مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره چهارم شماره اول. ص ۱۰-۱، ۱۳۹۶

طراحی سیستم پیش‌بینی بیماری قلبی- عروقی با استفاده از ماشین بردار پشتیبان

مریم سادات محمودی[1]*

• **دریافت مقاله:** ۹۶/۲/۱۱ • **پذیرش مقاله:** ۹۶/۳/۲۷

مقدمه: بیماری‌های قلبی- عروقی بزرگ‌ترین عامل مرگ‌ومیر در سراسر جهان است. طبق اعلام سازمان بهداشت جهانی در هر سال حدود ۱۲ میلیون مرگ‌ومیر در سراسر جهان بر اثر بیماری‌های قلبی -عروقی تخمین زده شده است. هدف اصلی از این مقاله، طراحی یک سیستم هوشمند به کمک کامپیوتر است که بتواند بیماری قلبی را در فرد مبتلا تشخیص دهد.

روش: در این مطالعه توصیفی- تحلیلی، داده‌های مربوط به ۲۷۰ نفر که شامل ۱۳ ویژگی می‌باشد مورد استفاده قرار گرفت. جهت تشخیص فرد مبتلا به بیماری قلبی از ترکیب سیستم فازی و طبقه‌بندی کننده ماشین بردار پشتیبان با استفاده از امکانات موجود در نرم افزار متلب پیاده‌سازی گردید و بر روی سیستم Core i5 تحت ویندوز ۷ شبیه‌سازی شد.

نتایج: در این تحقیق از تکنیک فازی و الگوریتم ماشین بردار پشتیبان جهت تشخیص بیماری قلبی استفاده شده که مؤثر واقع گردیده است. از آنجایی که در این مطالعه هدف، تشخیص درست و سریع می‌باشد؛ بنابراین تشخیص سریع، شانس نجات فرد را افزایش می‌دهد. همچنین معیارهای ارزیابی در این سیستم نرخ دسته‌بندی و حساسیت می‌باشد که عملکرد این سیستم بر اساس این شاخص‌ها به ترتیب ۸۵٪ و ۸۵/۸٪ به دست آمده است.

نتیجه‌گیری: با توجه به نتایج به دست آمده، مشاهده می‌شود که سیستم پیشنهادی با دقت نسبتاً بالایی، افراد مبتلا به بیماری قلبی- عروقی را تشخیص می‌دهد.

کلید واژه‌ها: بیماری قلبی- عروقی، سیستم فازی، طبقه‌بندی کننده ماشین بردار پشتیبان

• **ارجاع:** محمودی مریم‌سادات. طراحی سیستم پیش‌بینی بیماری قلبی- عروقی با استفاده از ماشین بردار پشتیبان. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۶؛ ۴(۱): ۱۰-۱.

۱. کارشناس ارشد علوم کامپیوتر، مربی، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور بیرجند، بیرجند، ایران.

* **نویسنده مسئول:** خراسان جنوبی، بیرجند، دانشگاه پیام نور مرکز بیرجند.

• **شماره تماس:** ۰۹۱۳۹۵۴۵۹۳۹

• **Email:** m_mahmoodi_64@yahoo.com

مقدمه

بیماری‌های قلبی از شایع‌ترین بیماری‌های مزمن و علت مرگ بزرگسالان در سراسر دنیا می‌باشد. وزارت بهداشت، درمان و آموزش پزشکی اعلام نموده است که ۳۳ تا ۳۸ درصد مرگ و میرها در کشور ناشی از بیماری‌های قلبی – عروقی است و ایران بالاترین آمار مرگ قلبی در جهان را دارا می‌باشد [۱،۲]. تغییر در سبک زندگی مردم باعث افزایش شیوع بیماری‌های قلبی در ایران است. شواهد موجود در تغییر سبک زندگی مردم نشان می‌دهد که شیوع بیماری قلبی– عروقی در ایران رو به افزایش است. تخمین زده می‌شود که در سال ۲۰۲۰، مرگ و میر ناشی از این بیماری‌ها به ۲۵ میلیون نفر افزایش یابد [۳].

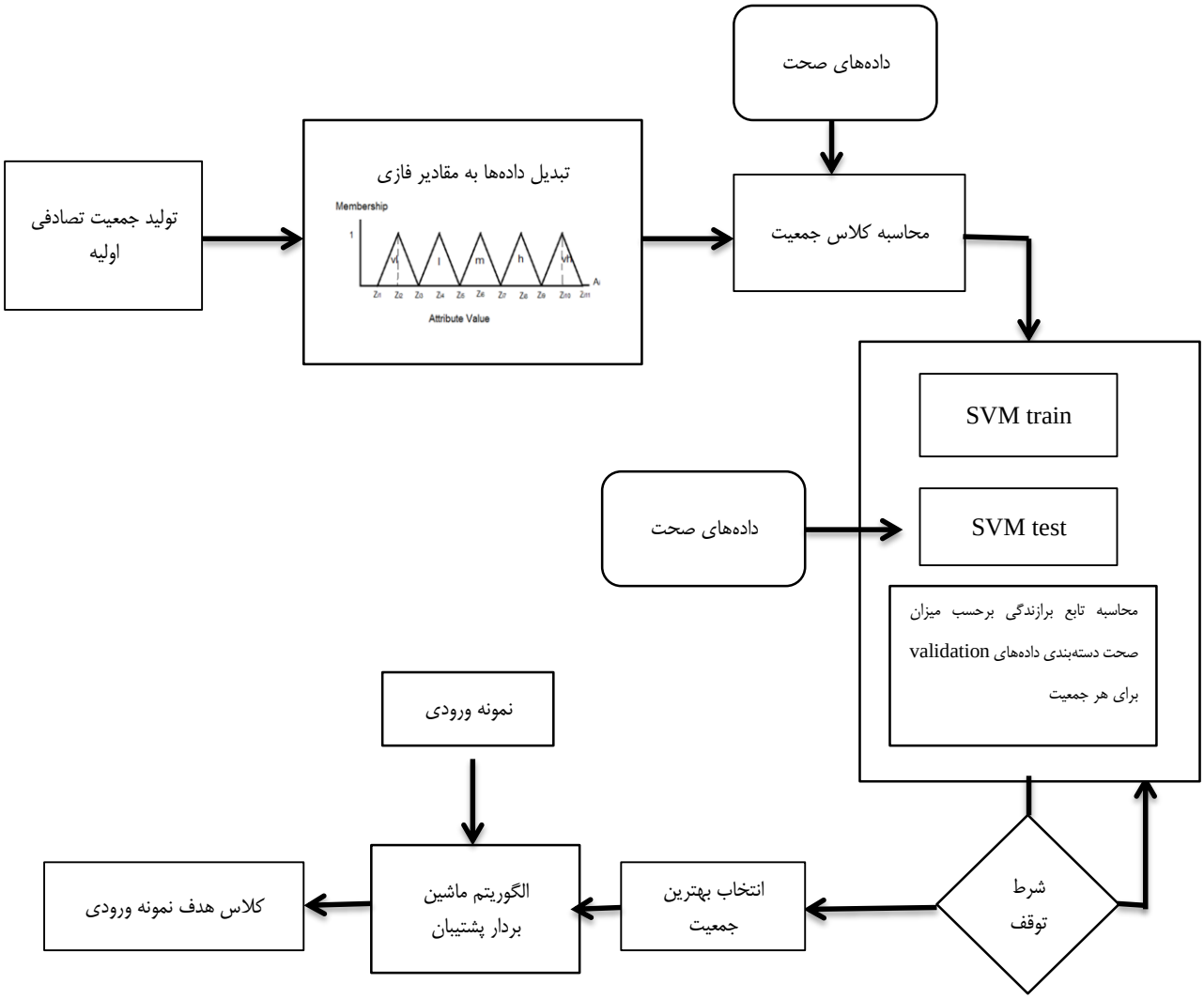
عوامل بسیاری در ایجاد، پیشرفت و شدت بیماری‌های قلبی تأثیر دارند. این عوامل را می‌توان به ۶ دسته تقسیم‌بندی نمود که عبارت‌اند از: عوامل زیستی و محیطی (مانند سن، جنسیت و استرس)، عادات زندگی (مانند ورزش، تغذیه، رعایت اصول بهداشتی)، عوامل خطر (مثل مصرف دخانیات، بالابودن سطح کلسترول خون، چاقی مفرط)، بیماری‌های زمینه‌ای (مثل بیماری ریوی انسدادی مزمن، عفونت، بیماری کلیوی مزمن، کم خونی، آرتریت)، عوامل روحی و شخصیتی (مثل خلق، مهارت‌های تطابقی، خودکارآمدی) و عوامل اجتماعی (مانند طبقه اجتماعی، داشتن خانواده و برخورداری از حمایت آن‌ها) [۴]. این علل به عنوان فاکتورهای خطر برای پیش‌بینی بیماری قلبی محسوب می‌شوند. بیش‌ترین نوع معمول از بیماری‌های قلبی، بیماری مادرزادی قلبی، نارسایی قلبی، فشارخون بالا، کاردیومیوپاتی، بیماری روماتیسمی قلب، تنگی ریوی و عروق کرونر بیماری عروق است. اگر عملکرد قلب به درستی انجام نشود و هر کدام از ناراحتی‌های قلبی بروز داده شود به بخش‌های دیگر بدن نیز تأثیر می‌گذارد و بنابراین لازم است که این بیماری سریع تشخیص داده شود [۵].

در سال‌های گذشته، پژوهش‌های زیادی در این زمینه انجام شده است. Banu و Jamala در [۶] از الگوریتم خوشه‌بندی فازی c-means جهت خوشه‌بندی بیماری قلبی استفاده شده است. Krishnaiah و همکاران [۷] از ترکیب سیستم‌های فازی و الگوریتم دسته‌بندی k-نزدیک‌ترین همسایه استفاده نموده است. میهمی و همکاران مطالعه‌ای تحت عنوان "طراحی یک سیستم خبره جهت شناسایی بیماری‌های قلبی با استفاده از الگوریتم فازی" انجام دادند و با کمک منطق فازی و با در نظر گرفتن ریسک فاکتورهای تأثیرگذار به پیش‌بینی بیماری قلبی پرداختند [۸]. در این سیستم، معیارهایی که برای تشخیص بیماری‌های قلبی توسط پزشک استفاده می‌شود به عنوان قوانین طراحی شده سیستم، جهت تصمیم‌گیری به سیستم داده شده و در مرحله بعد، این قوانین روی ریسک فاکتورها اعمال می‌شود. Suganya و Tamije Selvy جهت حذف داده‌های غیرقطعی، از سیستم فازی استفاده نمودند و یک رویکرد کارآمد برای استخراج الگوها از انبار داده‌های بیماران قلبی برای پیش‌بینی از حمله قلبی معرفی کرده‌اند [۹]. در مطالعه انجام شده توسط خسروانیان و آیت به منظور پیش‌بینی بیماری عروق کرونری قلب از شبکه عصبی استفاده شده است. آن‌ها در مطالعه خود نشان داده‌اند که شبکه‌های عصبی احتمالی، بهتر و قوی‌تر از سایر شبکه‌های عصبی در تشخیص بیماری عمل کرده‌اند [۱۰].

هدف از انجام تحقیق حاضر، طراحی سیستم هوشمند به کمک کامپیوتر جهت تشخیص فرد مبتلا به بیماری قلبی توسط سیستم فازی و الگوریتم ماشین پشتیبان است که در نهایت سیستم کامپیوتری تشخیص دهد که که فرد به بیماری قلبی مبتلا است یا خیر؟

روش

این مطالعه از نوع توصیفی–تحلیلی می‌باشد. در این بخش از کار که به عنوان روش کار و پیاده‌سازی شناخته می‌شود سعی شده با ارائه یک الگوریتم مناسب و اجرای آن بر روی داده‌ها هرچه بهتر افراد مبتلا به بیماری قلبی را تشخیص دهد. در ادامه فلوچارت الگوریتم پیشنهادی آمده است و به شرح مراحل پرداخته شده است (شکل ۱).

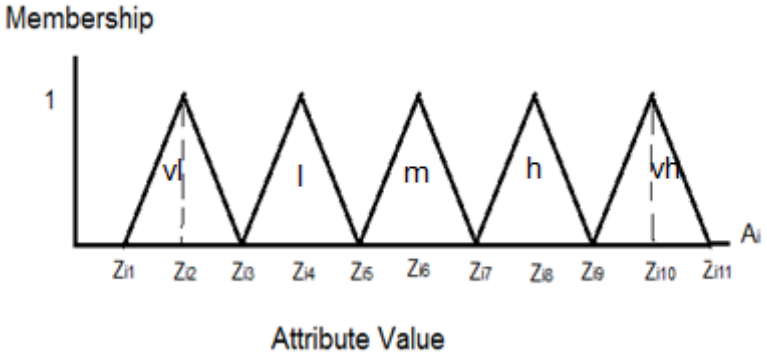


شکل ۱: فلوچارت مدل پیشنهادی

تبدیل داده‌های حقیقی به مقادیر فازی

در فرآیند تصمیم‌گیری برای تسهیل و درک بهتر کاربر از سطح مقدار یک صفت نسبت به دامنه مقادیر آن صفت، می‌توان اعداد حقیقی را به مقادیر زبانی تبدیل کرد. در این صورت کاربر درک شفاف‌تری از سطح مقدار یک صفت نسبت به دامنه مقادیر آن صفت خواهد داشت. در حقیقت بیان مقدار خصیصه‌ها به صورت مقادیر فازی به کاربر کمک می‌کند تا بدون توجه به مقادیر دامنه متغیر، به صورت تقریبی در مورد مقدار آن قضاوت کند. ساده‌ترین روش برای تبدیل اعداد حقیقی به مقادیر فازی، استفاده از نظرات فرد خبره است. این روش همیشه مقدور نیست، چرا که فرد خبره همیشه در دسترس نیست و ضریب اشتباهات نیز افزایش می‌یابد. روش دیگر استفاده از توابع عضویت می‌باشد. توابع عضویت متعددی از جمله توابع عضویت مثلثی و ذوزنقه‌ای وجود دارند که در اینجا از توابع عضویت مثلثی استفاده شده است.

در توابع عضویت مثلثی، دامنه هر متغیر (ویژگی) ورودی (A) به K بازه مثلثی منظم تقسیم می‌شود و هر بازه نشان‌دهنده یک مقدار زبانی است. این مقادیر زبانی عبارت‌اند از: very low (vl)، low (l)، medium (m)، high (h) و very high (vh)؛ بنابراین در این پژوهش مقدار K برابر ۵ در نظر گرفته شده است (شکل ۲).



شکل ۲: تابع عضویت فازی استفاده شده برای هر ویژگی

برای نمایش هر متغیر ورودی به ۱۱ نقطه عضویت Zi1، Zi2، ...، Zi11، نیاز می‌باشد، به طوری که Zi1 و Zi11 به ترتیب کم‌ترین و بیش‌ترین مقدار در هر متغیر ورودی است و بقیه نقاط در فواصل مساوی از هم قرار دارند.

در اینجا برای نمایش توابع عضویت هر ویژگی $\left(mf_{x_{ij}}\right)$ به شماره ضلع متغیر زبانی مربوطه نیاز است. اگر Ai کمتر یا مساوی با Zi2 باشد، مقدار تابع عضویت آن ویژگی (P) برابر ۱ می‌شود. اگر Ai کمتر یا مساوی با Zi3 باشد، مقدار تابع عضویت آن ویژگی، برابر ۲ شده و... در نهایت اگر Ai کمتر یا مساوی با Zi11 باشد، مقدار تابع عضویت آن، ۱۰ می‌شود؛ بنابراین مقدار تابع عضویت اولین ویژگی یکی از مقادیر ۱ تا ۱۰ می‌باشد. به همین صورت این مقدار برای تمامی ویژگی‌های داده‌ها محاسبه شده و همه داده‌ها به مقادیر فازی آن تبدیل می‌گردد.

در ادامه جهت محاسبه کلاس هر داده جدید، یک سطر داده تست با تمام سطرهای داده‌های آموزش مقایسه می‌شود و توابع عضویت جدیدی را محاسبه می‌نماید. بدین صورت که مقدار تابع عضویت اولین ویژگی داده تست $\left(mf_{t_{11}}\right)$ که در مرحله قبل محاسبه شده را با مقدار تابع عضویت ویژگی اول از اولین داده آموزش $\left(mf_{l_{11}}\right)$ مقایسه می‌کند و مقدار مشابهت را به عنوان یک مقدار عضویت بین ۰ تا ۱ ثبت می‌نماید$(S_{l_{11},t_{11}})$. این تابع روی دومین ویژگی داده تست و دومین ویژگی همان داده آموزش تا آخرین ویژگی داده تست و آموزش تکرار می‌شود. مقدار مشابهت بین ویژگی j ام از i امین داده آموزش $\left(l_{ij}\right)$ با ویژگی q ام از p امین داده تست$\left(t_{pq}\right)$ از رابطه ۱ محاسبه می‌گردد.

$$S\left(l_{ij},t_{pq}\right)=\frac{mf_{l_{ij}}-mf_{t_{pq}}}{mf_{t_{pq}}-10} \qquad (۱)$$

مقدار شباهت نهایی کل مشخصات سطر ۱ آموزش با سطر ۱ تست به صورت یک تابع عضویت بین ۰ تا ۱ به شکل زیر محاسبه می‌شود (رابطه ۲).

$$Mf(l_1,t_1)=\min\left[S_{l_{11},t_{11}},S_{l_{12},t_{12}},S_{l_{13},t_{13}},\dots,S_{l_{1m},t_{1m}}\right] \qquad (۲)$$

که در آن m تعداد ویژگی‌های هر داده است.

بدین ترتیب $Mf(l_{i=1:n},t_1)$ که بیان کننده همه مشابهت‌ها است را برای کلیه داده‌های آموزش باید محاسبه نمود.

(۳)

خروجی مربوط به کلاس اولین داده تست، کلاس داده آموزشی است که بیشترین مشابهت را دارد (رابطه ۳).

$$Out_{Class\ t_1}=Class\left\{Max_{i=1\ to\ n}\left[Mf(l_i,t_1)\right]\right\}$$

که در آن n، تعداد کل داده‌های آموزشی می‌باشد.

به همین صورت خروجی مربوط به کلیه داده‌های تست محاسبه می‌شود (رابطه ۴).

$$Out_{Class\ t_j}=Class\left\{Max_{i=1\ to\ n}\left[Mf(l_i,t_j)\right]\right\} \qquad (۴)$$

الگوریتم ماشین بردار پشتیبان

ماشین‌های بردار پشتیبانی (Support Vector Machine) SVMها از دقیق‌ترین و نیرومندترین الگوریتم‌های داده‌کاوی به شمار می‌روند. ماشین‌های بردار پشتیبان، در ابتدا توسط Vapnik در دهه ۹۰ میلادی توسعه داده شدند[۱۱].

SVMها به خانواده‌ای از مدل‌های خطی تعمیم‌یافته تعلق دارند. در این خانواده از مدل‌ها، بر اساس مقدار مربوط به ترکیب خطی خصیصه‌ها تصمیماتی در مورد دسته‌بندی و رگرسیون اتخاذ می‌شود. همچنین، گفته می‌شود که ماشین‌های بردار پشتیبان به روش‌های کرنل تعلق دارند.

SVMها، علاوه بر دارا بودن یک شالوده ریاضیاتی منسجم در تئوری یادگیری آماری، عملکرد بسیار خوب و موفقیت‌آمیزی را در کاربردهای عملی و واقعی از خود به نمایش گذاشته‌اند.

هدف ماشین‌های بردار پشتیبان دسته‌بندی دودویی با پیدا کردن بهترین خط از میان خطوط جداکننده با بیشینه نمودن حاشیه می‌باشد و تنها در زمانی که داده‌ها دو دسته باشند می‌توان از این روش استفاده نمود، زیرا در کل به‌صورت دودویی عمل نموده و قابلیت دسته‌بندی به دو گروه را دارد.

به همین منظور در ادامه روش پیشنهادی از الگوریتم جستجوی هارمونی و الگوریتم ماشین بردار پشتیبان جهت دسته‌بندی داده‌ها استفاده می‌شود. بدین صورت که ابتدا داده‌ها به سه دسته داده آموزش، تست و صحت تقسیم شده و با استفاده از این داده‌ها و تکرار الگوریتم هارمونی، ماشین بردار پشتیبان آموزش می‌بیند و سپس به دسته‌بندی داده‌ها می‌پردازد.

در ادامه مراحل بالا مورد بررسی قرار می‌گیرد:

۱- تولید جمعیت اولیه با مقادیر تصادفی

ابتدا N جمعیت اولیه، هر کدام به تعداد Npop کروموزوم با مقادیر تصادفی را تولید نموده و بر حسب شکل (۲) به مقادیر فازی مربوطه تبدیل می‌کند. سپس کلاس هر داده توسط رابطه (۴) و با استفاده از داده‌های صحت محاسبه می‌شود. بدین ترتیب جمعیت اولیه‌ای از کروموزوم‌ها برحسب داده‌های فازی تولید می‌گردد.

۲- محاسبه تابع برازندگی برحسب میزان صحت دسته‌بندی داده‌های validation برای هرجمعیت

در این مرحله میزان برازندگی هرجمعیت به کمک الگوریتم SVM ارزیابی می‌شود. بدین صورت که به کمک هر کدام از جمعیت‌ها و کلاس‌های نظیر آن‌ها، آموزش داده می‌شود و با کمک داده‌های validation مشخص می‌گردد که چه میزان از نمونه‌ها بر اساس تابع هدف درست دسته‌بندی شده‌اند؛ این میزان، دقت SVM را مشخص می‌کند و این دقت همان مقدار تابع برازندگی است.

۳- عملگر HMCR (Harmony Memory Considering Rate) و PAR (Pitch Adjusting Rate)

برای انجام عمل باز ترکیب دو جمعیتی را که دارای بهترین میزان برازندگی می‌باشند را به عنوان والد در نظر گرفته شده سپس فرزندان دو جمعیت را از نقطه برش (تعداد فرزندان) در دو جمعیت جا به جا می‌نماید.

در عملگر جهش (PAR) که با احتمال α صورت می‌پذیرد، بدین گونه عمل می‌گردد که شماره کلاس یک فرد را با شماره کلاس دیگری از مجموعه کلاس‌ها تعویض می‌شود.

٤- توقف الگوریتم

مجموعه فضای جمعیت را تا بدان جا به‌روز نموده که یا میزان دقت از یک حد آستانه مطلوب بیشتر شود و یا به تعداد تکرار مشخص برسد. در الگوریتم پیشنهادی، از تعداد ثابت تکرارها استفاده شده است.

۵- انتخاب بهترین جمعیت بر اساس بالاترین مقدار برازندگی

بعد از توقف الگوریتم، بهترین جمعیت که دارای بالاترین مقدار تابع برازندگی است را به عنوان جمعیت بهینه در جهت آموزش به الگوریتم SVM برای دسته‌بندی کردن داده‌های تست انتخاب می‌گردد.

در ماشین بردار پشتیبان، از تابع کرنل RBF (Radial Basis Function) (رابطه ۵) [۱۲] استفاده شده است. در این رابطه X_i و X_j به ترتیب بردار پشتیبان و رکورد داده تست می‌باشد. همچنین مقدار σ برابر ۱۰ در نظر گرفته شده است.

$$K(X_i, X_j) = exp(-\frac{\|X_i - X_j\|^2}{2\sigma^2}) \qquad (۵)$$

نتایج

در این تحقیق به منظور انجام مقایسه‌ای عادلانه برای تشخیص بیماری قلب، از مجموعه داده‌های یکسان در آموزش الگوریتم ماشین بردار پشتیبان استفاده شده است. برای این منظور از مجموعه داده Heart که از مخزن UCI (University of California, Irvine) [۱۳] که مرجعی برای یادگیری ماشین می‌باشد، انتخاب گردیده است. مجموعه داده Heart شامل ۱۳ ویژگی و ۲ کلاس می‌باشد. مجموعه کلاس‌ها به دو گروه مبتلا به بیماری قلبی و سالم تقسیم می‌شود که کلاس اول

شامل ۱۵۱ نفر افراد سالم و کلاس دوم شامل ۱۱۹ نفر بیمار می‌باشند. جدول ۱، ۱۳ ویژگی بیماری قلب را نشان می‌دهد.

جدول ۱: ویژگی‌های بیماری قلب

ویژگی	دسته‌بندی	
سن بیمار	مقدار پیوسته	
جنسیت	زن	۱
	مرد	۰
درد قفسه سینه (chest pain)	نوع معمولی	۱
	نوع معمولی آنژین	۲
	بدون آنژین	۳
	بدون علامت	۴
فشارخون در حال استراحت (Rest BP)	مقادیر پیوسته	
کلسترول سرم	مقدار پیوسته در mm hg	
قند خون ناشتا	بیشتر از ۱۲۰	۱
	کمتر از ۱۲۰	۰
نتیجه الکتروکاردیوگرافی	مقادیر ۰،۱ و ۲	
حداکثر ضربان قلب	مقادیر پیوسته	
آنژین ناشی از فعالیت	خیر	۱
	بله	۲
Old peak	مقادیر پیوسته	
Slope	شیب به بالا	۱
	مسطح	۲
	شیب به پایین	۳
تعداد عروق اصلی رنگی شده توسط فلوروسکوپی	مقادیر ۱-۴	
Thal(اسکن تالیوم)	نرمال	۳
	نقص ثابت	۶
	نقص برگشت‌پذیر	۷

متغیر هدف در این مطالعه، وجود یا عدم وجود بیماری قلبی است که در مورد هر کدام از افراد مورد بررسی یکی از این دو حالت ثبت گردیده است. مقدار این متغیر برابر با ۱، نشان‌دهنده وجود بیماری قلبی و صفر، نشان‌دهنده عدم وجود بیماری می‌باشد. به طور کلی برای بررسی میزان موفقیت و کارایی سیستم‌های دسته‌بندی و تشخیص بیماری‌ها، از ماتریس درهم‌ریختگی استفاده می‌شود. تحلیل‌های ماتریس درهم‌ریختگی در دسته-بندی و تشخیص بیماران منجر به ۴ حالت مثبت حقیقی (True Positive)، منفی حقیقی (True Negative)، مثبت کاذب (False Positive) و منفی کاذب (False Negative) می‌شود (جدول ۲).

جدول ۲: ماتریس درهم‌ریختگی

Nodule	Non-nodule
TP(True positive)	FP(False positive)
FN(False negative)	TN(True negative)

برای به دست آوردن مقادیر ذکرشده به شکل زیر عمل می‌کنیم:
TP، شامل تعداد نمونه‌های سرطانی است که سیستم آن‌ها را درست تشخیص داده است.
FP، شامل تعداد نمونه‌های سالمی است که سیستم آن‌ها را سرطانی تشخیص داده است.
FN، شامل تعداد نمونه‌های سرطانی است که سیستم آن‌ها را سالم تشخیص داده است.
TN، شامل تعداد نمونه‌های سالمی است که سیستم آن‌ها را سالم تشخیص داده است.
در این ماتریس عناصر قطر اصلی نشان دهنده تعداد مواردی است که به درستی دسته‌بندی شده و عناصر قطر فرعی مواردی می‌باشد که به درستی دسته‌بندی نشده‌اند؛ بنابراین با توجه به مقادیر به دست آمده از ماتریس درهم‌ریختگی چهار معیار را برای ارزیابی کارایی الگوریتم پیشنهادی می‌توان محاسبه نمود که عبارت‌اند از [۱۴]:

۱- نرخ دسته‌بندی
۲- نرخ صحت
۳- نرخ فراخوانی (نرخ حساسیت)
۴-F- سنجش

که هر کدام از روابط زیر محاسبه می‌شود.

$$Classification\ Rate = \frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (۶)$$

$$Precision = \frac{TP}{TP+FP} \quad (۷)$$

$$Recall = \frac{TP}{TP+FN} \quad (۸)$$

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (۹)$$

جهت پیاده‌سازی الگوریتم پیشنهادی و محاسبه معیارهای موردنظر از نرم افزار متلب و بر روی سیستم Core i5 تحت ویندوز ۷ استفاده شده است. جدول ۳ نتایج مربوط به الگوریتم پیشنهادی را نشان می‌دهد.

جدول ۳: نتایج به‌دست‌آمده از ارزیابی سیستم پیشنهادی

سیستم تشخیص بیماری قلبی	نتایج
نرخ دسته‌بندی	٪۸۵٪
نرخ صحت	٪ ۸۳/۵
نرخ فراخوانی (حساسیت)	۸۵٪ /۸
F- سنجش	٪ ۸۵

برای ارزیابی بیشتر الگوریتم پیشنهادی، از تکنیک‌های داده‌کاوی، جهت تشخیص فرد مبتلا به بیماری قلبی استفاده شده است و الگوریتم پیشنهادی را با استفاده از نرم‌افزار Weka با آن‌ها مقایسه می‌نماید. این نرم‌افزار مشتمل بر مجموعه‌ای از روش‌ها از قبیل دسته‌بندی، خوشه‌بندی و انتخاب ویژگی است. از مهم‌ترین الگوریتم‌های داده‌کاوی می‌توان به الگوریتم درخت تصمیم، شبکه عصبی مصنوعی، شبکه‌های بیزین و K-نزدیک‌ترین همسایه اشاره کرد. از مهم‌ترین الگوریتم‌های هر دسته می‌توان به درخت تصمیم C4.5، شبکه عصبی RBF، شبکه بیزین NaiveBayes و K-نزدیک‌ترین همسایه KNN اشاره کرد. جدول ۴ نیز نتایج به دست آمده از الگوریتم پیشنهادی را با این روش‌ها مقایسه می‌کند.

جدول ٤: مقایسه نتایج به دست آمده با سایر روش‌ها

نام الگوریتم	نرخ‌دسته‌بندی	نرخ‌صحت	نرخ فراخوانی	F-سنجش
RBF Network	٪ ۸۰/۷	۸۲٪/۲	٪ ۸۳/۳	۸۲٪/۸
NaiveBayes	٪ ۸۱/۹	٪ ۸۳	٪ ۸۴/۷	٪ ۸۳/۸
K-NN	٪ ۷۹	٪ ۷۸/۲	٪ ۸۶	٪ ۸۲
C4.5	٪ ۷۸	٪ ۸۱/۱	٪ ۸۰	۸۰٪ /۵
الگوریتم پیشنهادی	٪ ۸۵	۸۳٪ /۵	٪ ۸۵/۸	٪ ۸۵

بحث و نتیجه‌گیری

در این پژوهش یک سیستم تشخیص بیماری قلبی ارائه‌شده است. این سیستم برای دسته‌بندی افراد به‌عنوان بیمار قلبی و فرد سالم، از طبقه‌کننده ماشین بردار پشتیبان استفاده می‌نماید. با توجه به محاسبات انجام شده در این سیستم، نتایج قابل قبولی به دست آمده و دارای نرخ دسته بندی و نرخ حساسیت می‌باشد، که در مقایسه با نتایج سایر مطالعات انجام شده، دارای دقت قابل قبولی است.

در پژوهش‌های گذشته، Kumar و DeZhi یک الگوریتم پیش‌بینی بیماری قلبی ارائه داده‌اند [۱۵]؛ در مرحله اول با استفاده از تکنیک DM به انتخاب ویژگی می‌پردازد. سپس با استفاده از توابع عضویت فازی، قوانین مبتنی بر منطق فازی را تولید می‌نماید. در این پژوهش نشان داده شده است که الگوریتم پیشنهادی در مقایسه با سایر روش‌های داده‌کاوی، از قبیل شبکه عصبی و درخت تصمیم J48 نتایج بهتری را در بر گرفته است. نرخ دسته‌بندی در این مدل ۸۶٪ می‌باشد. Khatibi و Montazer نیز از یک موتور استنتاج ترکیبی جهت تشخیص بیماری قلبی استفاده نموده‌اند [۱۶]؛ در این روش در ابتدا توسط ترکیب نظریه Dempster-Shafer و تئوری مجموعه‌های فازی مجموعه قوانین فازی را تولید می‌نماید و سپس از روی این مجموعه قوانین، با نرخ دسته‌بندی ۹۱/۵۸٪ بیماری قلبی را شناسایی می‌کند. شفیعی و ابراهیمی [۱۷] از الگوریتم‌های مختلف بیوانفورماتیک از جمله درخت تصمیم، شبکه‌های عصبی، ماشین بردار پشتیبان، خوشه‌بندی و ...، برای پیش‌بینی بیماری قلبی استفاده نموده است. در مطالعه شفیعی و ابراهیمی از الگوریتم ویژگی وزنی، جهت انتخاب ویژگی استفاده نموده که در نهایت، وزنی بین ۰ و ۱ به هر ویژگی اختصاص داده شده است. این عمل، برای مجموعه داده‌های بزرگ مناسب بوده و اجازه می‌دهد ویژگی‌های مهم موردنیاز، به راحتی انتخاب شود. در این مطالعه روش شبکه عصبی با دقت ۸۰٪ و این دقت در الگوریتم‌های بیز ساده به ۸۵٪ رسیده است.

نتایج به‌دست‌آمده نشان می‌دهد که سیستم پیشنهادی نسبت به سایر مطالعات در تشخیص بیماری قلبی موفق بوده است؛ ولی هر چند که نرخ دسته‌بندی را تا حد قابل قبولی بهبود داده است؛ اما دارای خطای نسبتاً پایینی در شناسایی افراد مبتلا به بیماری قلبی بوده است.

افزایش دقت در شناسایی افراد مبتلا به بیماری قلبی به بزرگ‌تر بودن پایگاه داده‌ها بستگی دارد؛ بنابراین در تحقیقات آینده می‌توان با استفاده از پایگاه داده‌های بزرگ‌تر، تعداد رکوردها را افزایش داد و دقت الگوریتم را افزایش داد. همچنین می‌توان جهت کاهش زمان پردازش از الگوریتم‌های انتخاب ویژگی به منظور کاهش تعداد متغیرها استفاده نمود. از این روش می‌توان همچنین برای انواع دیگر بیماری‌ها و در مراکز آنکولوژی استفاده نمود.

References

1. Rafii F, Soleimani M, Seyedfatemi N. A model of patient participation with chronic in nursing care. Koomesh 2011; 12(3): 293-304. Persian

2. Langarizadeh M, Sadr-Ameli MA, Soleymani M. Development of vital monitoring decision support system for coronary care unit inpatients. J Health Adm 2017; 20(67):75-88. Persian

3. Eslami R, Sajadi S A, Farsi Z. Comparing the effect of peer education and orientation tour on the stress of patients candidate for coronary angiography in selected Hospital of Aja University of Medical Sciences. J Urmia Nurs Midwifery Fac 2015; 12(12):1119-27. Persian

4. Daliri N, Zakeri-Moghadam M. Aging and heart disease. Iranian Journal of Cardiovascular Nursing 2016; 5(1): 64-9. Persian

5. Sultana M, Haider A, Uddin MS. Analysis of Data Mining Techniques for Heart Disease Prediction. 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT); 2016 Sep 22-24; Dhaka, Bangladesh: IEEE; 2016.

6. Banu GR, Jamala JH. Predicting heart attack using fuzzy C Means clustering algorithm. International Journal of Latest Trends in Engineering and Technology 2015; 5(3): 439-43.

7. Krishnaiah V, Srinivas M, Narsimha G, Chandra N.S. Diagnosis of Heart Disease Patients Using Fuzzy

Classification Technique, International conference on Computer and Communications Technologies (ICCCT), India, 2014.
8. Maihami V, Khormehr A, Rahimi E. Designing an expert system for prediction of heart attack using fuzzy systems. Sci J Kurdistan Univ Med Sci 2016; 21: 118-31. Persian
9. Suganya S, Tamije Selvy P. A proficient heart disease prediction method using fuzzy-CART algorithm. International Journal of Scientific Engineering and Applied Science 2016; 2(1): 1-6.
10. Khosravanian A, Ayat SS. Presenting an intelligent system for diagnosis of coronary heart disease by using Probabilistic Neural Network. Health Inf Manage 2015; 12(1):3-13. Persian
11. Vapnik V. Statistical Learning Theory. NewYork: John Wiley & Sons; 1998.
12. Ravale U, Marathe N, Padiya P. Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function. Procedia Computer Science 2015; 45: 428-35.
13. Center for Machine Learning and Intelligent Systems. (UCI) [cited 2015 Feb 23] Machine Learning Repository, http://www.archive.ics.uci.edu.
14. Mahmoodi MS, Mahmoodi SA, Haghighi F, Mahmoodi SM. Determining the stage of breast cancer by data mining algorithms. Iran J Breast Dis 2014; 7(2): 36-44. Persian
15. Kumar K, DeZhi X. A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease. Global Journal of Computer Science and Technology 2014; 14(3): 16-22.
16. Khatibi V, Montazer GH. A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. Expert Syst Appl 2010: 13(12):8536–42.
17. Shafiee H, Ebrahimi M. Accurate prediction of coronary artery disease using bioinformatics algorithms. Qom Univ Med Sci J 2016; 14(4): 22-35. Persian

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2017; 4(1): 1-10

Designing a Heart Disease prediction System using Support Vector Machine

Mahmoodi Maryam Sadat[1]*

• **Received:** 4 May, 2017 • **Accepted:** 17 Jan, 2017

Introduction: Cardiovascular diseases are the leading cause of death worldwide. The world health Organization has estimated 12 million deaths per year worldwide, due to the cardiovascular diseases. The main aim of this study was to design a smart computer-aided system for the diagnosis of heart disease in patients.
Methods: In this study descriptive and analytical study, data of 270 people with 13 features were used. The fuzzy system and support vector machine classifier were combined using facilities of MATLAB software and were simulated by a system of core i5 and windows7, as the operating system, to diagnose patients with heart disease.
Results: Fuzzy technique and support vector machine algorithm that were used for the diagnosis of heart disease was efficient in rapid diagnosis and consequently increasing the patient chance for survival. Evaluation criteria in this system were rates of categorization and sensitivity and the system performance for these two indicators were respectively 85% and 85.8%.
Conclusion: According to the results, the proposed system can diagnose patients with cardiovascular diseases with a relatively high accuracy.

Keywords: Cardiovascular Disease, Fuzzy System, Support Vector Machine Classifier.

• **Citation:** Mahmoodi MS. Designing a Heart Disease prediction System using Support Vector Machine. Journal of Health and Biomedical Informatics 2017; 4(1): 1-10.

1. M.Sc. Computer Sciences, Computer Engineering Dept., Birjand Payame Noor University (PNU), Birjand, Iran.
***Correspondence**: Payame Noor University, Birjand, South Khorasan.
• **Tel:** 09139545939 • **Email:** m_mahmoodi_64@yahoo.com