A Simple Risk Model of the Incidence of Atrial Fibrillation and Coronary Artery Disease Using a Data Mining Algorithm: Risk Factor Prediction

Zeinali A[1], Rastegari A.A[2], Monajemi R[3], Khodadadi H[4]

Abstract

Purpose: Cardiovascular disease is one of the most important causes of death worldwide. Using data mining methods to create predictive models to identify people at risk to prevent complications from cardiovascular diseases will be very effective. The aim of this research is to predict the probability of infection in people with coronary heart disease and atrial fibrillation using support vector machine, neural network and decision tree algorithms based on factors affecting the disease.

Methods: This analytical research includes 300 records. The information required for this study was collected in 1400 using the records of patients admitted to Chamran and Khurshid hospitals in Isfahan. For data analysis, the information includes laboratory, demographic and family history sections using the CRISP method, the Cross Industry Standard Process for Data Mining (CRISP). Decision trees, neural networks and support vector machines are also used in the modeling section.

Results: The sensitivity and specificity in the neural network data mining algorithm are 87.5 and 71.11 respectively, 92.85 and 80 in the decision tree algorithm and 88.88 and 75 in the support vector machine. Therefore, the decision tree algorithm has a better performance for predicting the probability of heart and coronary artery diseases and atrial fibrillation. Also it was found that stress, high BMI, high blood pressure and type of job had the greatest effect on the occurrence of heart and coronary artery diseases and cardiac arrhythmias.

Conclusion: In the current study, the decision tree has the highest performance, so it can be used to determine the probability of coronary heart and vascular problems and atrial fibrillation.

Keywords: Heart disease, Coronary artery disease, Cardiac arrhythmia; Neural network, Support vector machine, Decision tree

Received: 2022.11.14 Accepted:2023.04.04

پیشگویی مدل ریسک فاکتورهای فیبریلاسیون دهلیزی و بیماری عروق کرونر با استفاده از الگوریتم های داده کاوی

آرزو زینلی[۱]، علی اصغر رستگاری[۲]، رامش منجمی[۳]، حامد خدادادی[۴]

هدف: مشکلات قلبی عروقی یکی از شایع ترین علل مرگ و میر در سراسر جهان است. استفاده از روش های داده کاوی برای ایجاد مدل های پیش بینی برای شناسایی افراد در معرض خطر برای جلوگیری از عوارض ناشی از بیماری های قلبی عروقی بسیار موثر خواهد بود. انگیزه اصلی این تحقیق پیش بینی احتمال عفونت در افراد مبتلا به بیماری عروق کرونر قلب و فیبریلاسیون دهلیزی با استفاده از الگوریتم های ماشین بردار پشتیبان، شبکه عصبی و درخت تصمیم بر اساس عوامل مؤثر بر بیماری است.

روش بررسی: این پژوهش از نوع تحلیلی است و پایگاه داده آن شامل ۳۰۰ رکورد می باشد. اطلاعات مورد نیاز این مطالعه در سال ۱۴۰۰ با استفاده از پرونده بیماران بستری در بیمارستان های چمران و خورشید اصفهان جمع آوری شد. برای تجزیه و تحلیل آن ها، اطلاعات شامل بخش های آزمایشگاهی، دموگرافیک و سابقه خانوادگی است که از روش کریسپ، فرآیندهای استاندارد صنعت متقابل برای داده کاوی (Cross Industry Standard Process for Data Mining; CRISP) استفاده می شود. همچنین در بخش مدل سازی از درخت های تصمیم، شبکه‌های عصبی و ماشین های بردار پشتیبان استفاده می شود.

یافته ها: بر طبق نتایج بدست آمده، حساسیت و ویژگی در الگوریتم داده کاوی شبکه عصبی به ترتیب (۷۱/۱۱،۸۷/۵)، در الگوریتم درخت تصمیم (۹۲/۸۵، ۸۰) و در ماشین بردار پشتیبان (۸۸/۸۸، ۷۵) بوده اند. لذا الگوریتم درخت تصمیم دارای عملکرد بهتری برای پیش بینی احتمال بیماری های قلبی و عروق کرونر و فیبریلاسیون دهلیزی است. همچنین با توجه به مطالعات انجام شده مشخص شد که استرس، اضافه وزن، فشار خون بالا و نوع شغل بیشترین تأثیر را در بروز بیماری های قلبی و عروق کرونر و آریتمی های قلبی داشتند.

نتیجه گیری: در مطالعه حاضر درخت تصمیم دارای بالاترین عملکرد است لذا می توان از آن برای تعیین احتمال مشکلات کرونری قلب و عروق و فیبریلاسیون دهلیزی استفاده کرد.

کلمات کلیدی: بیماری قلبی، بیماری عروق کرونر، آریتمی قلبی، شبکه عصبی، ماشین بردار پشتیبان، درخت تصمیم

نویسنده مسئول: علی اصغر رستگاری، aarastegari@gmail.com، ORCID: 0000-0003-3604-4243

آدرس: اصفهان، دانشگاه آزاد اسلامی واحد فلاورجان، گروه زیست شناسی

۱- دانشجوی دکتری رشته فیزیولوژی جانوری، گروه زیست شناسی، دانشگاه آزاد اسلامی واحد فلاورجان، اصفهان، ایران

۲-دانشیار گروه بیوشیمی سلولی مولکولی، دانشگاه آزاد اسلامی واحد فلاورجان، اصفهان، ایران

۳- استادیار گروه زیست شناسی، دانشگاه آزاد اسلامی واحد فلاورجان، اصفهان، ایران

۴- استادیار گروه مهندسی برق، دانشگاه آزاد اسلامی واحد خمینی شهر، اصفهان، ایران

مقدمه

بیماری های قلبی عروقی شایع ترین علت مرگ در بیشتر کشورهای جهان از جمله ایران و مهم ترین عامل از کار افتادگی است (۲، ۱). با وجود پیشرفت های سریع تشخیصی و درمانی هنوز یک سوم بیمارانی که دچار سکته-ی قلبی می شوند فوت می کنند و دو سوم آن ها زنده می مانند ولی هرگز بهبود کامل نمی یابند و به زندگی عادی برنمی گردند (۳). این بیماری ها، هزینه‌ی هنگفتی را بر نظام های بهداشتی درمانی کشور تحمیل می کنند. میزان بروز آنژین صدری، سکته قلبی و مرگ ناگهانی، از نظر اپیدمیولوژیک مورد بررسی قرار گرفته است که بر اساس عوامل خطر، سن، جنس و نژاد در سطوح فردی و بر اساس کشورها، مناطق جغرافیایی و طبقات اجتماعی در سطوح جمعیتی و در طول زمان متفاوت می باشند (۴). بیماری های عروق کرونر باعث مرگ و میر و ناتوانی در جمعیت ایران می گردند و تقریبا عامل ۵۰ درصد تمام مرگ ها در سال محسوب می شوند. در حالی که میزان مرگ و میر ناشی از بیماری های عروق کرونر قلبی بر حسب سن در کشور های توسعه یافته در حال کاهش است. شواهدی وجود دارد که نشان می دهد این بیماری در ایران رو به افزایش می باشد، به طوری که میزان آن بین ۲۰ تا ۴۰ درصد افزایش یافته است. به نظر می رسد تغییرات ایجاد شده درروند زندگی مانند مصرف زیاد غذای فرآوری شده و غنی از چربی های اشباع شده، سطح پایین فعالیت فیزیکی همراه با افزایش شیوع چاقی و دیابت نوع دو و افزایش فزاینده ی تنش و اضطراب منجر به افزایش پیشرونده در شیوع عوامل بیماری های قلبی عروقی شده است. زندگی انسان ها وابسته به عملکرد مناسب قلب است و اگر عملکرد قلب به شکل مناسب صورت نگیرد روی سایر قسمت های بدن از قبیل ذهن و کلیه و غیره اثر خواهد گذاشت. بیماری های قلبی که معمولا از آن به عنوان بیماری های شریان عروق کرونر نام برده می شود واژه ای با دامنه وسیع است که به هر نوع شرایطی که قلب تحت تاثیر قرار می دهد گفته می شود (۵). بیماری های شریان عروق کرونر بیماری مزمنی است که طی آن شریان کرونر به تدریج سفت و باریک می شود (۶). درحالی که بیشتر مردم مبتلا به بیماری قلبی نشانه هایی هم چون درد قفسه سینه و خستگی دارند ولی حدود پنجاه درصدشان تا زمان حمله قلبی هیچ نشانه ای ندارند (۷). آریتمی دهلیزی به صورت یک تاکی آریتمی فوق بطنی تعریف می گردد که با فعالیت دهلیزی غیر هماهنگ و متعاقب آن با نارسایی مکانیکی دهلیز مشخص می شد. فیبریلاسیون دهلیزی شایع ترین آریتمی قلبی (Cardiac Arrhythmia) است که میزان بروز آن رابطه مستقیمی با سن دارد. آمار ها نشان می دهد حدود یک سوم از بستری ها به علت اختلالات ریتم ناشی از این بیماری است که میزان ان طی ۲۰ سال اخیر ۶۶

درصد افزایش یافته است (۹، ۸). با توجه به شیوع بیماری های قلبی-عروقی در جهان، استفاده از روش های جدید در تحقیقات زیست پزشکی بسیار مورد توجه قرار گرفته است، زیرا با اینکه آنژیوگرافی دقیق ترین روش تشخیص بیماری-های عروق کرونر است، اما عوارض جانبی زیادی دارد و هزینه بر است، بنابراین با چنین عوارض جانبی، انگیزه بسیاری از محققان را برای استفاده از داده کاوی برای تشخیص برانگیخته است (۱۰).

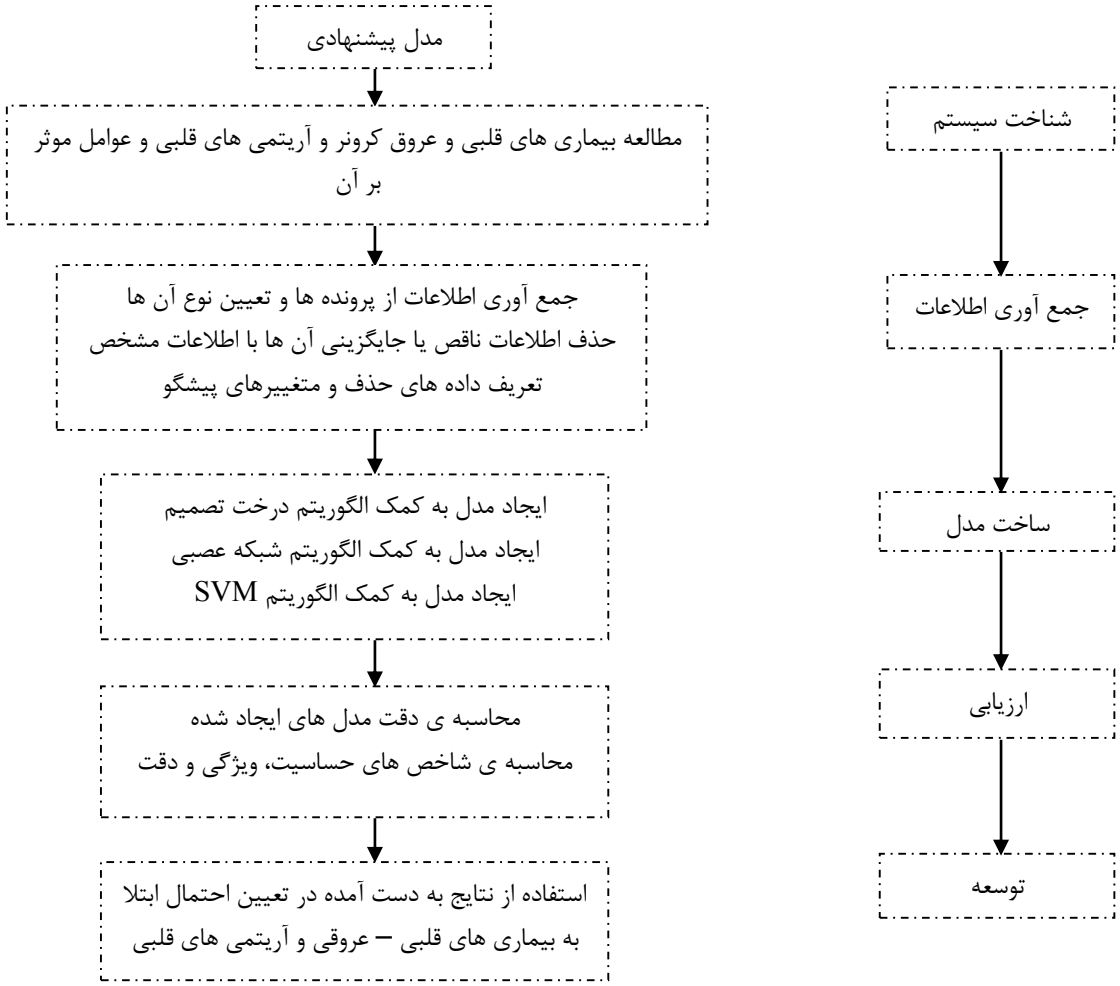
طبق نظر کارشناسان، دانش نوین داده کاوی از جمله دانش های در حال توسعه است که در سال های اخیر در تمامی عرصه ها جایگاه خود را تثبیت کرده است. به گونه ای که رشد ان در مقایسه با سایر دانش های برتر فزاینده است. استخراج اطلاعات از میان حجم انبوه داده های مرتبط با سوابق بیماری و پرونده‌های پزشکی افراد با استفاده از فرآیند داده کاوی می تواند منجر به شناسایی قوانین حاکم بر ایجاد بیماری ها، تشخیص، پیش بینی و درمان آن شود (۱۱). به طور کلی روش ها و الگوریتم هایی برای پیش بینی بیماری های قلبی مورد ارزیابی قرار گرفته اند که این الگوریتم ها بر روی مجموعه داده های متفاوت و تحت شرایط متنوعی آزمایش شده اند (۱۲).

در برخی از مطالعات درخت تصمیم و شبکه عصبی پرسپترون چندلایه، به عنوان بهترین الگوریتم های یادگیری ماشین در تشخیص بیماری های قلبی معرفی گردیدند(۱۳). در برخی دیگر الگوریتم ماشین بردار پشتیبان (Support Vector Machine; SVM) نسبت به مدل شبکه عصبی دقت و عملکرد بهتری در پیش بینی بیماری عروق کرونر قلب نشان داد و دارای حساسیت و صحت بالاتری بود (۱۴). در برخی نیز، از میان الگوریتم های داده کاوی، مدل شبکه عصبی با ساختار پرسپترون چند لایه بواسطه کلاس بندی بهتر مجموعه مشاهدات آزمون، بالاترین دقت را به خود اختصاص داده است. همچنین متغیرهای تعداد عروق بزرگ، کاهش استرس، نقص، درد قفسه سینه، اوج استرس، ضربان قلب، آنژین،جنسیت سن، ایستایی نوار قلب، فشار خون، قندخون و کلسترول سرم به ترتیب بیشترین اهمیت را از لحاظ مدل "شبکه عصبی با ساختار پرسپترون چند لایه" در پیش بینی متغیر پاسخ بیماری قلبی دارند (۱۶، ۱۵). به طور کلی در بسیاری از مطالعات به اثبات رسیده است که از الگوریتم های یادگیری ماشینی و تجزیه و تحلیل الگوریتم های داده کاوی مختلف می توان برای پیش بینی بیماری های قلبی استفاده نمود (۱۷-۲۳). با توجه به اهمیت الگوریتم های داده کاوی در پیش بینی بیماری های قلبی، مطالعه حاضر می کوشد تا استفاده از الگوریتم های داده کاوی را به طور خاص در پیشگویی مدل ریسک فاکتورهای فیبریلاسیون دهلیزی،آریتمی های قلبی و بیماری عروق کرونر مورد بررسی قرار دهد. لذا هدف اصلی ما در این بحث استفاده از الگوریتم های داده کاوی برای پیش بینی حملات قلبی است. با توجه به اهمیت الگوریتم های داده کاوی در پیش بینی بیماری های قلبی، مطالعه حاضر می کوشد تا استفاده از الگوریتم های داده- کاوی را به طور خاص در پیشگویی مدل ریسک فاکتورهای فیبریلاسیون دهلیزی،آریتمی های قلبی و بیماری عروق کرونر مورد بررسی قرار دهد. لذا هدف اصلی ما در این بحث استفاده از الگوریتم های داده کاوی برای پیش بینی حملات قلبی است.

روش بررسی

راه های مختلفی برای انجام الگوریتم های داده کاوی هست که یکی از آن ها روش پر بازده کریسپ(CRISP) می باشد (۲۴)، که این تحقیق بر پایه همین روش انجام شده است (نمودار ۱). این روش شامل شش مرحله شناخت، آماده سازی داده ها، مدل سازی، ارزیابی مدل، توسعه و پیاده سازی می باشد. این روش یک فر ایند چرخشی بوده و در برخی از مراحل امکان بازگشت به عقب وجود دارد (۲۴).

برای بررسی پیشینه تحقیق از مطالعات کتابخانه ای (رجوع به اسناد، مقالات، جستجو در پایگاه های علمی و...) استفاده گردید. در آغاز کار جهت استخراج داده ها، پرونده ۳۷۰ بیمار قلبی مراجعه کننده به بیمارستان های چمران و خورشید اصفهان در سال ۱۴۰۰ بررسی گردید، و تعداد ۷۰ رکورد از آن ها به علت عدم تکمیل عوامل مورد بررسی از پایگاه داده حذف گردید. روش جمع آوری داده ها به صورت سرشماری بود. از این تعداد پرونده مورد بررسی، ۱۳۸ نفر مبتلا به سکته قلبی و ۱۶۲ نفر مبتلا به نارسایی احتقانی بودند. جهت پیش بینی بیماری قلبی، متغیرهای کمی و کیفی موجود در این پرونده ها بررسی گردید و از آزمایشات و علائم مشترک موجود در پرونده ها ۳۲ متغیر کمی و ۷ متغیر کیفی موثر در بیماری استخراج شد. از



نمودار ۱ :گام های متدلوژی کریسپ

آنجایی که در بدو ورود بیمار به بیمارستان آزمایشاتی برای شناسایی بیماری احتمالی از بیماران گرفته می شود، و با استفاده از نظر خبرگان که مجموعه ای از پزشکان متخصص قلب و عروق و چند پرستار متخصص بخش آنژیو و بخش داخلی و جراحی قلب می باشند، متغیرهای اثر گذار انتخاب و متغیرهایی که تاثیر کمتری بر روی بیماری قلبی داشت، حذف گردیدند (ظرفیت کل جذب آهن در خون(Total Iron-Binding Caper; TIBC)، بررسی عملکرد غده تیروئید در دوران بارداری (TSH)، هورمون تیروئید تیروکسین (Thyroxin; T4)، هورمون تری یدوتیرونین (Triiodothyronine; T3)، (Anti-Thymocyte Globulin; ATG) و در پایان ۲۶ متغیر کمی و ۷ متغیر کیفی باقی و مورد بررسی قرار گرفتند (جداول ۱ و ۲).

فاکتورهای مشخص شده برای این مطالعه به دو قسمت متغیر هدف و متغیر پیشگو تقسیم بندی شدند. متغیر هدف ابتلا یا عدم ابتلا به بیماری و سایر متغیرها به عنوان متغیر پیشگو مورد استفاده قرار گرفتند. روش های داده کاوی متنوعی برای مدل سازی وجود دارد. در این مرحله با استفاده از تکنیک های داده کاوی به ارائه مدل پیش گویانه پرداخته شد. تجزیه و تحلیل داده ها و مدل سازی با استفاده از نرم افزار SPSS Clementine 0.12 انجام و در این مطالعه از روش های پیش گویانه در داده کاوی استفاده گردید. الگوریتم های مدل سازی مورد استفاده عبارتند از:

الگوریتم درخت تصمیم (Decision Tree; DT)؛ درخت تصمیم روشی محبوب و مرسوم در طبقه بندی است و امروزه در حوزه های مختلف به خصوص حوزه پزشکی از این الگوریتم به صورت گسترده استفاده می شود. ساختار درخت تصمیم در یادگیری ماشین، یک مدل پیش بینی کننده بوده و از پرکاربردترین روش های یادگیری ماشینی محسوب می‌شود که در آن، برگ ها نشان

دهند دسته بندی و شاخه‌ها و گره های میانی ویژگی های مختلف برای رسیدن به یک طبقه را نشان می دهند (۲۱).

شبکه عصبی پرسپترون چندلایه؛ یکی از محبوب ترین شبکه های عصبی پیش رو است. این شبکه ها دارای سه دسته لایه ورودی، مخفی و خروجی هستند. در شبکه های پرسپترون چندلایه، داده های ورودی به لایه ورودی وارد شده تا مقادیر ورودی به نرون های ورودی محاسبه شود. پس از آن برای اعمال پردازش، آن‌ها را به لایه های مخفی ارسال و در نهایت برای محاسبه مقدار خروجی شبکه، به نرون های لایه خروجی وارد می شوند. این شبکه بر مبنای الگوریتم پس انتشار خطا آموزش می بیند به این صورت که خروجی های محاسبه شده از شبکه با خروجی های واقعی مقایسه و مقدار خطا محاسبه و به عقب برگشت داده می شود تا خطاهای لایه های میانی هم محاسبه گردد. در انتها وزن های شبکه به روزرسانی و سعی می شود تابع هزینه خطا در گام های مختلف زمانی کمینه گردد (۱۹).

الگوریتم ماشین بردار پشتیبان(SVM)؛ دسته ای از الگوریتم- های یادگیری ماشینی با ناظر بوده که از آن ها برای طبقه بندی و رگرسیون بر پایه نظریه یادگیری آماری استفاده می شود. این الگوریتم هر داده را با توجه به طبقه شان به یک فضای جدید برده، به طوری که داده ها به صورت خطی (ابرصفحه) قابل دسته بندی باشند. سپس با یافتن خطوط پشتیبان (صفحات پشتیبان در فضای چندبعدی)، معادله خطی را که بتواند بیشترین فاصله بین دسته ها ایجاد کند، پیدا می کند (۱۷).

یافته ها

میزان فراوانی بیماری در رنج سنی ۶۶-۵۱ و در افراد با سابقه مصرف سیگار، سابقه دیابت، سابقه خانوادگی بیماری عروق کرونر و درد قفسه سینه بیشتر بود (جدول۱).

طبق نتایج از میان فاکتورهای مورد بررسی، زمان ترومبوپلاستین نسبی فعال شده(Activated Partial Thromboplastin Time; APTT)، پی تی، اوره، قند، کلسترول، نیتروژن اوره خون، کراتین، پتاسیم، هموگلوبین، کلسیم، فشارخون، آسپارتات آمینوترانسفراز، آلانین آمینوترانسفراز، تروپونین، کسرجهشی، کلسترول لیپوپروتئین، آنزیم های کبدی و اینترلوکین ها، بین دو گروه سالم و بیمار اختلافات معنی داری وجود داشت ($p\leq ۰/۰۵$). بین سایر فاکتورها اختلافات معنی داری وجود نداشت (جدول ۲).

ساخت مدل از الگوریتم های متنوعی مانند شبکه عصبی و درخت تصمیم و فاکتورهای ورودی متغیر هدف استفاده شده است. SVM یکی از روش های یادگیری با نظارت است که از آن برای طبقه بندی و رگرسیون استفاده می کنند (۲۵). در این الگوریتم ها، جنسیت، سن، سابقه‌ی مصرف سیگار، سابقه‌ی اعتیاد، سابقه‌ی فشار خون، سابقه‌ی چربی (کراتینین)Cr، BUN،(کلسیم) Ca، K(پتاسیم) خون، فاکتورهای قند و چربی، وزن، نوع شغل، میزان استرس، کبدی و فاکتورهای انعقادی به عنوان متغیرهای پیشگو تعیین شده و متغیر ابتلا یا عدم ابتلا به بیماری به عنوان فاکتور هدف مشخص شد و سپس اطلاعات به دست آمده از بیماران به دو بخش آموزش ۷۰ درصد و آزمون ۳۰ درصد تقسیم شدند. در مدل شبکه ی عصبی با چندین قسمت مواجه هستیم که عبارتند از لایه‌های ورودی، لایه های خروجی، لایه پنهان و گره (۲۷، ۲۶). هر گره، ورودی را دریافت کرده و آن را پردازش گرده و خروجی ایجاد می-کند و البته تشخیص اینکه آیا هر ورودی به گره خروجی خواهید رسید یا نه به وزن آن ورودی بستگی دارد (۲۰). در مورد درخت تصمیم باید گفت که ترکیب تعدادی استلزام منطقی (قانون اگر- یا ماشین بردار پشتیبان یک الگوریتم نظارت شده SVM آن گاه) است (۱۹). الگوریتم یادگیری ماشین است که هم برای مسائل طبقه‌بندی و هم مسائل رگرسیون قابل استفاده است (۲۸).

جهت ارزیابی مدل می باشد. بنابراین برای بررسی روایی مدل اطلاعات، به دو دسته آموزش ۷۵ درصد و آزمون ۳۰ درصد تقسیم شدند (۲۹). اطلاعات تست آزمودنی در ساخت مدل موثر هستند و اطلاعات بخش آزمون، مدل ساخته شده را مورد ارزیابی قرار می‌دهند. برای ارزیابی مدل ها می توان از ۴ شاخص دقت، حساسیت، شفافیت و صحت استفاده کرد. برای محاسبه ی این شاخص ها از ماتریس تداخلی یا اغتشاش استفاده می‌کنیم (جدول۳). در این جدول شاخص های دقت، حساسیت، شفافیت و صحت در الگوریتم‌های شبکه عصبی و درخت تصمیم مقایسه شده است.

نهایتا با استفاده از مدل های ارائه شده. موثرترین متغیرها در ابتلا به بیماری های قلبی-عروقی کرونرو فیبریلاسیون دهلیزی را می توان شناسایی کرد. با توجه به این که

جدول۱: توزیع فراوانی متغیرهای کیفی در افراد مورد مطالعه

متغیر		افراد سالم تعداد(درصد)	بیماران تعداد (درصد)
جنسیت	مرد	(۴۵/۴)۱۳۰	(۵۵/۶)۱۷۰
	زن	(۵۴/۶)۱۷۰	(۴۴/۴)۱۳۶
سن	۳۵-۵۰	(۵۰/۱)۱۵۴	(۲۰/۶۲)۷۱
	۵۱-۶۶	(۳۲/۵)۹۴	(۶۸/۳)۱۹۹
	۶۷-۸۰	(۱۷/۴)۵۲	(۱۱/۰۸)۳۶
سابقه خانوادگی	دارد	(۱۱/۲)۳۷	(۳۷/۹)۱۱۶
	ندارد	(۸۸/۸)۲۶۳	(۶۲/۱)۱۹۰
سابقه دیابت	دارد	(۰)۰	(۳۴/۰)۱۰۴
	ندارد	(۳۰۰)۱۰۰	(۶۶/۰)۲۰۲
سابقه مصرف سیگار	دارد	(۱۰)۳۰	(۴۰/۲)۱۲۳
	ندارد	(۹۰)۲۷۰	(۵۹/۸)۱۸۳
سابقه درد قفسه سینه	دارد	(۰)۰/۰۰	(۷۱/۴)۲۴۹
	ندارد	(۳۰۰)۱۰۰	(۶/۲۸)۵۷
استرس	دارد	(۰)۰/۰۰	(۵۳/۹)۱۶۵
	ندارد	(۳۰۰)۱۰۰	(۴۶/۱)۱۴۱

جدول۲: نتایج آزمون تی مستقل برای اساس متغیرهای مورد مطالعه

نام متغیر	افراد سالم میانگین±انحراف معیار	بیماران میانگین±انحراف معیار	p – مقدار*
زمان ترومبوپلاستین نسبی فعال شده*	۴/۳۲±۵۹۰/۴۹	۹/۴۵۷± ۳۲/۸۶	۰/۰۰۶
پی تی	۱/۱۵۶± ۱۲/۱۷	۴/۶۷۶± ۱۵/۵۰۳	۰/۰۰۱
اوره	۷/۳۲۸± ۱۵/۱۶	۱۶/۱۸۷± ۳۲/۷۴	۰/۰۷
قند ناشتا	۱۵/۶۱۵± ۹۰/۶۳	۵۹/۹۲۵± ۱۲۷/۷۸	۰/۰۰۰
قند	۱۸/۵۸۳± ۱۶۸/۹۵	۵۸/۲۲۸± ۱۵۵/۳۵۲	۰/۰۰۱
تری گلیسیرید	۶۵/۷۹۱± ۱۳۴/۸۷	۱۲۰/۲۹۱± ۱۸۸/۱۴۰	۰/۰۷۲
کلسترول	۲۲/۰۱۱± ۱۶۵/۹۰	۳۸/۱۵۴±۴۲۷/۸۳۳	۰/۰۳۲
لیپوپروتئین	۲۳/۳۱۰± ۷۰/۷۴	۲۵۷/۵۷± ۴۸/۷۵۱	۰/۰۰۱
نیتروژن اوره خون	۳/۷۱۵± ۱۳/۷۲۵	۵/۱۱۱± ۱۹/۴۰۵	۰/۰۳۲
کراتین	۰/۸۳۶± ۱/۲۸۹	۰/۷۸۵± ۱/۳۱۱	۰/۰۰۵
پتاسیم	۱/۰۲۷± ۴/۸۸۸	۰/۷۱۴± ۴/۰۴۴	۰/۰۴۲
سدیم	۳/۲۳۴± ۱۴۰/۰۷۵	۷/۹۱۴± ۱۳۴/۵۲۶	۰/۰۸۱
ای ن ار	۶/۱۹۱± ۱/۳۰۶	۱۴/۹۹۸± ۲/۴۸۶	۰/۱۲۰
هموگلوبین	۲/۱۳±۵۵۰/۱۵۸	۱/۸۸۳± ۱۳/۵۶۱	۰/۰۰۶
کلسیم	۱۳/۱۰۰±۵۸۰/۳۶۸	۲۵/۷۵۲± ۱۰۷/۲۲۸	۰/۰۰۳
فشارخون	۳۷/۴۳۸± ۴۵۶/۱۰۴	۵۸/۲۲۸± ۱۵۵/۳۵۲	۰/۰۰۱
آلکالین فسفاتاز	۲۷/۷۱۶± ۷۹/۶۴۵	۴۶/۵۶۵± ۲۶۲/۵۷	۰/۲۱۴
آسپارتات آمینوترانسفراز	۷/۲۳±۶۶۸/۱۷۸	۹/۶۸۴± ۲۸/۶۱۱	۰/۰۲۱
آلانین آمینوترانسفراز	۱۴/۶۷۹± ۲۹/۶۲۲	۱۱/۲۸۷± ۳۴/۶۱۴	۰/۰۲۳
تروپونین	۰/۰۵۷± ۰/۰۳۶	۱/۹۹۱± ۶/۹۴۲	۰/۰۰۲
کسر جهشی	۰/۰۴۳± ۰/۶۲۱	۰/۰۰۳± ۰/۳۹۵	۰/۰۰۵
کلسترول لیپوپروتئین	۵۳/۸۹۸± ۱۱۰/۷۶۹	۴۶۶/۶۰۲± ۱۱۹/۷۰۵	۰/۰۰۰
INT3	۲۱/۵۷۱± ۳۵/۴۱۹	-	۰/۰۰۷
INT6	۳/۹۳۵± ۴۱/۴۲۸	-	۰/۰۴۷
اینترلوکین ۳(IL3)	-	۰/۶۶۹± ۳۷/۲۷۰	۰/۰۰۰
اینترلوکین ۶(IL6)	-	۱/۱۳۵± ۴۳/۹۸۰	۰/۰۰۱

*سطح معناداری p<۰/۰۵، * Activated Partial Thromboplastin Time; APTT

جدول ۳: مقایسه شاخص های الگوریتم های درخت تصمیم، SVM و شبکه عصبی

معیارها	شبکه عصبی	SVM	درخت تصمیم
دقت	۸۱/۱۴	۸۲/۱۴	۸۲/۱۴
ویژگی	۷۱/۱۱	۷۵	۸۰
حساسیت	۸۷/۵	۸۸/۸۸	۹۲/۸۵
اختصاصی بودن	۷۵	۷۸/۲۶	۸۰/۲۸

ماشین بردار پشتیبانی (Support Vector Machines; SVM)

بیماری های قلبی-عروقی شایع ترین علت مرگ و میر در دنیا می باشند. اگر یک مجموعه در معرض خطر زیاد را تشخیص بدهیم و برنامه های غربالگری را برای آن اجرا کنیم به نتایج بیشتر و بهتر می رسیم. تکنیک های داده کاوی مورد استفاده در این مطالعه: درخت تصمیم، شبکه عصبی و ماشین بردار پشتیبان (SVM) بود که بالاترین درصد مربوط به درخت تصمیم می باشد. طبق گفته ی متخصصان قلب وعروق متغیرهای سن بالا، استرس، نوع شغل، سطح بالای چربی و قند و فشار خون بیشترین اثر را در ابتلا به بیماری قلبی عروق کرونر و آریتمی های قلبی دارد که در الگورتیم داده کاوی نیز این مطالب مورد تایید می باشد (۳۰).

بحث و نتیجه گیری

بیماری های قلبی عروقی شایع ترین علت مرگ در بیشتر کشورهای جهان از جمله ایران و مهم ترین عامل از کار افتادگی است. در این میان بکارگیری روش های داده کاوی برای ایجاد مدل های پیش بینی کننده جهت شناخت افراد در معرض خطر برای پیشگیری از افزایش عوارض ناشی از بیماری های قلبی-عروقی بسیار موثر خواهد بود. به گونه ای که از داده کاوی برای به دست آوردن روابط مفید بین عوامل خطر زا در بیماری های قلب و عروق استفاده می شود. لذا مطالعه پیشرو با هدف پیش بینی احتمال ابتلای افراد به بیماری های قلبی و عروق کرونر و فیبریلاسیون دهلیزی با استفاد از الگوریتم های: ماشین بردار پشتیبان (SVM)، شبکه عصبی و درخت تصمیم و بر اساس عوامل موثر بر ابتلا، انجام شد. بر طبق نتایج به دست آمده از سه الگورتیم داده کاوی شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان (SVM)؛ درخت تصمیم به واسطه مقادیر بالای شاخص های حساسیت، شفافیت، صحت و دقت، بالاترین صحت را داشته است و از میان فاکتورهای مورد بررسی، زمان ترومبوپلاستین نسبی فعال شده (APTT)، پی تی، اوره، قند، کلسترول، نیتروژن اوره خون، کراتین، پتاسیم، هموگلوبین، کلسیم، فشارخون، آسپارتات آمینوترانسفراز، آلانین آمینوترانسفراز، تروپونین، کسرجهشی، کلسترول لیپوپروتئین، آنزیم های کبدی و اینترلوکین ها، از عوامل بسیار موثر و پیش بین در بروز چنین مشکلاتی بوده است.

این یافته با نتایج مطالعات (۳۲، ۳۱، ۲۴، ۲۰، ۱۸، ۱۷) همراه است، آن ها نیز در مطالعات خود به اهمیت فاکتورهای مذکور در بروز بیماری های قلبی اشاره داشته-اند. به عنوان مثال اینترلوکین ها (Interleukin)، سیتوکین های (Cytokin) ساخته شده توسط انواع سلولهای سفید خون هستند که اغلب بر لنفوسیت (Lymphocytes) های دیگر موثر می باشند. این ترکیبات در سیستم ایمنی نقش مهمی دارند. تعداد آن ها بسیار زیاد است و با شماره مشخص می شوند مانند Interleukin 6. این پروتئین ها با تنوع زیادی که دارند عملکردهایی خارج از سیستم ایمنی هم دارند. اینترلوکین ۳ (Interleukin 3) یکی از اینترلوکین های مهم بدن است که از گلبول های سفید ترشح می شود و در پاسخ های التهابی و ایمنی نقش دارد. Interleukin3 از سلول های لنفوسیت تی کمک کننده، ائوزینوفیل (Eosinophil) ماست سل (Mast Cell) سلول کشنده طبیعی ترشح می شود و بر سلول های پیش ساز خونی و ماست سل موثر است. اصلی ترین کارکرد این سایتوکین تمایز و تولید سریع تر گویچه سرخ و گرانولوسیت (Granulocyte) و رها شدن هیستامین است.

جدول ۴: مقایسه نتایج مطالعات انجام شده

نویسندگان	تکنیک های داده کاوی	نوع شکایت بیمار	میزان صحت مدل	دستاوردها	فاکتورهای پیشگویی کننده
Shahid و Singh (۲)	Naïve bayas بهینه سازی شبکه های اجتماعی(SMO) * درخت تصمیم تقویت کنندگی وفقی*	بیماری عروق کرونر	۸۱/۰۵۹ درصد ۸۴/۸۶ درصد ۷۸/۲۴ درصد ۸۵/۸۹ درصد	دقت طبقه بندی به دست آمده ۸۳ درصد می باشد.	سن، وزن، جنس، دیابت، فشار خون، بیماری کلیوی، سیگار، قند خون، سدیم، پتاسیم، بیماری تیروئیدی، کراتینین
علیزاده و همکاران (۳۳)	C5.0 C.R Tree (درخت تصمیم) CHAID (الگوریتم مشخصه اسمی) QUEST* (الگوریتم دودویی گسترش)	بیماری قلبی	C.R Tree (درخت تصمیم) ۷۶/۰۴ درصد شبکه عصبی ۵۷/۲۹ درصد	آنژیوگرافی پر عارضه و هزینه بر می باشد، از روش های داده کاوی به پیش بینی پرداختند.	سن، فشار خون، وزن، جنس، کلسترول، شغل، سیگار، روش درمان
صفدری و همکاران (۱)	درخت تصمیم شبکه عصبی	انفاکتوس قلبی	درخت تصمیم ۹۳/۴ درصد	متغیرهای فشارخون بالا، چربی خون بالا و مصرف سیگار، بیشترین تاثیر را در ابتلا به آنفارکتوس قلبی دارا بودند. درخت تصمیم، قوانینی استخراج شده است که می تواند به عنوان الگویی در جهت پیشگویی احتمال ابتلا افراد به آنفارکتوس قلبی استفاده شود.	فشار خون، چربی خون، سابقه مصرف سیگار، تری گلیسیرید، کلسترول خون بالا (HDL)، کلسترول خون پایین(LDL)، Cr (کراتینین)، BUN
کاظمی و همکاران (۱۴)	شبکه عصبی چند لایه	بیماری قلبی	۸۳/۳ درصد	در این مطالعه متغیرهای تعداد عروق بزرگ، کاهش استرس، نقص، درد قفسه سینه، اوج استرس،آنژین، جنسیت، ایستایی نوار قلب، فشار خون، قند خون و کلسترول سرم بیشترین اهمیت را دارند.	تعداد عروق بزرگ، کاهش استرس، نقص، درد قفسه سینه، اوج استرس، آنژین، جنسیت، ایستایی نوار قلب، فشار خون، قند خون و کلسترول سرم
صباغ (۳۲)	C4.5 درخت تصمیم	بیماری عروق کرونر قلبی	۸۰/۲ درصد	به کمک درخت تصمیم، قوانینی استخراج شده است که می تواند به عنوان الگویی در جهت پیشگویی احتمال ابتلا افراد	سن، جنس، درد قفسه سینه، فسار خون، کلسترول، قند خون، آنژین
مظاهری و آشوری (۱۵)	بهینه سازی شبکه های اجتماعی Naïve bayes Bagging شبکه عصبی	بیماری عروق کرونر	۹۷/۲۲ درصد ۹۲/۰۹ درصد	عملکرد درخت تصمیم بهتر می باشد.	فاکتور هایGCG (پلی مورفیسم سه نوکلئوتیدی)، فاکتور های آزمایشگاهی، سن، جنس، دیابت، چاقی، فاکتور های تنفسی
Alonso و همکاران (۴)	طبقه بندی	آریتمی قلبی	۹۵ درصد	متغیرهایی که در پیشگویی اریتمی های قلبی در دسترس هستند را می توان به اندازه ی کافی پیش بینی کرد	قند خون، فشار خون، چربی خون،سن، جنس، عوامل محیطی، آنتروپی
Ilayaraja و Meyyappan (۳)	Apriori IMSIA کاوش قواعد وابستگی)	بیماری های قلبی		روش های داده کاوی کمک سریع به پزشک در تصمیم گیری های اضطراری برای نجات جان بیماران در معرض خطر ارایه می کند.	درد قفسه سینه، درد، حساسیت، ضعف، غش، آپنه، تنفسی، تاکی کاردی، برادی کاردی

*Tree Statically Effieient And Unbiased; QUEST ، *Bootsstrap Aggregating; Bagging ، *تقویت کنندگی وفقی(Ada boost)
*Low-Density Lipoprotein; LDL ،High-Density Lipoprotein; HDL، Social Media Optimization; SMO

Interleukin 6 از سلول های درشت خوار، لنفوسیت تی کمک کننده، لنفوسیت های بی و آستروسیت ترشح می شود و بر سلول های پلاسماسل و لنفوسیت موثر است. اصلی ترین کارکرد این سایتوکین تمایز سریع تر پلاسما سل ها و تولید بیشتر پادتن است.

بیماران عروق کرونری در مقابله با تنیدگی و فشارهای وارده بیشتر از افراد سالم از روش های مقابله ای با تنیدگی هیجان مدار استفاده می کنند و نیز تنیدگی منجر به افزایش در سطح بیومارکر ایمنی شناختی اینترلوکین-۶ میشود. طبق مطالعات و تحقیقات گذشته، عملکرد مدل-های طبقه بندی کننده ممکن است نتایج متفاوتی داشته باشند، که مقایسه این تفاوت ها و تشابهات با تحقیق فعلی در جدول ۴ ارائه شده است. در ادامه به تعدادی از مطالعات مشابه که نتایج همسویی با یافته های این پژوهش داشته اند پرداخته شده است. کاظمی و همکارانش (۱۵) در مطالعه ای تحت عنوان پیش بینی بیماری قلبی با استفاده از تکنیک داده کاوی شبکه عصبی دریافتند که با توجه به مدل های استفاده شده متغیرهای تعداد عروق بزرگ، کاهش استرس، نقص، درد قفسه سینه، اوج استرس، آنژین، جنسیت، ایستایی نوار قلب، فشار خون، قند خون و کلسترول سرم به ترتیب بیشترین اهمیت را از لحاظ مدل شبکه عصبی با ساختار پرسپتروم چند لایه در پیش بینی متغیر پاسخ بیماری قلبی دارند (۲۶). Alonso و همکارانش (۴) در تحقیقی با عنوان پیشگویی مدل ریسک فاکتورهای موثر بر آریتمی های قلبی در جمعیت متنوع جغرافیایی به این نتیجه رسیدند که متغیرهایی که در جمعیتی از ایالت متحده و اروپا در پیشگویی آریتمی های قلبی در دسترس هستند را می‌توان به اندازه ی کافی پیش بینی کرد(۴). در مطالعه ای دیگر تحت عنوان تشخیص بیماری عروق کرونر قلبی با استفاده از درخت توسط صباغ گل (۳۲) در رابطه با بیماری کرونر قلبی انجام شد، دریافت که در مقایسه با نتایج مطالعات انجام شده در حوزه داده کاوی بیماری قلبی، دقت به دست آمده الگوریتم پیشنهادی قابل قبول است (۴). صفدری و همکارانش (۱) در مقاله ای تحت عنوان مقایسه عملکرد درخت تصمیم گیری و شبکه عصبی در پیشگویی ابتلا به آنفارکتوس قلبی دریافتند که با ویژگیهای مشخص می توان تعیین کرد که احتمال ابتلا به آنفارکتوس قلبی چقدر است (۱).

در تحقیقی تحت عنوان پیش بینی روش درمان بیماری قلبی با استفاده از الگوریتم های داده کاوی مظاهری و همکارانش (۳۰) دریافتند که هدف این مطالعه ارائه مدلی برای پیش بینی روش درمانی مناسب بیماری قلبی به منظور کاهش هزینه های درمان و کیفیت ارائه خدمات بهتر به پزشکان می باشد. با توجه به قابل ملاحظه بودن خطرات اجرای روش تشخیص تهاجمی مانند، آنژیوگرافی و نیز حصول تجارب موفقیت آمیز داده کاوی در پزشکی، این مطالعه مدلی مبتنی بر تکنیک های داده کاوی ارائه نموده است. نقطه قابل بهبود مدل فوق ارائه سیستمی تصمیم یار جهت کمک به پزشکان برای افزایش صحت تشخیص روش درمان بیماری می باشد. Ilayaraja و Meyyappan (۳) در مطالعه پیش بینی خطر بیماری های قلبی از طریق داده کاوی های مکرر به این نتیجه رسیدند که از مجموع ۱۰۰۰ داده از پرونده بیماران مبتلا به بیماری مختلف قلبی پیش بینی میزان خطر به طور کارآمد امکان پذیر است (۳).

علیزاده و همکاران (۳۳) در مطالعه تحت عنوان های: تشخیص بیماری های عروق کرونر قلب با استفاده از داده کاوی بر اساس داده های آزمایشگاهی و فاکتوری اکولوژی و پیش بینی بیماری عروق کرونر با استفاده از الگوریتم های داده کاوی دریافتند که آیتم های اکولوژِی قلب و میزان برون ده قلبی بیشترین اثر را بربیماری های قلبی دارند و الگوریتم های داده کاوی دراین نتیجه بسیار موثرند (۳۳).

با توجه به آنچه بیان شد می توان اذعان نمود که، روش های داده کاوی در مدل سازی مناسب جهت پیش بینی امکان ابتلای افراد به بیماری های مختلف از جمله قلبی، مغزی و کلیوی استفاده شوند. در این تحقیق مدل درخت تصمیم بالاترین میزان صحت را دارا بوده و می تواند در برنامه های غربالگری جهت شناسایی افراد در معرض خطر استفاده شود. لذا پیشنهاد می شود که این مدل با متغیرهای بیشتر و در بازه زمانی بیشتر اجرا شده و زمانی که به روایی مورد تایید رسید، استفاده شود، هم چنین در تحقیقات بعدی از الگوریتم های دیگر داده کاوی و متغیرهای متفاوت تر استفاده و پس از تغییرات ضروری و رسیدن به سطح قابل قبول، مدلی کامل و مناسب برای پیش بینی ابتلا افراد به بیماری های عروق کرونر قلب و آریتمی های قلبی ارایه گردد.

انجام هر پژوهشی با محدودیت های مواجه است، مطالعه پیشرو نیز از این امر مستثنی نبوده و همزمانی انجام مطالعه با شیوع بیماری کرونا، دسترسی سخت به داده ها و عدم

همکاری لازم برخی از اعضای نمونه، از اهم محدودیت های انجام این مطالعه بوده اند.

سپاسگزاری

نویسندگان این مقاله تشکر و قدردانی خود را از تمام بزرگوارانی که در اجرای این پژوهش ما را یاری نمودند، می نمایند.این مقاله از پایان نامه ی اینجانب با عنوان پیش گویی مدل ریسک فاکتورهای بیماریهای عروق کرونر وفیبریلاسیون دهلیزی با استفاده از الگوریتم های داده کاوی در دانشگاه آزاد اسلامی واحد فلاورجان استخراج شده است و پس از بررسی با شناسه ی اخلاق IR.IAU.FALA.REC.1398.+ 52 مصوب گردیده است.

منابع

1. Safdari R, Ghazi Saeedi M, Gharooni M, Nasiri M, Argi G. Comparing performance of decision tree and neural network in predicting myocardial infarction. Journal of Paramedical Sciences & Rehabilitation 2014; 3(2): 26-35. [Persian]

2. Shahid AH, Singh MP. A novel approach for coronary artery disease diagnosis using hybrid particle swarm optimization based emotional neural network. Biocybernetics and Biomedical Engineering 2020; 40(4):1568-1568.

3. Ilayaraja M, Meyyappan T. Efficient data mining method to predict the risk of heart diseases through frequent itemsets. Procedia Computer Science. 2015; 1(70): 586-592.

4. Alonso A, Krijthe BP, Aspelund T, Stepas KA, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. JAm Heart Assoc 2013; 2(2): e000102.

5. Bahrami B, Shirvani MH. Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST) 2015; 2(2): 164-168.

6. Yaribeygi H, Taghipour H, Taghipour H. Prevalence of cardiovascular risk factors in patients undergoing CABG: brief report. Med J 2014; 72(8): 570-574. [Persian]

7. Daraei A, Hamidi H. An efficient predictive model for myocardial infarction using cost-sensitive j48 model. Iranian Journal of public health 2017;46(5): 68.

8. Murakami N, Tanno M, Kokubu N, Nishida J, et al. Distinct risk factors of atrial fibrillation in patients with and without coronary artery disease: a cross-sectional analysis of the BOREAS-CAG Registry data. Open Heart 2017; 4(1): e000573.

9. Sharifi M. Computational approaches to understand the adverse drug effect on potassium, sodium and calcium channels for predicting TdP cardiac arrhythmias. J Mol Graph Model 2017; 76: 152-160.

10. Vijayashree J, Sriman Narayana Iyengar NC. Heart disease prediction system using data mining and

hybrid intelligent techniques: A review. International Journal of Bio-Science and Bio-Technology 2016; 8(4):139-148.

11. Yadav C, Lade S, Suman MK. Predictive analysis for the diagnosis of coronary artery disease using association rule miningInternational Journal of Computer Applications 2014; 87(4): 0975-8887.

12. Veisi H, Qaidsharf HR, Ebrahimi M. Improving the efficiency of machine learning algorithms in heart disease diagnosis by optimizing data and features. Soft computing 2019; 8(1): 70-85. [Persian]

13. Ayatollahi, H., Gholamhosseini L, Salehi M. Comparing the performance of data mining algorithms in predicting coronary artery disease. Journal of Health and Biomedical Informatics 2017; 5(2): 252-264. [Persian]

14. Kazemi M, Mehdizadeh H, Shiri A. Heart disease forecast using neural network data mining technique. Journal of Ilam University of Medical Sciences 2017; 25(1): 20-32. [Persian]

15. Mazaheri S, Ashuri M, Bechari Z. Prediction of heart disease treatment method using data mining algorithms. Journal of Payavard Salamat 2017; 11(3); 287-296. [Persian]

16. Kumar MN, Koushik, KVS, Deepak K. Prediction of heart diseases using data mining and machine learning algorithms and tools. International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2018; 3(3), 887-898.

17. Salve I, Kumar D, Borikar A. Heart Disease Prediction using Data Mining Techniques. Journal of Algebraic Statistic 2022; 13(3): 2250-2259.

18. Huang PS, Tseng YH, Tsai CF, Chen J. et al. Artificial Intelligence-Enabled ECG Algorithm for the Prediction and Localization of Angiography-Proven Coronary Artery Disease. IJERPH 2022; 10(2): 10.3390/biomedicines10020394 .

19. Chen JIZ, Hengjinda P. Early prediction of coronary artery disease (CAD) by machine learning method-a comparative study. IRO Journals 2021; 3(01): 17-33.

20. Imamovic D, Babovic E, Bijedic N. Prediction of mortality in patients with cardiovascular disease using data mining methods IS(INFOTEH) 2020 (pp. 1-4).

21. Vindhya L, Beliray PA, Sravani CR, Divya DR. Prediction of Heart Disease Using Machine Learning Techniques. International Journal of Research in Engineering, Science and Management 2020; 3(8): 325-326.

22. Choi Y, Boo Y. Comparing logistic regression models with alternative machine learning methods to predict the risk of drug intoxication mortality. IJERPH 2020;17(3): 89.

23. Niaksu O. CRISP data mining methodology extension for medical domain. Baltic Journal of Modern Computing 2015; 3(2): 92.

24. Mathan K, Kumar PM, Panchatcharam P, Manogaran G, Varadharajan R. A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease. Design automation for embedded systems 2018; 22(3): 225-242.

25. Zabbah I, Hassaanzadeh M. The effect of continuous parameters on the diagnosis of coronary artery disease using artificial neural networks. Journal of Health Chimes 2017; 4(4): 29-39. [Persion]

26. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association 2017; 24(2): 361-370.

27. Cherfi A, Nouira K, Ferchichi A. Very fast C4. 5 decision tree algorithm. Applied Artificial Intelligence 2018; 32(2): 1-19.

28. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. Journal of clinical epidemiology 2013; 66(4): 398-407.

29. Rani KU. Analysis of heart diseases dataset using neural network approach. IJDKP 2011; 1(5):1-8.

30. Mazaheri S, Ashoori M, Bechari Z. A model to predict Heart disease treatment using data mining. Journal of Payavard Salamat 2017; 11(3): 287-296. [Persion]

31. Karimian B, Khadem Watan K, Seyed Mohammadzadeh MH, Alinejad V, et al. Investigating the relationship between the severity of coronary artery disease with uric acid and Fetuin A in patients undergoing coronary angiography. Journal of Medical Sciences Studies 2014; 26 (8): 671-663.

32. Sabbagh Gol H. Detection of Coronary Artery Disease Using C4.5 Decision Tree .Journal of Health and Biomedical Informatics Medical Informatics Research Center 2017; 3(4): 287-299. [Persian]

33. Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, et al. A data mining approach for diagnosis of coronary artery disease. Computer methods and programs in biomedicine 2013; 11(1):52-61. [Persian]