۶۰(۸۰)
سابقه مصرف سیگار		
مثبت	۲۵(۱۶/۷)	۷۲(۳۶)
منفی	۱۲۵(۸۳/۳)	۱۲۸(۶۴)
سابقه اعتیاد		
مثبت	۴(۲/۷)	۲۵(۱۲/۵)
منفی	۱۴۶(۹۷/۳)	۱۷۵(۸۷/۵)
سابقه فشارخون بالا[1]		
مثبت(۱۴۰=<میلی متر جیوه)	۷۶(۵۰/۷)	۸۷(۴۳/۵)
منفی(۱۳۹>= میلی متر جیوه)	۷۴(۴۹/۳)	۱۱۳(۵۶/۵)
سابقه بیماری دیابت		
مثبت	۳۶(۲۴)	۵۷(۲۸/۵)
منفی	۱۱۴(۷۶)	۱۴۳(۷۱/۵)
سابقه چربی خون بالا		
مثبت	۶۳(۴۲)	۷۸(۳۹)
منفی	۸۷(۵۸)	۱۲۲(۶۱)

۱- طبقه بندی سطوح فشار خون براساس راهنمای JNC(Joint National Committee) صورت گرفته است.

ج)مدل سازی

روشهای داده کاوی متنوعی برای مدلسازی وجود دارد. در این مرحله با استفاده از تکنیکهای داده‌کاوی به ارائه مدل پیشگویانه پرداخته شد. مدل سازی با استفاده از نرم افزار SPSS Clementine 12.0 انجام شد و در این مطالعه از روشهای پیش گویانه در داده کاوی استفاده گردید. در این مرحله الگوریتم های درخت تصمیم، شامل الگوریتم-های (C5, C&RT,[1] QUEST, CHAID) و شبکه عصبی با بکارگیری متغیرهای ورودی و تعیین متغیر هدف ایجاد شدند. برای ساخت مدل‌های درخت تصمیم و شبکه عصبی متغیرهای جنسیت، سن، سابقه مصرف سیگار، سابقه اعتیاد، سابقه فشارخون، سابقه چربی خون، فاکتورهای قندی و چربی، شاخص توده بدنی و گروه خونی به عنوان متغیرهای پیشگو تعیین شد و متغیر ابتلا یا عدم ابتلا به بیماری نیز به عنوان متغیر هدف تعیین گردید و در مرحله بعد داده ها به دو بخش آموزش گردید (۸۰درصد) و آزمون (۲۰درصد) تقسیم شدند. مدل شبکه عصبی دارای یک لایه ورودی، تعدادی لایه پنهان و یک لایه خروجی است. هر گره ورودی را دریافت کرده، آن را پردازش نموده و خروجی را تولید می کند. تعیین اینکه آیا هر ورودی به گره خروجی خواهد رسید یا خیر به وزن آن ورودی بستگی دارد (۲۴). یک درخت تصمیم ترکیب تعدادی استلزام منطقی (قانون اگر- آنگاه) است. درخت-های تصمیم تنها نمایشی از فرایند تصمیم گیری نیستند بلکه می توان از آنها در حل مسائل طبقه بندی نیز

[1]Quick Unbiased Efficient Statistical Tree

Archive of SID
www.SID.ir

جدول۳: نمونه ای از قوانین ایجاد شده توسط الگوریتم C5

ردیف	قوانین	احتمال ابتلا
۱	اگر جنسیت=مونث،سابقه مصرف سیگار=مثبت آنگاه احتمال ابتلا برابر است با	٪۸۳/۳۳
۲	اگر جنسیت=مونث، سابقه مصرف سیگار=منفی، سن فرد > ۷۷ سال باشد آنگاه احتمال ابتلا برابر است با	٪۸۷-۵۰
۳	اگر جنسیت=مونث، سابقه مصرف سیگار=منفی،سن فرد<= ۷۷ سال،سطح کراتینین <=۱/۴، شاخص BMI >=۲۹/۸ باشد آنگاه احتمال ابتلا برابر است با	٪۹۸/۶۳
۴	اگر جنسیت=مذکر ،گروه خونی فرد=-O و یا +A باشد آنگاه احتمال ابتلا برابر است با	٪۷۹/۴۲
۵	اگر جنسیت=مذکر ،گروه خونی فرد=-A ،سابقه چربی خون بالا=مثبت باشد آنگاه احتمال ابتلا برابر است با =	٪۷۹/۳۱
۶	اگر جنسیت=مذکر ،گروه خونی فرد=+AB ،سابقه فشارخون بالا=+ باشد آنگاه احتمال ابتلا برابر است با	٪۹۰
۷	اگر جنسیت=مذکر ،گروه خونی فرد=-AB باشد آنگاه احتمال ابتلا برابر است با	٪۵
۸	اگر جنسیت=مذکر ،گروه خونی فرد=+B ،میزان کراتینین> ۱،سابقه دیابت=منفی باشد آنگاه احتمال ابتلا برابر است با	٪۹۲/۰۷
۹	اگر جنسیت=مذکر ،گروه خونی فرد=-B باشد آنگاه احتمال ابتلا برابر است با	٪۱۶
۱۰	اگر جنسیت=مذکر ،گروه خونی فرد=+O، سابقه دیابت=مثبت باشد آنگاه احتمال ابتلا برابر است با	٪۹۱/۴۹
۱۱	اگر جنسیت=مذکر، گروه خونی فرد=+O، سابقه دیابت=منفی ،قند خون ناشتا >۱۲۱ باشد آنگاه احتمال ابتلا برابر است با	٪۸۲/۶۹
۱۲	اگر جنسیت=مذکر ،گروه خونی فرد=+O، سابقه دیابت=منفی ،قند خون ناشتا<۱۲۱، سابقه فشارخون=منفی، سابقه مصرف سیگار=مثبت باشد آنگاه احتمال ابتلا برابر است با	٪۵۲/۵۴
۱۳	اگر جنسیت=مذکر، گروه خونی فرد=+O، سابقه دیابت=منفی، قند خون ناشتا<۱۲۱، سابقه فشارخون=منفی، سابقه مصرف سیگار=مثبت، سن فرد <= ۵۳ سال باشد آنگاه احتمال ابتلا برابر است با	٪۷۹/۸۷
۱۴	اگر جنسیت=مذکر، گروه خونی فرد=+O، سابقه دیابت=منفی، قند خون ناشتا<۱۲۱،سابقه فشارخون=منفی،سابقه مصرف سیگار=مثبت ،سن فرد>۵۴ سال،سطح HDL >۳۳ باشد آنگاه احتمال ابتلا برابر است با	٪۹۵

استفاده کرد. معمولا مجموعه قوانین استخراج شده از درخت تصمیم، مهمترین اطلاعاتی است که از آنها بدست می آید. در مدل ایجاد شده در این نرم افزار جهت تقسیم شاخه ها از شاخص جینی[13] استفاده شده است (۲۵) .در این مطالعه معیار اطمینان برای قوانین ایجاد شده ۸۰ درصد تعیین شد. در جدول۳ مجموعه قوانین ایجاد شده توسط مدل C5 بیان شده است. دلیل انتخاب این مدل نیز به این جهت بود که با محاسبه شاخص‌های مورد نظر دارای بالاترین دقت در بین مدلهای اجرا شده بود. نحوه محاسبه شاخص‌ها در بخش ارزیابی ارائه شده است (جدول۳).

د) ارزیابی

در این مرحله پس از ایجاد مدل بایستی به ارزیابی مدل ایجاد شده پرداخت. برای بررسی صحت مدل داده‌ها به دو دسته آموزش (۸۰درصد) و آزمون (۲۰درصد) تقسیم شدند. داده‌های بخش آموزش مدل را می‌سازند و داده‌های بخش آزمون مدل ایجاد شده را مورد ارزیابی قرار می دهند. جهت ارزیابی مدل‌ها می‌توان از شاخص‌های حساسیت، ویژگی، دقت، ارزش اخباری مثبت و ارزش اخباری منفی استفاده کرد. جهت محاسبه این شاخص‌ها از ماتریس تداخلی ایجاد شده در محیط نرم افزار استفاده شد. در جدول۴ شاخص‌های مقایسه دقت الگوریتم شبکه شبکه عصبی و الگوریتم‌های درخت تصمیم ارائه شده است.

[2]Gini Index

Archive of SID
www.SID.ir

جدول۴: مقایسه دقت الگوریتم های درخت تصمیم و شبکه عصبی

معیارها	مدل				
	شبکه عصبی	درخت تصمیم			
		C5	C&RT	QUEST	CHAID
حساسیت = مثبت های واقعی / مثبت های واقعی + منفی های کاذب	٪۹۲	٪۹۶	٪۸۲/۶۷	۵۶	٪۸۳
ویژگی = منفی های واقعی / منفی های واقعی + مثبت های کاذب	٪۸۹/۵	٪۹۱/۵	٪۷۶/۵	٪۸۴/۵	٪۸۷
ارزش اخباری مثبت = مثبت های واقعی / مثبت های واقعی + مثبت های کاذب	٪۸۶/۲	٪۸۹/۴	٪۷۲/۵	٪۷۳	٪۶۲/۵
ارزش اخباری منفی = منفی های واقعی / منفی های واقعی + منفی های کاذب	٪۹۳/۲	٪۹۶/۸	٪۸۵/۵	٪۷۱/۹	٪۵۸/۳۳
دقت = منفی های واقعی +مثبت های واقعی / تعداد کل داده ها	٪۹۰/۵۷	٪۹۳/۴	٪۷۹/۱	٪۷۲/۲۸	٪۸۵/۷

ه) توسعه

ساخت مدل معمولا پایان پروژه نیست حتی اگر هدف مدل افزایش دانش است، دانش بدست آمده نیاز به سازماندهی دارد و نمایش به طریقی که کاربر بتواند از آن استفاده کند. دانش کشف شده باید سازماندهی شده و به شکل قابل ارائه برای دیگران درآید. ما در این مرحله بعد از ایجاد گزارشات لازم سعی کردیم توضیح دهیم که بر اساس مدلهای ایجاد شده تاثیرگذارترین فاکتورها در ابتلا فرد به آنفارکتوس قلبی کدامند. توجه به اینکه بیماریهای قلبی عروقی از جمله شایع ترین بیماری‌ها و علل مرگ محسوب می شوند، چنانچه بتوانیم یک مجموعه پرخطر را شناسایی و برنامه های غربالگری را برای آن اجرا کنیم کارایی برنامه بیشتر خواهد شد. علاوه بر این افراد یک جمعیت پرخطر ممکن است بیشتر راغب به شرکت در برنامه های غربالگری باشند به خصوص هنگامی که نتیجه تست (مدل ارائه شده) بر روی آنها مثبت باشد احتمال دارد بیشتر به توصیه های پزشکی گوش فرا دهند.

الگوریتم‌های مورد استفاده دراین مطالعه، الگوریتم C5 دارای بالاترین میزان دقت (۹۳/۴درصد) بود.در مرحله ارزیابی نظرات متخصصان حوزه نیز در مورد قوانین ایجاد شده اعمال می گردد. به این ترتیب که قوانین بدست آمده به متخصص مورد نظر ارائه شده و قوانینی که از نظر بالینی معتبر باشند به عنوان قوانین نهایی ارائه گردیدند. برخی از مهم‌ترین موارد این قوانین در زیر ارائه شده است. بنابراین طبق نظر متخصص قلب و عروق می‌توان گفت که ریسک فاکتورهای سن بالا، مصرف سیگار، فشارخون بالا، سطوح بالای کلسترول بیشترین تاثیر را در ابتلا به آنفارکتوس قلبی دارا هستند و این در حالی است که بر اساس مقایسه‌های انجام شده بر اساس اولویت بندی متغیرها توسط الگوریتم‌های مورد بررسی نیز این متغیرها جزو فاکتورهای اول قرار گرفته اند، که نشان از اهمیت این متغیرها در ابتلا به آنفارکتوس قلبی دارد.

برخی از مهم‌ترین قوانینی که از نظر پزشک متخصص قلب و عروق در این بررسی مورد تایید قرار گرفتند عبارتند از:

i. در ۷۱/۸۵ درصد از بیماران سابقه فشار خون و ابتلا به آنفارکتوس قلبی با هم مشاهده شدند.

ii. در ۵۵/۶۵ درصد از بیماران سابقه چربی خون و ابتلا به آنفارکتوس قلبی با هم مشاهده شدند.

iii. در ۴۵/۲۱ درصد از بیماران مصرف سیگار و ابتلا به آنفارکتوس قلبی با هم مشاهده شدند.

iv. در ۴۱/۶۴ درصد از بیماران سابقه دیابت و ابتلا به آنفارکتوس قلبی با هم مشاهده شدند.

v. در ۳۷/۱۲ درصد از بیماران گروه خونی A+ و ابتلا به آنفارکتوس قلبی با هم مشاهده شدند.

بحث و نتیجه گیری

در این تحقیق با استفاده از الگوریتم شبکه عصبی و درخت تصمیم به ارائه مدل و استخراج قوانین آن در راستای پیشگویی احتمال ابتلا به آنفارکتوس قلبی

Archive of SID

www.SID.ir

جدول ۵: مقایسه نتایج مطالعات انجام شده در حوزه داده کاوی در بیماری قلبی

نویسندگان و سال ارائه تحقیق	الگوریتم های مورد استفاده	نوع بیماری	دقت(نوع) مدل نهایی	یافته ها	متغیرهای پیشگویی کننده
Christine(1998)	رگرسیون لجستیک ،درخت طبقه بندی	آنفارکتوس قلبی	٪۸۱ (درخت طبقه بندی)	عملکرد بهتر درخت تصمیم در پیشگویی ابتلا به آنفارکتوس قلبی	سن،سابقه خانوادگی بیماری قلبی، مصرف سیگار، درد در ناحیه قفسه سینه، فشار خون بالا، دیابت، تعریق شبانه، استفراغ، جنسیت و ...
Biglarian(2004)	شبکه عصبی مصنوعی، رگرسیون لجستیک	پیوند عروق کرونری قلب	٪۹۹/۳۳ (شبکه عصبی مصنوعی)	عملکرد بهتر شبکه عصبی مصنوعی در پیش بینی کننده های مهم مرگ و میر درون بیمارستانی پس از جراحی قلب باز	سن، شاخص توده بدنی، کلسترول، تری گلیسرید، فشارخون، مصرف سیگار، دیابت، چربی خون، سابقه بیماری قلبی و...
Mehrabi(2007)	شبکه عصبی پرسپترون چند لایه، شبکه عصبی با تابع مبتنی بر شعاع	نارسایی احتقانی قلب و بیماری مزمن انسدادی ریه	٪۹۳ (شبکه عصبی با تابع مبتنی بر شعاع)	توان بالای مدل شبکه عصبی با تابع مبتنی بر شعاع در افتراق دو بیماری CHF و COPD	سن، جنس، تپش قلب، درد ناحیه قفسه سینه، تنگی نفس، ارتوپنه، ورم، خستگی، سیانوز،سرفه،خلط خونی،کاهش وزن و...
Kajabadi	درخت تصمیم	بیماری عروق کرونر	محاسبه نشده	عوامل تاثیر گذار عمده بر بروز بیماری قلبی مشخص شده اند.	چربی، فاکتورهای خونی، فاکتورهای چاقی، متغیرهای قندی، متغیرهای عمومی(سن، جنس، استعمال دخانیات و..)، آپولیپوپروتئین ها، فاکتور التهابی CRP
Karaolis(2010)	درخت تصمیم C4.5	آنفارکتوس قلبی،پیوند عروق کرونر قلبی	٪۶۶ (درخت تصمیم C4.5)	عوامل تاثیر گذار عمده بر بروز آنفارکتوس قلبی مشخص شده اند.	جنسیت، سن، فشارخون بالا، چربی خون بالا، مصرف سیگار، سطح کلسترول، دیابت و...
Jyoti (2011)	شبکه بیز درخت تصمیم ،شبکه عصبی مصنوعی	بیماری قلبی	٪۸۹ (درخت تصمیم)	ایجاد قوانینی جهت یافتن ارتباط بین متغیرها	جنسیت، سن، درد قفسه سینه، فشارخون بالا، قند خون ناشتا، سطح کلسترول، مصرف سیگار، شاخص توده بدنی و...
Mohammadpour (2011)	شبکه عصبی مصنوعی	بیماری عروق کرونری قلب	٪۹۶ (شبکه عصبی مصنوعی)	طبقه بندی صحیح بیماران نیازمند آنژیوگرافی و دارو درمانی	سن، شاخص توده بدنی،کراتینین، کلسترول تام، تری گلیسرید، سابقه مصرف سیگار ،سابقه فشارخون، سابقه دیابت، سابقه بیماری قلبی،ن تیجه تست ورزش و ...

پرداختیم. بهترین نتایج از الگوریتم درخت تصمیم C5 بدست آمد که دقت آن ۹۳/۴ درصد بود. بیشترین فاکتورهای تاثیرگذار در ابتلا افراد سن بالا، سابقه فشار خون بالا و چربی خون بالا و مصرف سیگار بودند. با استفاده از قوانین بدست آمده برای یک فرد جدید با داشتن متغیرهای مشخص،می توان تعیین کرد که احتمال ابتلا وی به آنفارکتوس قلبی چقدر خواهد بود. در جدول ۵ به مقایسه نتایج پژوهش‌های مشابه با پژوهش حاضر می پردازیم.

مطابق مطالعات گذشته، عملکرد مدلهای طبقه بندی-کننده ممکن است بر روی پایگاه‌های داده مختلف نتایج متفاوتی داشته باشد. برای مثال Karaolis و همکارانش در مطالعه‌ای تحت عنوان "پیش‌گویی ابتلا به بیماری عروق کرونر با استفاده از درخت تصمیم‌گیری" از الگوریتم درخت تصمیم برای ارزیابی ریسک فاکتورهای بیماری عروق کرونر استفاده کردند. آنها ریسک فاکتورهای ابتلا را به دو دسته کلی، ریسک فاکتورهای قبل از رویداد و ریسک فاکتورهای بعد از رویداد تقسیم بندی کرده است (۲۵). یافته های آن مشابه با قوانین استخراج شده از الگوریتم درخت تصمیم در مطالعه حاضر می باشد.

در بررسی انجام شده توسط Jyoti جهت پیش بینی ابتلا به بیماری قلبی مدل ارائه شده توسط درخت تصمیم دارای بالاترین میزان دقت (۸۹درصد) بوده است (۲۶). این در حالی است که درخت تصمیم ارائه شده در این مطالعه از دقت بالاتری برخوردار بوده است. تفاوت مشاهده شده را می‌توان از تعداد بیشتر متغیرهای مورد

Archive of SID
www.SID.ir

بررسی در مطالعه اشاره شده دانست. در بررسی انجام شده توسط Kurt عملکرد رگرسیون لجستیک، درخت تصمیم CART و شبکه عصبی در پیش‌گویی بیماری عروق انجام شده و در این مطالعه مدل شبکه عصبی به عنوان بهترین مدل معرفی شده است (۲۷).

در مطالعه محمد پور به کاربرد شبکه عصبی در ارزیابی بیماری عروق کرونر قلب پرداخته و حساسیت مدل بدست آمده ۹۶٪ بدست آمد، که نشان از توان بالای این مدل در تشخیص سریعتر بیمارانی است که نیازمند اقدامات تشخیصی و درمانی هستند. حساسیت بالای مدل ارائه شده در این مطالعه را می‌توان به کارگیری متغیرهای مفیدی چون نتیجه تست ورزش و نتیجه اکو و همچنین تعیین تعداد نرونهای کمتر در لایه میانی شبکه عصبی بیان کرد (۲۸).

در پژوهش Christine به مقایسه عملکرد رگرسیون لجستیک و چند الگوریتم از درخت تصمیم در تعیین ابتلا به آنفارکتوس قلبی پرداخته شده است و مدل درخت تصمیم با حساسیت ۸۱٪ به عنوان مدل مناسبی جهت پیشگویی معرفی شده است (۲۹). در مطالعه‌ای که به مقایسه عملکرد شبکه‌های عصبی پرسپترون چند لایه و با تابع مبتنی بر شعاع در سیستم‌های تصمیم یار بالینی برای تفکیک در بیماری نارسایی قلبی و انسداد مزمن ریه پرداخته شده است، مدل شبکه عصبی با حساسیت ۹۳٪ به عنوان مدل برتر جهت ایجاد تمایز در تشخیص بیماران مبتلا به نارسایی قلبی و انسداد مزمن ریه معرفی شده است (۳۰). در مقاله‌ای که با عنوان کاربرد شبکه عصبی مصنوعی در تعیین پیش بینی کننده‌های مهم مرگ و میر درون بیمارستانی پس از جراحی قلب باز و مقایسه آن با رگرسیون لجستیک در بیمارستان شریعتی تهران انجام شد، شبکه عصبی مصنوعی با دقت ۹۹/۳۳٪ عملکرد بهتری داشت (۳۱).

در مطالعه‌ای با عنوان داده‌کاوی پزشکی، راهکاری جهت کشف روابط بین عوامل خطرزای قلبی به ارائه مدلی برای برقراری ارتباط بین عوامل خطرزای قلبی پرداخته است، که از بین متغیرهای مورد استفاده در این مطالعه، فاکتورهایی مانند LDL، سطح کلسترول، سن شاخص توده بدن، APOB، تری گلیسرید، APOBA، APOB/APOA و استعمال دخانیات بوده است (۳۲). که تاثیر این متغیرها در بروز بیماری قلبی در مطالعه حاضر نیز تایید شده است.

تکنیک‌های داده‌کاوی می تواند در طراحی مدل‌های مناسب جهت پیش‌گویی امکان ابتلای افراد به بیماریهای خاص استفاده شوند و در این مطالعه مدل C5 دارای بالاترین میزان دقت بوده و می تواند در برنامه‌های غربالگری جهت شناسایی افراد در معرض خطر استفاده شود. پیشنهاد می شود که این مدل با مجموعه داده‌های بیشتر و در بازه زمانی طولانی‌تر اجرا شده و پس از رسیدن به سطح دقت مطلوب در برنامه‌های غربالگری مورد استفاده قرار گیرد. پیشنهاد می‌شود در مطالعاتی بعدی دقت این مدل در داده‌های سایر مراکز درمانی و یا سایر مقاطع زمانی مورد ارزیابی قرار گیرد و پس از ایجاد تغییرات ضروری و رسیدن به سطح مطلوب دقت، در انتخاب مدلی جامع و مناسب جهت پیشگویی احتمال ابتلا افراد به آنفارکتوس قلبی اقدام شود.

سپاسگزاری

این مقاله برگرفته از پایان نامه کارشناسی ارشد اینجانب می باشد، بدین وسیله از پرسنل محترم واحد مدارک پزشکی بیمارستان شهید رجایی تهران کمال تشکر و قدردانی را دارم.

منابع

1. Sezavar SH, Valizade M, Moradi M, Rahbar M.H. [review of early myocardial infarction and its risk factor in patients hospitalized in Rasool Akram hospital in Tehran]. Hormozgan Medical Journal 2010; 4(2): 156-163. [Persian]

2. Fauci A, Braunwald E, Kasper D, Hauser S, Longo D. Harrison's Principles of Internal Medicine, 17th Edition: New York; 2008:71-73.

3. Mohammadi F, Taherian A, Hosseini MA Rahgozar M. [Effect of Home-Based Cardiac Rehabilitation on Quality of Life in the Patient with Myocardial Infarction]. rehabilitation journal 2006 ; 7(7): 11-19.[Persian]

4. Gavish D, Leibovitz E, Elly I, Shargorodsky M, Zimlichman R.[Follow-up in a lipid clinic improves the management of risk factors in

Archive of SID

www.SID.ir

Cardiovascular disease patients]. The Israel Medical Association Journal 2002; 4(9): 694-697.

5. Gavish D, Leibovitz E, Elly I, Shargorodsky M, Zimlichman R.[Follow-up in a lipid clinic improves the management of risk factors in cardiovascular disease patients]. The Israel Medical Association Journal 2002; 4(9): 694-697.

6. Smith DL, Fehling PC, Frisch A, Haller JM, Winke M, Dailey aMW. [The Prevalence of Cardiovascular Disease Risk Factors and Obesity in Firefighters]. Journal of Obesity 2012:2-9.

7. Boo S, Erika Sivarajan Froelicher. [Cardiovascular Risk Factors and 10-year Risk for Coronary Heart Disease in Korean Women]. Asian Nursing Research 2012; 6(3): 1-8.

8. Soni J, Ansari U, Sharma D, Soni S.[Predictive Data Mining for Medical Diagnoses: An Overview of Heart Disease Prediction].International Journal of Computer Applications 2011; 17(8): 85-93.

9. Subbalakshmi G, Road Y. [Decision Support in Heart Disease Prediction System using Naive Bayes]. Indian Journal of Computer Science and Engineering 2011; 2(2): 183-195.

10. Fayyad M, Piatetsky G, Uthurusamy R, Smyth P. [Advances in Knowledge Discovery and Data Mining]. American Association of artificial intelligence 1996; 17(3): 37-54.

11. Lavrac N. [Selected techniques for data mining in medicine]. Artificial Intelligence in Medicine 1999; 16(1): 3-23.

12. Huanga M, Chenb M, Leea S. [Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis]. Expert Systems with Applications 2007; 32(3): 856–867.

13. Alizadeh S, Ghazanfari M, Teimorpour B .Data Mining and Knowledge Discovery, Publication of Iran University of Science and Technology: 2nd ed 2011: 63-65.

14. Collins GS, Mallett S, Omar O. [Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting]. BMC Medicine 2011; 5(3): 146-159.

15. Ashwinkumar M, Anandakumar R. [Ethical and Legal Issues for Medical Data Mining].international Journal of Computer Applications 2010.

16. Cooper G, Aliferis F, Ambrosinoa R, Aronisb B, Buchananb RC, Michael J. Fined CG, et al.[An evaluation of machine-learning methods for predicting pneumonia mortality]. Artificial Intelligence in Medicine 1996; 9(4): 107-138.

17. Olivas R. [DecisionTrees,A Primer for Decision-making Professionals] 2007; 9(4): 4-6. Available from:www.lumenaut.com/download/decision_tree_primer.

18. Colombet I, Ruelland A, Chatellier G, Gueyffier F, Degoulet P, Christine M. [Models to predict cardiovascular risk: comparison of CART, Multilayer perception and logistic regression]. Proc AMIA Symp 2000:156-160.

19. Tsien C, Fraster H. [using classification tree and logistic regression in myocardial infarction]. Proceeding of the 9th world congress on medical informatics; 1998.

20. Abdelghani B, Erhan G. [Predicting Breast Cancer Survivability Using Data Mining Techniques]. Washington University 2005; 2(2): 53-62.

21. Stark K, Pfeiffer D. [The application of non-parametric techniques to solve classification problems in complex data sets in veterinary epidemiology-An example]. Intelligent Data Analysis 2006; 3(1): 23-25.

22. Ordonez C. [Comparing Association rules and Decision Trees for Disease Prediction].intelligent data analysis 2011; 15(2): 3-9.

23. Jianxin C, Yanwei X, Guangcheng X, Jianqiang Y, Dongbin Z. [A Comparison of Four Data Mining Models: Bayes, Neural Network, SVM and decision Trees in Identifying Syndromes in Coronary Heart Disease]. Springer- Verlag Berlin Heidelberg 2007:1274–1279.

24. IBM SPSS Modeler CRISP-DM Guide. Available from:www.spss.ch/upload/1107356429_CrispDM1

Archive of SID
www.SID.ir

25. Margaret R, Kraft C, Desouza A. [Data Mining in Healthcare Information Systems. Case Study of a Veterans' Administration Spinal Cord Injury Population].Proceedings of the 36th Hawaii International Conference on System Sciences 2003.

26. IBM SPSS Modeler 15 Applications guide. Available from: ftp://ftp.software.ibm.com/software/analytics/spss

27. Karaolis M, Moutiris J, Papaconstantinou L, Pattichis C S. [Association Rule Analysis for the Assessment of the Risk of Coronary Heart Events]. Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2009.

28. Jyoti S, Ujma A, Dipesh S,Sunita S. [Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction]. International Journal of Computer Applications 2011; 17(8): 35-43.

29. Kurt I, Ture M, Kurum T. [Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease]. Expert Systems with Applications 2008; 34(3): 366–374.

30. Mohammadpour A, Esmaeili M, Ghaemian A, Esmaeili J. [Application of Artificial Neural Network for Assessing Coronary Artery Disease]. Journal Mazand Univ Med Sci 2012; 22(86): 9-17. [Persian]

31. Mehrabi S, Maghsodlo M, Arab Ali Baik H, Normand R, Nowzari Y. [Using Multilayer Perceptron neural network and Radial Basis Function in Clinical Diagnostic Decision Support System for differentiated between Congestive Heart Failure and Chronic Obstructive Pulmonary Disorder].Medical physics journal 2007; 12(3). [Persian]

32. Biglarian A, Babaee R, Azmie R. [Application of Artificial Neural Network Model in Determining Important Predictors of In Hospital Mortality After Coronary Artery Bypass Graft Surgery, and it's Comparison with Logistic Regression Model]. Modarres J Med Sci 2004; 7(1): 23-30. [Persian]

33. Kajabadi A, Saraee M, Asgari S. [Medical Data Mining: An Approach to Discovery Relationships among Cardiovascular Risk Factors]. Available from: www. civilica. Com/Paper- IDMC03-IDMC03_ 065.

Archive of SID

www.SID.ir