

Original Article

Diagnosis of Heart Disease Using Binary Grasshopper Optimization Algorithm and K-Nearest Neighbors

Neda dezhaloud¹ , Farhad Soleimanian Gharehchopogh² 

¹ Msc Student, Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran.

² Assistant Professor, Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran.

ARTICLE INFO

Corresponding Author:

Farhad Soleimanian Gharehchopogh

e-mail addresses:

bonab.farhad@gmail.com

Received: 07/July/2020

Modified: 12/Sep/2020

Accepted: 20/Sep/2020

Available online: 07/Nov/2020

Keywords:

Heart Disease Detection

Binary grasshopper optimization Algorithm

K-Nearest Neighbor

Classification

ABSTRACT


Introduction: The heart is one of the main organs of the human body, and its unhealthiness is an important factor in human mortality. Heart disease may be asymptomatic, but medical tests can predict and diagnose it. Diagnosis of heart disease requires extensive experience of specialist physicians. The aim of this study is to help physicians diagnose heart disease based on hybrid Binary Grasshopper Optimization (BGO) Algorithm and K-Nearest Neighbors (KNN). The BGO algorithm is used for feature selection (FS), and the KNN is used for classification.

Methods: In this study, the medical records of 270 patients in the field of heart disease with 13 features were evaluated. The number of patients is equal to 120 and the absence of disease is equal to 150, so the data set is balanced. Patient information is taken from the standard UCI (University of California, Irvine) database. The evaluation of the proposed model has been done in MATLAB simulation.

Results: According to the evaluations, the accuracy was 89.82%, the sensitivity was 89.61%, and the specificity was 90.41%, which are acceptable compared to the results of previous studies in the field of heart disease. Also, the percentage of accuracy of the proposed method based on 7 features (Age, Sex, Chest Pain, BP, Electrocardiographic, Angina, and Thallium) is equal to 90.35%.

Conclusion: According to the results of this study, for the diagnosis of heart disease, the proposed method has been more effective in diagnosing the disease and selecting important features in comparison with previous methods.

تشخیص بیماری قلبی با استفاده از الگوریتم بهینه‌سازی ملخ دودویی و k نزدیک‌ترین همسایه

ندا دژآلود^۱، فرهاد سلیمانیان قره چپق^{۲*} 

^۱ دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران.
^۲ استادیار، گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران.

اطلاعات مقاله

نویسنده مسئول:

فرهاد سلیمانیان قره چپق
رایانامه:

bonab.farhad@gmail.com

وصول مقاله: ۹۹/۰۴/۱۷

اصلاح نهایی: ۹۹/۰۶/۲۲

پذیرش نهایی: ۹۹/۰۶/۳۰

انتشار آنلاین: ۹۹/۰۸/۱۷

واژه‌های کلیدی:

تشخیص بیماری قلبی
الگوریتم بهینه‌سازی ملخ دودویی
الگوریتم k نزدیک‌ترین همسایه
طبقه‌بندی

چکیده

مقدمه: قلب یکی از ارگان‌های اصلی بدن انسان است و سالم نبودن آن عامل مهمی در مرگ‌ومیر انسان‌ها است. بیماری قلبی ممکن است بدون علامت باشد اما می‌توان از طریق آزمایش‌های پزشکی این نوع بیماری را پیش‌بینی و تشخیص داد. تشخیص بیماری قلبی به تجربیات زیاد پزشکان متخصص نیاز دارد. هدف مطالعه حاضر، تشخیص بیماری قلبی به‌منظور کمک به پزشکان بر مبنای ترکیب الگوریتم بهینه‌سازی ملخ دودویی و k نزدیک‌ترین همسایه است. از الگوریتم بهینه‌سازی ملخ دودویی برای انتخاب ویژگی‌ها و از الگوریتم K نزدیک‌ترین همسایه برای طبقه‌بندی استفاده شده است.

روش‌ها: این مطالعه از نوع توصیفی-تحلیلی بود. در این مطالعه، پرونده پزشکی ۲۷۰ بیمار در حوزه بیماری قلبی با تعداد ۱۳ ویژگی بررسی شد. تعداد بیماران مبتلا برابر با ۱۲۰ و فقدان بیماری برابر با ۱۵۰ بود، لذا مجموعه داده در حالت متوازن است. اطلاعات بیماران از پایگاه داده استاندارد UCI استخراج شد. ارزیابی مدل پیشنهادی در شبیه‌سازی MATLAB انجام شد.

یافته‌ها: با توجه به ارزیابی‌های انجام شده بر روی روش پیشنهادی، درصد صحت برابر با ۸۹/۸، درصد حساسیت برابر با ۸۹/۶ و درصد ویژگی برابر با ۹۰/۴ به دست آمد که در مقایسه با نتایج مطالعات انجام شده در حوزه بیماری قلبی، دقت به‌دست‌آمده روش پیشنهادی، قابل قبول است. همچنین، درصد صحت روش پیشنهادی بر مبنای هفت ویژگی (Age, Sex, Chest Pain, BP, Electrocardiographic, Angina, Thallium) برابر با ۹۰/۳ درصد به‌دست آمد.

نتیجه‌گیری: با توجه به نتایج حاصل از این مطالعه، برای تشخیص بیماری قلبی، روش پیشنهادی در تشخیص بیماری و انتخاب ویژگی‌های مهم نسبت به روش‌های قبلی بهتر عمل کرده است.

اعتبارسنجی و بررسی دقت مدل ساخته شده استفاده می‌شود. [۱۲] کاظمی و همکاران [۱۳] برای تشخیص بیماری قلبی از فن داده‌کاوی شبکه عصبی پرسپترون چندلایه استفاده کرده‌اند. نتایج نشان داده است که دقت تشخیص در مرحله آموزش برابر با ۸۷/۷۵ درصد و برای مجموعه رکورد‌های آزمون برابر با ۸۳/۳۳ درصد بوده است. هروی و ستایشی [۱۴] مدلی با عنوان «تشخیص هوشمند و سریع بیماری قلبی بر اساس هم‌افزایی شبکه‌های عصبی خطی و روش رگرسیون منطقی (Logic Regression)» پیشنهاد داده‌اند. میانگین خطای مشاهده‌شده روی کل مجموعه‌ی داده‌ها برابر با ۱۱/۱۱ درصد بوده است. صباغ گل [۱۵] مدلی با عنوان «تشخیص بیماری عروق کرونر قلبی با استفاده از درخت تصمیم C4.5» به منظور تشخیص بیماری قلبی پیشنهاد داده است. درصد صحت برابر با ۷۲/۶ و درصد حساسیت و درصد ویژگی به ترتیب برابر با ۸۴/۱ و ۶۱/۸ بوده‌اند. محمودی [۱۶] مدلی با عنوان «طراحی سیستم پیش‌بینی بیماری قلبی - عروقی با استفاده از ماشین بردار پشتیبان» پیشنهاد داده است. معیارهای صحت و حساسیت به ترتیب برابر با ۸۵ و ۸۵/۵ بوده است. نچپال و همکاران [۱۷] برای انتخاب ویژگی از الگوریتم جستجوی گرانشی استفاده کرده‌اند. نتایج نشان داده است که درصد صحت مدل الگوریتم جستجوی گرانشی (یکی از الگوریتم‌های فرا ابتکاری) با k نزدیک‌ترین همسایه (یکی از الگوریتم‌های یادگیری ماشین برای طبقه‌بندی) برابر با ۶۴/۸ و الگوریتم بهینه‌سازی اجتماع ذرات (یکی از الگوریتم‌های فرا ابتکاری) با k نزدیک‌ترین همسایه برابر با ۸۲/۶ بوده است. ذباح و همکاران [۱۸] مدلی با عنوان «بررسی تأثیر پارامترهای پیوسته در تشخیص بیماری عروق کرونر قلبی با استفاده از شبکه‌های عصبی مصنوعی» برای تشخیص بیماری قلبی پیشنهاد داده‌اند. درصد صحت شبکه عصبی پرسپترون چندلایه برابر با ۸۸ بوده است. ایده این مطالعه ارائه روشی مبتنی بر انتخاب ویژگی‌های مهم در تشخیص بیماری قلبی با استفاده از الگوریتم بهینه‌سازی ملخ دودویی [۱۹] است که هدف آن بالا بردن قابلیت تعمیم طبقه‌بندی k نزدیک‌ترین همسایه [۲۰] به منظور افزایش طبقه‌بندی داده‌های بیماری قلبی

قلب یکی از ارگان‌های اصلی بدن انسان است و کارکرد نادرست آن عامل مهمی در مرگ‌ومیر انسان‌ها است. [۱] بیماری‌های قلبی دارای انواع مختلفی می‌باشند که آسیب‌های اساسی به شریان‌های قلب، عملکرد دریچه قلب، عضله قلب و ضربان قلب وارد می‌نمایند. [۲] طبق گزارش سازمان بهداشت جهانی، بیماری‌های قلبی یکی از شایع‌ترین دلایل مرگ در جهان است. طبق آمارهای سال ۲۰۱۶ تعداد ۵۶/۹ میلیون مرگ در سراسر جهان بنا به ۱۰ دلیل اتفاق افتاده‌اند. بیماری قلبی بزرگ‌ترین کشنده در جهان است که در مجموع ۱۵/۲ میلیون مرگ در سال ۲۰۱۶ بر مبنای بیماری قلبی گزارش شده است. [۳] طبق گزارش‌های سازمان بهداشت جهانی استفاده از روش‌های دقیق و معاینه دوره‌ای کارآمد برای تشخیص بیماری‌های قلبی بسیار مهم است. [۴]

بیماری قلبی ممکن است بدون علامت باشد اما می‌توان این موارد را از طریق الگوریتم‌های نظارتی تشخیص داد. در الگوریتم‌های نظارتی، از نمونه‌هایی با برچسب مشخص برای آموزش الگوریتم استفاده می‌شود. [۵] به الگوریتم آموزش داده می‌شود که چگونه نمونه بیمار و سالم را طبقه‌بندی کند. پس از آموزش، الگوریتم می‌تواند داده‌های جدید بدون برچسب را طبقه‌بندی کند تا مشخص کند نمونه جدید متعلق به نمونه بیمار یا سالم است. [۵] کشف کردن الگوها و دانش نهفته در داده‌های بیماران می‌تواند به پزشکان مراکز درمانی در جهت ارتقاء و بهبود فرآیندهای تشخیص کمک شایانی کند و بدین ترتیب آن‌ها را در تصمیم‌گیری بهتر و مهیا کردن دوره درمانی مناسب برای بیماران هدایت نماید. [۸-۶]

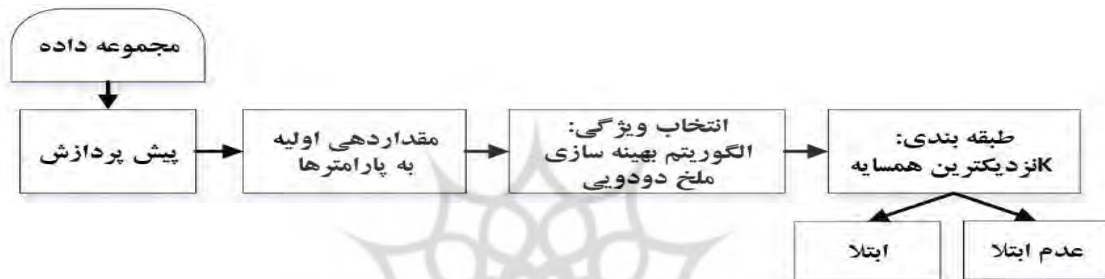
الگوریتم‌های طبقه‌بندی در مرحله آموزش بر مبنای نمونه‌های از قبل دسته‌بندی شده، مجموعه‌ای از قواعد تصمیم‌گیری را منظور تفکیک صحیح نمونه‌ها تولید می‌نمایند. [۹، ۱۰] در الگوریتم‌های طبقه‌بندی مجموعه داده اصلی به دو مجموعه داده با عنوان مجموعه داده‌های آموزشی و آزمایشی تقسیم می‌شود. [۱۱] با استفاده از مجموعه داده آموزشی مدل تشخیص ساخته می‌شود و از مجموعه داده آزمایشی برای

دارد. ابتدا عملیات پیش پردازش به منظور نرمال سازی داده ها انجام شد. سپس، توسط الگوریتم بهینه سازی ملخ دودویی، انتخاب ویژگی بر روی نمونه ها انجام و توسط الگوریتم k نزدیک ترین همسایه عمل طبقه بندی شد. برای شبیه سازی مدل پیشنهادی از نرم افزار MATLAB ۲۰۱۷ استفاده شده است و از ۸۰ درصد داده ها برای مرحله آموزش و ۲۰ درصد باقیمانده برای مرحله آزمون استفاده شده است. در شکل های یک و دو بلوک دیاگرام و روند نمای روش پیشنهادی نشان داده شده است.

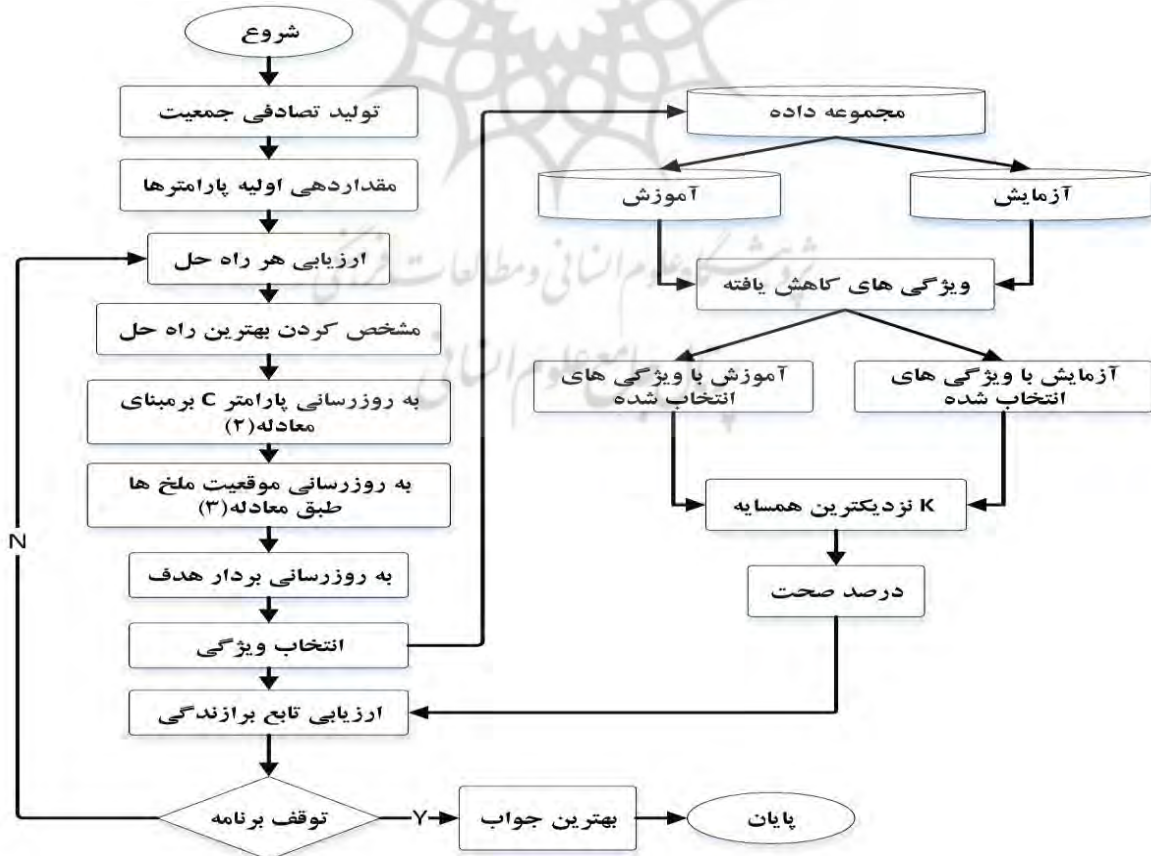
است. عدم آگاهی بیماران از بیماری قلبی یک مانع بزرگ برای به حداکثر رساندن بهبود این بیماران در طول دوره بیماری است. هدف اصلی در این مطالعه استفاده از الگوریتم بهینه سازی ملخ دودویی بر روی مجموعه داده استاندارد بیماران قلبی (۲۷۰ نمونه با ۱۳ ویژگی) برگرفته از UCI [۲۱] است که به منظور تشخیص ویژگی های مهم بکار گرفته می شود.

روند روش پیشنهادی

مطالعه حاضر، وظیفه تشخیص بیماری قلبی بر مبنای الگوریتم بهینه سازی ملخ دودویی و k نزدیک ترین همسایه را بر عهده



شکل ۱: بلاک دیاگرام روش پیشنهادی



شکل ۲: روند نمای روش پیشنهادی

نداشتن بیماری قلبی بر اساس مقدار ویژگی‌ها اشاره دارد که مقدار یک بیانگر عدم ابتلا و مقدار دو بیانگر ابتلا است. جدول شماره یک نشان‌دهنده ویژگی‌های پایگاه داده قلب مورد استفاده است.

(۱) گردآوری مجموعه داده

مجموعه داده قلب، از پایگاه داده استاندارد UCI گرفته شده است که شامل ۲۷۰ نمونه (۱۲۰ نفر بیمار و ۱۵۰ نفر عدم بیماری) و ۱۳ ویژگی (ویژگی ۱۴م مربوط به نوع کلاس یعنی بیمار یا عدم بیماری است) است. [۲۱] این پایگاه داده یک ویژگی کلاس (تشخیص) دارد که به وجود یا وجود

جدول ۱: ویژگی‌ها و مقدار آن‌ها در مجموعه داده قلب [۲۱]

ویژگی‌ها	توضیحات	مجموعه مقادیر	فراوانی	درصد فراوانی
Age	سن	۷۶ تا ۲۹	کمتر از ۵۳ سال	۵۰/۲۳
			بیشتر از ۵۳ سال	۴۹/۷۷
Sex	جنسیت	۱: مرد ۰: زن	۱	۶۳/۱
			۰	۳۶/۹۹
Chest Pain Type	میزان درد قفسه سینه	۱: آنژین معمولی ۲: آنژین غیر معمولی ۳: بدون درد ۴: بدون علامت	۱	۲۲
			۲	۲۱/۴۶
			۳	۳۵/۱۶
			۴	۳۳/۷۹
Resting Blood Pressure (BP)	فشارخون در زمان استراحت	۲۰۰ تا ۹۴	≥ 140	۷۰/۳۲
			< 140	۲۹/۶۸
Serum Cholesterol in mg/dl (SC)	کلسترول خون	۵۶۴ تا ۱۲۶	مطلوب (زیر ۲۰۰)	۱۶/۸۹
			بین ۲۰۰-۲۳۹	۳۲/۸۸
			خطرناک (بالای ۲۴۰)	۵۰/۲۳
Fasting Blood Sugar (FBS) > 120 mg/dl	قند خون ناشتا	۱: دارد ۰: ندارد	۱	۸۷/۶۷
			۰	۱۲/۳۳
Resting Electrocardiographic Results (values 0,1,2)	نوار قلب	۰: نرمال ۱: موج غیر قلبی ۲: موج افزایش مقطعی	۰	۵۳/۹
			۱	۰/۵
			۲	۴۵/۶
Maximum Heart Rate (HR) Achieved	حداکثر ضربان قلب	۷۱ تا ۲۰۲	> 100	۹۸/۶۳
			< 100	۱/۳۷
Exercise Induced Angina	آنژین ناشی از ورزش	۱: بلی ۰: خیر	۱	۲۲
			۰	۷۸

۱/۶۷	≥ 0.8	۰ تا ۶/۲	کاهش موج ST (موج سینوسی) ایجاد شده	old peak = ST (Sinus Tachycardia) depression induced by exercise relative to rest
۳۲/۹	> 0.8			
۸۵/۵	۱	۱: بالا رفتن	بیان کننده شیب قطعه ST در زمان	The Slope of the Peak Exercise ST Segment
۳۶	۲	۲: یکنواخت	حداکثر ورزش	
۵/۵	۳	۳: پایین اومدن		
۷۱/۲	۰	۰ تا ۳	تعداد عروق اصلی: این صفت بیانگر تعداد رگ‌هایی که در فلوروسکوپی دیده می‌شود.	Number of Major Vessels (0-3) Colored by Fluoroscopy
۱۸/۳	۱			
۶/۵	۲			
۳	۳			
۶۹/۴	۳	۳: ضایعه ثابت	اسکن تالیوم است که شامل ۳ مقدار ضایعه ثابت، نرمال و ضایعه قابل برگشت است	Thallium
۴/۲	۶	۶: ضایعه نرمال		
۲۶/۴	۷	۷: ضایعه قابل برگشت		
۵۵/۶	۱	۱ (عدم ابتلا)، ۲ (ابتلا)	کلاس	Class

$$N_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

۴) انتخاب ویژگی

در این مطالعه، یک روش دودویی از الگوریتم بهینه‌سازی ملخ برای مسئله انتخاب ویژگی استفاده شد. برای تبدیل الگوریتم بهینه‌سازی ملخ به حالت دودویی از تابع سیگموئید طبق معادله (۴) استفاده شده است. در الگوریتم‌های فرا ابتکاری برای تبدیل حالت پیوسته به گسسته از تابع سیگموئید استفاده می‌شود. [۱۱] در مسائل بهینه‌سازی، راه‌حل‌های بهینه در فضای جستجوی D بعدی توسط عامل‌های گروه جستجو می‌شوند و هر عامل موجود در گروه مانند یک نقطه یا موقعیتی در فضای D-بعدی به نظر می‌رسد. عامل i^{th} موجود در گروه بیانگر یک بردار D بعدی $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ است. در مسائل بهینه‌سازی پیوسته توسط بردار x_i مقادیر واقعی را می‌گیرد، اما در مسائل بهینه‌سازی دودویی مقادیر صفر یا یک را می‌گیرد.

۲) مقداردهی اولیه به پارامترها

در این قسمت پارامترهای اولیه برای الگوریتم بهینه‌سازی ملخ دودویی مقداردهی شد. در الگوریتم بهینه‌سازی ملخ دودویی، تعداد اولیه ملخ‌ها و همچنین تعداد تکرار مقداردهی اولیه می‌شود.

۳) پیش‌پردازش داده‌ها

نکته اساسی که باید در آموزش مدل پیشنهادی مورد توجه قرار گیرد نرمال‌سازی داده‌ها قبل از استفاده در مدل است. اصولاً وارد کردن داده‌ها به صورت خام منجر به کاهش صحت می‌شود [۱۴]، در مطالعه حاضر از نرمال‌سازی min-max (استاندارد کردن داده‌ها در یک بازه مشخص) استفاده شده است که مقایسه بین داده‌ها را تسهیل می‌کند و مطابق با معادله (۱) تعریف می‌شود. داده کمینه (x_{min}) و داده بیشینه (x_{max}) به ترتیب کمترین و بیشترین مقدار در هر متغیر (x_i) می‌باشد.

دام است و C یک ثابت کاهش برای کم کردن منطقه دافعه و جاذبه است. پارامتر d_{ij} فاصله بین ملخ i با ملخ j را نشان می‌دهد. پارامترهای X_i و X_j به ملخ i و j اشاره می‌کنند. پارامتر S یک تابع برای تعریف فشار نیروی اجتماعی است که به صورت $S(r) = fe^{-r} - e^{-r}$ تعریف می‌شود که در آن f (مقدار برابر با ۰/۵) نشان‌دهنده شدت جاذبه و I نشان‌دهنده طول مقیاس جاذبه (مقدار برابر با ۱/۵) است.

۴-۵) تابع سیگموئید: برای تبدیل حالت پیوسته به دودویی

$$T(\Delta X_i) = \frac{1}{1 + e^{-\Delta X_i}} \quad (۴)$$

۵-۵) به روزرسانی بردار هدف در صورتی که ملخی با برازندگی بهتر پیدا شود. پارامتر X_{best} و X_i بیانگر بهترین بردار و بردارهای همسایه هستند و پارامتر X'_i بیانگر بردار جدید است.

$$X'_i = \{X_{best} \oplus X_i\} \quad (۵)$$

به منظور جابه‌جایی مقدار بردار از به روزرسانی بردار استفاده می‌شود. بدین معنی که تنوع در بین جمعیت ایجاد شود و فضای جستجو به طور کامل کاوش شود. برای تولید بردارهای جدید از عملگر XOR استفاده می‌شود. در شکل (۳) تولید بردار جدید بر مبنای به روزرسانی نشان داده شده است [۱۹].

ویژگی‌ها	F1	F2	F3	F4	F5	F6	F7	F8	F9
X_{best}	۱	۰	۱	۰	۰	۱	۱	۰	۱
X_i	۱	۱	۰	۱	۰	۰	۱	۱	۱
X'_i	۰	۱	۱	۱	۰	۱	۰	۱	۰

شکل ۳: تولید بردار جدید بر مبنای به روزرسانی

۶) برگرداندن بهترین راه حل (Target) به عنوان جواب نهایی
 ۷) ارزیابی برازندگی راه حل‌ها (معادله (۶)). پارامتر accuracy به معنای دقت طبقه‌بندی (تعداد نمونه‌های درست طبقه‌بندی شده / تعداد کل نمونه‌ها)، یک عامل وزنی است که مقدار آن در فضای [۰،۱] است و در این

در روش پیشنهادی، بردارهای دودویی به طول N تولید می‌شوند، به طوری که N تعداد ویژگی‌ها و ابعاد فضای جستجو را دربر دارد. مراحل روش پیشنهادی به منظور انتخاب ویژگی به شرح زیر است:

۱) مقداردهی اولیه پارامترها: تعیین مقدار پارامترهای حد پایین (Cmin) و حد بالا (Cmax) برای جستجو در فضای مسئله و حداکثر تعداد دور

۲) تولید تصادفی جمعیت ملخ‌ها در فضای جستجو

۳) ارزیابی راه حل‌ها در جمعیت

۴) مشخص کردن بهترین راه حل (راه حل هدف (Target))

۵) تکرار برنامه تا زمان اتمام شرط (تعداد تکرار)

۵-۱) به روزرسانی ضرب C با استفاده از معادله (۲). این مکانیزم باعث افزایش تعداد تعاملات در بهره‌برداری (در هر تکرار بهترین راه حل در همسایگی راه حل بهینه جاری که امکان بهبود دارد یافت می‌شود) می‌گردد. پارامتر i بیانگر تعداد تکرار جاری و پارامتر L بیانگر تعداد کل تکرارهای الگوریتم است.

$$C = Cmax - i \frac{Cmax - Cmin}{L} \quad (۲)$$

۵-۲) نرمال‌سازی فاصله بین ملخ‌ها

۵-۳) به روزرسانی موقعیت هر ملخ (معادله (۳)). در یک فضای جستجوی دودویی، ممکن نیست موقعیت ملخ‌ها با افزودن مقادیر به روزرسانی شود، زیرا بردارهای موقعیت ملخ‌ها فقط می‌توانند شامل صفر یا یک باشند. استفاده از توابع انتقال یکی از ساده‌ترین روش‌ها برای تبدیل یک الگوریتم از حالت پیوسته به دودویی است. در تابع انتقال سیگموئید، احتمال به روزرسانی راه حل‌ها به صورت دودویی از صفر به یک و برعکس تعریف می‌شود. در الگوریتم بهینه‌سازی ملخ دودویی، تغییر موقعیت ذرات بر اساس معادله (۳) تعریف می‌شود.

$$\Delta X = c \left(\sum_{j=1, j \neq i}^N c \frac{Ub_d - Lb_d}{2} S(|x_j^d - x_i^d|) \frac{x_j - x_i}{d_{ij}} \right) \quad (۳)$$

در معادله (۳) Ub_d حد بالا (مقدار بازه صفر و یک) در بعد dam است و Lb_d حد پایین (پایین در بازه صفر و یک) در بعد

آن‌ها مثبت بوده و الگوریتم دسته‌بندی دسته آن‌ها را به اشتباه منفی تشخیص داده است.

جدول ۲: معیارهای ارزیابی

معادلات	توضیحات
$\text{دقت} = \frac{TP}{TP + FP}$	دقت
$\text{حساسیت} = \frac{TP}{TP + FN}$	حساسیت
$\text{ویژگی} = \frac{TN}{TN + FP}$	ویژگی
$F\text{-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$	F-Measure
$\text{صحت} = \frac{TP + TN}{TP + TN + FP + FN}$	صحت

دقت نشان‌دهنده تعداد نمونه‌های درست به تعداد کل نمونه‌ها است. حساسیت یا بازخوانی یعنی تعداد نمونه‌های مثبتی که بر تمام نمونه‌های واقعاً مثبت به درستی تشخیص داده شده‌اند. ویژگی، بیانگر تعداد نمونه‌های سالم است که بیمار نیستند به عبارت دیگر نسبتی از افراد سالم که نتیجه آزمایش آن‌ها منفی بوده است. درصد صحت یک روش طبقه‌بندی بر روی مجموعه داده‌های آموزشی، درصد مشاهداتی از مجموعه آموزش است که به درستی توسط روش مورد استفاده طبقه‌بندی شده است. معیار F-Measure بر مبنای ترکیب دقت و حساسیت (بازخوانی) به دست می‌آید.

۱۱-۱-۱-۱

در جدول شماره سه، نتایج روش پیشنهادی در مرحله آزمون بر مبنای تعداد تکرار (تعداد تکرار برای الگوریتم بهینه‌سازی ملخ دودویی)، بدون انتخاب ویژگی و تعداد k نشان داده شده است. تعداد تکرار، بیانگر تعداد چرخه الگوریتم بهینه‌سازی ملخ دودویی است که بهترین ویژگی‌ها را به ترتیب در فضای مسئله کشف نماید و سپس ویژگی‌های انتخاب شده به الگوریتم k نزدیک‌ترین همسایه داده می‌شوند. تعداد K بیانگر تعداد همسایگان تأثیرگذار در تصمیم‌گیری است. تعداد تکرارها در روش پیشنهادی برابر با ۵۰ تا ۲۵۰ می‌باشند.

مطالعه مقدار آن برابر با ۰/۵ است. پارامتر L_f تعداد ویژگی‌های انتخاب شده و L_t تعداد کل ویژگی‌ها است.

$$\text{Fitness} = \text{maximize}(\text{accuracy} + w_f \times (1 - \frac{L_f}{L_t})) \quad (۶)$$

۸) پایان: اتمام تکرار (مرحله انتخاب و ویژگی انجام شده است و بهترین ویژگی‌ها نمایش داده می‌شوند).

۵) طبقه‌بندی

در این الگوریتم یک نمونه آموزشی از بردارها انتخاب $X_j = (x_{j1}, \dots, x_{jd}), j = 1, 2, \dots, N$ می‌شود. پارامتر d به تعداد ویژگی‌هایی که انتخاب شده‌اند اشاره دارد و x_{jk} نمونه j را بر روی ویژگی k نشان می‌دهد. در طبقه‌بندی نزدیک‌ترین همسایه، نمونه ناشناخته $Z_j = (z_{j1}, \dots, z_{jd})$ بر اساس فاصله اقلیدسی (فاصله معمولی دونقطه) با استفاده از معادله (۷) طبقه‌بندی می‌شود. پس از محاسبه فاصله اقلیدسی، Z_i درون کلاسی قرار می‌گیرد که به نمونه‌های آموزشی آن کلاس نزدیک‌تر باشد.

$$D_{ij} = \sqrt{\sum_{k=1}^d \|z_{ik} - x_{jk}\|^2} \quad (۷)$$

۶) معیارهای ارزیابی

در این مطالعه، به منظور ارزیابی مدل پیشنهادی از پنج معیار دقت، بازخوانی (حساسیت)، ویژگی، و F-Measure و صحت (Accuracy) طبق جدول شماره دو استفاده شده است. پارامترهای درست مثبت (TP)، درست منفی (TN)، کاذب مثبت (FP)، کاذب منفی (FN) از پارامترهای اصلی برای معیارهای ارزیابی هستند. درست مثبت، تعداد نمونه‌های واقعاً مثبتی می‌باشند که به درستی تشخیص داده شده‌اند. درست منفی، تعداد نمونه‌های واقعاً منفی می‌باشند که به درستی تشخیص داده شده‌اند. پارامتر کاذب مثبت بیانگر تعداد نمونه‌های است که مبتلابه بیماری نبوده‌اند، اما نتایج آزمایش آن‌ها را به غلط، مثبت تشخیص داده است. پارامتر کاذب منفی، بیانگر تعداد رکوردهایی است که دسته واقعی

k در الگوریتم k نزدیکترین همسایه) برابر با سه است. البته مدل پیشنهادی با تکرارهای بالاتری هم ارزیابی شد، اما بهترین جوابها در ۲۵۰ بار تکرار بودند و لذا از بیان و تعریف تکرارهای دیگر اجتناب شده است.

همچنین تعداد k (تعداد k در الگوریتم k نزدیکترین همسایه) در روش پیشنهادی در محدوده سه تا پنج است. این محدوده به این دلیل انتخاب شده است که بهترین جوابها را شامل می شود. طبق نتایج به دست آمده، بهترین مقدار برای k (تعداد

جدول ۳: نتایج مدل پیشنهادی بر مبنای تعداد تکرار و بدون انتخاب ویژگی

تعداد تکرار	تعداد k	دقت	حساسیت	ویژگی	F-Measure	صحت
۵۰	۳	۸۱/۳	۸۱/۵	۸۳/۱	۸۱/۴	۸۱/۸
	۴	۸۰/۱	۸۰/۲	۸۲/۵	۸۰/۲	۸۰/۶
	۵	۷۹/۸	۸۰/۰	۸۱/۱	۷۹/۹	۸۰/۲
۱۰۰	۳	۸۴/۲	۸۴/۵	۸۵/۰	۸۴/۳	۸۴/۷
	۴	۸۳/۹	۸۳/۶	۸۴/۶	۸۳/۷	۸۴/۲
	۵	۸۳/۱	۸۳/۳	۸۴/۵	۸۳/۲	۸۳/۵
۱۵۰	۳	۸۴/۴	۸۴/۳	۸۵/۱	۸۴/۲	۸۴/۷
	۴	۸۳/۷	۸۳/۵	۸۴/۲	۸۳/۶	۸۴/۱
	۵	۸۳/۱	۸۳/۴	۸۵/۱	۸۳/۲	۸۳/۶
۲۰۰	۳	۸۷/۲	۸۷/۴	۸۸/۳	۸۷/۳	۸۷/۵
	۴	۸۶/۳	۸۶/۵	۸۶/۵	۸۶/۳	۸۶/۹
	۵	۸۵/۹	۸۶/۰	۸۷/۹	۸۶/۴	۸۶/۳
۲۵۰	۳	۸۹/۴	۸۹/۶	۹۰/۴	۸۹/۳	۸۹/۸
	۴	۸۹/۱	۸۹/۲	۹۰/۱	۸۹/۱	۸۹/۳
	۵	۸۸/۱	۸۸/۲	۸۹/۲	۸۹/۱	۸۸/۶

با تعداد ویژگی های کمتری انجام می شود. برای مثال اگر تعداد ویژگی ها برابر با هفت (سن، جنس، میزان درد قفسه سینه، فشارخون در زمان استراحت، نوار قلب، آنزیم ناشی از ورزش، اسکن تالیوم) باشند آنگاه درصد صحت برابر با ۹۰/۳ درصد است و اگر تعداد ویژگی ها برابر با ۱۰ باشند آنگاه درصد صحت برابر با ۹۰/۲ درصد است.

در جدول شماره چهار، نتایج مدل پیشنهادی بر مبنای انتخاب ویژگی نشان داده شده است. ترتیب ویژگی های انتخاب شده بر مبنای جدول شماره یک است، مثلاً ویژگی یک مربوط به سن است. نتایج جدول شماره سه بر مبنای ۲۵۰ بار تکرار حاصل شده است. اگر تعداد ویژگی ها کمتر باشند درصد صحت مدل پیشنهادی بیشتر است. به این دلیل که طبقه بندی

جدول ۴: نتایج مدل پیشنهادی بر مبنای انتخاب ویژگی

ردیف	تعداد ویژگی	ویژگی های انتخاب شده از مجموعه داده	دقت	حساسیت	F-Measure	صحت
۱	۷	Age, Sex, Chest Pain, BP, SC, Electrocardiographic, Angina	۹۰/۱	۹۰/۱	۹۰/۰	۹۰/۳
۲	۷	Age, Sex, Chest Pain, SC, FBS, Angina, Vessels	۹۰/۱	۸۹/۹	۸۹/۹	۹۰/۳
۳	۷	Age, BP, FBS, Electrocardiographic, HR, ST, ST Segment	۹۰/۲	۹۰/۲	۹۰/۱	۹۰/۳

۹۰/۳	۹۰/۲	۹۰/۲	۹۰/۱	Age, Sex, Chest Pain, BP, Electrocardiographic, Angina, Thallium	۷	۴
۹۰/۲	۹۰/۰	۹۰/۱	۹۰/۰	Sex, Chest Pain, BP, SC, Electrocardiographic, Angina, ST Segment, Vessels	۸	۵
۹۰/۲	۹۰/۰	۹۰/۰	۸۹/۹	Age, Chest Pain, BP, SC, FBS, HR, ST Segment, Vessels	۸	۶
۹۰/۲	۹۰/۱	۹۰/۱	۹۰/۱	Age, Chest Pain, BP, FBS, Electrocardiographic, ST, ST Segment, Thallium	۸	۷
۹۰/۲	۸۹/۹	۹۰/۱	۸۹/۸	BP, SC, FBS, Electrocardiographic, HR, Angina, ST, ST Segment, Vessels, Thallium	۹	۸
۹۰/۲	۸۹/۵	۸۹/۳	۸۹/۷	BP, SC, Electrocardiographic, HR, Angina, ST, ST Segment, Vessels, Thallium	۹	۹
۹۰/۲	۹۰/۱	۹۰/۲	۹۰/۰	Age, SC, FBS, Electrocardiographic, HR, Angina, ST, ST Segment, Thallium	۹	۱۰
۹۰/۲	۹۰/۰	۹۰/۰	۸۹/۹	Age, Sex, SC, FBS, Electrocardiographic, HR, Angina, ST, Vessels, Thallium	۱۰	۱۱
۹۰/۱	۸۹/۷	۸۹/۹۲	۸۹/۸۱	Age, Chest Pain, BP, FBS, Electrocardiographic, HR, ST, ST Segment, Vessels, Thallium	۱۰	۱۲
۹۰/۲	۹۰/۰	۹۰/۱	۹۰/۰	Sex, Chest Pain, BP, SC, FBS, Angina, ST, ST Segment, Vessels, Thallium	۱۰	۱۳
۹۰/۱	۸۹/۵	۸۹/۴	۸۹/۶	Age, Sex, Chest Pain, BP, SC, FBS, Electrocardiographic, HR, ST, ST Segment, Thallium	۱۱	۱۴
۹۰/۱	۹۰/۰	۹۰/۱	۸۹/۵	Sex, Chest Pain, BP, SC, FBS, Electrocardiographic, HR, Angina, ST, ST Segment, Vessels	۱۱	۱۵
۹۰/۱	۸۹/۳	۸۹/۲	۸۹/۳	Age, Sex, Chest Pain, BP, SC, FBS, Electrocardiographic, HR, Angina, ST, ST Segment,	۱۱	۱۶
۹۰/۰	۸۹/۴	۸۹/۴	۸۹/۶	Age, Sex, Chest Pain, SC, FBS, Electrocardiographic, HR, Angina, ST, ST Segment, Vessels, Thallium	۱۲	۱۷
۸۹/۹	۸۹/۹	۹۰/۰	۸۹/۳	Age, Sex, Chest Pain, BP, SC, FBS, Electrocardiographic, Angina, ST, ST Segment, Vessels, Thallium	۱۲	۱۸
۸۹/۸	۸۹/۳	۸۹/۴	۸۹/۱	همه ویژگی‌ها	۱۳	۱۹

درخت تصمیم‌گیری C4.5 [۱۵]، ماشین بردار پشتیبان [۱۶]، الگوریتم جستجوی گرانشی با k نزدیک‌ترین همسایه [۱۷]، الگوریتم بهینه‌سازی اجتماع ذرات با k نزدیک‌ترین همسایه [۱۷]، شبکه عصبی مصنوعی چندلایه [۱۸]، الگوریتم سیستم ایمنی مصنوعی [۲۲]، الگوریتم بهینه‌سازی اجتماع سالپ

در جدول شماره پنج مقایسه روش پیشنهادی با روش‌های دیگر بر مبنای معیار درصد صحت نشان داده شده است. مقایسه‌ها بر مبنای تمام ویژگی‌ها انجام شده است. در مقایسه از مدل‌های شبکه عصبی با ساختار پرسپترون چندلایه [۱۳]، رگرسیون منطقی و شبکه عصبی پرسپترون تک لایه [۱۴]،

[۲۳]، الگوریتم ژنتیک [۲۳]، الگوریتم بهینه‌سازی جنگل تصادفی [۲۴] و الگوریتم سیاه‌چاله باینری [۲۵] استفاده شد.

جدول ۵: مقایسه روش پیشنهادی با روش‌های دیگر بر مبنای کل

ویژگی‌ها

رفرنس	مدل	صحت
[۱۳]	شبکه عصبی با ساختار پرسپترون چندلایه	۸۳/۳
[۱۴]	رگرسیون منطقی و شبکه عصبی پرسپترون تک لایه	۹۰/۰
[۱۵]	درخت تصمیم‌گیری C4.5	۷۲/۶
[۱۶]	ماشین بردار پشتیبان	۸۳/۵
[۱۷]	الگوریتم جستجوی گرانشی با k نزدیک‌ترین همسایه	۶۴/۸
	الگوریتم بهینه‌سازی اجتماع ذرات با k نزدیک‌ترین همسایه	۸۲/۶
[۱۸]	شبکه عصبی مصنوعی چندلایه	۸۸/۰
[۲۲]	الگوریتم سیستم ایمنی مصنوعی	۸۲/۵
[۲۳]	الگوریتم بهینه‌سازی اجتماع سالپ	۸۸/۱
	الگوریتم ژنتیک	۸۴/۴
[۲۴]	الگوریتم بهینه‌سازی جنگل تصادفی	۸۵/۱
[۲۵]	الگوریتم سیاه‌چاله باینری	۸۲/۹
-	مدل پیشنهادی	۸۹/۸

تک لایه برای تشخیص بیماری قلبی استفاده کرده‌اند. در مدل رگرسیون منطقی و شبکه عصبی پرسپترون تک لایه، مقادیر اولیه اوزان و بایاس به صورت تصادفی انتخاب شده‌اند و لذا این امکان وجود داشته است که مقادیر خروجی و میزان دقت نتایج در هر اجرا کمی تغییر نماید. باین حال، میانگین به دست آمده از نتایج چندین بار اجرای برنامه برابر با ۹۰ درصد بوده است. بر اساس مطالعه صباغ گل [۱۵] به ترتیب متغیرهای سطح بالای کلسترول، جنس، سن بالا، بالا بودن ماکزیم ضربان قلب، اسکن تالیوم بالاتر از سه و نوار قلب غیرفرمال بیشترین تأثیر را در ابتلا به بیماری قلبی داشته‌اند. در ۹۲ درصد از بیماران بالای ۵۰ سال که دارای کلسترول بالاتر از ۲۰۰ واحد بوده‌اند ابتلا به بیماری قلبی مشاهده شده است. نچپال و همکاران [۱۷] از الگوریتم جستجوی گرانشی به منظور انتخاب ویژگی‌های بیماری قلبی استفاده کرده‌اند. در ابتدا، درصد صحت مدل‌ها آن‌ها بدون انتخاب ویژگی برابر با ۶۴/۸ درصد بوده اما بعد از انتخاب ویژگی درصد قابل توجهی از بهبود (در حدود ۱۸ درصد) در دقت تشخیص حاصل شده است. نتایج مندرج در جدول شماره چهار نشان می‌دهد که انتخاب ویژگی بر دقت تشخیص مؤثر است. مثلاً در جدول شماره چهار، عملیات انتخاب ویژگی برای هفت ویژگی در چهار حالت انجام شده است؛ اما حالت با هفت ویژگی (سن، جنس، میزان درد قفسه سینه، فشارخون در زمان استراحت، نوار قلب، آنژین ناشی از ورزش، اسکن تالیوم) دارای درصد صحت بیشتری در مقایسه با دیگر حالات است. همچنین برای نه ویژگی، سه حالت مختلف از انتخاب ویژگی وجود دارد که درصد صحت حالت با ویژگی‌های (میزان درد قفسه سینه، فشارخون در زمان استراحت، کلسترول خون، نوار قلب، حداکثر ضربان قلب، آنژین ناشی از ورزش، کاهش موج ST، موج ST در زمان حداکثر ورزش، تعداد عروق اصلی) برابر با ۹۰/۲ درصد است که در مقایسه با دو حالت دیگر بیشتر است.

ارزیابی بر مبنای انتخاب ویژگی نشان داد که درصد صحت روش پیشنهادی با هفت ویژگی برابر با ۹۰/۳ درصد و درصد حساسیت برابر با ۹۰/۲ بوده است. همچنین، درصد صحت مدل پیشنهادی با تمام ویژگی‌ها برابر با ۸۹/۸ بوده است. روش پیشنهادی در مقایسه با شبکه عصبی مصنوعی چندلایه [۱۴] و ماشین بردار پشتیبان [۱۶] از کارایی بالاتری برخوردار بوده است و حتی خروجی روش پیشنهادی در هر بار اجرا متفاوت و دارای درصد صحت بهتری بوده است.

کازمی و همکاران [۱۳] با استفاده از شبکه عصبی مصنوعی چندلایه توانسته‌اند مهم‌ترین ویژگی‌ها در بیماری قلبی را تشخیص دهند. درصد صحت به دست آمده توسط شبکه عصبی مصنوعی چندلایه برابر با ۸۳/۳ بوده است. هروی و ستایشی [۱۴] از رگرسیون منطقی و شبکه عصبی پرسپترون

گردد. در این پژوهش بر مبنای الگوریتم بهینه‌سازی ملخ دودویی و k نزدیک‌ترین همسایه روشی هوشمند برای تشخیص بیماری قلبی پیشنهاد داده شد.

از مدل پیشنهادی می‌توان به منظور تشخیص بیماری قلبی برای افراد در معرض خطر استفاده کرد. در ضمن، نتایج این مطالعه بر روی مجموعه داده‌های استاندارد UCI انجام شد؛ اما با بررسی نتایج این مطالعه و کشف ویژگی‌های مهم می‌توان مجموعه داده‌های واقعی را از مراکز درمانی و بهداشتی تهیه نمود و افراد مشکوک به بیماری قلبی را بر مبنای استدلال ویژگی‌ها، تشخیص داد.

تشکر و قدردانی: این مقاله حاصل بخشی از پایان نامه با عنوان شناسایی عوامل تأثیرگذار در سیستم پرونده سلامت الکترونیکی به منظور تشخیص بیماری با استفاده از الگوریتم‌های نظارتی در مقطع کارشناسی ارشد، در سال ۱۳۹۹، در دانشگاه آزاد اسلامی واحد ارومیه است.

هدف کاربردی این پژوهش بررسی امکان استفاده از سیستم‌های هوشمند تشخیص بیماری قلبی، تحلیل داده‌های بیماران توسط پزشکان به منظور تشخیص ویژگی‌های مهم بیماری قلبی مهیا بود. با استفاده از الگوریتم‌های نظارتی می‌توان داده‌های بیماران قلبی را به صورت هوشمندانه تحلیل کرد و اطلاعات لازم در خصوص بیماری را به بیماران پیشنهاد داد. مزیت‌های سیستم‌های هوشمند تشخیص بیماری می‌تواند منجر به افزایش دقت تشخیص بیماری، تصمیم‌گیری دقیق درباره نوع عارضه بیماری، کاهش هزینه و افزایش کارایی، کاهش اشتباهات درمانی و مراقبت‌های پزشکی

ملاحظات اخلاقی

رعایت دستورالعمل‌های اخلاقی: در مطالعه حاضر به تمامی ملاحظات اخلاقی از جمله شرط امانت و صداقت توجه شده است.

حمایت مالی: پژوهش حاضر تحت حمایت مالی هیچ سازمانی نبوده است.

تضاد منافع: نویسندگان اعلام می‌دارند که در این پژوهش هیچ‌گونه تضاد منافی وجود ندارد.

References

1. Shekar KC, Chandra P, Rao KV. A framework for automatic detection of heart diseases using dynamic deep neural activation functions. Journal of Ambient Intelligence and Humanized Computing. 2020; 1-15.
2. Kausar N, Palaniappan S, Samir BB, Abdullah A, Dey N. (2016) Systematic Analysis of Applied Data Mining Based Optimization Algorithms in Clinical Attribute Extraction and Classification for Diagnosis of Cardiac Patients. In: Hassanien AE, Grosan C, Fahmy Tolba M. (eds) Applications of Intelligent Optimization in Biology and Medicine. Intelligent Systems Reference Library. Springer. Cham; 2015. p. 217-231.
3. Publication bias in clinical research [Internet]. 2018 [2018 May 24; cited 2020 Sep 05]; Available from: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
4. Jayaraman V, Sultana HP. Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. Journal of Ambient Intelligence and Humanized Computing. 2019; 1-15.
5. Amin MS, Chiam YK, Varathan KD. Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics. 2019; 36: 82-93.
6. Parks R, Wigand RT, Othmani MB, Serhier Z, Bouhaddou O. Electronic health records implementation in Morocco: Challenges of silo efforts and recommendations for improvements. International Journal of Medical Informatics. 2019; 129: 430-437.
7. Rudolf JW, Dighe AS. Decision Support Tools within the Electronic Health Record. Clinics in Laboratory Medicine. 2019; 39(2): 197-213.

8. Liu L, Zhao S, Chen H, Wang A. A new machine learning method for identifying Alzheimer's disease. *Simulation Modelling Practice and Theory*. 2020; 99:1-20.
9. Ullah R, Khan S, Ali H, Chaudhary II, Ahmad I. A comparative study of machine learning classifiers for risk prediction of asthma disease. *Photodiagnosis and Photodynamic Therapy*. 2019; 28: 292-296.
10. Shayanfar H, Gharehchopogh FS. Farmland fertility: A new metaheuristic algorithm for solving continuous optimization problems. *Applied Soft Computing*. 2018; 71: 728-746.
11. Gharehchopogh FS, Mousavi SK. A Decision Support System for Diagnosis of Diabetes and Hepatitis, based on the Combination of Particle Swarm Optimization and Firefly Algorithm. *Journal of Health and Biomedical Informatics*. 2019; 6 (1) :32-45. **[In Persian]**.
12. Ardani S, Soleimanian Gharehchopogh F. Diagnosing Liver Disease using Firefly Algorithm based on Adaboost. *Journal of Health Administration*. 2019; 22 (1) :61-77. **[In Persian]**.
13. Kazemi M, Mehdizadeh H, Shiri A. Heart disease forecast using neural network data mining technique. *sjimu*. 2017; 25 (1) :20-32. **[In Persian]**.
14. Heravi M, Setayeshi S. Intelligent and fast recognition of heart disease based on synergy of linear neural network and logistic regression model. *J Mazandaran Univ Med Sci*. 2014; 24 (112) :78-87. **[In Persian]**
15. Sabbagh_Gol H. Detection of Coronary Artery Disease Using C4.5 Decision Tree. *Journal of Health and Biomedical Informatics*. 2017; 3 (4) :287-299. **[In Persian]**
16. Mahmoodi MS. Designing a Heart Disease prediction System using Support Vector Machine. *Journal of Health and Biomedical Informatics*. 2017; 4 (1) :1-10. **[In Persian]**
17. Nagpal S, Arora S, Dey S, Shreya. Feature Selection using Gravitational Search Algorithm for Biomedical Data. *Procedia Computer Science*. 2017; 115: 258-265.
18. Zabbah I, Hassaanazadeh M, kohjani Z. The Effect of Continuous Parameters on the Diagnosis of Coronary Artery Disease Using Artificial Neural Networks. *Journal of Torbat Heydariyeh University of Medical Sciences (Journal of Health Chimes)*. 2017; 4 (4) :29-39. **[In Persian]**
19. Mafarja M, Aljarah I, Faris H, Hammouri AI, Al-Zoubi AM, Mirjalili S. Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications*. 20149; 117: 267-286.
20. Martin B. Instance-Based Learning: Nearest Neighbour with Generalisation [PhD thesis]. Hamilton: New Zealand; University of Waikato; 1995.
21. Publication bias in clinical research [Internet]. 2004 [Updated 2004; cited 2020 Sep 05]; Available from: [http://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](http://archive.ics.uci.edu/ml/datasets/statlog+(heart))
22. Wang Y, Li T. Local feature selection based on artificial immune system for classification. *Applied Soft Computing*. 2020; 87: 1-27.
23. Tubishat M, Idris N, Shuib L, Abushariah MAM, Mirjalili S. Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Systems with Applications*. 2020; 145:1-23.
24. Ghaemi M, Feizi-Derakhshi MR. Feature selection using Forest Optimization Algorithm. *Pattern Recognition*. 2016; 60: 121-129.
25. Pashaei E, Aydin N. Binary black hole algorithm for feature selection and classification on biological data. *Applied Soft Computing*. 2017; 56: 94-106.