

ORIGINAL ARTICLE

Intelligent and fast recognition of heart disease based on synergy of linear neural network and logistic regression model

Mojtaba Heravi¹,
Saeed Setayeshi²

¹ MSc, Department of Knowledge Engineering and Decision Sciences, Department of Computer, School of Management and Computer Engineering, University of Economic Sciences, Tehran, Iran

² Associate Professor, Department of Medical Radiation Engineering, School of Physics and Nuclear Engineering, Amirkabir University of Technology, Tehran, Iran

(Received September 29, 2013; Accepted April 13, 2014)

Abstract

Background and purpose: Diseases have been the greatest threat for human being along the history. Heart disease (HD) has gained special attention in medical studies. Recently studying on classification and diagnosis of HD as a key topic and a lot of researches have been done in order to increase precise and reduce error in this type of decisions. With development of intelligent learning systems, these systems have played a great role in reducing the error of decision support systems (DSS).

Materials and methods: In this study, a simple hybrid model of logistic regression and single-layer perceptron neural network was presented which was trained with four-different learning rules (separately). The model for improving the classification and patterns recognition of HD has been used on clinical data of 270 patients from the Cleveland Clinic (UCI website). This method has been used in statistical data normalization and detection of noisy data, network training with only 20% of the data exist was performed. The model has been implemented in MATLAB.

Results: The mean-error of the proposed model on the total dataset was 11.11%, which was achieved a significant improvement compared to recent similar methods. In addition, the results showed that the proposed approach was very capable in dealing with noise in the data.

Conclusion: The results clearly showed that the linear proposed technique had large impact on reducing the error in the classification and identification of patients more accurately in a shorter time than conventional methods and complex nonlinear. The method can help physicians for early detection of disease or as a DSS.

Keywords: Heart diseases diagnosis, biomedical engineering, pattern recognition, machine learning, artificial neural network, single layer perceptron, logistic regression

J Mazand Univ Med Sci 2014; 24(112): 78-87 (Persian).

تشخیص هوشمند و سریع بیماری قلبی بر اساس هم‌افزایی شبکه‌های عصبی خطی و روش رگرسیون منطقی

مجتبی هروی^۱

سعید ستایشی^۲

چکیده

سابقه و هدف: در طول تاریخ، بیماری‌ها بزرگ‌ترین تهدید برای بشر به شمار می‌روند. در این میان بیماری‌های قلبی از توجه بیشتری در مطالعات پژوهشی برخوردارند. در سال‌های اخیر، دسته‌بندی و تشخیص امراض قلبی به عنوان یک مبحث کلیدی بررسی شده و تحقیقات بسیاری در زمینه افزایش دقت و کاهش خطای در این گونه تصمیم‌گیری‌ها صورت گرفته است. با ایجاد سیستم‌های هوشمند یادگیر، این سیستم‌ها در پشتیبانی از تصمیمات و کاهش خطای نقش بزرگی را ایفا کرده‌اند.

مواد و روش‌ها: در این تحقیق، یک مدل تکیبی ساده از رگرسیون منطقی و شبکه عصبی پرسپترون تک لایه ارایه شده است که با چهار قانون یادگیری مختلف (به صورت مجزا) آموزش می‌یابد. این مدل برای بهبود دسته‌بندی و بازناسی الگوهای بیماری قلب، روی اطلاعات بالینی ۷۰ بیمار از کلینیک کلیولند (Cleveland Clinic) (سایت UCI) استفاده شد. این روش از نرم‌افزار سازی آماری بهره می‌گیرد و پس از تشخیص داده‌های مخدوش، آموزش شبکه فقط با ۲۰ درصد از داده‌های موجود انجام می‌شود. پیاده‌سازی مدل در نرم‌افزار Matlab صورت گرفته است.

یافته‌ها: میانگین خطای مشاهده شده مدل پیشنهادی روی کل مجموعه داده‌ها ۱۱/۱۱ درصد به دست آمد که بهبود قابل ملاحظه‌ای را نسبت به روش‌های مشابه اخیر نشان می‌دهد. همچنین، یافته‌ها نشان داد که مدل پیشنهادی در برخورد با اختلالات موجود در داده‌ها بسیار توانمند عمل می‌کند.

استنتاج: تکیک خطی مطرح شده تأثیر زیادی بر کاهش خطای در دسته‌بندی و شناسایی افراد بیمار نسبت به روش‌های متداول و غیر خطی پیچیده با دقتی بیشتر و در زمانی کمتر داشته است. این روش برای تشخیص زود هنگام بیماری و یا به عنوان یک سیستم پشتیبان تصمیم می‌تواند به پژوهشکار یاری رساند.

واژه‌های کلیدی: تشخیص بیماری‌های قلبی، مهندسی پزشکی، بازناسی الگو، یادگیری ماشین، شبکه عصبی مصنوعی، پرسپترون تک لایه، رگرسیون منطقی

علل مهم مرگ و میر در دنیا به شمار می‌روند (۳-۱). در ایران نیز

مطابق نمودار شماره ۱، بیشترین تعداد مرگ و میر را بیماری‌های قلبی-عروقی تشکیل داده‌اند.

پژوهش‌های اخیر نشان داده‌اند که بیماری‌های قلبی تحت یک وضعیت خاص اتفاق نمی‌افتد، بلکه شرایط زیادی در به وجود آمدن این گونه بیماری‌ها نقش دارند و همچنان دلایل اساسی بیماری‌های قلبی مهم می‌باشند (۴-۷).

مقدمه

در طول تاریخ حیات بشری، بیماری‌ها بزرگ‌ترین تهدید برای بشر به شمار رفته‌اند که تأثیرات محربی نیز روی زندگی انسان‌ها داشته‌اند. در این میان، بیماری قلب از توجه بیشتری در تحقیقات پژوهشی برخوردار بوده است. بر طبق یافته مرکز آمار و بیولوژی سرطان استرالیا و همچنین بر اساس گزارش دفتر آمار و اطلاعات جمعیتی و مهاجرت ایران، بیماری‌های قلبی یکی از دو

E-mail: mojtaba.heravi@ues.ac.ir

مؤلف مسئول: مجتبی هروی - تهران: خیابان نمسی، دانشگاه علوم اقتصادی، دانشکده مهندسی کامپیوتر و مدیریت، گروه کامپیوتر

۱. کارشناس ارشد مهندسی دانش و علوم تصمیم، گروه کامپیوتر، دانشکده مهندسی کامپیوتر و مدیریت، دانشگاه علوم اقتصادی تهران، تهران، و عضو باشگاه پژوهشگران جوان و نخبگان، دانشگاه آزاد اسلامی، واحد قزوین، ایران

۲. دانشیار، گروه مهندسی پرتو پژوهشکی، دانشکده مهندسی هسته‌ای و فیزیک، دانشگاه صنعتی امیرکبیر، تهران، ایران

۳. تاریخ دریافت: ۱۳۹۲/۷/۷ تاریخ ارجاع جهت اصلاحات: ۱۳۹۲/۱/۲۴ تاریخ تصویب: ۱۳۹۳/۱/۲۴

نام سیستم پیش‌بینی عارضه‌یابی قلب هوشمند (IHDPS) Intelligent heart disease prediction system کمک روش‌های داده کاوی ارایه داده‌اند (۱۵). Avci و Turkoglu یک سیستم هوشمند بر مبنای استفاده از دیدگاه Genetic-support GSVM یا vector machine (برای دسته‌بندی سیگنال‌های داپلر دریچه‌های قلب ارایه داده‌اند) (۱۶).

در این مطالعه، با توجه به مهم و شایع بودن بیماری قلب، هدف این بود که با استفاده از یک روش غیر تهاجمی، مدلی از ترکیب رگرسیون منطقی دو مقداره و شبکه عصبی پرسترون تک لایه به عنوان یک سیستم پشتیبان تصمیم هوشمند یا تصمیم‌گیر مستقل با پاسخگویی سریع و دقت بالا تولید شود که بتواند از روی اطلاعات و علایم بالینی فرد مشکوک به بیماری قلبی، ابتلاء یا عدم ابتلاء او را تشخیص دهد. همچنین در مقابل روش‌های تهاجمی از خطر کمتری نیز برخوردار باشد.

مواد و روش‌ها

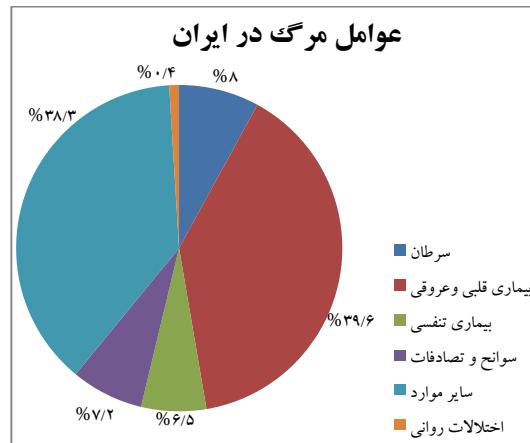
تحقیق حاضر یک مطالعه تشخیصی بود که بر اساس متغیرهای ورودی، به تصمیم‌گیری در مورد وضعیت وجود یا عدم وجود بیماری قلبی پرداخت. الگوریتم پیشنهادی ما از دو فاز اصلی تشکیل شده است: فاز اول، شامل بخش‌هایی برای آماده‌سازی داده‌ها قبل از انجام فرایند آموزش و دسته‌بندی است و فاز دوم، مربوط به بخش آموزش ماشین و دسته‌بندی داده‌ها برای بازشناسی و تفکیک ورودی‌های جدید از هم بر اساس شبکه عصبی می‌باشد. طرح کلی الگوریتم در تصویر شماره ۱ نشان داده شده است.

۱. فاز پیش پردازش

۱-۱. نرم‌افزاری

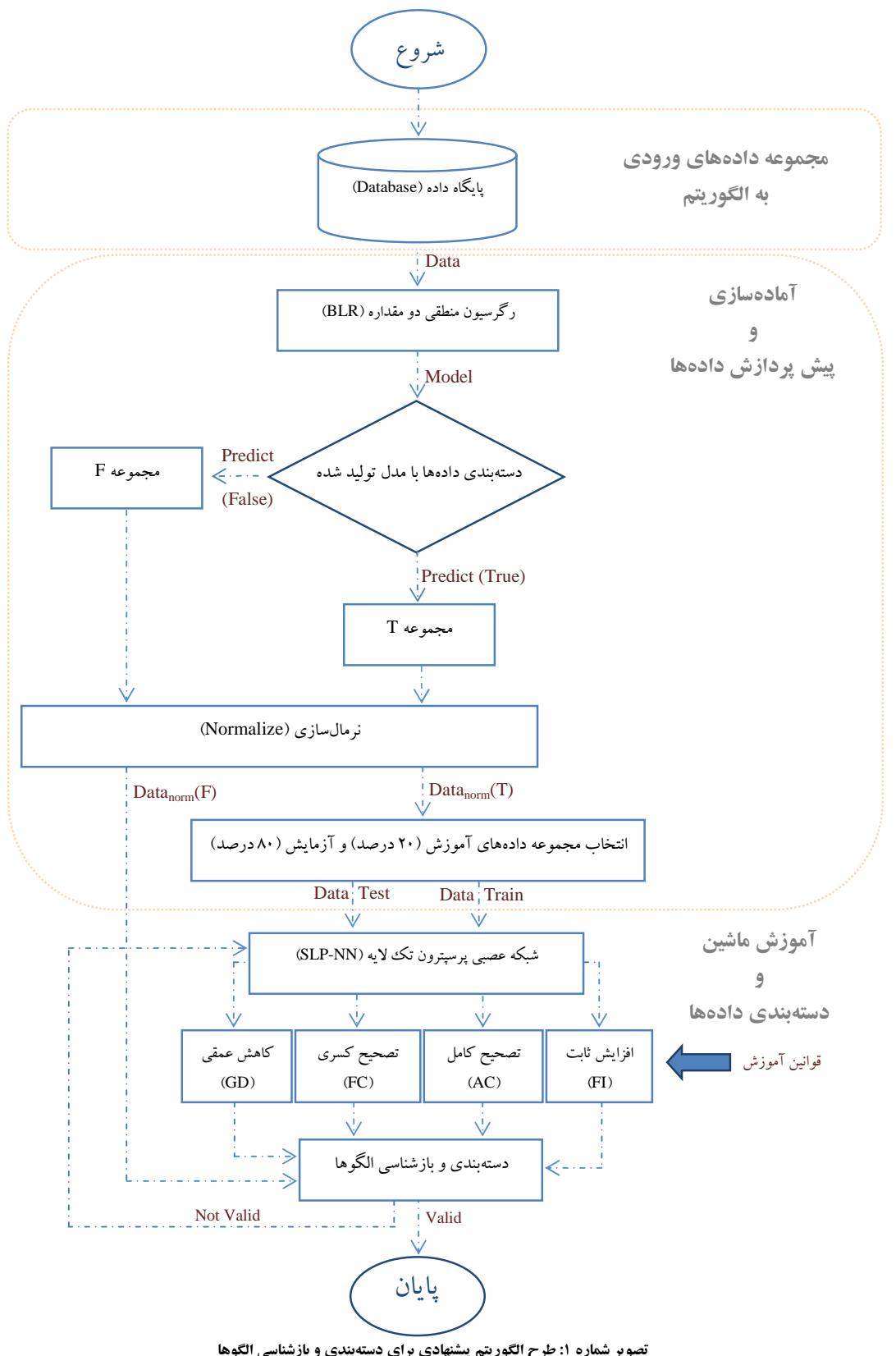
امروزه، در بیشتر مسایلی که با آن‌ها مواجه می‌شویم، مقادیر ورودی در مقیاس‌های یکسانی نیست و این مسئله، تأثیر منفی در روند همگرایی و افزایش زمان و تعداد دفعات تکرار اعمال آموزش خواهد گذاشت. پس بایستی داده‌ها را به نحو

در علوم پزشکی، چون شناسایی فرد بیمار از روی علایم بالینی صورت می‌پذیرد (به طور متداول)، در نتیجه نیاز به افراد خبره (پزشک معالج) در این زمینه حائز اهمیت است. همچنین افراد خبره همیشه و همه جا در دسترس نیستند و امکان بروز اشتباه در فرد خبره (هنگام زیاد بودن علایم یا متشابه بودن آن‌ها با انواع دیگر بیماری) بیشتر از یک سیستم یادگیر با حافظه نگهداری پیشنهادی بیمار است. سرعت و دقت پردازش اطلاعات در یک سیستم یادگیر نیز خیلی بیشتر از یک فرد خبره در شرایط حساس و پرتنش می‌باشد. این موارد به روشنی بر ضرورت وجود ماشینی تأکید می‌کند که بتواند در موقع حساس - که با کوچک‌ترین اشتباه پزشک ممکن است بیماری به درستی تشخیص داده نشود و جان بیمار به خطر بیفتد - در تصمیم‌گیری به پزشک کمک نماید.



نمودار شماره ۱: مقایسه عوامل متفاوت منجر به مرگ در ایران (۲۰۱۲)

در سال‌های اخیر، از مفاهیم هوش محاسباتی ماشین در یافتن روابط بین بیماری‌های مختلف و ویژگی‌های بیماران استفاده شده است (۱۲-۱۴). به طور مثال Eberhart و همکاران، یک شبکه عصبی تطبیقی را برای دسته‌بندی الگوهای الکتروکاردیوگراف چند کاناله به کار برده‌اند (۱۳). Cios و همکاران از یک شبکه عصبی برای یافتن بیماری‌های قلبی توسط تصاویر اکوکاردیوگرافی استفاده کرده‌اند (۱۴). Palaniappan و Awang یک نمونه عارضه‌یابی کامپیوتری (Computer-aided diagnostics CAD) یا



که از تعریف قانون پارتو گرفته شده است (۲۰، ۲۱). در این قانون، به ترتیب ۲۰ درصد جامعه آماری (داده‌ها) را اقلیت حیاتی و اکثربیت کم اهمیت تشکیل می‌دهند. به عبارتی، اگر ۲۰ درصد از داده‌های مهم به درستی در آموزش شرکت یابند، امکان پیش‌بینی صحیح ۸۰ درصد از کل داده‌ها میسر خواهد شد. مجموعه دیگری که در روند آموزش استفاده شد، مجموعه اعتبارسنجی بود که در هر مرحله از آموزش، میزان بهبود دقت و میزان کاهش خطای توسط آن بررسی شد. در مدل پیشنهادی برای جلوگیری از زیاده برآش، تدبیری اتخاذ شد که اگر در طول یادگیری میزان خطای مجموعه اعتبارسنجی، ۷ مرتبه به صورت متوالی افزایش پیدا می‌کرد، یادگیری متوقف می‌شد و بیشترین مقدار دقت پیش از توقف، به عنوان پاسخ به خروجی فرستاده می‌شد.

۲. فاز آموزش با شبکه عصبی پرسپترون تک لایه شبکه‌های عصبی به عنوان یک ابزار قدرتمند برای تعزیز و تحلیل پایگاه داده‌ها در مسایلی همچون دسته‌بندی، تشخیص، شناسایی و پشتیبانی از تصمیمات به کار گرفته می‌شوند (۲۲-۲۴). شبکه‌های عصبی توانایی یادگیری، مدل کردن، توسعه و بررسی دسته‌بندی‌های الگوها در یک پدیده بدون نیاز به هیچ دانش قبلی برای درک توزیع داده‌ها را دارا می‌باشند و به راحتی قابلیت انطباق با مسایلی مانند انواع تشخیص پزشکی و پیش‌بینی بیماری در افراد را دارند و علاوه بر سرعت بخشیدن در پاسخ‌گویی، از دقت بالایی نیز برخوردارند (۲۳-۲۶). شبکه‌های متفاوتی برای دسته‌بندی و مدل کردن مسایل به کار گرفته شده‌اند (۲۴).

شبکه عصبی مصنوعی پرسپترون تک لایه جزو ساده‌ترین شبکه‌های خطی محسوب می‌شود (۲۷). معادله این شبکه برای یک الگو در رابطه ۴ ییان شده است. فرض کنید یک مجموعه از الگوهای آموزشی $\{X_1, \dots, X_n\}$ داریم که هر بردار آن به یکی از دو کلاس ۱ و ۰۲ انتساب داده شده‌اند و N تعداد الگوهای ورودی می‌باشد. هدف پیدا کردن یک بردار وزن (در حالت تک نرونی)، $(W_1, \dots, W_i, \dots, W_R)$ و مقدار

مؤثری نرمال کرد و تغییر داد تا این مشکلات برطرف شود. به این منظور، برای نرمالیزه کردن داده‌ها از روش نرمالیزاسیون آماری جهت همگن کردن داده‌ها استفاده شد که از رابطه زیر تبعیت می‌کند:

$$\text{Data}_{\text{norm}} = \frac{\text{Data} - \text{Mean}_{\text{Data}}}{\text{Std}_{\text{Data}}} \quad (\text{رابطه ۱})$$

که در آن Data همان داده‌های خام ورودی و Mean و Std به ترتیب بیانگر میانگین داده‌ها و انحراف معیار داده‌ها در هر خصیصه می‌باشند.

۱-۲. رگرسیون منطقی دو مقادیره

یکی از کاربردهای مهم رگرسیون، بررسی وابستگی‌های موجود در داده‌ها برای پیش‌بینی و به دست آوردن رفتار یک متغیر (Y)، بر حسب سایر متغیرها (X) می‌باشد (۱۷-۱۹). حال این رفتار می‌تواند خطی یا به شکل منحنی باشد. حالت خاصی از رگرسیون، رگرسیون منطقی دو مقادیره است که یک تکیک غیر خطی با یادگیری نظارتی، ساختاری آماری و بدون هوشمندی برای پیش‌بینی احتمالاتی یک اتفاق می‌باشد. اگر معادله آن به صورت زیر در نظر گرفته شود:

$$Y = \beta \cdot X + \alpha \quad (\text{رابطه ۲})$$

که در آن β ضرایب تأثیر متغیرهای ورودی (X) و α مقدار ثابت (بایاس) می‌باشد، برای پیدا کردن احتمال وقوع هر کدام از کلاس‌ها، از رابطه زیر استفاده می‌گردد:

$$P = \frac{e^Y}{1 + e^Y} \quad (\text{رابطه ۳})$$

$$\begin{cases} \text{if } 0 \leq P \leq 0.5 \Rightarrow \text{Class 1} \\ \text{if } 0.5 \leq P \leq 1 \Rightarrow \text{Class 2} \end{cases}$$

به طوری که اگر $0.5 \leq P \leq 1$ باشد، الگو متعلق به کلاس اول و اگر $0 < P \leq 0.5$ باشد، الگو متعلق به کلاس دوم می‌شود.

۱-۳. نحوه انتخاب مجموعه داده‌های آموزش و آزمایش و روند یادگیری

داده‌های آموزش بیشترین تأثیر گذاری را روی یادگیری ماشین دارند. روشی که ما برای تفکیک این داده‌ها به دو دسته آموزش و آزمایش پیشنهاد می‌کنیم، انتخاب ۲۰ به ۸۰ است

رابطه ۱۰ به دست می‌آید. در قانون تصحیح کسری، میزان تأثیر پارامتر یادگیری با متغیر λ مشخص می‌شود که مقداری بین [۰-۲] است. رابطه ۱۱ نیز، معادله تغییر یا عدم تغییر اوزان و ترم بایاس در قانون کاهش عمقی است. با ذکر این موضوع که E مقدار میانگین مربع خطای است، ∇E مقدار گرادیان آن می‌باشد و نرخ پارامتر یادگیری (c) مقداری بین [۰-۲] است.

۳. داده‌های زیستی مورد بررسی
داده‌های مورد بررسی، از اطلاعات کلینیک کلیولند (Cleveland Clinic) موجود در سایت UCI جمع‌آوری شد و در حیطه تشخیص بیماری قلب بود (۲۹). این داده‌ها مربوط به ۲۷۰ نمونه بود که ۱۳ خصیصه از اطلاعات بالینی هر نمونه به همراه کلاس آن‌ها را شامل می‌شد (۲۹). این خصیصه‌ها شامل موارد زیر می‌باشند:

- سن
- جنسیت: مرد یا زن
- نوع درد قفسه سینه: شامل چهار وضعیت، آنژین صدری متداول (Angina)، آنژین صدری نابهنجار (Abnang)، درد غیر آنژینی (Notang) و بدون علامت (Asympt).
- فشار خون: در حالت استراحت
- میزان کلسترول خون (mg/dl)
- میزان قند خون ناشتا (mg/dl): بیشتر از ۱۲۰ میلی گرم به دسی لیتر است یا خیر.
- وضعیت نوار قلب در حالت استراحت: شامل چهار وضعیت، عادی (Normal)، غیر طبیعی (Abnormal)، اختلال در موج ST-T، هیپرتروفی بطن (Hypertrophy) و حداقل ضربان قلب
- میزان درد قلبی هنگام فعالیت: ناشی از آنژین پارامتر ST: میزان افت موج ST در حالت فعالیت نسبت به حالت استراحت
- شب ST در اوج فعالیت: شامل سه وضعیت، شب صعودی، مسطح و شب نزولی

آستانه‌ای یا ترم بایاس θ است که R تعداد ورودی‌های هر الگوی بعد آن می‌باشد $\{X_1, \dots, X_R\}$ ، همچنین Threshold یک تابع محرك است که در تنظیم مقدار خروجی (f) تأثیر دارد. در حالت کلی، اگر $f < 0$ باشد، الگو مربوط به کلاس ۰ و اگر $f > 0$ باشد، الگو در کلاس ۱ جای می‌گیرد.

$$f = \text{threshold}(\sum_{i=1}^R w_i \cdot x_i, \theta) \quad (\text{رابطه ۴})$$

برای آموزش این شبکه، از چهار قانون یادگیری استفاده شد که شامل قوانین یادگیری افزایش ثابت (Fixed increment) یا Absolute correction کامل (AC) یا (FI)، تصحیح کسری (Fractional correction) یا (FC) یا Gradient descent (GD) می‌باشند (۲۸). قانونی که بهترین نتیجه را بدهد، به عنوان خروجی نهایی بازگردانده می‌شود. رابطه‌های این چهار قانون با تغییراتی که در اوزان و مقدار ترم بایاس به وجود می‌آورند، عمل یادگیری را انجام می‌دهند. اگر فرض کنیم:

$$y(j) = (x_1, x_2, \dots, x_R, 1) \quad (\text{رابطه ۵})$$

$$W = (w_1, w_2, \dots, w_R, \Theta) \quad (\text{رابطه ۶})$$

$$y(j) \rightarrow -y(j) \text{ for all } y(j) \in \Omega_2 \quad (\text{رابطه ۷})$$

که $y(j)$ نگاشتی از الگوی X_j می‌باشد، مقدار عدد یک در نگاشت صورت گرفته، مربوط به ترم بایاس اضافه شده به بردار وزن است. آن گاه خواهیم داشت:

$$w(k+1) = \quad (\text{رابطه ۸})$$

$$\begin{cases} w(k) + cy(k) & \text{if } w(k) \cdot y(k) \leq 0 \\ w(k) & \text{if } w(k) \cdot y(k) > 0 \end{cases}$$

$$c = \frac{\|w(k) \cdot y(k)\|}{\|y(k)\|^2} \quad (\text{رابطه ۹})$$

$$c = \lambda \frac{\|w(k) \cdot y(k)\|}{\|y(k)\|^2} \quad (\text{رابطه ۱۰})$$

$$w(k+1) = \quad (\text{رابطه ۱۱})$$

$$\begin{cases} w(k) - c\nabla E & \text{if } w(k) \cdot y(k) \leq 0 \\ w(k) & \text{if } w(k) \cdot y(k) > 0 \end{cases}$$

رابطه ۸، معادله کلی برای تغییر یا عدم تغییر اوزان و ترم بایاس در سه قانون اول است. با توجه به این مطلب که در قانون افزایش ثابت، نرخ پارامتر یادگیری (c) مقداری بین [۰-۲] دارد. همچنین نرخ پارامتر یادگیری (c) در قانون تصحیح کامل و قانون تصحیح کسری، به ترتیب مقدارشان از رابطه ۹ و

۴ و ۵ روند بهبود دقت در هنگام آموزش با چهار روش یادگیری مطرح شده، ترسیم شده است.

جدول شماره ۱: مقدار ضرایب تأثیر و ترم بایاس بهینه (β)	
B	متغیر (ویرگی)ها
-۰/۰۱۷	سن
۱/۵۴۹	جنسیت
۰/۷۰۱	نوع درد قدسه سینه
۰/۰۲۵	فشار خون
۰/۰۰۷	میزان کلسترول (mg/dl)
-۰/۷۶۶	میزان قند خون ناشتا (mg/dl)
۰/۳۰۴	وضعيت نوار قلب در حالت استراحت
-۰/۰۲۲	حداکثر ضربان قلب
۰/۸۱۶	میزان درد قلبی هنگام فعالیت
۰/۳۴۴	پارامتر ST
۰/۴۵۱	شیب ST در اوج فعالیت
۱/۱۶۱	تعداد فلوروسکوپی
۰/۷۱۳	شرایط قلب بیمار
-۹/۵۵۰	مقدار ثابت (ترم بایاس)

جدول شماره ۲: مقایسه نتایج واقعی مدل با نتایج مطلوب موجود

		مشاهدات		کلاس	درست (درصد)	پیش‌بینی
		۱	۲			
۸۸/۶۶	۱۷	۱۳۳	۱	کلاس		
۸۰/۸۳	۹۷	۲۳	۲	کلاس		
۸۵/۱۸				درصد کل		
بررس ارزشی برابر ۰/۵ می‌باشد.						

مشاهده می‌شود که مقدار خروجی یادگیری با افزایش ثابت نتیجه بهتری را تولید می‌کند و از شناسایی خوبی برخوردار است. در مدل پیشنهادی، چون مقدار اولیه اوزان و ترم بایاس به صورت تصادفی انتخاب می‌شوند، امکان این که مقدار خروجی و میزان دقت نتایج در هر اجرا کمی تغییر کنند، وجود دارد. با این حال، میانگین به دست آمده از نتایج چندین بار اجرای برنامه روی هر یک از ترکیب‌های مختلف داده‌ها (شامل ۱۰ ترکیب مختلف)، در مجموعه آموزش و آزمایش نشان داده شد.

- تعداد فلوروسکوپی: تعداد فلوروسکوپی‌های گرفته شده از رگ‌های اصلی به صورت رنگی
- شرایط قلب بیمار: شامل سه وضعیت، عادی، نقص ثابت و نقص برگشت‌پذیر

همچنین این پایگاه داده، افراد را در دو کلاس شامل محتمل به بیماری قلبی و غیر محتمل به بیماری قلبی قرار می‌دهد. هیچ کدام از داده‌هایی مورد بررسی، مقادیر گمشده و یا نامشخص نداشتند؛ یعنی تمامی اطلاعات نمونه‌ها موجود و مشخص بودند.

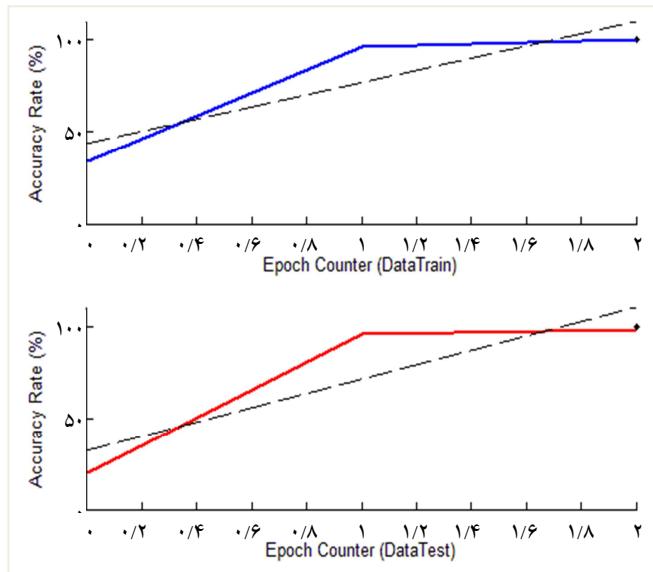
یافته‌ها

روند پیاده‌سازی صورت گرفته توسط نرم‌افزار Matlab و یافته‌های به دست آمده به شرح زیر می‌باشد. ابتدا با استفاده از روشی که در بخش ۲-۱ مطرح شد، ضرایب تأثیر و مقدار ثابت بهینه به دست آمد (جدول شماره ۱). فرمول مدل ریاضی تقریبی پدیده نیز در رابطه ۱۲ نوشته شده است. پس از آن، با مدل تولید شده، کل نمونه‌ها دسته‌بندی شدند. برای ارزیابی دقت و راستی آزمایی، نتایج پیش‌بینی مدل (نتایج واقعی) با مشاهدات حاصل از پدیده (نتایج مطلوب) مقایسه شدند (جدول شماره ۲).

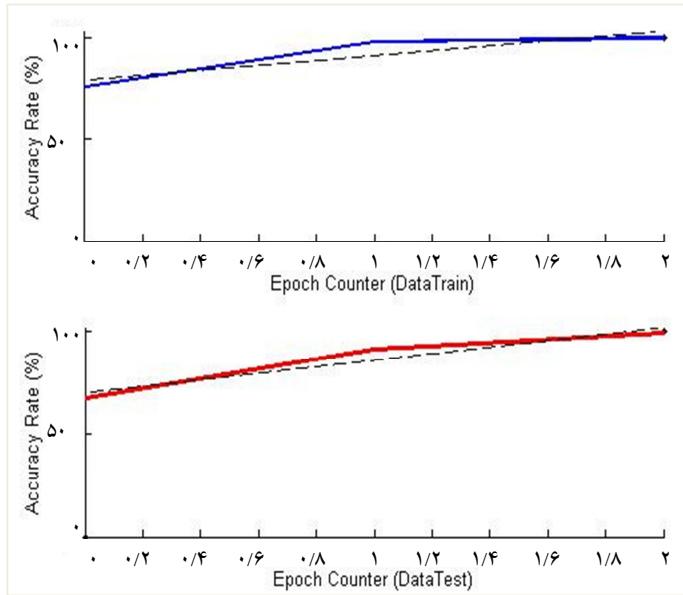
(رابطه ۱۲) $Y = -0/017X_1 + 1/549X_2 + 0/713X_3 - 9/550$
 نمونه از کل نمونه‌ها به درستی و ۴۰ نمونه باقی نمانده به اشتباه در کلاس دیگر پیش‌بینی شدند (جدول شماره ۲). این دو دسته به ترتیب مجموعه T و مجموعه F نام‌گذاری شدند. این جداسازی داده‌ها به دو دسته F و T به این خاطر صورت پذیرفت که داده‌هایی که از حساسیت یا اختلال بیشتری نسبت به مابقی داده‌ها برخوردار هستند و در فرایند یادگیری امکان تولید خطرا افزایش می‌دهند، از چرخه آموزش کنار گذاشته شوند. با رابطه ۱، داده‌های دو دسته نرمال شدند و سپس دسته T بر اساس مفاهیم بخش ۳-۱ به صورت تصادفی به دو دسته آموزش و آزمایش برای ورود به شبکه مطرح شده در بخش ۲، تفکیک گردیدند. در جدول شماره ۳ اطلاعات کلی مربوط به پارامترهای ورودی و نتایج خروجی به دست آمده بعد از اتمام آموزش آمده است. همچنین در نمودارهای شماره ۳، ۲، ۱

جدول شماره ۳: نتایج به دست آمده بعد از اتمام آموزش

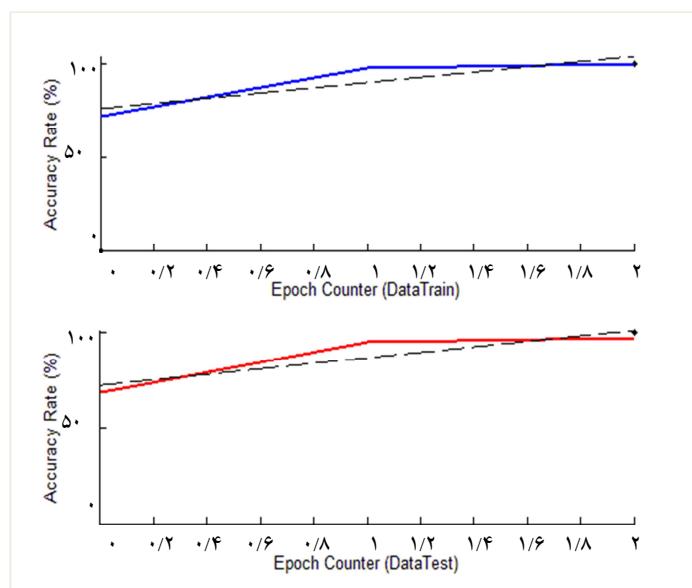
روش یادگیری (Rule)	نرخ پارامتر ورودی	تعداد تکرار (Epoch)	دقت بازشناسی نمونه (درصد) (Recognition accuracy)	مجموعه آزمایش	آموزش	تعداد شناسایی (Recognition accuracy)	غلط درست
Fixed Inc.	-	۳	۹۸/۳۳	۹۸/۵	۱۰۰	۲۴۰	۳۰
AbsoluteCor.	-	۳	۹۸/۸	۱۲/۵	۱۰۰	۲۳۳	۳۷
Fractional Cor.	۲	۳	۹۷/۲۲	۲۷/۵	۱۰۰	۲۳۶	۳۴
Gradient Des.	۱	۴	۹۵/۰۰	۲۷/۰۰	۱۰۰	۲۳۲	۳۸



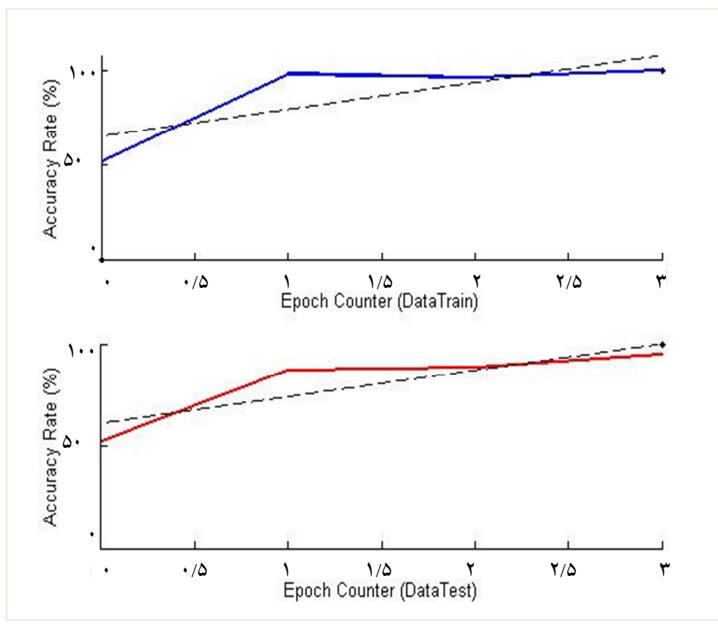
نمودار شماره ۲: روند افزایش دقت در قانون افزایش ثابت



نمودار شماره ۳: روند افزایش دقت در قانون تصحیح کامل



نمودار شماره ۴: روند افزایش دقت در قانون تصویح کسری



نمودار شماره ۵: روند افزایش دقت در قانون کاهش عمقی

است که علاوه بر صرف زمان بیشتر، هوشمندی سیستم نیز زیر سوال بود. مدل پیشنهادی، یک الگوریتم خطی و ساده با دقت و سرعت تحلیل بالاتر است که تنها با ۲۰ درصد از داده‌ها قابل آموزش و اجرا است. بین این روش با روش‌های مطرح دیگر در جدول شماره ۴ مقایسه‌ای صورت گرفته است که نتایج آن ادعای مدل پیشنهاد شده را ثابت کرده است.

بحث

در بیشتر کارهایی که در زمینه پزشکی صورت گرفته است، از الگوریتم‌های غیر خطی و پیچیده استفاده شده است و این مسأله باعث افزایش زمان پاسخگویی سیستم می‌شود. همچنین یادگیری از روی تعداد زیادی الگو به عنوان مجموعه آموزش (بیش از ۷۰ درصد) در این الگوریتم‌ها باعث شده

جدول شماره ۴: مقایسه نتایج الگوریتم‌های مختلف ارایه شده

عنوان کامل	الگوریتم یادگیری	عنوان کوتاه	در سال	میانگین خطای (درصد)	میانگین دقت (درصد)	بیشترین دقت (درصد)	توضیحات
Inverted hierarchical neuro-fuzzy binary space partitioning	(۳۰) HNFB-1	۲۰۰۶	۱۶/۲۶	۸۳/۷۴	۸۷/۹۱	تعداد قانون تولید شده برابر ۳۱	
artificial neural network and fuzzy neural network	(۳۱) ANN-FNN	۲۰۰۸	۱۳/۲	۸۶/۸	۸۷/۴	-	
genetic algorithms and support vector machines	(۳۲) GA-SVM	۲۰۰۹	۱۴/۸۵	۸۵/۱۵	۸۵/۸۱	با معیار همبستگی و اندازه جمعیت برابر ۱۶	
Principal component analysis and differential evolution classifier	(۳۳) PCA-DE	۲۰۱۰	۱۲/۳۷	۸۷/۶۳	۹۱/۸۶	-	
gradient descent with momentum and adaptive gain	(۳۴) GDM/AG	۲۰۱۰	۱۲/۱۲	۸۷/۸۸	۹۰/۳	در الگوریتم شبکه عصبی پس انتشار	
Rotation forest ensemble classifier	RF ensemble (۳۵)	۲۰۱۱	۱۹/۵۱	۸۰/۴۹	۸۰/۴۹	شامل ۳۰ الگوریتم یادگیری ماشین	
Artificial photosynthesis and phototropism mechanism and artificial neural network	APPM-ANN (۳۶)	۲۰۱۲	۱۴/۵۲	۸۵/۴۸	۹۴/۸۲	-	
motivated feature selection and computer feature selection	(۳۷) MFS-CFS	۲۰۱۳	۱۴/۴۸	۸۵/۵۲	۹۱/۰۵	با تغییر روی تعداد اعضای کلاس‌ها	
binary logistic regression and single layer perceptron neural network	BLR-SLPNN	-	۱۱/۱۱	۸۸/۸۹	۹۰/۰۰	روش پیشنهادی	

در خود پیش از بحرانی شدن شرایط داشته باشد و سپس برای تشخیص دقیق‌تر به پزشک متخصص مراجعه کند. حتی یک پزشک برای حصول اطمینان از تشخیص خود (بیماری قلبی) نیز می‌تواند از این مدل بهره ببرد.

در واقع، هدف اصلی تولید این مدل این بود که با توسعه و استفاده همگانی از این مدل، هر فردی بتواند تنها با داشتن اطلاعات بالینی و آزمایشگاهی خود، بدون نیاز به پزشک، یک تشخیص اولیه به نسبت دقیق در مورد بروز بیماری قلبی

References

1. Australian Bureau of Statistics. Causes of Death, Australia [Online]. [cited 2009]; Available from: URL: <http://abs.gov.au/ausstats/abs@.nsf/Products/696C1CF9601E4D8DCA25788400127BF0?>
2. Benjamin King RJ, Robins MW. Cancer Biology. New York, NY: Pearson/Prentice Hall; 2006.
3. Shabestan News Agency. Report of death rates about 2012 year, have a record of heart disease [Online]. [cited 2012 Aug 11]; Available from: URL: <http://www.shabestan.ir/NSite/FullStory/News/?Serv=0&Id=161273&Mode>
4. Chilnick LD. Heart Disease: an essential guide for the newly diagnosed. Boston, MA: Da Capo Press; 2008.
5. Health Central. Heart Disease. Definition [Online]. [cited 2010]; Available from: URL: <http://www.healthcentral.com/encyclopedia/408/458.html#Definition of Heart Disease>
6. King L. Taking on Heart Disease. New York, NY: Rodale; 2004.
7. Silverstein A, Silverstein Nunn L. Heart Disease. Washington, DC: Twenty-First Century Books; 2006.
8. Huang Z, Li J, Su H, Watts GS, Chen H. Large-scale regulatory network analysis from microarray data: modified Bayesian network learning and association rule mining. Decision Support Systems 2007; 43(4): 1207-25.
9. Ishibuchi H, Kuwajima I, Nojima Y. Prescreening of candidate rules using association rule mining and pareto-optimality in genetic rule selection. Lecture Notes in Computer Science 2007; 4693: 509-16.
10. Karabatak M, Cevdet Ince M. An expert system for detection of breast cancer based on association

- rules and neural network. *Expert Systems with Applications* 2009; 36(2): 3465-9.
11. Mi Shin A, Hee Lee I, Ho Lee G, Joon Park H, Seop Park H, Yoon K, et al. Diagnostic analysis of patients with essential hypertension using association rule mining. *Healthc Inform Res* 2010; 16(2): 77-81.
 12. Wang Z, Hoy WE. Is the Framingham coronary heart disease absolute risk function applicable to Aboriginal people? *Med J Aust* 2005; 182(2): 66-9.
 13. Eberhart RC, Dobbins RW, Webber WRS. CaseNet: a neural network tool for EEG waveform classification. Minneapolis, MN: IEEE Xplore; 1989. p. 60-8.
 14. Cios KJ, Chen K, Langenderfer RA. Use of neural networks in detecting cardiac diseases from echocardiographic images. *IEEE Eng Med Biol Mag* 1990; 9(3): 58-60.
 15. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. Washington, DC: IEEE Computer Society; 2008. p. 108-15.
 16. Avci E, Turkoglu I. An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. *Expert Systems with Applications* 2009; 36(2): 2873-8.
 17. Inoue T, Taguri M, Hoshi M. Prediction of nitrogen oxide concentration by a regression model. *Atmospheric Environment* 1986; 20(12): 2325-37.
 18. Inoue T, Hoshi M, Taguri M. Stability of two prediction schemes for hourly nitrogen oxide concentrations by a regression model. *Atmospheric Environment* 1987; 21(4): 920-42.
 19. Andrés Sánchez J. Calculating insurance claim reserves with fuzzy regression. *Fuzzy Sets and Systems* 2006; 157(23): 3091-108.
 20. Juran JM. Universals in Management Planning and Controlling. TPOK/Juran Institute 1954; 1: 1-13.
 21. Pareto V, Bongiorno A, Livingston A, Rogers JH. Trattato di sociologia generale. London, UK: Jonathan Cape Ltd; 1935.
 22. Horn W. Artificial Intelligence in Medicine. New York, NY: Springer; 1999. p. 201-4.
 23. Zhang GP. Neural networks for classification: a survey. *IEEE Systems, Man, and Cybernetics Society* 2000; 30(4): 451-62.
 24. Bishop CH. Neural Networks for Pattern Recognition. New York, NY: Oxford University Press, Inc; 1995.
 25. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995; 346(8983): 1135-8.
 26. Haykin S. *Neural Networks: A comprehensive foundation*. 2nd ed. New Jersey, NJ: Prentice Hall; 1999.
 27. Webb AR. *Statistical pattern recognition*. 2nd ed. Hoboken, NJ: Wiley; 2002.
 28. Nilsson NJ. *The mathematical foundations of learning machines*. San Francisco, CA: Morgan Kaufmann Publishers; 1965.
 29. Center for Machine Learning and Intelligent Systems. Cleveland heart disease data details [Online]. [cited 2010 Feb]; Available from: URL: <http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease.names>
 30. Goncalves LB, Vellasco MMBR, Pacheco MAC, de Souza FJ. Inverted hierarchical neuro-fuzzy BSP system: a novel neuro-fuzzy model for pattern classification and rule extraction in databases. *IEEE Systems, Man, and Cybernetics Society* 2006; 36(2): 236-48.
 31. Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications* 2008; 35(1-2): 82-9.
 32. Tan KC, Teoh EJ, Yu Q, Goh KC. A hybrid evolutionary algorithm for attribute selection in data mining. *Expert Systems with Applications* 2009; 36(4): 8616-30.
 33. Luukka P, Lampinen J. A Classification method based on principal component analysis and differential evolution algorithm applied for prediction diagnosis from clinical EMR heart data sets. *Computational Intelligence in Optimization* 2010; 7: 236-83.
 34. Nawi NM, Ghazali R, Mohd Salleh MN. The development of improved back-propagation neural networks algorithm for predicting patients with heart disease. *Information Computing and Applications* 2010; 6377: 317-24.
 35. Ozciit A, Gulten A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed* 2011; 104(3): 443-51.
 36. Cui Z, Yang C, Sanyal S. Training artificial neural networks using APPM. *International Journal of Wireless and Mobile Computing* 2012; 5: 168-74.
 37. Nahar J, Imam T, Tickle KS, Phoebe Chen Y. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Systems with Applications* 2013; 40(1): 96-104.