مقاله پژوهشی

مقدمه: بیماری‌های ایسکمیک قلبی یکی از شایع‌ترین بیماری‌های ایست که هم‌اکنون مرگ‌آور بالایی را در جهان به خود اختصاص می‌دهد. بیماری‌های ایسکمیک قلب به دلیل تغییرات در سطح شریان‌های قلبی به وجود می‌آیند که تأمین کننده خون قلب هستند. این امر به‌ویژه در جلوگیری از توده‌های استیوانفیک بیماری و تغییر در کلیه‌های زندگی آن ها می‌تواند سبب کاهش ماغر و بیمار افزایش در اندازه جسمی گردد.

روش بررسی: مطالعه سلامت مردم یزد (باس) به بررسی سلامت و بیماری‌های نمونه تصادفی 1000 نفر از مردم یزد در حالی که در جدال و مبارزه با کرونا و در پیشگیری از امید به آینده بود. مطالعه با استفاده از روش بیشترین تأثیر را در افزایش حساسیت دسته‌بندی تهیه و کشف زیر‌گروه CN2 با استفاده از میزان مورد است. این بیماری‌ها بیشترین تأثیر را در افزایش حساسیت دسته‌بندی تهیه و کشف زیر‌گروه CN2 با استفاده از میزان مورد است.

نتایج: پاسخگویی دانش داد که برای غربالگری افراد مستعد به بیماری‌های ایسکمیک قلبی با استفاده از تولید نمونه افراد مبتلا به روش CN2-SD مطرح می‌شود.

تئیه‌گری: با بررسی مدل‌های روش‌های غربالگری افراد مستعد بیماری‌های ایسکمیک قلبی کارایی مناسبی دارد و به‌ویژه در افراد اولیه که به دنبال مراقبت اضافی از سلامتی می‌شود، بسیاری از هنگامی که مشابه استفاده از تهیه و کشف زیر‌گروه CN2 با استفاده از میزان مورد است.

واژه‌های کلیدی: داده‌کاپی، پایش سلامت، بیوپنی بیماری‌های ایسکمیک قلبی، میزان سازی داده‌های گروهی، پایش سلامت، داده‌کاپی، بیماری‌های ایسکمیک قلبی، میزان سازی داده‌های گروهی.

ارجاع: طبیعی مهدی مجد، سعادت جو فاطمه، مریبای مسعود. مدل پیش‌بینی بیماری‌های ایسکمیک قلبی با استفاده از داده‌کاپی داده‌های مطالعه سلامت مردم یزد (باس). مجله علمی پژوهشی دانشگاه علوم پزشکی شهید صدوقی یزد، 1398، 13 (2)، صفحات 1346-1350.
این مقاله مربوط به استفاده از مدل‌های آماری برای تشخیص حمله قلبی می‌باشد.

با استفاده از داده‌های جمع‌آوری‌شده از مطالعه سلامت مردم یزد (یازد)، اطلاعات سلامت و بیماری بیش از هزار نفر جمع‌آوری و نتیجه‌گیری از آن تهیه شده است. در این مطالعه، عوامل ایمنی و دیگر عواملی که ممکن است به رخپاشندگی مبتلا شدن به حمله قلبی تاثیر بگذارند، مورد بررسی قرار گرفته است. مدل‌های آماری مانند مدل دو‌لایه‌ای میلی‌پرسی، مدل دو‌لایه‌ای چند‌طبقه نسبت به سایر مدل‌ها بهتر عمل می‌کنند.

در این مقاله، مدل‌های آماری مانند Support Vector Machine (SVM) و Multi-Layer Perceptron (MLP) مورد بررسی قرار گرفته که نتایج نشان می‌دهد که سایر مدل‌ها با استفاده از این داده‌ها، ممکن است بهتر عمل کنند.

در عمل پزشکی تعیین یک محدوده مشخص برای سالم و یا بیمار بودن شرایط و یا نیاز به درمان پیشنهاد می‌شود و تشخیص در بیماری‌ها از موارد مبهم و غیرقابل انجام می‌شود. در این مقاله، با استفاده از داده‌های جمع‌آوری‌شده، مدل‌های آماری مانند SVM و MLP مورد بررسی قرار می‌گیرند.

در نهایت، نتایج نشان می‌دهد که مدل‌های آماری می‌توانند در تشخیص حمله قلبی مؤثر باشند و در پیشگیری از این بیماری ممکن است کاربردی داشته باشند.

در نهایت، نتایج نشان می‌دهد که مدل‌های آماری می‌توانند در تشخیص حمله قلبی مؤثر باشند و در پیشگیری از این بیماری ممکن است کاربردی داشته باشند.
به کار گرفته شد و این نتیجه به دست آمد که
masahin بردار پشتیبان با دقت ۸۴/۹ درصد بهترین روش برای
پیش بینی بیماری های قلبی عروق است (۷). در مقاله بهداد
میرزاپور و همکاران با اینکه یک دسته بند مثبت بر دقت
تصمیم و شکه عصبی و الگوی حکم جمعی برندگان برای
پیش بینی بیماری های قلبی از داده های UCI با ۱۴ ویژگی
استفاده شده است. بهترین دقت برای تصمیم با مقدار
۸۵/۳ درصد بوده است (۸). در خصوص تجربیات نظری
پیش بینی قلبی با استفاده از شکه عصبی و یا سیستم عصبی فاژی
تحلیل انگشتی است. تعداد ۱۳ تاریخی مؤثر در بیماری
بهصورت تجريبي و یک هکم از آنها به آنها. عمل
پیش بینی از روی داده ها توسط شبکه عصبی فاژی و شکه
عصبی مصنوعی به ترتیب با دقت ۷۸ و ۷۵/۳ درصد صورت
پذیرفت (۹).

کومران سن و همکاران با استفاده از روش شبکه عصبی فاژی
Neuro-Fuzzy Integrated Approach تجربیات دوستلی است. نتیجه به آنالیز مجموعه داده
Two Level UCI بررسی به عمل آمده این بود که به نتایی ۹ عامل در بیماری
Coronary Artery Disease CAD باکری برای انجام شده از حداکثر کمتر و دقت بالا بوده
CN2 است (۲). در مقایسه با بخش اصلی الگوی صورت و درخت تصمیم
گیری الگوی CN2 نتیجه بوده است. الگوی CN2 روز داده های UCI
است با دقت ۹۳/۷ درصد بیماری را تشخیص دهد (۱۰). در بخش مناسبی بیماری
از روی داده های طرح سلامت و
CHD تغذیه کننده که این شده است. در این مقاله روش ترکیبی از
درخت تصمیم و منطق فازی راه گیری گردیده که مبنی به افزایش
دقت و حساسیت در تشخیص شده است. دقت پیش بینی در
روش ترکیبی فوق برای ۶۹/۵ درصد بوده است (۱۱).

روش پرسی

همانطور که قبل بیان شد اندر تعقیبی انجام شده در
واستا ی پیش بینی بیماری قلبی بوده است. در این مقاله معیار
رتبه دوم روش شبکه عصبی چند لایه (MLP) با دقت ۹۱
درصد بود. پیش بینی الگوی C5 با دقت ۸۷/۹ درصد قرار دارد.
برای ارزیابی روش های داده کاوی از روش - 10 Fold cross-
Validation استفاده شده است. در مقاله پیش بینی بیماری عروق کرون موجب کمک شبکه عصبی
و گروینگ متنی بر دقت رگرسیون و دستی‌بندی (۵) از ۱۲۳۸ نفر بیماری عروق
کرون و ۹۱۵ نفر مبتلا شمل ۹ فاکتور خطر بیماری قلبی
استفاده شده است. این نمونه‌ها از مرکز قلب تهران و با
استفاده از آنتی‌ویروسی فیس بیست و دی ۱۳۳۸ درصد
بدست آمده است. به حذف ۵ تغییر انجام آن با ۴ فاکتور
خطر بیماری مجدداً همان دقت حاصل شده است که نشان
می‌دهد تکنیک‌های گروینگ متنی به کاهش پیچیدگی مدل
کمک می‌کند.

در مقاله راسل داس و همکاران از داده‌های بیماری قلبی
University of California, Irvine UCI
خبره استفاده شده بود. در این مقاله ۱۲۳۸ نفر از ۷۷ مین با کمک
انتخاب شده و بعد از حذف داده‌های گم شده به تعداد
۱۳۹۷ نمونه نسبت به طراحی سیستم خبره با استفاده از شبکه‌های
Aقدم گردد. Artificial Neural Network
اصبعی مصنوعی
اس است که بیشترین دقت آن ۸۹/۲ درصد بوده است (۲). در
طراحی یک سیستم خبره برای تشخیص بیماری قلبی از
۱۹ صفت مجموعه داده UCI استفاده شده است. آن مقاله به
اس است که بیشترین دقت آن ۸۹/۲ درصد بوده است (۲). در
طراحی یک سیستم خبره برای تشخیص بیماری قلبی از
۱۹ صفت مجموعه داده UCI استفاده شده است. آن مقاله به
اس است که بیشترین دقت آن ۸۹/۲ درصد بوده است (۲). در
طراحی یک سیستم خبره برای تشخیص بیماری قلبی از
۱۹ صفت مجموعه داده UCI استفاده شده است. آن مقاله به
نمونه‌گیری مطالعه‌ی پایه‌ی به صورت خوش‌های انجام‌شده است.

در مرحله‌ی پیشین، نمونه‌گیری دست‌بندی بر حسب کد پستی و خوش‌های شریعتی صورت گرفته است. سپس، ۲۰۰ خوشه‌ی ۵۰ نفری انتخاب شدند. این طرح یک مطالعه‌ی آن‌لاین است که در مدت چهار روز انجام‌شده است. جامعه‌آمیزی مطالعه‌ای افراد بالای ۲۰ تا ۶۹ سال ساکن شهرستان پرداخته‌اند. روش نمون‌گیری این مطالعه‌ی سطح‌های مختلف بوده و دوینج بوده است: نمونه‌گیری دست‌بندی بر حسب کد پستی و خوش‌های شهری است. در این طرح، بر اساس فهرست برداری خانوار سال ۱۳۹۶، سرخوشی‌ها انتخاب و با حکم زنده است. راست نسبت به‌بک‌سیر نرسیدن نپرست نامه اقدام شده و خوان‌هاوردی بعده بانگت انتخاب کردی. درسرخوری که دیک پلاک چند خوان‌های موجود داشته (مثل مجتمع‌های مسکونی)، از واحد اول شروع و بعد به‌و‌حدودی بعدی مراحلش دارد. درسرخوری که پلی از پلاک نفر واحد شرایط مجدیده محل بوده با همه افراد ۲۰ تا ۶۹ سال مصاحبه صورت گرفته است (ولی در هر گروه سنی ده ساله فقط یک نفر از هر آدرس) تا امکان بررسی تجمع‌های اصلی و تحت شویاندا خوشه‌ی پرسته‌گری، اخذ رضایت آگاهانه و رعایت اصول اخلاقی آورده شد. روش پرسته‌گری به‌صورت زبان‌ انتخابی قوام و مضمون بوده و در بررسی و ارزیابی افکار و دیدگاه‌های فردی نشان داده شده است.
روش های دسته‌بندی استفاده از مشاهده‌های بی‌درد و شیب‌بانمی‌باید، این روش معنی‌برنگانه‌تری که در آن دو دسته داده‌ها وجود دارد، استفاده می‌شود. در این الگوریتم دسته‌بندی، مجموعه مورد انتخاب تا زمان در نظر گرفته شود، در مثال، یک مورد ممکن است از یک ترکیب تا زمان استفاده شود. در این الگوریتم، می‌تواند در صورتی که دو دسته داده می‌باشد، به یک ضرورت دسته‌بندی کنند، هفته یافتن دو مورد مشابه یا مربوط به هم می‌تواند شایع باشد. ناوتایی است که مشترکی فاصله را دارد. دوین دسته بندی استفاده شده در این مقاله الگوریتم بیزی می‌باشد. این الگوریتم تحت نام لاگستیک از جمله بزرگ و مستقل، شبکه ای‌ها و شکله باور شده است. همه این نام جا برگرفته از این که در قاعد قدرتی گیره جهت دسته‌بندی داده‌ها است. این الگوریتم بر اساس مجموعه پارامترها و تغییر آنکه بیشینی نتایج بر پایه احتمال را فراهم می‌کند. (۱۲).

روش استفاده شده است: الگوریتم آدابست Logistic Regression روش رگرسیون لجستیک مبتنی بر آمار است که به‌دیگری مانند با نظرات استفاده می‌شود. الگوریتم یادگیری به نظرات به‌روش گفتگو و در مهارت امکان‌پذیر بودن توان داشته که در زمان آموزش، نتیجه‌های به‌روش ممکن است در حساب باشند. در این الگوریتم محاسبه می‌شود که مقدار خروجی آن به‌جای مقدارهای دیگر است و تابع می‌تواند به‌صورت خلاصه یک‌گونه پایداری شود. وقتی خروجی نباید گوس آن با توزیع ابتدایی چاپ‌گری شود برای مثال دسته‌بندی قابل استفاده است. رگرسیون لجستیک حالت خاص پیاده‌سازی رگرسیون با توزیع برنولی یا تابع گاما خروجی آن را می‌تواند به‌صورت خلاصه یک‌گونه پایداری شود. وقتی خروجی نباید گوس آن با (۱۳) شکل‌های معمولی بر اساس مغز انسان با جای‌گیری پیامدار می‌شود و به‌صورت قرم‌ساده ای از ورودی و خروجی‌ها است. تجزیه‌بندی یک‌گونه را می‌توان از طریق شکهبندی ریاضی معمولی مثلاً فرض کرده در این الگوریتم مورد استفاده قرار گرفته است. در این الگوریتم می‌تواند در موارد یک یا چند متغیر، ممکن است از این الگوریتم به‌صورت می‌توانند استفاده کنند. (۱۱) دسته‌بندی الگوریتم قواعد CN2 CN2 Rule Inducer در سال ۱۹۸۹ و واس ایرانی که در آن دو دسته داده‌ها وجود دارد، استفاده می‌شود. در این الگوریتم دسته‌بندی، مجموعه مورد انتخاب تا زمان در نظر گرفته شود، در مثال، یک مورد ممکن است از یک ترکیب تا زمان استفاده شود. در این الگوریتم، می‌تواند در صورتی که دو دسته داده می‌باشد، به یک ضرورت دسته‌بندی کنند، هفته یافتن دو مورد مشابه یا مربوط به هم می‌تواند شایع باشد. ناوتایی است که مشترکی فاصله را دارد. دوین دسته بندی استفاده شده در این مقاله الگوریتم بیزی می‌باشد. این الگوریتم تحت نام لاگستیک از جمله بزرگ و مستقل، شبکه ای‌ها و شکله باور شده است. همه این نام جا برگرفته از این که در قاعد قدرتی گیره جهت دسته‌بندی داده‌ها است. این الگوریتم بر اساس مجموعه پارامترها و تغییر آنکه بیشینی نتایج بر پایه احتمال را فراهم می‌کند. (۱۲).

روش استفاده شده است: الگوریتم آدابست Adaboost روی روشی است که به‌روش دقت الگوریتم یادگیری می‌باشد. روند کلی این الگوریتم دین‌سوزی است که مجموعه‌ای وزن‌دار در تعداد زیادی دسته‌بندی کننده ضعیف به‌عنوان دسته‌بندی کننده نهایی انتخاب می‌شود. به‌دعارت دیگر آدابست با استفاده از مجموعه ای‌ها در دسته‌بندی کننده ضعیف می‌تواند متجر بی‌که دسته‌بندی کندنه قدرت‌مرد گردد.

از آنجا که این روش می‌تواند دسته نهایی ضعیف را نقوچمی نماید برای دسته‌بندی داده‌های نیازدار مناسب نمی‌باشد. اگرای داده‌های نامتنوان عملکرد بیشتری دارد (۱۶). یکی از مسائل مهم در زمینه داده کاوی، مسئله دسته‌بندی مجموعه داده‌های نامتنوان است. اصل مجموعه‌ای داده نامتنوان و عملکرد گفتگوی داده کاوی در موارد یک کلاس از نامتنوان عملکرد مشکل ضعیفی دارد. در یک مسئله تحقیق پرشنگ نمی‌توان به‌مقایسه با نمی‌توان به‌مقایسه این مجموعه هم‌مدخل در اقلیت یافتن دارد. ولی هدف از
آن هزینه در مجموعه داده‌ها وجود ندارد (18). در این مقاله از روش سطح داده جهت توانایی سازی داده‌ها استفاده شده است. دانشگاه که در مرحله پایگیری دسته‌بندی تولید می‌شود، می‌باشد در مرحله ارزیابی مورد تحلیل بر پایه یک کارایی گروهی یادگیرنده دسته‌بندی، را نیز مشخص کرد. این معماری را می‌توان که برای مجموعه داده‌های آموزشی در مرحله یادگیری و هم برای مجموعه گروه‌های آزمایشی در مرحله ارزیابی محاسبه نمود. در اینجا مفهوم ماتریس در هم ریختگی بایان می‌شود. این ماتریس چگونگی عملکرد الگوریتم دسته‌بندی را با توجه به مجموعه داده ورودی به تفکیک این ویژگی مسائل دسته‌بندی مطلق جدول 1 نمایش می‌دهد.

جدول 1: ماتریس در ریختگی

<table>
<thead>
<tr>
<th>وضعیت</th>
<th>پیش‌بینی</th>
<th>واقعی</th>
</tr>
</thead>
<tbody>
<tr>
<td>پیش‌بینی</td>
<td>Positive</td>
<td>TP</td>
</tr>
<tr>
<td>واقعی</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>TP+FP</td>
<td></td>
</tr>
</tbody>
</table>

روش سطح الگوریتم

1. روش پردازش قبل از دسته‌بندی، موج مجموعه داده های نامتوان می‌شود.
2. روش سطح الگوریتم

این روش که یکی از روش‌های بازبینی در سطح الگوریتم است با نگرش در الگوریتم دسته‌بندی به نوع مسئله عدم توانایی مرتفع می‌شود. این روش در این مقاله از دسته‌بندی است و هم‌زمان با توجه به استفاده از الگوریتم ایجاد خواهد کرد. مهمترین نقطه این روش، تعیین هزینه دسته‌بندی نادرست است که عوامل اطلاعاتی در مورد استفاده از داده‌های پیش‌بینی یافتن نموتهای بیماری است. به این ترتیب کل داستان نسبت به سایر نمونه‌ها دارای اولویت پیش‌بینی است. در اغلب الگوریتم‌های دسته‌بندی، تیپ‌ها تیپ‌ها به جهت کلاس‌های است که پیش‌بینی نمونه‌ها دارای اولویت هستند. از این روش توانایی کمی در برش گویی صحیح داده‌های کلاس اقلیت از خود نشان می‌دهد. در نتیجه، قوانین دسته‌بندی سبب می‌شود تا داده‌های کلاس اقلیت، نادرست دسته‌بندی شوند. برای حل این مشکل باید از روش‌های موجود به داده‌های نامتوان استفاده نمود. تکنیک‌های منعی برای حل سیستم از ارتقاء با کلاس نامتوان به استفاده است که در سه مرحله روش سطح داده، روش سطح الگوریتم و روش حساس به هزینه تقسیم می‌شوند. در ادامه به توصیف هرکدام از این روش‌ها خواهد پرداخت.

سید محمود رضا طباطبایی و همکاران

دوره بیست و هفتم، شماره سه، خرداد 1398

مجله دانشگاه علوم پزشکی و خدمات بهداشتی – دارویی شهید صدوقی پردیس

1361
ROC Receiver Operating Characteristic curve: The area under the ROC curve (AUC) is a measure of model performance. A perfect classifier would have an AUC of 1, while a classifier with no discriminatory power would have an AUC of 0.5. Cross-Validation is a technique used to estimate the performance of a model on unseen data.

Detection Rate (DR): The fraction of true positives that are identified by the test.

False Alarm Rate (FAR): The fraction of true negatives that are incorrectly predicted as positive.

Precision: True positive rate / (true positive rate + false positive rate)

Recall: True positive rate / (true positive rate + false negative rate)

Accuracy: (true positive + true negative) / (true positive + true negative + false positive + false negative)

Sensitivity: True positive rate / (true positive rate + false negative rate)

Specificity: True negative rate / (true negative rate + false positive rate)

Error Rate: False positive rate / (false positive rate + true negative rate)

False Alarm Rate: False positive rate / (false positive rate + true negative rate)

Precision: True positive / (true positive + false positive)

Recall: True positive / (true positive + false negative)

Accuracy: (true positive + true negative) / (true positive + true negative + false positive + false negative)

Sensitivity: True positive / (true positive + false negative)

Specificity: True negative / (true negative + false positive)

Error Rate: False positive / (false positive + true negative)

False Alarm Rate: False positive / (false positive + true negative)

Detection Rate (DR): True positive / (true positive + false negative)

False Alarm Rate (FAR): False positive / (false positive + true negative)
شید محمدضا طباطبایی و همکاران

نتایج

در مطالعه پای سکه که از داده های آن برای آن پژوهش استفاده شده، جامعه نمونه گیری، ساکنی شرکت‌ها بر اساس میانگین که به صورت تصادفی خوشه ای نمونه گیری شده و نسبت افراد سال های 10 یا بیشتر به سال بر اساس 100 می‌باشد.

برای بررسی تعادل بین دو کلاس نسبت به متر مربع در سطح داده و با روش افایی کلساقیت به روش پیوسته استریب افایی است. این کلاس ده برابر افایی پایه می‌باشد. لازم به ذکر است میزان افایی کلساقیت نمونه کلاس بی نیاز به تعداد نمونه سال و در راستا متر مربع سازی می‌تواند به اندازه‌ای‌ها انجام می‌شود و قابل تنظیم می‌باشد. کلیه عوامل جایی نمونه ها در توجه کلیه اطلاعات و تعیین‌های Bootstrap در این مقاله با توجه به چک و یافتن ترغیبی افایی و جایه‌گذاری Bootstrap، با استفاده از روش Bootstrap، تعیین نمونه خود کلاس ادامه بردار می‌شود، تعداد نمونه موجود در این کلاس در برقرار افایی پایه است. لازم به ذکر است میزان افایی یکی از عوامل تعیین کلاس نیز می‌باشد. به تعداد نمونه ها و در راستا متر مربع سازی داده‌ها انجام می‌شود و قابل تنظیم می‌باشد.

کلیه عوامل جایی نمونه ها در توجه کلیه اطلاعات و تعیین‌های Bootstrap در این مقاله با توجه به چک و یافتن ترغیبی افایی و جایه‌گذاری Bootstrap، با استفاده از روش Bootstrap، تعیین نمونه خود کلاس ادامه بردار می‌شود، تعداد نمونه موجود در این کلاس در برقرار افایی پایه است. لازم به ذکر است میزان افایی یکی از عوامل تعیین کلاس نیز می‌باشد. به تعداد نمونه ها و در راستا متر مربع سازی داده‌ها انجام می‌شود و قابل تنظیم می‌باشد.

نقطه کلیدی بررسی کلیه کلاس با استفاده از bootstrap، تعیین نمونه خود کلاس ادامه بردار می‌شود، تعداد نمونه موجود در این کلاس در برقرار افایی پایه است. لازم به ذکر است میزان افایی یکی از عوامل تعیین کلاس نیز می‌باشد. به تعداد نمونه ها و در راستا متر مربع سازی داده‌ها انجام می‌شود و قابل تنظیم می‌باشد.

ملاحظات اخلاقی

برپوزیل این تحقیق نسبت دانشگاه علم و هنر تابیه شده است.

IR.SSU.REC.1397.246837
جدول ۳. ماتریس درهمیختگی حاصل از ارزیابی با ۱۰-fold تا ۱۲۹۴

<table>
<thead>
<tr>
<th>آداپوست</th>
<th>بیمار سال</th>
<th>بیمار</th>
<th>نتایج، سطح پذیرفتن</th>
<th>نمودار نام دستگاه</th>
</tr>
</thead>
<tbody>
<tr>
<td>بیمار</td>
<td>حساسیت</td>
<td>حساسیت</td>
<td>دقت</td>
<td>نمودار</td>
</tr>
<tr>
<td>بیمار سال</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
</tr>
<tr>
<td>بیمار</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
</tr>
<tr>
<td>بیمار</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
</tr>
<tr>
<td>بیمار</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
</tr>
</tbody>
</table>

۱۰-Fold cross-validation

جدول ۴. نتایج ارزیابی با CN2

<table>
<thead>
<tr>
<th>ادابوست</th>
<th>بیمار سال</th>
<th>بیمار</th>
<th>نتایج، سطح پذیرفتن</th>
<th>نمودار نام دستگاه</th>
</tr>
</thead>
<tbody>
<tr>
<td>بیمار</td>
<td>حساسیت</td>
<td>حساسیت</td>
<td>دقت</td>
<td>نمودار</td>
</tr>
<tr>
<td>بیمار سال</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
</tr>
<tr>
<td>بیمار</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
</tr>
<tr>
<td>بیمار</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
<td>19/7/04</td>
</tr>
<tr>
<td>بیمار</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
<td>0/8/05</td>
</tr>
</tbody>
</table>

Downloaded from jssu.ssu.ac.ir at 5:13 IRDT on Saturday August 28th 2021
سید محمدضا طباطبایی و همکاران

جدول ۱: ماتریس درهم پرخیصی حاصل از آزمایی با کل داده‌ها

<table>
<thead>
<tr>
<th>پیشینه</th>
<th>بردار پشتیبان</th>
<th>بیمار سالم</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>واقعی بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>مجموع</th>
</tr>
</thead>
<tbody>
<tr>
<td>بیمار</td>
<td>۵۶۷</td>
<td>۱۷۴</td>
<td>۷۳۸</td>
<td>۷۳۸</td>
<td>۷۳۸</td>
<td>۷۳۸</td>
<td>۷۳۸</td>
<td>۷۳۸</td>
</tr>
<tr>
<td>بیمار</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
<td>۴۹۷</td>
</tr>
<tr>
<td>مجموع</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
<td>۱۸۳۸</td>
</tr>
</tbody>
</table>

آداوسیت نرگسون لجستیک

<table>
<thead>
<tr>
<th>پیشینه</th>
<th>واقعی بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>بیمار بیمار</th>
<th>مجموع</th>
</tr>
</thead>
<tbody>
<tr>
<td>بیمار</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
<td>۴۴۹</td>
</tr>
<tr>
<td>بیمار</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
<td>۱۸۱۱</td>
</tr>
<tr>
<td>مجموع</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
<td>۱۳۸۲</td>
</tr>
</tbody>
</table>

جدول ۲: نتیجه آزمایی با کل داده‌ها

<table>
<thead>
<tr>
<th>حسابت</th>
<th>دقت</th>
<th>سطح نمودار</th>
<th>دسته‌بند</th>
<th>سطح نمودار</th>
<th>دقت</th>
<th>حسابت</th>
</tr>
</thead>
<tbody>
<tr>
<td>ماسیون پشتیبان</td>
<td>۰/۴۶۲</td>
<td>۰/۷۲۷</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۲۷</td>
<td>۰/۴۶۲</td>
</tr>
<tr>
<td>بردار</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
</tr>
<tr>
<td>شیب عصبی</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
</tr>
<tr>
<td>اتفاق فاقد</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
</tr>
<tr>
<td>بیز</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
<td>۰/۷۳۵</td>
</tr>
<tr>
<td>نرگسون لجستیک</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
<td>۰/۷۵۵</td>
</tr>
<tr>
<td>اداوسیت</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
<td>۰/۷۴۴</td>
</tr>
<tr>
<td>درخت تصمیم</td>
<td>۰/۷۴۳</td>
<td>۰/۷۴۳</td>
<td>۰/۷۴۳</td>
<td>۰/۷۴۳</td>
<td>۰/۷۴۳</td>
<td>۰/۷۴۳</td>
</tr>
<tr>
<td>نزدیکترین همسایه</td>
<td>۰/۷۴۵</td>
<td>۰/۷۴۵</td>
<td>۰/۷۴۵</td>
<td>۰/۷۴۵</td>
<td>۰/۷۴۵</td>
<td>۰/۷۴۵</td>
</tr>
</tbody>
</table>
جاه انجام عمل غربالگری بیماری مستند بیماری کرونا
لیبی از داده‌های طرح پاس استفاده شده است. بدیل
نمونه‌برداری بودن نسبت افراد سالم به بیماران لیبی با تقویت
داده‌های کلسی افراد سالم به روش بون استفاده نسبت به
متوان افراد سالم به داده‌های اقدام شده است. نتیجه آزموز و غربالگری
مستندنده‌ها به روش 10fold cross validation
که روش‌های مختلف ارزیابی تأثیر بسزایی در نتایج این گونه
روش متوان‌سازی دارند و مقاله‌ها را مقالاتی که از داده‌های
استفاده کرده‌اند دقیقاً تقریباً مشابه را نشان
UCI می‌دهند. اما بدیل این که توزیع داده‌های متناسب با

بحث
از مقادیر ماتریس در هر یک از تغییرات و روابط بین شده می‌توان
 مقادیر معیارهای ارزیابی را محاسبه کرد. در اینجا از سه معیار
 دقت، حساسیت و سطح زیست نمودار ROC استفاده شده
است. این مقادیر برای دسته‌بندی‌های استفاده شده در جدول

1 نمودار

از متغیرها

دسته‌بندی می‌شود شبکه‌های مبتنی بر ارزیابی با کل داده‌ها

نتیجه‌گیری

جاه انجام عمل غربالگری بیماری مستند بیماری کرونا
لیبی از داده‌های طرح پاس استفاده شده است. بدیل
نمونه‌برداری بودن نسبت افراد سالم به بیماران لیبی با تقویت
داده‌های کلسی افراد سالم به روش بون استفاده نسبت به
متوان افراد سالم به داده‌های اقدام شده است. نتیجه آزموز و غربالگری
مستندنده‌ها به روش 10fold cross validation
که روش‌های مختلف ارزیابی تأثیر بسزایی در نتایج این گونه
روش متوان‌سازی دارند و مقاله‌ها را مقالاتی که از داده‌های
استفاده کرده‌اند دقیقاً تقریباً مشابه را نشان
UCI می‌دهند. اما بدیل این که توزیع داده‌های متناسب با

در بیشینین بیماری‌های عروق کرونفلق با استفاده از داده‌های

شکل 1 نمودار ROC
References:


4. Xing Y, Wang J, Zhao Z. Combination data mining methods with new medical data to predicting


8- Mirzaei, B, Reza R. A Decision Based Modeling and Neural Network Based Modeling Model and Bird Motion Detection Algorithm for Heart Disease. 2015, 2nd IEEE National Conference of Technology, Energy and Data on Electrical and Computer Engineering, Kermanshah, 2016. [Persain]


16- Molahosseini A, Amirkhani H, Rahmati M, Using Adaboost Classifier Composition And Genetic Algorithms In Cryogenic Analysis Methods. 15th
International Annual Conference of the Computer Society of Iran, Tehran, 2009.


The prediction model for cardiovascular disease using Yazd's health study data (YaHS)

Seyed Mohammad Reza Tabatabaei Nodoushan¹, Fatemeh Saadatjoo², Masoud Mirzaei³

Introduction: Ischemic heart disease is one of the most common diseases, which has led to high mortality rates all over the world. This disease is caused by narrowing or blockage of coronary arteries, which are the provider of blood to the heart. Identifying the people susceptible to this disease and bringing changes in their lifestyles has been said to reduce the related mortality rates and increase the patient's longevity.

Methods: Yazd people Health Study (YaHS) was conducted on a random sample of 10,000 people living in the city of Yazd, Iran in the years 2014-15 for a general health and disease survey. These data were first balanced by bootstrapping technique due to their unbalanced nature. Next, classification methods were used in the training phase. Various classifiers, such as artificial neural network, rule inducer, regression, and AdaBoost were used in order to evaluate the proposed method with two scenarios.

Results: The results showed that the screening of the people susceptible to ischemic heart disease had the most significant effect on increasing the sensitivity of the discovery classifier of CN2 subgroup through using balanced data by bootstrapping method followed by their analysis for the purpose of producing a sample of the patients. This classifier proved to have the potential for detecting 83.6% of the people susceptible to this disease.

Conclusion: Therefore, it can be concluded that data mining methods are effective in screening for susceptible people with ischemic heart disease. This method can be compared with other traditional screening methods in that it is more cost-effective and faster.

Keywords: Data mining, Health monitoring, Prediction, Ischemic heart disease, Data balancing, Rule induction CN2-SD.


¹Department of Computer Engineering, Science and Arts University, Yazd, Iran.
²Department of Computer Engineering, Science and Arts University, Yazd, Iran.
³Department of Epidemiology, Yazd Cardiovascular Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

*Corresponding author: Tel: 09134509917, email: masoudmirzaei@yahoo.com