Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]

Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

Evaluation of Important Factors in Identifying Asymptomatic Carotid Artery Stenosis in order to Prevent Stroke by Using Data Mining Tools

Farzad Firouzi Jahantigh[1*], Razieh Alizadeh[2]

[1]Department of Industrial Engineering, University of Sistan and Baluchestan, Zahedan, Iran

[2]Department of Industrial Engineering, Payame Noor University, Tehran, Iran

Article Info:

Received: **14 Sep 2015** ***Accepted*****: 16 Jul 2016**

ABSTRACT

Introduction: Asymptomatic carotid artery stenosis is one of the factors that causes stroke. Other factors such as high blood pressure, cardiac diseases, smoking, diabetes, and physical inactivity may also cause the disease. Understanding and identifying the factors that cause carotid artery stenosis will help in prevention of acute stroke. Using data mining techniques, this study was aimed to discover the rules and relations that are effective in identifying asymptomatic carotid artery stenosis. ***Materials and Methods:*** To find the best approach, logistic regression (LR), genetic algorithm (GA), and chi-square test were used to predict carotid artery stenosis in patients. ***Results:*** 372 participants, 173 women (% 46.5) and 199 men (% 53.5), with an average age of 70.74± 5.29 were investigated. The results showed gender, smoking, coronary artery disease, high blood pressure, inactivity, prevention of pregnancy by medication, uremia (excessive amounts of urea and other nitrogenous compounds in the blood), and pulse rate environment are the significant risk factors for asymptomatic carotid artery. In addition, GA was a better method for this approach compared to LR. ***Conclusion:*** Our study revealed that coronary artery disease and hypertension are important factors in predicting and prognosis of asymptomatic carotid artery stenosis.

Key words:

1. Data Mining
2. Carotid Arteries
3. Chi-Square Distribution

***Corresponding Author:** Farzad Firouzi Jahantigh

E-mail: *firouzi@eng.usb.ac.ir*

doi: 10.18869/acadpub.shefa.4.4.1

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

بررسی عوامل مهم در شناسایی تنگی شریان کاروتید بدون علامت جهت پیشگیری از سکتهٔ مغزی با استفاده از ابزار داده کاوی

فرزاد فیروزی جهانتیغ[۱*]، راضیه علیزاده[۲]

[۱] گروه مهندسی صنایع، دانشگاه سیستان و بلوچستان، زاهدان، ایران

[۲] گروه مهندسی صنایع، دانشگاه پیام نور، تهران، ایران

اطلاعات مقاله:

تاریخ دریافت: ۲۳ شهریور ۱۳۹۴ **تاریخ پذیرش:** ۲۶ تیر ۱۳۹۵

چکیده

مقدمه: تنگی شریان کاروتید بدون علامت یکی از عواملی است که سکتهٔ مغزی ایجاد می‌کند. عوامل دیگر مانند فشار خون بالا، بیماری‌های قلبی، استعمال دخانیات، دیابت و عدم فعالیت بدنی ممکن است همچنین موجب بیماری شوند. فهمیدن و تشخیص عواملی که موجب تنگی شریان کاروتید می‌شوند در پیشگیری از سکتهٔ مغزی حاد، کمک خواهد کرد. با استفاده از روش‌های داده کاوی، این مطالعه به‌منظور کشف قوانین و روابطی که در شناسایی تنگی شریان کاروتید بدون علامت مؤثر هستند انجام شد. **مواد و روش‌ها:** برای پیدا کردن بهترین روش، رگرسیون لجستیک، الگوریتم ژنتیک و آزمون مجذور کای به‌منظور پیش‌بینی تنگی شریان کاروتید در بیماران استفاده شدند. **یافته‌ها:** ۳۷۲ شرکت‌کننده، ۱۷۳ زن (۴۶/۵ درصد) و ۱۹۹ مرد (۵۳/۵ درصد) با میانگین سنی ۵/۲۹ ± ۷۰/۷۴ مورد بررسی قرار گرفتند. نتایج نشان داد جنسیت، استعمال دخانیات، بیماری عروق کرونر، فشارخون بالا، بی تحرکی، پیشگیری از بارداری به وسیلهٔ دارو، اورمی (مقادیر بیش از حد اوره و سایر ترکیبات نیتروژنی در خون) و میزان نبض محیطی عوامل خطر معنی‌داری برای شریان کاروتید بدون علامت هستند. علاوه بر این الگوریتم ژنتیک در مقایسه با رگرسیون لجستیک یک روش بهتری برای این رویکرد بود. **نتیجه‌گیری:** مطالعهٔ ما نشان داد که بیماری عروق کرونر و فشار خون بالا عوامل مهمی در پیش‌بینی و پیش‌آگهی تنگی شریان کاروتید بدون علامت هستند.

کلید واژه‌ها:

۱. داده‌کاوی
۲. شریان‌های کاروتید
۳. توزیع مربع کای

* **نویسنده مسئول:** فرزاد فیروزی جهانتیغ

آدرس الکترونیکی: *firouzi@eng.usb.ac.ir*

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

مقدمه

سکتهٔ مغزی سومین علت شایع مرگ و میر پس از بیماری‌های عروق کرونر و سرطان و همچنین یکی از علل اصلی از کارافتادگی است (۱) و از طرفی مهم‌ترین علت ناتوانی‌های بلندمدت در کشورهای توسعه یافته است (۳-۱). حادثهٔ مغزی -عروقی معمولاً سکته (CVA)[1] نامیده می‌شود؛ که به معنای کاهش ناگهانی تأمین جریان خون مغز یا در اثر خونریزی مغزی انسان است و به وسیلهٔ اختلال در رگ‌های خونی ورودی مغز به وجود می‌آید. در نتیجه فقدان اکسیژن سبب می‌شود که بافت مغز آسیب ببیند یا حتی نابود شود (۲). سکتهٔ مغزی تأثیرات شدیدی بر کیفیت زندگی می‌گذارد و این اثرات به وسیلهٔ سازمان جهانی بهداشت تحت عناوین ضایعه، ناتوانی و معلولیت طبقه‌بندی می‌شود.

مطالعات نشان داده است که در ۱۵ میلیون حادثهٔ سکتهٔ مغزی که هر ساله در سراسر جهان رخ می‌دهد، ۵ میلیون منتج به مرگ و میر می‌شود و بیش از ۵ میلیون با عوامل عصبی دایمی و وابستگی زندگی می‌کنند. با گذشت هر دهه از زندگی، خطر سکتهٔ مغزی افزایش پیدا می‌کند و این امر باعث افزایش تعداد معلولیت‌ها نسبت به‌کل جمعیت خواهد بود (۴، ۳).

در کشور ایران در طول یک سال، ۴۳ هزار نفر دچار سکتهٔ مغزی می‌شوند، بر طبق آمار جهانی نیز به ازای هر ۱۰۰ هزار نفر حدود ۵۰ تا ۱۰۰ نفر دچار سکتهٔ مغزی می‌شوند، از آنجا که سکته‌های قلبی و مغزی به دنبال گرفتگی عروق ایجاد می‌شوند، از این رو در یک خانواده قرار می‌گیرند و بیماری آترواسکلروز یا تصلب شرایین اینجا معنا می‌گیرد (۵).

با توجه به مطالعات انجام شده، تنگی عروق کاروتید می‌تواند نقش مهمی در سکتهٔ مغزی داشته باشد که بررسی عوامل تأثیرگذار در تنگی عروق کاروتید به جهت کاهش سکتهٔ مغزی می‌تواند مؤثر باشد. عوامل متعددی برای تنگی کاروتید وجود دارد. به‌عنوان مثال فشارخون بالا، عوارض قلبی، مصرف دخانیات، دیابت، عدم تحرک همچنین سن و جنسیت نیز در ارتباط با خطر تنگی کاروتید مرتبط هستند (۷، ۶). دانستن و تشخیص عواملی که به تنگی کاروتید بدون علامت منجر می‌شود در طراحی روشی برای پیشگیری از آن کمک خواهد کرد و استفاده از روش‌های آماری و داده کاوی ممکن است به تعیین عوامل خطر تنگی کاروتید کمک قابل توجهی کند. امروزه استفاده از فنون داده کاوی در راستای مطالعه در زمینهٔ سلامت، قابل تأمل و قابل ملاحظه است (۹، ۸)

یکی از پرکاربردترین موارد استفادهٔ داده کاوی در زمینهٔ پزشکی می‌باشد که در سال‌های اخیر تحقیقات و مطالعات زیادی پیرامون آن انجام شده است. وقتی تعداد پارامترها در تشخیص بیماری زیاد باشد، ممکن است تشخیص بیماری حتی برای یک متخصص خبرهٔ پزشکی نیز به‌سختی امکان‌پذیر باشد. به همین دلیل در چند دههٔ اخیر ابزار تشخیص کامپیوتری با هدف کمک به پزشک مورد استفاده قرار می‌گیرد تا به نحوی بی‌نظمی را از داده‌ها خارج کند و همچنین خطاهای احتمالی ناشی از خستگی و یا بی‌تجربگی فرد خبره را کاهش دهد و داده‌های پزشکی را در زمان کمتر و با جزئیات بیشتر در اختیار پزشک قرار دهد. مؤسسهٔ فناوری ماساچوست[2] داده کاوی را یکی از ده فناوری برتری می‌داند که نقش چشمگیری در تحول جهان خواهد داشت (۱۰). امروزه بخش سلامت بیشترین نیاز را به داده کاوی پیدا کرده است و حرکت از پزشکی سنتی به سمت پزشکی مبتنی بر شواهد از جمله مواردی است که می‌تواند مؤکد این امر باشد (۱۱).

در مطالعهٔ انجام شده توسط شیرانی و همکارانش شیوع تنگی کاروتید قابل توجه (بیشتر از ۶۰ درصد) در ۱۰۴۵ بیمار کاندید بایپاس عروق کرونر با متوسط سنی ۶۰ سال را ۶/۹۹ درصد و در بیمارانی با سن بیشتر از ۶۰ سال ۱۲/۵ درصد گزارش کرده است (۱۲).

تحقیقات متعددی برای پیش‌بینی تنگی شریان کاروتید انجام شده است به طور مثال می‌توان به موارد زیر اشاره کرد:

صالحی عمران و همکارانش در ۱/۳ درصد از بیماران با درگیری کرونر، تنگی بیشتر از ۶۰ درصد کاروتید را گزارش کردند. گرچه در هیچ‌یک از بیماران با تنگی کاروتید پس از عمل قلب، حادثهٔ مغزی رخ نداد اما ۴/۱ درصد بیمارانی که درگیری کاروتید نداشتند پس از عمل، دچار حوادث عروقی -مغزی شدند. در این مطالعه، جنسیت مؤنث، فشارخون، بیماری عروق محیطی و سن، از عوامل خطر بروز تنگی کاروتید گزارش شدند. این محققین پیشنهاد کردند که داپلر قبل از عمل برای این‌گونه بیماران لازم است (۵).

تار زمانی و همکاران مطالعه‌ای را در بیمارستان شهید مدنی تبریز به هدف ارزیابی وضعیت کاروتید در بیماران کاندید عمل بایپس کرونر انجام دادند. شرایین کاروتید داخلی راست و چپ به ترتیب در ۲/ ۴۳و ۴۲/۱ درصد دارای پلاک بودند. در مجموع ۹/۲ درصد از بیماران، تنگی قابل توجه در کاروتید داشتند. در این مطالعه گرچه تشکیل پلاک، ارتباطی با جنسیت نشان نداده بود ولی در مردان، تنگی کاروتید چپ شایع‌تر دیده شد. همچنین وجود پلاک در هر دو کاروتید با افزایش سن، همراهی نشان داده بود. این محققین چنین استنباط

[1] Cerebral vascular accident
[2] Massachusetts institute of technology

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

کردند که در بیماران ایرانی کاندید عمل الکتیو بایپس کرونر[3]، شیوع اختلال کاروتید بسیار کم است (۱۳).

عوامل مختلف بیماری در این تحقیق عبارت‌اند از ۱- دیابت یا بیماری قند، یک اختلال متابولیک سوخت و سازی در بدن است ۲- هیپرلیپیدمی[4] یا هایپرلیپوپروتئینما به معنی افزایش چربی خون است ۳- سندرم اورمی، از عوارض نارسایی مزمن کلیه است، این سندرم بیانگر تأثیر افزایش اوره و کراتینین خون به ارگان‌های مختلف بدن است ۴- فشارخون بالا[5] یا پرفشاری خون که گاهی به آن پرفشاری شریانی گفته می‌شود، یک بیماری مزمن است که در آن فشار خون در شریان‌ها بالا می‌رود ۵- کم‌تحرکی که فقط به استراحت در تخت اطلاق نمی‌شود، در واقع اگر کمتر از پنج هزار قدم در روز یعنی کمتر از سه کیلومتر در روز پیاده‌روی می‌کنید، زندگی بی‌تحرکی دارید و در این صورت خطر بیماری‌ها شما را تهدید خواهد کرد۶ - چاقی وضعیت پزشکی است که بافت چربی بیش از حد طبیعی در بدن فرد انباشته شده باشد. شایع‌ترین روش برای تخمین چاقی استفاده از شاخص تودهٔ بدن (BMI)[6] است. شاخصی که با تقسیم کردن وزن شخص برحسب کیلوگرم بر توان دوم قد وی برحسب متر به دست میاید. طبق تعریف هنگامی‌که شاخص تودهٔ بدن از ۲۳۰ کیلوگرم بر متر بیشتر باشد، آن فرد چاق محسوب می‌شود ۷- داروهای خوراکی که مقدار مصرفی قرص‌ها، شربت‌ها و دیگر داروهایی که فرد مصرف می‌کند می‌باشد ۸- اورمی درنتیجهٔ سوخت و ساز مواد پروتئینی در بدن ماده‌ای سمی به نام اوره به وجود می‌آید که این ماده به وسیلهٔ کلیه‌ها دفع می‌شود. در اثر نارسائی کلیه‌ها این ماده خوب دفع نشده و میزانش در خون بالا می‌رود و شخص دچار اورمی می‌گردد (میزان طبیعی اورهٔ خون بین ۳۰ تا ۴۰ میلی‌گرم در ۱۰۰ سی‌سی است) ۹- بیماری متابولیک[7] به مجموعه‌ای از شرایط گفته می‌شود که عبارت هستند از: فشار خون بالا، سطوح افزایش‌یافتهٔ انسولین در خون، چربی انباشتهٔ اضافی در اطراف شکم و سطوح افزایش‌یافتهٔ چربی خون، بنا به گفتهٔ محققان سندروم متابولیک گروهی از عوامل خطر شامل چربی دور کمر، چاقی شکمی، فشارخون بالا، دیابت، تری گلیسرید بالا و HDL (چربی خوب) پایین است که با هم رخ می‌دهند و خطر ابتلاء به بیماری‌های قلبی، سکته و دیابت را افزایش می‌دهند.

هدف از این پژوهش شناسایی تنگی کاروتید با استفاده از روش‌های داده کاوی در مجموعهٔ داده‌های مورد استفاده است. علاوه بر این ما به مقایسهٔ روش‌های مختلف با هدف رسیدن به بهترین نتایج در مجموعهٔ داده‌ها می‌پردازیم. برای این منظور از روش‌های آماری در داده کاوی، الگوریتم ژنتیک (GA)[8]، رگرسیون لجستیک باینری[9] و آزمون مجذور کای برای مجموع داده‌های بیماران استفاده شده و در آخر هم برای هر روش نتایج نسبی ارائه شده است.

مواد و روش‌ها

این پژوهش از دیدگاه شیوه، توصیفی و از دیدگاه هدف، یک پژوهش کاربردی است که به صورت مقطعی[10] در بیمارستان امام علی زاهدان در سال ۱۳۹۳ انجام گرفته است. داده‌های مورد بررسی عوامل خطر در بیماری تنگی کاروتید هستند که عبارتند از عامل جنسیت، استعمال دخانیات، بیماری عروق کرونر، فشارخون بالا، کم‌تحرکی، پیشگیری از بارداری خوراکی، اورمی، نبض محیطی و بیماری‌های متابولیک که در این تحقیق مورد استفاده قرار گرفتند.

نحوهٔ گردآوری داده‌ها در تحقیق از طریق مصاحبه بوده و به روش تصادفی با ۳۷۲ نفر از افرادی که برای آزمایش‌های دوره‌ای به بیمارستان مراجعه کرده بودند صورت گرفته است. مصاحبه را یکی از روش‌هایی دانسته‌اند که امکان دریافت پاسخ در آن بیش از روش‌های دیگر است، زیرا در هنگام مصاحبه امکان تحریک آزمودنی برای دادن پاسخ وجود دارد و نیز می‌توان در صورت ابهام، با توضیح موضوع را روشن ساخت. شرایط ورود به مطالعه بر اساس افرادی که برای آزمایش‌های دوره‌ای در سال ۱۳۹۳ به بیمارستان امام علی زاهدان مراجعه کرده بودند و بدون علامت بیماری و بالای ۶۵ سال سن داشتند انتخاب شدند.

پس از آزمایش سونوگرافی داپلرکه برای بررسی اثر عوامل خطر سکتهٔ مغزی انجام شد پزشک از بیماران می‌خواهد تا سابقهٔ درمانی خود را در رابطه با عوامل خطر تنگی کاروتید بدون علامت ارائه دهند.

تجزیه و تحلیل آماری

در این تحقیق نحوهٔ نمایش داده‌ها از طریق جداول داده‌ها و توضیح در متن می‌باشد.

هرگاه احتمال وقوع اتفاقی به صورت تصادفی بسیار اندک بوده باشد و به عبارت دیگر به احتمال زیاد دارای دلیلی غیر از شانس بوده است، آن را به لحاظ آماری معنی‌دار گویند.

مقدار *P*: برابر است با حداکثر خطا در تصمیم‌گیری بر اساس آزمون فرضیه‌ها، به عبارت دیگر حداقل درجهٔ اطمینان به صحت آزمون برابر خواهد بود با *P*-۱. در گروه پزشکی معمولاً مقدار قابل قبول، *P*<۰/۰۵ لحاظ می‌شود، یعنی مقدار فوق در آزمون فرضیه‌ها بیانگر وجود اختلاف آماری، وجود رابطه یا همبستگی، وجود

[3] Coronary artery bypass
[4] Hyperlipidemia
[5] Hypertention
[6] Body mass index
[7] Metabolic syndrome
[8] Genetic algorithm
[9] Binary logistic regression
[10] Cross-sectional

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

تأثیر و غیره می‌باشد. (سطح معنی‌داری در این مطالعه ۰/۰۵ در نظر گرفته شده است).

عوامل تنگی عروق کاروتید به طور جداگانه با استفاده از توزیع مربع کای، رگرسیون لجستیک[11] و مدل‌های الگوریتم ژنتیک باینری و چند متغیره مورد بررسی قرار گرفت.

برای تجزیه و تحلیل اطلاعات در این تحقیق از برنامهٔ نرم‌افزاری SPSS ویرایش ۱۰/۱ و نرم‌افزارهای داده کاوی استفاده شد.

از آزمون توزیع مربع کای برای سنجش متغیرهای عامل جنسیت، استعمال دخانیات، بیماری عروق کرونر، فشارخون بالا، کم‌تحرکی، پیشگیری از بارداری خوراکی، اورمی و نبض محیطی استفاده شد. از رگرسیون لجستیک برای سنجش متغیرهای استعمال دخانیات، بیماری عروق کرونر و فشارخون بالا استفاده شد. از الگوریتم ژنتیک برای سنجش متغیرهای فشارخون بالا، کم‌تحرکی و بیماری عروق کرونر استفاده شد.

راه‌های متناوبی برای تقسیم داده‌ها در مجموعه‌ها وجود دارد و روشی که اینجا به کار برده‌ایم در خط تقاضای توزیع آماری است مانند تقسیم مستمر داده به دو دسته با مقدار متوسط آن. ما در حال حاضر عملیات و راهکارهای متناسب و مناسب را به شرح زیر یکپارچه کردیم: ۱- حساسیت بالا ۲- ویژگی عالی ۳-Roc[12] (AUC)[13] بزرگ.

سطح زیر منحنی Roc (AUC)، بیانگر مقدار پیش‌بینی سیستم از طریق توصیف توانایی آن در تخمین درست وقایع رخ داده و عدم تخمین آن است. ایده‌آل‌ترین مدل، بیشترین سطح زیر منحنی را دارد. AUC را از منحنی Roc می‌توان با معادلهٔ زیر اندازه‌گیری کرد، وقتی (ویژگی -۱) = t و Roc (t) حساسیت است:

$$AUC = \int_0^1 Roc\,(t)dt \quad (۱)$$

هدف اصلی شناسایی مؤثرترین عوامل برای تشخیص و پیش‌بینی بیماری است. در مؤثرترین پیش‌بینی، صحت پیش‌بینی از مدل همیشه در حال افزایش است.

در اینجا ما نمونه‌ای از شبه کد را برای الگوریتم هستهٔ ژنتیکی ارائه می‌کنیم:

مقداردهی اولیهٔ یک جمعیت از کروموزوم N به طور تصادفی (فرضیهٔ N)
محاسبهٔ سازگاری هر فرضیه
انتخاب بهترین ایفاء کننده (عملکرد)
درخواست عملگر جهشی و عملگر تقاطعی از کروموزوم N به کروموزوم top باید بررسی شود.
بررسی کنید آیا عملکرد بهبود یافته است اگر هیچ بهبودی در نسل گذشتهٔ X نیست متوقف کنید.

در غیر این صورت بازگشت به گام ۲

سیستم تعداد نمونه‌هایی که توسط این فرضیه تحت پوشش هستند را می‌شمارد (تعداد مثبت). منطق مشابه برای پیدا کردن تعداد منفی نیز عملی است که در تعداد کل نمونه‌های این فرضیه تحت پوشش نیستند. (کل نمونه منهای تعداد مثبت)؛ و مثبت کاذب که تعداد کل نمونه‌هایی است که فرضیه اشتباهاً مثبت فرض کرده بود. با استفاده از این مقادیر تمام فرضیه‌ها مورد بررسی قرار گرفتند تا بهترین فرضیه را که تابع برازندگی را بیشینه می‌سازد، پیدا کنیم. در این مورد حداکثر تعداد مثبت و حداقل تعداد منفی است.

با استفاده از الگوریتم ژنتیک، دو روش ساده معرفی شد:

روش اول: ۱ = STATE (بیماری): بهترین فرضیه برای توضیح دادن ۱= STATE، یعنی امکان ابتلاء به بیماری ۱= CAD است.

هنگامی‌که وجود بیماری عروق کرونر (CAD)[14] به‌عنوان شاخص ابتلاء یا وجود بیماری مورد بررسی قرار گرفت حساسیت آن ۰/۴۳، ویژگی ۰/۸۵ و ROC AUC، ۶۳/۸٪ بود.

روش دوم: ۰= STATE (سالم): بهترین فرضیه که عدم ابتلاء یا وجود بیماری را توضیح می‌دهد (۱ = STATE برای حالت بیماری) است: ۰ = CAD، ۰ = فشارخون بالا و ۰ = کم‌تحرکی.

حساسیت قانون دوم، ۰/۳۳ و ویژگی به تنهایی ۰/۹۲ و ROC AUC، ۶۲٪ است.

از این قوانین می‌توان برای پیش‌بینی بیماری‌ها استفاده کرد. به‌عنوان مثال استفاده از روش یک؛ هنگامی‌که یک بیمار جدید می‌رسد اگر او دارای سابقهٔ CAD باشد، او به احتمال زیاد به تنگی شریان کاروتید مجانبی با حساسیت و ویژگی بیان شده مبتلا است.

در تجزیه و تحلیل الگوریتم ژنتیک، اندازهٔ جمعیت به طور پیش فرض ۲۵۶= N (که تعداد کل داده‌ها است) و top-scoring، NTOP=۶۴ انتخاب شد. برای تولید نسل جدید از سه عملگر ژنتیکی مرسوم تقاطع، جهش و تولید مجدد استفاده می‌شود. عملگر جهش به طور تصادفی تغییر یک یا چند جنبه از فرضیه و عملگر تقاطع ادغام دو فرضیه به شکل دو فرضیهٔ جدید است. مقدار پیش فرض برای نرخ تقاطع ۶۰٪ و برای نرخ جهش ۲٪ در هر فرضیه بود.

مشخصات عملکردی دریافت‌کننده

برای مقایسهٔ این عملکردها با مدل‌های آماری دیگر، کسری از کل موارد ذکرشده، حساسیت، دقت، ناحیهٔ زیر منحنی (AUC) و مشخصات عملکردی دریافت‌کننده (ROC) ارائه شد.

[11] Logistic regression
[12] Receiver operating characteristic
[13] Area under the curve
[14] Coronary artery disease

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

یافته‌ها

از بین ۳۷۲ نفر شرکت‌کننده، ۱۷۳ نفر، زن (۴۶/۵ درصد) و ۱۹۹ نفر، مرد (۵۳/۵ درصد) بودند که میانگین سن شرکت‌کنندگان در این تحقیق ۵/۲۹ ± ۷۰/۷۴ بود. ویژگی‌های جمعیت‌شناختی و بالینی از شرکت‌کنندگان را می‌توان در جدول ۱ یافت.

جدول ۱- ویژگی‌های بیمار.

عوامل	بله (درصد) تعداد	خیر (درصد) تعداد
سابقهٔ بیماری دیابت	۱۰۲(۲۷/۴)	۲۷۰(۷۲/۶)
استعمال دخانیات	۱۴۵(۳۹)	۲۲۷(۶۱)
هیپرلیپیدمی	۱۹۶(۵۲/۸)	۱۷۵(۴۷/۲)
بیماری عروق کرونر	۱۲۴(۳۳/۳)	۲۴۸(۶۶/۷)
فشارخون بالا	۲۱۵(۵۷/۸)	۱۵۷(۴۲/۲)
زندگی کم‌تحرک	۱۱۷(۳۱/۵۷)	۲۵۵(۶۸/۵)
چاقی	۲۵۲(۶۷/۸)	۱۲۰(۳۲/۳)
داروهای خوراکی	۱۷(۴/۶)	۳۵۵(۹۵/۴)
اورمی	۲۶(۷)	۳۴۶(۹۳)
بیماری متابولیک	۱۳(۳/۵)	۳۵۹(۹۶/۵)

نتایج توزیع مربع کای

عوامل تنگی بدون علامت عروق کاروتید به طور جداگانه با استفاده از آزمون مجذور کای مورد بررسی قرار گرفت و مقدار سطح معنی‌داری در جدول ۲ ذکر شده است. نتایج آزمون مجذور کای نشان داد که عامل جنسیت، مصرف دخانیات، بیماری عروق کرونر فشارخون بالا، بی‌تحرکی، پیشگیری از بارداری به وسیلهٔ دارو، اورمی (وجود مقادیر زیاد اوره و دیگر ترکیبات نیتروژنی در خون) و سابقهٔ نبض محیطی از عوامل معنی‌دار برای شریان کاروتید بدون علایم بودند (سطح معنی‌داری در این مطالعه ۰/۰۵ در نظر گرفته شده است).

نتایج تجزیه و تحلیل رگرسیون لجستیک

متغیرهایی که در آن‌ها مقدار $P<0/05$ هستند در تجزیه و تحلیل تک متغیره قرار دارند. جنسیت ($P<0/05$) مصرف دخانیات ($P<0/05$)، بیماری عروق کرونر ($P<0/001$)، فشارخون بالا ($P<0/01$)، کم‌تحرکی ($P<0/05$)، پیشگیری از بارداری به وسیلهٔ دارو ($P<0/05$)، اورمی ($P<0/05$)، میزان نبض محیطی ($P<0/05$) انتخاب شدند و پس از آن توسط مدل رگرسیون لجستیک چند متغیره با استفاده از نسبت احتمال به روش گام به گام ارزیابی شد.

جدول ۲- نتایج توزیع مربع کای. * نشان‌دهندهٔ $P<0/05$، ** نشان‌دهندهٔ $P<0/01$ و *** نشان‌دهندهٔ $P<0/001$ می‌باشد.

متغیر	χ^2	سطح معنی‌داری
جنسیت	۵/۲۲	۰/۰۲۲*
سابقهٔ بیماری دیابت	۰/۰۱۲	۰/۹۰۳
استعمال دخانیات	۶/۲۳	۰/۰۱۳*
هیپرلیپیدمی	۲/۰۷۹	۰/۱۴۹
بیماری عروق کرونر (CAD)	۲۸/۵۷	۰/۰۰۰۱***
فشارخون بالا	۹/۴۰۱	۰/۰۰۲**
زندگی کم‌تحرک	۶/۲۹	۰/۰۱۲*
چاقی	۱/۸۶	۰/۳۹۵
قرص‌های خوراکی پیشگیری	۴/۸۴	۰/۰۲۸*
اورمی	۴/۲۶	۰/۰۳۹*
بیماری متابولیک	۰/۰۵۸	۰/۸۱۰
بیماری‌های خونی	۰/۰۴۸	۰/۸۲۶
سابقهٔ جراحی	۰/۱۰۶	۰/۷۴۵
نبض محیطی	۴/۳۹	۰/۰۳۶*
بیماری قلبی	۱/۶۱	۰/۲۰۴

با توجه به رگرسیون لجستیک، سن، بیماری عروق کرونر و سابقهٔ فشارخون بالا در پیش‌بینی تنگی عروق کاروتید بدون علامت مهم هستند با بررسی سطح زیر منحنی Roc (AUC)، که بیانگر مقدار پیش‌بینی سیستم از طریق توصیف توانایی آن در تخمین درست وقایع رخ داده است به این نتایج دست می یابیم: ۰/۷۲۸ ($P<0/001$، ۰/۷۸۲-۰/۶۷۵ :95%CI) که CI فاصلهٔ اطمینان[15] است.

عوامل مهم در جدول ۳ مطابق با نسبت شانس ((B) exponents(EXP) یا امید ریاضی فهرست شده است. به دلیل دشوار بودن گزارش B بهتر است مقدار Exp(B) به عنوان ضرایب رگرسیونی هر یک از متغیرهای وارد شده به مدل استفاده شود. آزمون استنباط (نیکویی برازش یا همان سنجش درستی مدل) همان مدل هوسمر لئیمززهاو[16] است که حاصل یک χ^2 از هر ۳/۴۴ داشت که یک مقدار جزئی بود نتایج به دست آمده نشان داد؛ جنسیت ($P=0/022$) مصرف دخانیات ($P=0/013$)، بیماری عروق کرونر ($P=0/0001$)، فشارخون بالا ($P=0/002$)، بی‌تحرکی ($P=0/012$)، پیشگیری از بارداری به وسیلهٔ دارو ($P=0/028$)، اورمی (وجود مقادیر زیاد اوره

[15] Confidence interval
[16] Hosmer–Lemeshow (H–L)

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

و دیگر ترکیبات نیتروژنی در خون)-(P=۰/۰۳۹) و سابقهٔ نبض محیطی (P=۰/۰۳۶)، عوامل مهم برای شریان کاروتید بدون علایم بودند (P>۰/۰۵) و توصیه می‌کند که مدل باید با داده‌ها به خوبی و به‌درستی مطابقت کند؛ به عبارت دیگر، فرضیهٔ صفر یک مدل خوب که با داده‌ها مطابقت دارد قابل دفاع است .مقایسهٔ GA با مدل‌های آماری دیگر در جدول ۴ آمده است (ROC ناحیهٔ بین ۰ و ۱ داده است).

جدول ۳- نمایش نتایج از تجزیه و تحلیل رگرسیون لجستیک چند متغیره برای تنگی بدون علامت عروق کاروتید.

متغیر	نمرهٔ والد	df[a]	سطح معنی‌داری	نسبت شانس	با فاصلهٔ اطمینان ۹۵ درصد	
					بالا	پایین
سن	۴/۹۱	۱	۰/۰۰۸	۰/۰۶۵	۱/۱۱	۱/۰۱
استعمال دخانیات	۵/۶۲	۱	۰/۰۱۸	۰/۵۴۴	۰/۹	۰/۳۲
بیماری عروق کرونر (۱)	۲۰/۲۹	۱	۰/۰۰۰۱	۰/۲۶۴	۰/۴۶	۰/۱۴
فشارخون (۱)	۱۰/۱۶	۱	۰/۰۰۱	۰/۴۵۵	۰/۷۳	۰/۲۸

جدول ۴- مقایسهٔ عوامل مهم برای روش‌های متعدد (n/e: غیر واجد شرایط).

متغیر	روش	AUC	۹۵٪CI	سطح معنی‌داری
جنس، سیگار کشیدن، بیماری عروق کرونر، فشارخون بالا، کم‌تحرکی، پیشگیری از بارداری خوراکی، اورمی، نبض محیطی	آزمون مربع کای	n/e	n/e	P<۰/۰۵
سن، استعمال دخانیات، بیماری عروق کرونر، فشارخون بالا	رگرسیون لجستیک	۰/۷۲۸	۰/۶۷-۰/۷۸	P=۰/۰۰۰۱
روش ۱ State (۱) CAD = ۱	الگوریتم ژنتیک	۰/۶۳۸	۰/۵۹-۰/۶۹	P<۰/۰۵
روش ۲ State (۰) CAD = ۰ فشارخون بالا = ۰، کم‌تحرکی = ۰	الگوریتم ژنتیک	۰/۶۲	۰/۵۵-۰/۶۸	P<۰/۰۵

بحث و نتیجه‌گیری

در همهٔ مدل‌های آماری در این پژوهش بیماری شریان‌های کرونری و عامل فشارخون بالا به‌عنوان شاخص‌های مهم تنگی عروق کاروتید بدون علامت، برای همهٔ بیماران سالمند بدون علامت با تشخیص‌های مختلف انتخاب شدند (۱۴، ۱۵)

رگرسیون لجستیک برای این گروه نشان داد که سن، مصرف دخانیات، بیماری شریان‌های کرونری و فشارخون بالا پارامترهای مهم آماری با ۰/۷۲۸ ارزش AUC هستند. با این حال، محاسبهٔ فرمول رگرسیون لجستیک در عمل روزانه پیچیده و سخت است. برای تفسیر کردن رگرسیون لجستیک می‌توان از دستگاه‌های پشتیبانی و تصمیم گیرندهٔ محاسباتی محور استفاده کرد اما از آن نمی‌توان یک روش ساده برای کمک به پزشکان فراهم کرد.

تجزیه و تحلیل توسط الگوریتم ژنتیک دو روش ساده برای پیش‌بینی تنگی مجرا به وجود می‌آورد. در این روش‌ها، ما حساسیت ۰/۴۳ و ویژگی ۰/۸۵ و ROC AUC ۰/۶۴ را به دست آوردیم. AUC در الگوریتم ژنتیک کمتر از مقداری است که توسط رگرسیون لجستیک به دست آمده است.

روش دو الگوریتم ژنتیک به طور خلاصه می‌تواند چنین بیان شود: اگر بیماری شریان‌های کرونری صفر و فشارخون بالا صفر و کم‌تحرکی صفر باشد در این صورت می‌توان به سلامتی کامل و عدم ابتلاء به بیماری رسید.

در مقایسه با رگرسیون لجستیک، الگوریتم ژنتیک به محاسبات بیشتر نیاز ندارد. استفاده از الگوریتم ژنتیک مزیت دیگری هم دارد. این سیستم امکان استفاده از عملیات و راهکارهای مختلف و مناسب در کار داده کاوی پزشکی دارد. اغلب پزشکان به تجهیزات مختلفی برای حساسیت‌های بیشتر یا صفات ویژه نیاز دارند و سیستم

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]
Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

داده کاوی الگوریتم ژنتیک، انعطاف‌پذیری کافی برای متناسب کردن راهکارها و عملیات مختلف بر روی این تجهیزات را دارد. همچنین همانند مطالعات اخیر، در این مجموعه داده‌ها، جنسیت به‌عنوان تنها عامل خطرساز ابتلاء، به وسیلهٔ توزیع مربع کای شناخته شد (۱۶).

در مطالعهٔ مشابه دیگر از سابقهٔ حوادث عروقی مغز، فشارخون، مصرف سیگار، سن بالای ۶۰ سال، جنس مؤنث و بیماری شدید عروق کرونر به‌عنوان عوامل خطر مرتبط به ایجاد تنگی قابل توجه کاروتید یاد شده است که با نتایج ما همخوانی دارد (۱۸، ۱۷).

Jacobowitz و همکاران ۳۹۴ فرد را به طور معمول تحت سونوگرافی داپلر عروق کاروتید قرار دادند و میزان شیوع تنگی بیشتر یا مساوی ۵۰ درصد را در جمعیت عادی ۶/۹ درصد بیان نمودند ضمناً در این مطالعه به ارزیابی تأثیر عوامل خطر قلبی عروقی از قبیل سیگار، فشارخون، بیماری‌های قلبی همراه و هیپرکلسترولمی بر بروز[17] موارد جدید تنگی در پیگیری چند ساله جمعیت فوق پرداختند و طبق نتایج گزارش شده میزان بروز بدون حضور عامل خطر ۸/۱ درصد، در صورت وجود یک عامل خطر ۸/۵ درصد و در حضور دو و سه عامل خطر به ترتیب به ۵/۱۳ و ۷/۱۶ درصد افزایش می‌دهد (۲۰، ۱۹).

Masabni و همکارانش به بررسی این موضوع پرداختند که «آیا غربالگری تنگی کاروتید بعد از عمل باعث کاهش سکتهٔ مغزی ضمن عمل در بیماران تحت پیوند عروق کرونر می‌شود یا خیر؟» و نشان دادند که آزمایش قبل از عمل شریان کاروتید، با شناسایی بیماران مبتلا به تنگی شریان کاروتید حاد و تغییر مدیریت آن‌ها، عوارض عصبی و روانی بعد از عمل را در بیمارانی که تحت عمل CABG[18] هستند، کاهش می‌دهد (۲۱).

به نظر می‌رسد این مطالعه اولین مطالعه است که به بررسی استفاده از الگوریتم ژنتیک در پیش‌بینی تنگی کاروتید بدون علامت پرداخته است. خواستهٔ ما تولید کردن روش ساده است که می‌تواند توسط پزشکان بیان و بحث شود. این نتایج اولیه دلگرم کننده هستند و مطالعات بعدی با گروه بزرگ‌تر بیمار می‌تواند در نظر گرفته شود. ما روش‌های تجزیه و تحلیل داده‌های تک متغیره و چند متغیره را در یک کار داده کاوی پزشکی به کار بردیم. دو روش چند متغیره به نتایج بهتر از روش‌های تک متغیره رسید. رگرسیون لجستیک بازده ROC AUC بزرگتری نسبت به الگوریتم ژنتیک دارد. با این حال الگوریتم ژنتیک جایگزین خوبی برای رگرسیون لجستیک است زیرا الگوریتم ژنتیک روش‌های ساده‌ای را تولید می‌کند که می‌تواند در عمل روزانه استفاده شود.

روش الگوریتم ژنتیک باید برای پیش‌بینی بیماری شریان‌های کرونری بررسی‌های بیشتری داشته باشد. همان‌طور که مطالعات ما نشان می‌دهد بیماری شریان‌های کرونری و فشارخون بالا شاخص‌های مهم برای پیش‌بینی و پیش‌آگهی از تنگی شریان کاروتید بدون علامت هستند. ما بر این باوریم مطالعات شبیه به این می‌تواند برای کمک به سازمان بهداشت و درمان برای تولید دستورالعمل‌های بهتر و به‌عنوان یک پایگاه دانش برای سازمان مورد استفاده قرار گیرد.

منابع

1. Devroey D, Buntinx F, Van Casteren V. Trends in stroke attack rates and mortality form a 16 year nationwide registration. Am Heart Assoc. 2002; 106: S 752.

2. Vrdoljak D, Rumboldt M. Quality of life after stroke in croatian patients. Coll Antropol. 2008; 32(2): 355-9.

3. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al. Executive summary: heart disease and stroke statistics-2012 update a report from the American heart association. Circulation. 2012; 125(1): 188-97.

4. Murabito JM, Evans JC, Larson MG, Nieto K, Levy D, Wilson PW. The ankle-brachial index in the elderly and risk of stroke, coronary disease, and death: the framingham study. Arch Intern Med. 2003; 163(16): 1939-42.

5. Salehiomran A, Shirani S, Karimi A, Ahmadi H, Marzban M, Movahedi N, et al. Screening of carotid artery stenosis in coronary artery bypass grafting patients. J Tehran Heart Cent. 2010; 5(1): 25-8.

6. Inzitari D, Eliasziw M, Gates P, Sharpe BL, Chan RK, Meldrum HE, et al. The causes and risk of stroke in patients with asymptomatic internal-carotid-artery stenosis. N Engl J Med. 2000; 342(23): 1693-700.

7. Mathiesen EB, Joakimsen O, Bønaa KH. Prevalence of and risk factors associated with carotid artery stenosis: the Tromsø Study. Cerebrovasc Dis. 2001; 12(1): 44-51.

8. Wang B, Wong CM, Wan F, Mak PU, Mak P-I, Vai MI. Trial pruning based on genetic algorithm for single-trial EEG classification. Comput Electr Eng. 2012; 38(1): 35-44.

9. Austin PC, Lee DS, Steyerberg EW, Tu JV. Regression trees for predicting mortality in patients with

[17] Incidence
[18] Coronary artery bypass grafting

www.SID.ir

Archive of SID

[DOI: 10.18869/acadpub.shefa.4.4.1]

Downloaded from shefayekhatam.ir at 11:18 IRST on Saturday February 18th 2017

cardiovascular disease: What improvement is achieved by using ensemble-based methods? Biom J. 2012; 54(5): 657-73.

10. moghadasi H, Hoseini A, Asadi F, jahanbakhsh M. Data mining and its applications in health. Health Inf Manag. 2012; 9(2): 297-304.

11. Cheng T-H, Wei C-P, Tseng VS. Feature selection for medical data mining: comparisons of expert judgment and automatic approaches. 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06). 2006; p. 165-70.

12. Shirani S, Boroumand MA, Abbasi SH, Maghsoodi N, Shakiba M, Karimi A, et al. The role of carotid artery screening before coronary artery bypass graft surgery. J Tehran Heart Cent. 2006; 1(1): 29-32.

13. Tarzamni MK, Afrasyabi A, Farhoodi M, Karimi F, Farhang S. Low prevalence of significant carotid artery disease in Iranian patients undergoing elective coronary artery bypass. Cardiovasc Ultrasound. 2007; 5(3): 1 Doi: 10.1186/1476-7120-5-3.

14. Pandian JD. Recent concepts in the management of extracranial carotid stenosis: carotid endarterectomy versus carotid artery stenting. Neurol India. 2011; 59(3): 376-82.

15. Er O, Tanrikulu AC, Abakay A, Temurtas F. An approach based on probabilistic neural network for diagnosis of Mesothelioma's disease. Comput Electr Eng. 2012; 38(1): 75-81.

16. Rockman CB, Jacobowitz GR, Gagne PJ, Adelman MA, Lamparello PJ, Landis R, et al. Focused screening for occult carotid artery disease: patients with known heart disease are at high risk. J Vasc Surg. 2004; 39(1): 44-50.

17. Berens ES, Kouchoukos NT, Murphy SF, Wareing TH. Preoperative carotid artery screening in elderly patients undergoing cardiac surgery. J Vasc Surg. 1992; 15(2): 313-23.

18. Sabeti S, Schillinger M, Mlekusch W, Willfort A, Haumer M, Nachtmann T, et al. Quantification of internal carotid artery stenosis with duplex us: comparative analysis of different flow velocity criteria 1. Radiology. 2004; 232(2): 431-9.

19. Jacobowitz GR, Rockman CB, Gagne PJ, Adelman MA, Lamparello PJ, Landis R, et al. A model for predicting occult carotid artery stenosis: screening is justified in a selected population. J Vasc Surg. 2003; 38(4): 705-9.

20. Neale ML, Chambers JL, Kelly AT, Connard S, Lawton MA, Roche J, et al. Reappraisal of duplex criteria to assess significant carotid stenosis with special reference to reports from the North American symptomatic carotid endarterectomy trial and the european carotid surgery trial. J Vasc Surg. 1994; 20(4): 642-9.

21. Masabni K, Raza S, Blackstone EH, Gornik HL, Sabik JF. Does preoperative carotid stenosis screening reduce perioperative stroke in patients undergoing coronary artery bypass grafting? J Thorac Cardiovasc Surg. 2015; 149(5): 1253-60.

www.SID.ir