

## تشخیص بیماری قلبی عروق کرونر با سیستم هوشمند ترکیبی بر اساس الگوریتم نهنگ، شبیه‌ساز تبرید و ماشین بردار پشتیبان\*

زینب حسنی<sup>۱</sup>

مهین خسروی مشیزی<sup>۲</sup>

### چکیده

در سال‌های اخیر، الگوریتم‌های یادگیری ماشین به طور گسترده در تشخیص و درمان به موقع بیماری‌ها نقش بسزایی را ایفا می‌کنند. علاوه بر این، تشخیص بیماری در مراحل آغازین آن، در بهبود بیماری و کاهش هزینه‌های درمانی بیمار بسیار مؤثر است. بیماری قلبی یکی از دلایل اصلی مرگ در جهان شناخته شده است. مطالعات زیادی برای تشخیص بیماری و طراحی سیستم هوشمند و کارا انجام شده است. در این مقاله، الگوریتم ترکیبی نهنگ و شبیه‌ساز تبرید برای شناسایی عوامل مؤثر در تشخیص بیماری ارائه شده است و الگوریتم ماشین بردار پشتیبان برای طبقه‌بندی مؤثر بیماری لحاظ شده است. رویکرد پیشنهادی با مجموعه داده بیماری قلبی کلیولند در پایگاه داده UCI ارزیابی شده است. الگوریتم پیشنهادی با صحت ۸۷/۷۸ درصد با تعداد ویژگی کمتر توانسته بیماری را تشخیص دهد. نتایج حاصل برتری روش پیشنهادی را نشان می‌دهد و همچنین رویکرد پیشنهادی می‌تواند پزشکان را در تشخیص درست و در مراحل اولیه بیماری یاری رساند.

**کلمات کلیدی:** الگوریتم بهینه‌سازی نهنگ (WOA)، الگوریتم ترکیبی، الگوریتم شبیه‌ساز تبرید (SA)، بیماری قلبی عروق کرونر، ماشین بردار پشتیبان (SVM).

\* تاریخ دریافت: ۹۷/۸/۲۵؛ تاریخ پذیرش: ۹۷/۱۰/۲۶.

<sup>۱</sup> مربی گروه کامپیوتر، دانشکده فنی و علوم پایه، دانشگاه کوثر، بجنورد، ایران (نویسنده مسئول) hassani@kub.ac.ir

<sup>۲</sup> دانشجو کارشناسی ارشد مهندسی کامپیوتر، دانشکده فنی و علوم پایه، دانشگاه کوثر، بجنورد، ایران

mahin.khosravi2324@gmail.com

## مقدمه

در طول تاریخ بیماری‌ها بزرگترین تهدید برای بشر بشمار می‌روند. در این بین بیماری قلبی از اهمیت بیشتری در مطالعات پزشکی برخوردار است. بیماری قلبی یکی از شایع‌ترین انواع بیماری‌ها است که بیشترین آمار مرگ و میر را به خود اختصاص داده است (وانگ و همکاران، ۲۰۱۴). بیماری قلبی عروقی کرونر یکی از شایع‌ترین نوع بیماری‌های قلبی است که بیشترین آمار مرگ و میر را به خود اختصاص داده است. در این بیماری سرخرگ‌های تاجی تنگ و باریک می‌شوند (استنوسیس) و عضلات قلب از رسیدن خون و اکسیژن کافی محروم می‌شوند. در این صورت ممکن است در حالت استراحت مشکلی برای فرد رخ ندهد اما وقتی که قلب مجبور باشد کار بیشتری را انجام دهد، سرخرگ‌ها نمی‌توانند به این عضلات خون و اکسیژن کافی برسانند و در نتیجه فرد دچار درد سینه و آئزین قلبی می‌شود و در حالت‌های وخیم‌تر اگر به علت مسدود شدن سرخرگ تاجی توسط یک لخته خون، جلوی خون‌رسانی آن گرفته شده باشد، قسمتی از عضلات قلب دیگر به آن خون نمی‌رسد و در نتیجه فرد خواهد مرد و به سکت قلبی می‌انجامد (باچان و همکاران، ۲۰۱۷). چندین روش تشخیص بیماری قلبی کرونری وجود دارد مثل تست ورزش، الکتروکاردیوگرافی و آنژیوگرافی که بهترین روش تشخیص آن آنژیوگرافی است. روش‌های گفته شده محدودیت‌هایی دارد، عدم دقت کافی در تشخیص که این باعث می‌شود پزشکان ترغیب در استفاده از روش مبتنی بر رایانه شوند (دولت آبادی و همکاران، ۲۰۱۷).

مطالعات گسترده‌ای برای پیش‌بینی و تشخیص بیماری قلبی توسط پژوهشگران صورت گرفته است. خمفیل و بانجینگ (۲۰۱۰) برای تشخیص بیماری قلبی درخت تصمیم‌گیری، با لجستیک رگرسیون و شبکه عصبی را مطالعه کرده‌اند که شبکه عصبی با صحت ۸۰/۲ درصد بالاترین عملکرد را برای تشخیص بیماری داشته است. خسرویان و آیت (۲۰۱۵) برای پیش‌بینی بیماری عروق کرونری قلب شبکه عصبی را مطالعه کرده‌اند. آن‌ها در مطالعه خود نشان داده‌اند که شبکه‌های عصبی احتمالی، عملکرد بهتری نسبت به

سایر شبکه‌های عصبی در تشخیص بیماری دارد. پوریه و همکاران (۲۰۱۷) طبقه‌بندی بیزین، درخت تصمیم، الگوریتم نزدیکترین همسایه KNN، ماشین بردار پشتیبان SVM، شبکه عصبی MLP و RBF برای پیش‌بینی بیماری قلبی را بررسی کرده‌اند. نتایج آن‌ها نشان می‌دهد که ماشین بردار پشتیبان SVM ۸۴/۱۵ درصد بالاترین صحت را داشته است. در ژانگ و همکاران (۲۰۱۷) درخت تصمیم C4.5، ماشین بردار پشتیبان SVM، طبقه‌بندی بیزین و AdaBoost مطالعه شده است. عملکرد تصادفی جنگل C4.5 با صحت ۸۹/۶ درصد، SVM با صحت ۸۹/۲ درصد، طبقه‌بند Bayes با صحت ۸۵/۲ درصد، AdaBoost با صحت ۸۲/۸ درصد برای پیش‌بینی بیماری قلبی و عروقی به دست آمده است.

در این مقاله، الگوریتم ترکیبی نهنگ و شبیه‌ساز تبرید برای تشخیص بیماری قلبی کرونر معرفی شده است و از ماشین بردار پشتیبان برای طبقه‌بندی مؤثر داده بیماری استفاده شده است. در ادامه، در بخش دوم ماشین بردار پشتیبان، الگوریتم نهنگ و شبیه‌ساز تبرید شرح داده می‌شود. روش پیشنهادی و یافته‌ها به ترتیب در بخش‌های سوم و چهارم ارائه می‌شود، در پایان مقاله نیز نتیجه‌گیری ارائه می‌شود.

## روش‌شناسی پژوهش

در این مقاله، شناسایی عوامل مؤثر بر بیماری قلبی کرونری با الگوریتم ترکیبی نهنگ و شبیه‌ساز تبرید مطالعه شده است. از ماشین بردار پشتیبان برای طبقه‌بندی بیماری قلبی کرونر با هدف تشخیص صحیح بیماری استفاده شده است. برای ارزیابی روش پیشنهادی مجموعه داده‌های بیماری قلبی کرونری با ۱۳ ویژگی از ۲۷۰ نمونه در پایگاه داده استاندارد UCI انتخاب شده است. در ادامه به معرفی الگوریتم ماشین بردار پشتیبان، الگوریتم بهینه‌سازی نهنگ، شبیه‌ساز تبرید پرداخته می‌شود.

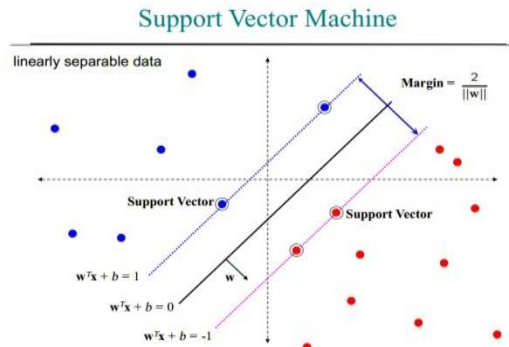
## الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یکی از الگوریتم‌های یادگیری تحت نظارت در یادگیری ماشین است که در سال ۱۹۶۳ توسط واپنیک ارائه شده است (واپنیک، ۱۹۹۵؛

بورجس، ۱۹۹۸). ماشین بردار پشتیبان را می‌توان برای طبقه‌بندی اشیاء یا تشخیص الگو برای کلاس‌های خاص لحاظ کرد که دو کلاس  $\{+1 \text{ و } -1\}$  را با استفاده از یک مرز خطی از هم جدا می‌کند. تعدادی از نقاط آموزشی که فاصله کمتری تا مرز تصمیم‌گیری دارند به عنوان ماشین بردار پشتیبان تعریف شده و برای محاسبه مرز تصمیم‌گیری دو کلاس متمایز، از تکنیک حاشیه بهینه استفاده می‌شود (واپنیک و چرووننکیس، ۱۹۹۱). در این تکنیک، مرز خطی بین دو کلاس به نحوی محاسبه می‌شود که نمونه‌های کلاس  $+1$  در یک طرف و نمونه‌های کلاس  $-1$  در طرف دیگر وجود دارند. همچنین مرز تصمیم‌گیری باید طوری باشد که فاصله نزدیک‌ترین نمونه آموزشی هر دو کلاس نسبت به مرز حداکثر شود. شکل ۱ نحوه قرار گرفتن صفحات در ماشین بردار پشتیبان را نشان می‌دهد. ماشین بردار پشتیبان به منظور طبقه‌بندی داده‌ها با پیچیدگی بالا، داده‌ها را به وسیله تابع کرنل  $(\varphi)$  به فضای با ابعاد خیلی بالاتر می‌برد. تابع کرنل داده‌ها را از فضای ورودی به فضای با ابعاد بالاتر به گونه‌ای تصویر می‌کند که امکان جداسازی خطی داده‌ها در آن فضا وجود داشته باشد. سپس داده‌های آموزشی انتقال داده شده توسط یک فراصفحه بهینه در آن فضا جداسازی می‌کند. در واقع، بدون هیچ تغییری در ماهیت مسئله بهینه‌سازی با بکارگیری تابع انتقال  $\varphi$  داده‌ها به فضایی با ابعاد بالاتر نگاشت می‌شوند تا در این فضا سطح تصمیم‌گیری به صورت خطی تعیین شود (کومان و کوپریوا، ۲۰۰۶). در این مطالعه از کرنل RBF در طبقه‌بندی ماشین بردار پشتیبان استفاده شده است.

$$K(x, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad \text{رابطه (۱)}$$

دلیل این امر توان این تابع کرنل در آنالیز داده‌ها با ابعاد بالا در مقایسه با سایر کرنل‌ها می‌باشد.



شکل ۱. قرار گرفتن صفحات در ماشین بردار پشتیبان

### الگوریتم بهینه‌سازی نهنگ

الگوریتم بهینه‌سازی نهنگ (WOA) یکی از الگوریتم‌های فراابتکاری است که توسط میرجلیلی در سال ۲۰۱۶ پیشنهاد شده است (میرجلیلی و لوئیس، ۲۰۱۶). این الگوریتم از مکانیزم شکار نهنگ‌ها الهام گرفته است که نهنگ‌ها می‌توانند موقعیت طعمه‌ها را به خوبی تشخیص دهند و آن‌ها را محاصره کنند. در الگوریتم، موقعیت مکانی دور زدن و حرکت ماریچی و مکانیزم‌های شکار تصادفی نهنگ شبیه‌سازی شده است. این مدل شامل سه مرحله محاصره شکار، حمله جابجایی و جستجوی شکار است.

#### محاصره شکار: از آنجا که موقعیت طراحی بهینه در فضای جستجو در ابتدا معلوم

نیست، الگوریتم WOA فرض می‌کند که بهترین راه‌حل نامزد فعلی، طعمه هدف است یا نزدیک به حالت بهینه است. پس از آن که بهترین عامل جستجو تعریف شد، عامل‌های دیگر جستجو به این ترتیب سعی می‌کنند موقعیت‌های خود را نسبت به بهترین عامل جستجو بروزرسانی کنند. این رفتار با معادلات زیر نشان داده شده است (میرجلیلی و لوئیس، ۲۰۱۶):

$$D = |C \cdot \vec{X}^*(t) - \vec{X}(t)| \quad \text{رابطه (۲)}$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot D \quad \text{رابطه (۳)}$$

t تکرار اخیر را نشان می‌دهد،  $A^{\rightarrow}$  و  $C^{\rightarrow}$  بردار ضرایب هستند،  $X^*$  بردار موقعیت بهترین راه‌حلی است که تاکنون به دست آمده،  $X^{\rightarrow}$  بردار موقعیت،  $\|$  مقدار قدرمطلق است و ضرب درایه به درایه است. لازم به ذکر است که اگر یک راه حل بهتر وجود داشته باشد،  $X^*$  باید در هر تکرار بروزرسانی شود.

بردارها  $\vec{A}$  و  $C^{\rightarrow}$  به صورت زیر محاسبه می‌شوند:

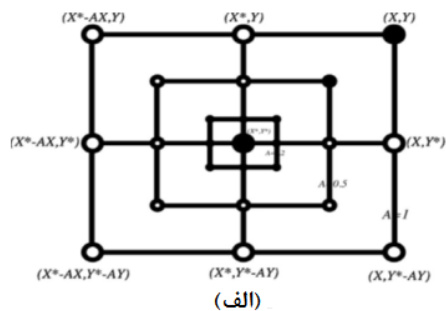
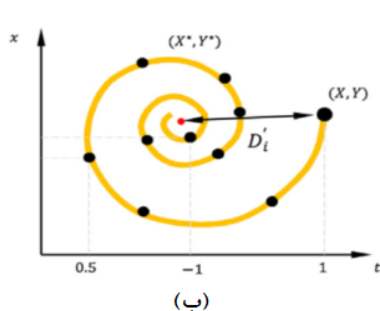
$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad \text{رابطه (۴)}$$

$$\vec{C} = 2 \cdot \vec{r} \quad \text{رابطه (۵)}$$

که a در طول تکرارها (در هر دو مرحله اکتشاف و بهره‌برداری) به طور خطی از ۲ به ۰ کاهش می‌یابد و r برداری تصادفی در [۰،۱] است.

**حمله جابجایی:** این مرحله، مرحله بهره‌برداری الگوریتم فراابتکاری است. به منظور مدل‌سازی ریاضیاتی رفتار حمله جابجایی نهنگ‌ها، دو رویکرد به شرح زیر طراحی شده است:

**مکانیزم محاصره انقباضی:** این رفتار با کاهش مقدار  $\vec{a}$  در رابطه (۴) به دست می‌آید. توجه داشته باشید که محدوده نوسان  $\vec{A}$  نیز توسط  $\vec{a}$  کاهش می‌یابد.



شکل ۲. جستجو در فضای دوبعدی (الف)، جستجو در فضای سه‌بعدی (ب)

به عبارت دیگر،  $\vec{A}$  یک مقدار تصادفی در فاصله  $[-a, a]$  است که در آن a در طول تکرار از ۲ به ۰ کاهش می‌یابد. با قرار دادن مقادیر تصادفی برای  $\vec{A}$  در  $[-1, 1]$ ، موقعیت جدیدی از یک عامل جستجو می‌تواند در هر نقطه بین موقعیت اصلی عامل و موقعیت

بهترین عامل فعلی تعریف شود. شکل ۲ (الف) موقعیت‌های احتمالی را از  $(X, Y)$  تا  $(X^*, Y^*)$  نشان می‌دهد که می‌تواند به وسیله  $0 \leq A \leq 1$  در یک فضای دو بعدی به دست آید.

**به روز رسانی موقعیت حلزونی:** همان‌طور که در شکل ۲ دیده می‌شود، ابتدا فاصله بین نهنگ قرار گرفته در مختصات  $(X, Y)$  و طعمه موجود در  $(X^*, Y^*)$  محاسبه می‌شود. سپس یک معادله مارپیچی بین موقعیت نهنگ و شکار ایجاد می‌شود تا حرکت حلزونی شکل نهنگ تقلید کند که در رابطه (۶) نمایش داده شده است.

$$\vec{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (\text{رابطه ۶})$$

که  $|\vec{X}(t) - \vec{X}^*(t)| = D'$  به فاصله  $\vec{X}(t)$  تا طعمه اشاره دارد (بهترین راه‌حل به دست آمده تاکنون)،  $b$  مقداری ثابت برای تعریف شکل مارپیچ لگاریتمی است،  $(l)$  عددی تصادفی در بازه  $[-1, 1]$  است و  $(.)$  عمل ضرب درایه به درایه است.

لازم به ذکر است که نهنگ‌ها، حول طعمه در امتداد یک دایره انقباضی و هم‌زمان در مسیر مارپیچی شکلی به شنا در می‌آیند. جهت مدلسازی این رفتار هم‌زمان، فرض شده است که نهنگ‌ها با احتمال ۵۰ درصد از بین مکانیزم محاصره انقباضی یا مدل مارپیچی یکی را انتخاب می‌کنند تا موقعیت نهنگ‌ها در طول بهینه‌سازی بروزرسانی شود. مدل ریاضی به این صورت است:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (\text{رابطه ۷})$$

که در آن  $p$  عددی تصادفی است.

**جستجوی شکار:** علاوه بر روش حمله جابجایی، نهنگ طعمه را به صورت تصادفی جستجو می‌کند که می‌تواند برای جستجوی شکار (مرحله اکتشاف) به کار رود. در حقیقت نهنگ بر اساس  $\vec{A}$  همان روش بر مبنای تغییر بردار موقعیت یکدیگر، جستجو را به صورت تصادفی انجام می‌دهد. بنابراین،  $\vec{A}$  را با مقادیر تصادفی بزرگتر از ۱ یا کمتر از ۱- به کار می‌بریم تا عامل جستجو را وادار به دور شدن از یک نهنگ مرجع کنیم. در مقایسه با مرحله بهره‌برداری، برای بروزرسانی موقعیت یک عامل جستجو در مرحله اکتشافی، به جای یافتن

بهترین عامل جستجو تاکنون، از یک عامل جستجوی انتخاب شده به صورت تصادفی استفاده می‌کنیم. این مکانیزم بر اکتشاف تأکید دارد و به الگوریتم Woa اجازه می‌دهد یک جستجو سراسری را انجام دهد.

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad \text{رابطه ۸}$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad \text{رابطه ۹}$$

$\vec{X}_{rand}$  یک بردار موقعیت تصادفی (یک نهنگ تصادفی) است که از جمعیت فعلی انتخاب شده است.

```

Generate Initial Population X i (i = 1, 2, ..., n)
Calculate the fitness of each solution
X * = the best search agent
while (t < Max_Iteration)
    for each solution
        Update a, A, C, l, and p
        Update a, A, C, l, and p
        if 1 (p < 0.5)
            if 2 (|A| < + 1)
                Update the position of the current solution by Eq. (3)
            else if 2 (|A| > + 1)
                Select a random search agent ()
        Update the position of the current search agent by the Eq. (9)
    end if 2
    else if 1 (p ≥ 0.5)
        Update the position of the current search by the Eq. (6)
    end if 1
end for
Check if any solution goes beyond the search space and amend it
Calculate the fitness of each solution
Update X* if there is a better solution t = t + 1
end while
return X*
    
```

شکل ۳. شبه کد الگوریتم نهنگ

### الگوریتم شبیه‌ساز تبرید

الگوریتم شبیه‌ساز تبرید (SA)، الگوریتم شبیه‌ساز فرابتکاری ساده و اثربخش در حل مسائل بهینه‌سازی است. منشأ الگوریتم تبرید شبیه‌سازی شده، کارهای کریک پاتریک، کرنی و همکارانشان در سال‌های ۱۹۸۳ و ۱۹۸۵ است (پاتریک و همکاران، ۱۹۸۳). آن‌ها برای حل مسائل پیچیده بهینه‌سازی، روشی مبتنی بر تکنیک تبرید شبیه‌ساز پیشنهاد کردند.



الگوریتم تبرید شبیه‌سازی شده برای حل یک مسئله بهینه‌سازی، ابتدا از یک جواب اولیه شروع می‌کند و سپس در یک حلقه تکرار به جواب‌های همسایه حرکت می‌کند. اگر جواب همسایه بهتر از حالت فعلی باشد، الگوریتم آن را به عنوان جواب فعلی قرار می‌دهد، در غیر این صورت الگوریتم آن جواب را با احتمال  $P = e^{-\theta/T}$  که احتمال بولترمن نامیده می‌شود به عنوان جواب فعلی می‌پذیرد که  $\theta$  تفاوت بین تابع هدف جواب فعلی و جواب همسایه و  $T$  یک پارامتر به نام دما است. در هر دما، چندین تکرار اجرا می‌شود و سپس دما به آرامی کاهش پیدا می‌کند. در گام‌های اول دما خیلی بالا قرار داده می‌شود تا احتمال پذیرش بیشتری برای جواب‌های بدتر وجود داشته باشد. با کاهش تدریجی دما، در گام‌های پایانی احتمال کمتری برای پذیرش جواب‌های بدتر وجود خواهد داشت و بنابراین الگوریتم به سمت یک جواب خوب همگرا می‌شود.

در این کار ابتدا حالت خلوص  $2 * |N|$  است که  $|N|$  به صورت قراردادی نشان‌دهنده تعداد ویژگی‌های هر مجموعه داده و برنامه خنک‌کننده به عنوان  $T = 93/0 * T$  محاسبه می‌شود (جانسن و چن، ۲۰۰۳). انتخاب مسابقات تورنمنت یک مکانیزم انتخاب ساده و آسان برای پیاده‌سازی است که توسط گولدربرج و همکاران بیان شده است (گلدبرگ و کرب، ۱۹۸۹). این یکی از بهترین مکانیزم‌های انتخاب در الگوریتم‌های تکاملی است (سانچا و آندیتا، ۲۰۱۶). در انتخاب مسابقات راه‌حل‌های  $N$  باید به طور تصادفی از جمعیت انتخاب شوند. این راه‌حل‌ها با یکدیگر مقایسه می‌شوند و مسابقات برای تعیین برنده قرار می‌گیرد. این مسابقات شامل یک عدد تصادفی بین ۰ و ۱ می‌شود، سپس با یک احتمال مقایسه می‌شود که یک مکانیزم مناسب برای تنظیم فشار انتخاب (معمولاً ۰/۵ تنظیم می‌شود) است اگر مقدار تصادفی بزرگ‌تر باشد، راه‌حل با بالاترین ارزش انتخاب خواهد شد و راه‌حل ضعیف انتخاب نمی‌شود. این ویژگی در مسابقات انتخاب فرصتی برای بیشترین راه‌حل انتخاب می‌کند که تنوع راه‌حل‌های انتخاب شده را حفظ می‌کند (سانچا و آندیتا، ۲۰۱۶).

### الگوریتم ترکیبی WOA-SA برای تشخیص بیماری قلبی عروق کرونر

انتخاب روش رویکرد پیشنهادی برای حل یک مسئله بهینه‌سازی باینری است که راه‌حل یک بردار یک بعدی با مقادیر صفر و یک (باینری) است. الگوریتم ترکیبی WOA-SA عوامل مؤثر در بیماری قلبی کرونر را شناسایی می‌کند که نتیجه حاصل از آن یک بردار به طول تعداد ویژگی‌ها با مقادیر صفر و یک است و مقدار ۱ نماینگر ویژگی انتخاب شده است و مقدار ۰ به معنای ویژگی انتخاب نشده است. هر راه‌حل بر اساس عملکرد پیشنهاد شده مورد ارزیابی قرار می‌گیرد که به طبقه‌بند SVM بستگی دارد.

در الگوریتم WOA می‌توان به محاسبه فاصله بین عامل تحقیق و بهترین نهنگ شناخته شده پرداخت. با استفاده از یک الگوریتم جستجو محلی مناسب برای جستجو همسایگی در اطراف نهنگ، بهترین راه‌حل شناخته شود تا نتایج را بهبود بخشد. علاوه بر این، از آنجا که اکتشاف در الگوریتم WOA بستگی به تغییر موقعیت هر جستجوی عامل با توجه به یک راه‌حل تصادفی است، با استفاده از مکانیزم انتخابی مختلف مانند انتخاب مسابقات ممکن است توانایی اکتشاف در الگوریتم را بهبود داد. این به این معنی است که انتخاب مسابقات شانس بیشتری برای تصمیم‌گیری‌های ناچیز دارد که در طول فرآیند جستجو انتخاب می‌شود و بسته به انتخاب، توانایی تنوع الگوریتم WOA را بهبود می‌بخشد. الگوریتم WOA یک الگوریتم بهینه‌سازی است که بدون در نظر گرفتن راه‌حل فعلی از اپراتور نامشخص برای بهره‌برداری استفاده می‌کند. ما این اپراتور را با یک الگوریتم جستجو محلی جایگزین می‌کنیم که راه‌حل را به عنوان حالت اولیه آن لحاظ می‌کند و الگوریتم بر اساس آن جستجو را ادامه می‌دهد و راه‌حل اصلی را با راه‌حل بهبود یافته جایگزین کرده است. این رویکرد نشان‌دهنده ترکیبی بین جستجوی سراسری (WOA) و جستجو محلی (SA) است. در رویکرد پیشنهادی، مدل ترکیبی (LTH) بین دو الگوریتم در نظر گرفته شده است. در (LTH)، الگوریتم SA در الگوریتم WOA برای بهترین راه‌حل در همسایگی لحاظ می‌شود و راه‌حل تصادفی انتخاب شده (جایگزین معادله ۹) و بهترین راه‌حل همسایه در جستجو (جایگزین معادله ۳) تعبیه شده است. این فرآیند توانایی

استفاده از الگوریتم WOA را بهبود می‌بخشد. الگوریتم SA به عنوان یک اپراتور در الگوریتم WOA قرار می‌گیرد.

انتخاب ویژگی (FS) به عنوان یک مسئله بهینه‌سازی در نظر گرفته می‌شود که در آن راه‌حل‌ها به مقادیر [۰,۱] محدود می‌شوند. بنابراین، الگوریتم ترکیبی WOA-SA برای حل مسائل بهینه‌سازی ارائه شده است. در این روش، یک بردار صفر و یک برای نشان دادن راه‌حل مسئله مورد استفاده قرار می‌گیرد که در آن بردار صفر نشان دهنده ویژگی انتخاب نشده و عنصر یک متناظر با ویژگی انتخاب شده است. اندازه بردار نیز برابر با تعداد ویژگی‌های مجموعه داده بیماری قلبی کرونری است. در این مطالعه، یک روش انتخاب ویژگی مبتنی بر الگوریتم ترکیبی WOA-SA ارائه شده است. طبقه‌بند SVM به عنوان تابع ارزیاب ویژگی‌های انتخابی به کار می‌رود. مراحل روش پیشنهادی به شرح زیر است:

۱. **پیش‌پردازش داده‌ها:** برای به دست آوردن نتایج کارا تر نیازمند داده‌هایی در محدوده یکسان هستیم از این رو، ابتدا فرایند نرمال‌سازی مطابق با رابطه ۱۰ صورت می‌پذیرد.

$$X = (X - X_{MIN}) / (X_{MAX} - X_{MIN}) \quad \text{رابطه ۱۰}$$

که در آن  $X_{MAX}$  و  $X_{MIN}$  به ترتیب به بیشترین مقدار و کمترین مقدار برای هر مشخصه می‌باشند.

۲. **مقداردهی اولیه پارامترهای الگوریتم:** در این مرحله مقادیر اولیه پارامتر از

جمله تعداد تکرار و تعداد جمعیت نهنگ‌ها مقداردهی می‌شود.

۳. محاسبه تابع شایستگی نهنگ‌ها با فراخوانی SVM و داده‌های ورودی

۴. تشخیص بیماری با الگوریتم ترکیبی WOA-SA

معیار توقف الگوریتم ترکیبی، برابر با تعداد تکرار است. برای ارزیابی بهترین ویژگی‌های انتخاب شده توسط طبقه‌بندی کننده SVM از روش اعتبارسنجی K-fold استفاده می‌شود، در این نوع اعتبارسنجی داده‌ها به K زیرمجموعه افراز می‌شوند، از این K زیرمجموعه K-1 دسته داده برای آموزش و اعتبارسنجی و باقیمانده داده‌ها (یک دسته)

برای تست استفاده می‌شود. این روال k بار تکرار می‌شود. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می‌شود.

میزان مقدار تابع شایستگی برابر با مقدار صحت کلاس‌بندی الگوریتم ماشین بردار پشتیبان است. در این مطالعه، برای مقایسه و بررسی نتایج حاصل شده، شاخص ارزیابی صحت<sup>۱</sup>، دقت<sup>۲</sup>، ویژگی<sup>۳</sup> در نظر گرفته شده است.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{رابطه (۱۱)}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{رابطه (۱۲)}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad \text{رابطه (۱۳)}$$

که TP، FP، TN و FN به ترتیب تعداد مثبت حقیقی، مثبت کاذب، منفی حقیقی و منفی کاذب هستند.

در روش پیشنهادی، مجموعه داده‌های استاندارد بیماری قلب کلیند از پایگاه داده یادگیری ماشین UCI<sup>۴</sup> انتخاب شده است. در ابتدا ۲۷۰ نمونه با ۷۶ ویژگی متمایز با مقادیر حقیقی جمع‌آوری شده است در نهایت مجموعه داده‌ها با ۱۳ ویژگی انتخاب شده است. جدول ۱، ۱۳ ویژگی بیماری قلبی عروق کرونر را نشان می‌دهد.

جدول ۱. شرح مجموعه داده‌های کلیند

ویژگی	نوع ویژگی
سن	عددی
جنسیت	اگر مرد=۱ اگر زن=۰
درد قفسه سینه	آنژین معمولی=۱ آنژین ناپیچک=۲ درد غیر زاویه درد=۳
فشار خون	عددی
کلسترول	عددی

<sup>1</sup> Accuracy

<sup>2</sup> Precision

<sup>3</sup> Specificity

<sup>4</sup> <http://archive.ics.uci.edu>

ویژگی	نوع ویژگی
قند خون	اگر بله = ۱ اگر نه = ۰
الکتروکاردیوگرام	اختلال در موج ST-T = ۱ هیپوتروفی LV = ۲
حداکثر ضربان قلب	عددی
ورزش	اگر بله = ۱ اگر نه = ۰
افسردگی ST	عددی
تعداد عروق اصلی	عددی

### یافته‌های پژوهش

الگوریتم پیشنهادی با مجموعه داده بیماری قلبی با نرم افزار متلب ۲۰۱۵ پیاده سازی شده که با تعداد تکرارها و تعداد جمعیت‌های مختلف اجرا شده است. در کل ۱۳ ویژگی با ۲۷۰ نمونه افراد بیمار و غیربیمار داریم. الگوریتم ترکیبی بر اساس بهترین راه حل، تعداد ویژگی‌های مؤثر بیماری را شناسایی می کند. جدول ۲ نتایج حاصل از الگوریتم پیشنهادی را نشان می دهد. الگوریتم پیشنهادی ۱۰ ویژگی مؤثر را با صحت ۸۷/۷۸ درصد شناسایی کرده است. جدول ۳ مقایسه بین الگوریتم پیشنهادی با مطالعات پیشین داده بیماری قلبی کرونر است که بالاتر بودن نتیجه حاصل شده، برتری روش پیشنهادی را نشان می دهد.

جدول ۲. نتایج الگوریتم پیشنهادی با مجموعه داده‌های بیماری قلبی

بهرترین انتخاب	ویژگی	دقت	صحت	تعداد نهنگ	تعداد تکرار
۱۰	۸۹/۵۱	۸۸/۶۷	۸۷/۷۸	۱۰	۲۵
۱۰	۸۶/۰۹	۹۰/۶۶	۸۶/۲۹	۱۵	
۱۰	۸۸/۵۱	۸۹/۳۳	۸۷/۴۰	۱۰	۴۰
۹	۸۸/۸۶	۸۶/۶۶	۸۶/۲۹	۱۵	

جدول ۳. مقایسه الگوریتم SVM\_WOASA با مطالعات پیشین

روش	صحت	مرجع
SVM	۸۵/۱۸	خمفیل و بونجینگ (۲۰۱۰)
KNN	۸۲/۹۶	پوریه و همکاران (۲۰۱۷)
ANN	۷۳/۳۳	ژو و همکاران (۲۰۱۷)
WOASA_SVM	۸۷/۷۸	[روش پیشنهادی]

### نتیجه‌گیری و پیشنهادها

در این مقاله، انتخاب ویژگی با الگوریتم ماشین بردار پشتیبان و الگوریتم ترکیبی WOA\_SA پیشنهاد شده است. رویکرد پیشنهادی با مجموعه داده‌های بیماری قلبی پایگاه داده UCI ارزیابی شده است. بیماری قلبی بیشترین آمار مرگ و میر را دارد که با آزمایشات عینی قابل تشخیص نیست و همچنین برای بیمار محدودیت‌های را در بردارد. هدف اصلی روش پیشنهادی انتخاب کمترین ویژگی با حداکثر تشخیص بیماری است که صحت بالاتری نسبت به استفاده از تمام ویژگی‌های مجموعه داده‌ها می‌توان به دست آورد. الگوریتم پیشنهادی ۱۰ ویژگی موثر با ۸۷/۷۸ درصد شناسایی کرده است. که در مقایسه با نتایج پیشین بهبود گسترده‌ای را داشته است.

## منابع

- Khemphila and V. Boonjing. (2010). Comparing Performances of Logistic Regression, Decision trees, and Neural Networks for Classifying Heart Disease Patients. 2010 IEEE International Conference on Computer Information Systems and Industrial Management Systems, pp. 193-199.
- Buchan, K., Filannino, M., Uzuner, O. (2017). Automatic prediction of coronary artery disease from clinical narratives. Journal of biomedical informatics, Vol.72, pp.23-32.
- Burges C. (1998). A tutorial on support vector machines for pattern recognition1. Data Mining and Knowledge Discovery, vol.2, pp. 121 – 167.
- Center for Machine Learning and Intelligent Systems. Cleveland heart disease data details, Available from: URL <http://archive.ics.uci.edu/ml/machine-learningdatabases/heart-disease/heart-disease-Names>
- Cortes C. Vapnik V. (1995). Support-vector networks. Machine Learning, 20, pp. 273–297.
- D. Goldberg, K. Deb, B. Korb. (1989). Messy genetic algorithms: motivation, analysis, and first results, Complex Syst. 3, pp.493–530.
- Davari Dolatabadi, A., Esmael Zadeh, S., Mohammadzadeh, B. (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimised SVM. Vol.138, pp. 117-126.
- G. Sanchita, D. Anindita, et al. (2016). Evolutionary algorithm based techniques to handle big data, in: P.B.S. Mishra, et al. (Eds.), Techniques and Environments for Big Data Analysis: Parallel, Cloud, and Grid Computing, Springer International publishing: Cham, pp. 113–158.
- Huang T. Kecman V. Kopriva I. (2006). Kernel based algorithms for mining huge data sets, supervised, semi-supervised, and unsupervised learning, Springer-Verlag, Berlin, Heidelberg.
- Khosravani A, Ayat SS. (2015). Presenting an intelligent system for diagnosis of coronary heart disease by using Probabilistic Neural Network. Health Inf Manage; 12(1), pp.3-13.
- R. Jensen, Q. Shen. (2003). Finding Rough Set Reducts with Ant Colony Optimization, in: Proceedings of the 2003 UK Workshop on Computational Intelligence, pp. 15–22.
- S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. (1983). Optimization by simulated annealing, Science 220 (4598), pp.671–680.
- S. Mirjalili, A. Lewis. (2016). The whale optimization algorithm, Adv. Eng. Softw. Vol. 95, pp. 51–67.
- S. Pouriyeh, S. Vahid, G. Sannino, G. D. Pietro and H. Arabnia, J. Gutierrez. (2017). A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease,” IEEE Symposium on Computers and Communication.
- S. Xu, Z. Zhang, D. Wang, J. Hu, X. Duan and T. Zhu. (2017). Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework. International Conference on Big Data Analysis.
- V. Vapnik and A. Chervonenkis. (1991). The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognition and Image Analysis, vol. 1, no. 3, pp. 283-305.
- Wong, N.D. (2014). Epidemiological studies of CHD and the evolution of preventive cardiology. Nat. Rev. Cardiol. Vol.11, pp.276-289.

### استناد به این مقاله:

حسنی، زینب، خسروی مشیزی، مهین. (۱۳۹۷). «تشخیص بیماری قلبی عروق کرونر با سیستم هوشمند ترکیبی بر اساس الگوریتم نهنگ، شبیه‌ساز تبرید و ماشین بردار پشتیبان». مدیریت مهندسی و رایانش نرم، ۶ (۲)، ۱۶۷-۱۸۱.