Kharazmi University

Mathematical Research
Year 2026, Volume 12, Issue 2, pp. 22–42

Print ISSN: 2588-2546
Online ISSN: 2588-2554
DOI: 10.22034/mmr.2026.11745

Optimizing Twin Support Vector Machine for Breast Cancer Classification Using a Hybrid Metaheuristic Algorithm Based on Multi-Agent Reinforcement Learning

Ayda Rahimi[(1)] and Omid SolaymaniFard[(2)] [1]

[(1)] Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran
[(2)] Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Received: 14 April 2026 Accepted: 25 August 2026 Published online: 23 September 2026

Abstract: Breast cancer diagnosis requires accurate classification methods to support early detection and improve clinical outcomes. This study aims to propose a hybrid optimization framework to enhance Twin Support Vector Machine (Twin SVM) performance for breast cancer classification. The framework integrates four marine-inspired metaheuristic algorithms — Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), Salp Swarm Algorithm (SSA), and Marine Predators Algorithm (MPA) — as cooperative structure. Q-Learning dynamically controls the exploration-exploitation balance by adaptively selecting the most effective search operator at each iteration, preventing premature convergence and improving global search capability. Experimental results on two benchmark breast cancer datasets demonstrate that the proposed method achieves superior accuracy, precision, and recall compared to conventional Twin SVM and single-algorithm optimizers.

Keywords: Twin Support Vector Machine, Breast Cancer Classification, Hybrid Metaheuristic Algorithm, Multi Agent Reinforcement Learning.

©2026 Kharazmi University, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0 license) (http://creativecommons.org/licenses/by-nc/4.0/).

[1]Corresponding author
E-mail addresses: (Ayda Rahimi) ayda.rahimi@mail.um.ac.ir, (Omid SolaymaniFard) soleimani@ferdowsi.um.ac.ir

پژوهش‌های ریاضی
سال ۱۴۰۵، دوره 12، شماره 2، ص ۲۲–۴۲
شاپا چاپی: ۲۵۴۶-۲۵۸۸
شاپا الکترونیکی: ۲۵۵۴-۲۵۸۸
DOI: 10.22034/mmr.2026.11745

بهینه‌سازی ماشین بردار پشتیبان دوقلو در طبقه‌بندی سرطان پستان به کمک یک الگوریتم فرااکتشافی ترکیبی مبتنی بر یادگیری تقویتی چند عاملی

آیدا رحیمی [۱] و امید سلیمانی‌فرد [۲] ۲

[۱] گروه ریاضی کاربردی، دانشکدهٔ علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران
[۲] گروه ریاضی کابردی، دانشکدهٔ علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران

تاریخ دریافت: ۱۴۰۵/۰۱/۲۵ تاریخ پذیرش: ۱۴۰۵/۰۶/۰۳ تاریخ انتشار: ۱۴۰۵/۰۷/۰۱

چکیده: تشخیص سرطان پستان نیازمند روش‌های طبقه‌بندی دقیق برای شناسایی زودهنگام و بهبود پیامدهای بالینی است. این پژوهش قصد دارد یک چارچوب بهینه‌سازی ترکیبی برای ارتقای عملکرد ماشین بردار پشتیبان دوقلو (Twin SVM) در طبقه‌بندی سرطان پستان پیشنهاد بدهد. این چارچوب چهار الگوریتم فرا اکتشافی الهام‌گرفته از اکوسیستم دریایی — PSO، WOA، SSA و MPA — را به‌عنوان عامل‌های مکمل در قالب یک ساختار ادغام می‌کند. الگوریتم Q-Learning با انتخاب تطبیقی مؤثرترین عملگر جستجو در هر تکرار، تعادل بین کاوش و بهره‌برداری را به‌صورت پویا مدیریت کرده و از همگرایی زودهنگام جلوگیری می‌نماید. نتایج آزمایشی روی مجموعه داده‌های معیار سرطان پستان نشان می‌دهد که روش پیشنهادی از نظر دقت کلی، صحت و بازخوانی نسبت به Twin SVM متداول و بهینه‌سازهای تک‌الگوریتمی برتری قابل‌توجهی دارد.

واژه‌های کلیدی: ماشین بردار پشتیبان دوقلو، طبقه‌بندی سرطان پستان، الگوریتم فرااکتشافی ترکیبی، یادگیری تقویتی چندعاملی.

۱ مقدمه

سرطان پستان به عنوان یکی از شایع‌ترین و در عین حال پیچیده‌ترین انواع سرطان در میان زنان، چالشی جدی برای نظام سلامت جهانی به شمار می‌آید. تشخیص زودهنگام و دقیق این بیماری که به دو دسته عمده خوش‌خیم (Benign)

۲نویسنده مسئول مقاله
(Ayda Rahimi) ayda.rahimi@mail.um.ac.ir, (Omid SolaymaniFard) soleimani@ferdowsi.um.ac.ir

و بدخیم (Malignant) تقسیم می‌شود، نقشی حیاتی در افزایش نرخ بقا و بهبود کیفیت زندگی بیماران ایفا می‌کند[۱]. در دهه‌های اخیر، با پیشرفت فناوری‌های تصویر برداری پزشکی و ظهور کلان داده‌ها، استفاده از روش‌های هوش مصنوعی و یادگیری ماشین به ابزاری قدرتمند برای تحلیل داده‌های پزشکی و پشتیبانی از تصمیم‌گیری بالینی مبدل شده است[۲]. این الگوریتم‌ها قادرند الگوهای پیچیده و نامشهود در داده‌ها را شناسایی کرده و مدل‌های پیش‌بینی کننده‌ای با دقت بالا ارائه دهند[۳–۶].

در میان انبوه روش‌های طبقه‌بندی، ماشین بردار پشتیبان (SVM) به دلیل مبانی نظری استوار و توانایی بالا در تعمیم پذیری، به طور گسترده در مسائل تشخیص پزشکی مورد استفاده قرار گرفته است[۷]. با این حال ماشین بردار پشتیبان کلاسیک علاوه بر مواجهه با محدودیت‌هایی همچون حساسیت بالا به پارامترها و عدم توانایی در مدیریت بهینه‌ی داده‌های نامتوازن با چالش‌هایی نظیر پیچیدگی محاسباتی قابل توجه در مجموعه داده‌های بزرگ که ناشی از حل یک مسئله برنامه‌ریزی درجه دوم (QPP) است، نیز روبروست. برای رفع این چالش‌ها ماشین بردار پشتیبان دوقلو (Twin SVM) معرفی شد که با حل دو مسئله‌ی بهینه‌سازی کوچک‌تر به جای یک مسئله بزرگ، سرعت محاسبات را به طور چشم‌گیری افزایش می‌دهد و دقت طبقه‌بندی را بهبود می‌بخشد[۸].

با وجود برتری محاسباتی، کارایی ماشین بردار پشتیبان دوقلو نیز همانند نسخه کلاسیک آن، به شدت به تنظیم بهینه هایپرپارامترهای خود، از جمله پارامتر جریمه و پارامترهای تابع هسته، وابسته است. انتخاب این پارامترها خود یک مسئله بهینه سازی پیچیده است که حل آن با روش‌های سنتی مانند جستجوی شبکه‌ای (Grid Search) ناکارآمد و زمان‌بر است. اینجاست که الگوریتم‌های فرااکتشافی (Metaheuristic Algorithms) به عنوان ابزارهایی قدرتمند برای جستجو در فضاهای پیچیده و یافتن تنظیمات بهینه برای پارامترها، راهگشا بوده‌اند.

الگوریتم‌های فرااکتشافی مبتنی بر رفتار جمعی موجودات زنده نظیر بهینه سازی ازدحام ذرات (PSO) [۹] ، الگوریتم ازدحام سالپ ها (SSA) [۱۰]، الگوریتم شکارچیان دریایی (MPA) [۱۱] و الگوریتم بهینه‌سازی نهنگ‌ها (WOA) [۱۲] در سال‌های اخیر کاربردهای موفقی در حل مسائل بهینه‌سازی پیچیده داشته‌اند.

با این حال، قضیه No Free Lunch Theorem بیان می‌دارد که هیچ الگوریتم فرااکتشافی واحدی وجود ندارد که برای تمام مسائل بهینه‌سازی بهترین عملکرد را داشته باشد. هر الگوریتم نقاط قوت و ضعف خود را در توازن میان دو فاز اصلی جستجو، یعنی کاوش (Exploration) و بهره‌برداری (Exploitation) داراست. به بیان دیگر هر یک از این الگوریتم‌ها دارای نقاط قوت و ضعف خاصی هستند؛ برای مثال، PSO در کاوش سریع فضای جستجو توانمند است اما ممکن است در بهره‌برداری محلی ضعیف عمل کند، در حالی که WOA در بهره‌برداری دقیق‌تر است اما ممکن است در مراحل اولیه کاوش کافی نداشته باشد. این مشاهدات، انگیزه این پژوهش برای ارائه یک چارچوب نوین و هوشمند است.

در این مقاله یک اکوسیستم دریایی پیشنهاد داده می‌شود که از قابلیت چهار الگوریتم الهام گرفته از طبیعت یعنی PSO، SSA، MPA و WOA به صورت همزمان بهره می‌برد. برای مدیریت هوشمندانه این اکوسیستم و ایجاد توازن پویا میان کاوش و بهره‌برداری، از یادگیری تقویتی چندعاملی (Multi Agent Reinforcement Learning) مبتنی بر الگوریتم Q-Learning استفاده می‌شود. این عامل هوشمند در هر مرحله از فرآیند بهینه سازی، بر اساس وضعیت فعلی جستجو، تصمیم می‌گیرد که کدام یک از چهار الگوریتم برای هدایت جمعیت به سمت پاسخ بهتر، فعال شود.

هدف اصلی این پژوهش، توسعه یک سیستم هوشمند تشخیص سرطان پستان است که با ترکیب ماشین بردار پشتیبان دوقلو و الگوریتم فرااکتشافی ترکیبی مبتنی بر Q-Learning، دقت، حساسیت، ویژگی و سرعت تشخیص را به طور همزمان بهبود بخشد. این رویکرد نه تنها به انتخاب بهینه پارامترهای ماشین برار پشتیبان دوقلو کمک می کند، بلکه با مدیریت هوشمند فرآیند جستجو، از گیر افتادن در بهینه‌های محلی جلوگیری می‌نماید. نوآوری اصلی این کار در سه جنبه خلاصه می‌شود: (۱) طراحی یک اکوسیستم فرااکتشافی متشکل از الگوریتم‌های متنوع برای پوشش جامع فضای جستجو، (۲) استفاده از یادگیری تقویتی به عنوان یک مکانیزم تصمیم گیری سطح بالا برای مدیریت پویای فرآیند بهینه سازی و (۳) بکارگیری این چارچوب یکپارچه برای حل یک مسئله مهم در حوزه سلامت.

۲ مفاهیم اولیه

بنا به آنچه در بخش پیشین گفته شد رویکرد پیشنهادی این مطالعه با بهره‌گیری از روش ماشین بردار پشتیبان دوقلو و استفاده از مفهوم یادگیری تقویتی چندعاملی (MARL) برای ایجاد یک اکوسیستم بر پایه الگوریتم‌های فرااکتشافی منتخب پایه‌ریزی شده است. برای این منظور، این بخش اختصاص می‌یابد به ارائه مروری کوتاه و مختصر بر مبانی

نظری و تشریح ساختار روش ماشین بردار پشتیبان دوقلو، یادگیری تقویتی چندعاملی و الگوریتم‌های فرااکتشافی (PSO)، (SSA)، (MPA) و (WOA).

۱.۲ مروری بر ماشین بردار پشتیبان دوقلو

ماشین بردار پشتیبان یکی از روش‌های شناخته‌شده در یادگیری نظارت‌شده برای مسائل طبقه‌بندی دودویی است. ایده اصلی این روش بر یافتن یک مرز تصمیم‌گیری بهینه است که بتواند با بیشترین فاصله ممکن، دو کلاس داده را از یکدیگر جدا کند. این فاصله که به آن «حاشیه» گفته می‌شود، نقش مهمی در افزایش قدرت تعمیم‌پذیری مدل دارد. با وجود دقت مناسب ماشین بردار پشتیبان در بسیاری از کاربردها، فرآیند آموزش آن به‌ویژه زمانی که حجم داده‌ها زیاد است معمولاً مستلزم حل یک مسئله برنامه‌ریزی درجه دوم با ابعاد نسبتاً بزرگ است که می‌تواند از نظر محاسباتی زمان‌بر و پرهزینه باشد. به‌منظور کاهش این هزینه محاسباتی، ماشین بردار پشتیبان دوقلو به‌عنوان نسخه‌ای توسعه‌یافته و کارآمدتر از ماشین بردار پشتیبان ارائه شد. برخلاف ماشین بردار پشتیبان استاندارد که تنها یک ابرصفحه تصمیم‌گیرنده را تعیین می‌کند، در ماشین بردار پشتیبان دوقلو دو ابرصفحه غیرموازی به‌صورت همزمان محاسبه می‌شوند. هر یک از این ابرصفحه‌ها به یکی از کلاس‌ها نزدیک‌تر بوده و در عین حال تلاش می‌کند تا حد امکان از داده‌های کلاس مقابل فاصله بگیرد. نکته کلیدی در این روش آن است که مسئله بزرگ برنامه‌ریزی درجه دوم در ماشین بردار پشتیبان به دو مسئله کوچک‌تر تجزیه می‌شود. این رویکرد باعث کاهش پیچیدگی محاسباتی و در نتیجه کاهش زمان آموزش مدل می‌شود[۸].

فرض کنید مجموعه داده‌های آموزشی به صورت $S = \{(x_i, y_i)\}_{i=1}^{l}$ در اختیار باشد، که در آن $x_i \in \mathbb{R}^d$ بردار ویژگی و $y_i \in \{-1, +1\}$ برچسب کلاس نمونه iام است. داده‌های متعلق به کلاس مثبت و منفی به‌ترتیب با ماتریس‌های $A \in \mathbb{R}^{m_1 \times d}$ و $B \in \mathbb{R}^{m_2 \times d}$ نمایش داده می‌شوند و ماتریس کل داده‌ها به صورت $D = [A; B]$ تعریف می‌گردد.

در چارچوب ماشین بردار پشتیبان دوقلو، دو ابرصفحه غیرموازی که به طور همزمان محاسبه می‌شوند؛ در حالت کلی، به کمک توابع تصمیم

$$f_1(x) = K(x, D^T)w_1 + b_1 = 0 \tag{۱.۲}$$

و

$$f_2(x) = K(x, D^T)w_2 + b_2 = 0 \tag{۲.۲}$$

بیان می‌شوند که در آن $K(\cdot, \cdot)$ تابع هسته و w_i, b_i پارامترهای ابرصفحه iام هستند. در حالت خطی با انتخاب $K(x, y) = x^T y$ که در آن D ماتریس واحد است، روابط (۱.۲) و (۲.۲) به فرم $x^T w_i + b_i = 0$ تبدیل می‌شوند.

فرم پرایمال

ابرصفحه اول که به کلاس A نزدیک است، از طریق حل مسئله بهینه‌سازی

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} \left\| K(A, D^T)w_1 + e_1 b_1 \right\|^2 + C_1 e_2^T \xi \\ \text{s.t.} \quad & -\left(K(B, D^T)w_1 + e_2 b_1\right) + \xi \geq e_2, \quad \xi \geq 0 \end{aligned} \tag{۳.۲}$$

و به طور مشابه، ابرصفحه دوم که به کلاس B نزدیک است، از حل مسئله

$$\begin{aligned} \min_{w_2, b_2, \eta} \quad & \frac{1}{2} \left\| K(B, D^T)w_2 + e_2 b_2 \right\|^2 + C_2 e_1^T \eta \\ \text{s.t.} \quad & \left(K(A, D^T)w_2 + e_1 b_2\right) + \eta \geq e_1, \qquad \eta \geq 0 \end{aligned} \tag{۴.۲}$$

حاصل می‌شوند.

فرم دوآل

برای استخراج فرم دوآل، ماتریس‌های

$$H = \left[K(A, D^T)\ \ e_1\right], \qquad G = \left[K(B, D^T)\ \ e_2\right] \tag{۵.۲}$$

و بردارهای پارامتر

$$u_1 = \begin{bmatrix} w_1 \\ b_1 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} w_2 \\ b_2 \end{bmatrix} \tag{۶.۲}$$

در نظر گرفته می‌شوند.

با اعمال شرایط کاروش–کان–تاکر، فرم دوآل مسئله (۳.۲) به شکل

$$\begin{aligned} \max_{\alpha} \quad & e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH + \delta I)^{-1}G^T\alpha \\ \text{s.t.} \quad & 0 \le \alpha \le C_1 \end{aligned} \tag{۷.۲}$$

به‌دست می‌آید و پارامترهای ابرصفحه اول از رابطه

$$u_1 = -(H^TH + \delta I)^{-1}G^T\alpha \tag{۸.۲}$$

محاسبه می‌شوند.

به طور مشابه، فرم دوآل مسئله (۴.۲) برابر است با

$$\begin{aligned} \max_{\beta} \quad & e_1^T\beta - \frac{1}{2}\beta^T H(G^TG + \delta I)^{-1}H^T\beta \\ \text{s.t.} \quad & 0 \le \beta \le C_2 \end{aligned} \tag{۹.۲}$$

و پارامترهای ابرصفحه دوم از رابطه

$$u_2 = (G^TG + \delta I)^{-1}H^T\beta \tag{۱۰.۲}$$

به دست می‌آیند.

پس از تعیین دو ابرصفحه (H,G)، یک نمونه جدید x به کلاسی تخصیص داده می‌شود که فاصله هندسی آن در فضای ویژگی تا ابرصفحه متناظر کمتر باشد. در چارچوب هسته‌ای، قاعده تصمیم به صورت

$$\text{Class}(i) = \arg\min_{i=1,2} \left|K(x^T, D^T)u_i + b_i\right|, \qquad for\ \ i = 1, 2 \tag{۱۱.۲}$$

تعریف می‌شود (خواننده می تواند برای جزئیات بیشتر مرجع [۱۳] را مطالعه کند).

۱.۱.۲ تابع هسته

استفاده از ماشین بردار پشتیبان، برای حل مسائل طبقه‌بندی در بسیاری از کاربردهای عملی، مستلزم انتخاب یک تابع هسته مناسب است. در اغلب مسائل دنیای واقعی، داده‌ها در فضای ویژگی اولیه به‌صورت خطی تفکیک‌پذیر نیستند، بنابراین نگاشت آن‌ها به فضایی با ابعاد بالاتر برای جداسازی خطی ضروری است. با این حال، افزایش پیچیدگی محاسباتی یک چالش اساسی به شمار می‌رود به همین منظور، برای کاهش این پیچیدگی از تابع هسته بهره گرفته می‌شود. از نظر تئوری، هرتابعی که بتواند در شرط مرسر (Mercer) صدق کند[۱۴]، می‌تواند به عنوان تابع هسته الگوریتم ماشین بردار پشتیبان مورد استفاده قرار گیرد. با این وجود، انتخاب نوع تابع هسته تأثیر مستقیمی بر عملکرد طبقه‌بند دارد و می‌تواند منجر به رفتارهای متفاوت در مدل شود. بنابراین، انتخاب صحیح تابع هسته و تنظیم مناسب پارامترهای آن، نقش تعیین‌کننده‌ای در بهبود ساختار طبقه‌بند و افزایش دقت مدل ایفا می‌کند[۲].

به طور کلی، مهم‌ترین توابع هسته مورد استفاده در پژوهش‌های مرتبط عبارت اند از:

۱) هسته خطی

این تابع ساده‌ترین نوع بوده و برای مسائل طبقه‌بندی خطی مورد استفاده قرار می‌گیرد. تابع هسته خطی به صورت

$$K(x_i, x_j) = x_i \cdot x_j \tag{۱۲.۲}$$

تعریف می‌شود که در این حالت، طبقه‌بند حاصل، یک مرز تصمیم خطی در فضای ویژگی اولیه ایجاد می‌کند.

۲) هسته چندجمله‌ای

این تابع هسته امکان مدل‌سازی روابط غیرخطی را از طریق ایجاد یک طبقه‌بند چندجمله‌ای با درجه d فراهم می‌کند و به صورت

$$K(x_i, x_j) = (x_i \cdot x_j + ۱)^d \tag{۱۳.۲}$$

تعریف می‌شود که در آن d درجه چندجمله‌ای است و نقش مهمی در پیچیدگی مدل ایفا می‌کند.

۳) هسته شعاعی

این تابع هسته که با نام هسته گوسی نیز شناخته می‌شود بر اساس فاصله اقلیدسی بین نمونه‌ها تعریف شده و به فرم

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^۲}{\sigma^۲}\right) \tag{۱۴.۲}$$

بیان می‌گردد. در این حالت، طبقه‌بند حاصل با روش رایج هسته شعاعی تفاوت دارد؛ زیرا برای هر تابع پایه، یک بردار پشتیبان متناظر به مرکز هر کدام از توابع وجود دارد و وزن‌های خروجی به صورت خودکار توسط الگوریتم تعیین می‌شوند.

۴) هسته سیگموئید

تابع سیگموئید نیز می‌تواند به عنوان هسته مورد استفاده قرار گیرد که با رابطه

$$K(x, x_i) = \tanh\left(v(x \cdot x_i) + c\right) \tag{۱۵.۲}$$

تعریف می‌شود.

در این پژوهش از هسته پایه شعاعی استفاده می‌شود. از آنجا که داده‌های مورد استفاده در این پژوهش دارای ساختار غیرخطی هستند، استفاده از هسته پایه شعاعی امکان مدل‌سازی بهتر این روابط را فراهم می‌کند و در نتیجه به بهبود دقت طبقه‌بندی کمک می‌نماید.

۲.۲ یادگیری تقویتی چندعاملی

یادگیری تقویتی یکی از شاخه‌های هوش مصنوعی است که ساختار آن شباهت زیادی به یادگیری انسان دارد. در این مدل، ما با یک عامل سروکار داریم که در یک محیط قرار می‌گیرد و باید از طریق آزمون و خطا، بهترین تصمیم‌ها را بگیرد[۱۵، ۱۶]. در اینجا عامل با انجام هر اقدام، یک بازخورد از محیط دریافت می‌کند. اگر اقدامش درست باشد، پاداش می‌گیرد و اگر اشتباه باشد، جریمه می‌شود(شکل۱) . نکته کلیدی در این روش، آینده‌نگری است؛ یعنی سیستم فقط به دنبال سود سریع و لحظه‌ای نیست بلکه یاد می‌گیرد تصمیماتی بگیرد که شاید در ابتدا پاداش کمی داشته باشند اما در نهایت منجر به بهترین نتیجه ممکن می‌شوند. با این حال چارچوب فوق بیانگر حالت پایه و تک‌عاملی یادگیری تقویتی است در حالی‌که در بسیاری از مسائل پیچیده ـ از جمله مسئله مورد بررسی در این پژوهش ـ تصمیم‌گیری به‌صورت هم‌زمان توسط چندین عامل انجام می‌شود که رفتار هر یک می‌تواند بر عملکرد سایرین و بر وضعیت کلی محیط اثرگذار باشد؛ از این‌رو، در این مطالعه از رویکرد یادگیری تقویتی چندعاملی استفاده می‌شود که در آن مجموعه‌ای از عامل‌ها در یک محیط مشترک تعامل کرده و هر عامل بر اساس سیاست تصمیم‌گیری مستقل خود عمل می‌کند در حالی‌که نتیجه نهایی حاصل برهم‌کنش جمعی تصمیم‌های تمامی عامل‌ها خواهد بود(شکل ۲).

۱.۲.۲ اجزای اصلی مسئله یادگیری تقویتی

در مدل‌سازی یادگیری تقویتی، معمولاً اجزای زیر تعریف می‌شوند[۱۶]:

۱) محیط (Environment)
محیط سامانه‌ای است که عامل یا مجموعه‌ای از عامل‌ها در آن قرار دارند و با آن تعامل می‌کنند. این سامانه در واکنش به تصمیم‌های اتخاذشده، پیامد متناظر با آن تصمیم را به‌صورت پاداش یا جریمه بازمی‌گرداند. در واقع، محیط چارچوب پویای مسئله را تعیین می‌کند و مشخص می‌سازد که هر تصمیم در یک وضعیت مشخص، به چه نتیجه‌ای منجر خواهد شد.

۲) عمل (Action) عمل تصمیمی است که عامل در هر لحظه بر اساس وضعیت جاری اتخاذ می‌کند. مجموعه تمام عمل‌ها را با نماد A و عمل انتخاب‌شده در زمان t را با A_t نمایش می‌دهیم. فضای عمل‌ها بسته به ماهیت مسئله می‌تواند گسسته یا پیوسته تعریف شوند.

۳) حالت (State)
حالت مجموعه‌ای از اطلاعات است که وضعیت فعلی محیط را توصیف می‌کند و اطلاعاتی را در اختیار عامل قرار می‌دهد تا بر اساس آن تصمیم‌گیری کند. متغیر S مجموعه حالات ممکن را نشان می‌دهد و در زمان t، وضعیت سیستم با S_t نمایش داده می‌شود. حالت می‌تواند شامل هرگونه اطلاعات مرتبط با وضعیت جاری سیستم باشد و بسته به نوع مسئله، به‌صورت گسسته یا پیوسته تعریف شود. تغییر وضعیت سیستم به حالت فعلی و عمل انتخاب‌شده وابسته است و احتمال انتقال به حالت بعدی با توزیع شرطی

$$P(S_{t+1} \mid S_t, A_t). \tag{۱۶.۲}$$

توصیف می‌شود.

۴) پاداش (Reward) پاداش (R_t)، بازخوردی عددی و دارای مقدار مثبت یا منفی است که در زمان t و در پاسخ به عمل اتخاذشده، از طرف محیط، تولید و به عامل ارائه می‌شود. این کمیت میزان مطلوبیت پیامد حاصل از تصمیم اتخاذشده را بیان می‌کند. به‌صورت کلی، پاداش می‌تواند تابعی از حالت فعلی، عمل انجام‌شده و حالت بعدی باشد و به شکل

$$R_t = R(S_t, A_t, S_{t+1}). \tag{۱۷.۲}$$

تعریف می‌شود.

۵) سیاست (Policy)

سیاست قاعده‌ای است که تعیین می‌کند عامل در هر وضعیت چه عملی را انتخاب کند. اگر سیاست به‌صورت تصادفی تعریف شود، برای هر وضعیت احتمال انتخاب هر یک از عمل‌های ممکن را معین می‌کند؛ یعنی مشخص می‌کند با چه احتمالی هر عمل در یک وضعیت خاص انتخاب می‌شود. در مقابل، در سیاست قطعی، برای هر وضعیت تنها یک عمل مشخص در نظر گرفته می‌شود و عامل همواره همان عمل را در آن وضعیت انتخاب می‌کند. هدف در یادگیری تقویتی، یافتن سیاستی است که مجموع پاداش‌های مورد انتظار را در طول زمان بیشینه کند.

در حالت کلی، یک سیاست تصادفی به شکل

$$\pi(a \mid s) = P(A_t = a \mid S_t = s), \tag{۱۸.۲}$$

و در حالت قطعی، سیاست به‌صورت تابعی از وضعیت

$$A_t = \pi(S_t). \tag{۱۹.۲}$$

تعریف می‌شوند.

۶) تابع ارزش (Value Function)

تابع ارزش برای ارزیابی کیفیت یک سیاست به کار گرفته می‌شود. این تابع بیان می‌کند اگر سیستم از وضعیت s_t شروع کرده و مطابق سیاست مشخصی عمل کند، در آینده چه میزان پاداش مورد انتظار دریافت خواهد شد. به‌عبارت دیگر، تابع ارزش معیار سنجش عملکرد یک سیاست در افق زمانی آینده است. نسخه بهینه این تابع از طریق بیشینه‌سازی روی تمامی سیاست‌های ممکن به‌دست می‌آید. تابع ارزش بهینه به شکل

$$V_t^*(s_t) = \max_{\pi \in \Pi} V_t^{\pi}(s_t) = \max_{\pi \in \Pi} \mathbb{E}_{\pi} \left[\sum_{k=۰}^{T-t} \gamma^k R_{t+k} \;\middle|\; S_t = s_t \right] \tag{۲۰.۲}$$

تعریف می‌شود که در آن، $\gamma \in [۰, ۱]$ نشان‌دهنده ضریب تنزیل است و اهمیت پاداش‌های آینده را نسبت به پاداش‌های فوری تعیین می‌کند. هرچه مقدار γ به صفر نزدیک‌تر باشد، عامل پاداش‌های آنی را ترجیح می‌دهد (دید کوتاه‌مدت) و هرچه به یک نزدیک‌تر باشد، به پاداش‌های دوردست در آینده بهای بیشتری می‌دهد (دید بلندمدت).

۲.۲.۲ الگوریتم Q-Learning

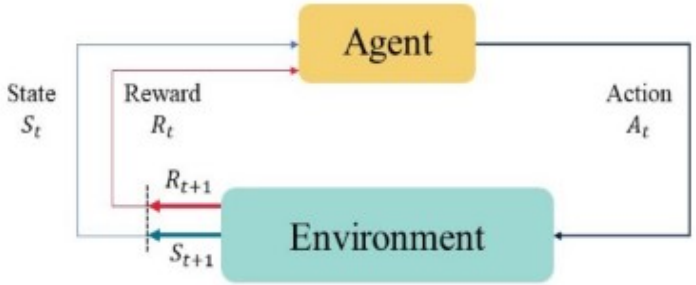
هرچند تابع ارزش رکن اصلی یادگیری تقویتی است، اما حتی با وجود در اختیار داشتن مدلی دقیق از محیط، یافتن بهترین استراتژی از طریق محاسبات مستقیم معمولاً بسیار پیچیده است. به همین دلیل، از روش یادگیری تفاضل زمانی (TD) استفاده می‌شود. در این روش، عامل منتظر نمی‌ماند تا تمام مراحل به پایان برسد، بلکه در هر گام، تخمین قبلی خود را بر اساس آنچه مشاهده کرده اصلاح می‌کند. این رویکرد بدون نیاز به دانستن فرمول‌های پیچیده محیط، می‌تواند راهکارهای بهینه را پیدا کند. در این میان، الگوریتم Q-Learning یکی از محبوب‌ترین روش‌ها در این حوزه است. این الگوریتم به جای تخمین کلی، مستقیماً به سراغ برآورد تابع Q می‌رود. ویژگی کلیدی Q-Learning این است که برای اجرا، نیازی به اطلاعات قبلی از احتمالات محیط ندارد و می‌تواند سیاست‌های بهینه را مستقیماً از دل داده‌های یک سیستم بیرون بکشد

$$Q_{t+۱}(S_t, A_t) = Q_t(S_t, A_t) + \alpha[R_t + \gamma \max_a Q_t(S_{t+۱}, a) - Q_t(S_t, A_t)] \tag{۲۱.۲}$$

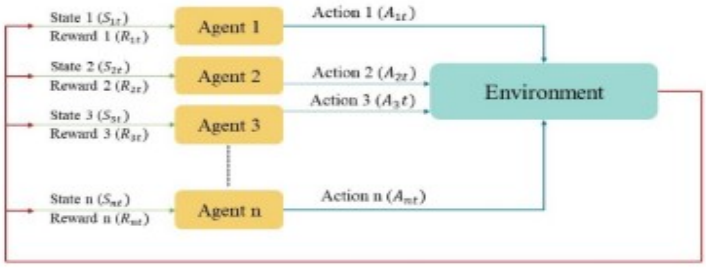
در واقع، با این روش کافی است وضعیت فعلی را بدانیم تا بتوانیم سیاست بهینه را به کمک رابطه‌ی

$$\pi^*(s) = \arg\max_a Q(s, a) \tag{۲۲.۲}$$

برای رسیدن به بهترین نتیجه در بلندمدت انتخاب کنیم.



شکل ۱: طرح‌واره‌ی یادگیری تقویتی .



شکل ۲: طرح‌واره‌ی یادگیری تقویتی چندعاملی .

۳.۲ الگوریتم‌های فرااکتشافی

الگوریتم‌های فرااکتشافی دسته‌ای از الگوریتم‌های بهینه‌سازی هستند که برای حل مسائل پیچیده‌ای که روش‌های کلاسیک در آن کارایی ندارند، طراحی شده‌اند. این الگوریتم‌ها از فرآیندهای طبیعی مانند تکامل، زیست‌شناسی یا فیزیک الهام می‌گیرند و بدنبال یافتن جواب‌های قابل قبول در یک زمان منطقی هستند. در پژوهش حاضر، از الگوریتم‌های PSO، SSA، MPA و WOA استفاده می‌شود که در ادامه به به معرفی مختصر روند کار و فرمول‌های هر یک خواهیم پرداخت.

PSO

به عنوان یکی از روش‌های فرااکتشافی، بر اساس رفتار جمعی موجودات زنده مانند پرندگان یا ماهی‌ها در یافتن منابع غذایی بنا شده است[۹]. در این الگوریتم، مجموعه‌ای از ذرات که هر یک نمایانگر یک جواب در فضای جستجو هستند، به طور همزمان به سمت جواب بهینه حرکت می‌کنند. تکامل موقعیت و سرعت هر ذره تحت تأثیر دو عامل اصلی هدایت می‌شود: بهترین موقعیت کشف شده توسط خود ذره $(pbest)$ و بهترین موقعیت کشف شده توسط هر ذره در کل ازدحام $(gbest)$. این رویکرد ترکیبی، به ذرات اجازه می‌دهد تا از تجربیات فردی و جمعی برای دستیابی به جواب بهینه بهره‌مند شوند.

فرمول به‌روز رسانی سرعت ذره v در گام زمانی $t+1$ به صورت

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (pbest_i - x_i^t) + c_2 \cdot r_2 \cdot (gbest - x_i^t) \qquad (۲۳.۲)$$

است.

در رابطه (۲۳.۲)، پارامترها و ضرایب عملکردی الگوریتم نقش کلیدی در هدایت ذرات ایفا می‌کنند. به طوری که x_i^t و v_i^t به ترتیب نشان‌دهنده موقعیت مکانی و سرعت ذره iـام در گام زمانی t هستند. در این میان، پارامتر w که تحت عنوان وزن لختی شناخته می‌شود، وظیفه کنترل میزان تمایل ذره به حفظ سرعت قبلی خود را بر عهده دارد مقادیر بزرگ‌تر این ضریب توانایی کاوش سراسری الگوریتم را افزایش می‌دهد در حالی که مقادیر کوچک‌تر آن بر جستجوی محلی و همگرایی دقیق‌تر تمرکز دارد. از سوی دیگر، حرکت ذرات تحت تأثیر دو مؤلفه یادگیری به نام‌های c_1 و c_2 هدایت می‌شود. ضریب c_1 یا همان مؤلفه یادگیری فردی، میزان تأثیر بهترین تجربه شخصی ذره $(pbest_i)$

را در گام بعدی تعیین کرده و میزان الگوبرداری ذره از بهترین موقعیت کشف شده توسط خود را نشان می‌دهد. در مقابل، ضریب $c_۲$ به عنوان مؤلفه یادگیری جمعی یا اجتماعی، میزان اثرگذاری بهترین تجربه کل توده ذرات ($gbest$) را مشخص می‌کند و بیانگر تمایل ذره به حرکت به سمت بهترین نقطه کشف‌شده توسط گروه است. در نهایت به منظور حفظ پویایی، القای ماهیت تصادفی به فرآیند جستجو و همچنین جلوگیری از همگرایی زودهنگام ذرات در محلی نامناسب از دو عدد تصادفی $r_۱$ و $r_۲$ با توزیع یکنواخت در بازه $[۰, ۱]$ استفاده می‌شود.

سپس موقعیت جدید هر ذره x_i با استفاده از سرعت به‌روز رسانی شده به صورت

$$x_i^{t+۱} = x_i^t + v_i^{t+۱} \tag{۲۴.۲}$$

محاسبه می‌شود.

SSA

یک روش فرااکتشافی نوین است که با الهام از رفتار زیستی سالپ‌ها در اقیانوس طراحی شده است[۱۰]. ویژگی متمایز این موجودات، تشکیل زنجیره‌های سالپی برای بهبود حرکت و جستجوی غذا است که در این الگوریتم به صورت ریاضی مدل‌سازی می‌شود. در این ساختار، جمعیت به دو گروه پیشرو (Leader) و پیرو (Followers) تقسیم می‌شود؛ سالپ رهبر در رأس زنجیره قرار داشته و وظیفه هدایت گروه به سمت منبع غذا را بر عهده دارد، در حالی که سالپ‌های پیرو به صورت متوالی از یکدیگر تبعیت کرده و انسجام زنجیره را حفظ می‌کنند.

موقعیت سالپ رهبر بر اساس موقعیت منبع غذا (F_j)، پارامتر کنترلی $c_۱$ و اعداد تصادفی $c_۲$ و $c_۳$ در بازه $[۰, ۱]$ به شکل

$$x_j^{۱} = \begin{cases} F_j + c_۱((ub_j - lb_j)c_۲ + lb_j) & c_۳ \geq ۰/۵ \\ F_j - c_۱((ub_j - lb_j)c_۲ + lb_j) & c_۳ < ۰/۵ \end{cases} \tag{۲۵.۲}$$

به‌روز رسانی می‌شود.

$c_۱$ به عنوان مهم‌ترین پارامتر در این الگوریتم عمل کرده زیرا تعادل بین فاز اکتشاف و استخراج را حفظ می‌کند و به صورت

$$c_۱ = ۲e^{-\left(\frac{۴l}{L}\right)^۲} \tag{۲۶.۲}$$

تعریف می‌شود که در آن l تکرار فعلی و L حداکثر تعداد تکرارها است.

همچنین معادله بروزرسانی موقعیت سالپ پیرو iـام در بعد jـام

$$x_j^i = \frac{۱}{۲}(x_j^i + x_j^{i-۱}), \quad i \geq ۲ \tag{۲۷.۲}$$

می‌باشد.

MPA

این الگوریتم به طور خاص بر اساس رفتار تعقیب و گریز بین شکارچی و طعمه، که در لایه‌های مختلف اقیانوس رخ می‌دهد، طراحی شده است به طوری که جمعیت اولیه‌ای از شکارچیان و طعمه‌ها را تولید می‌کند و در هر تکرار، موقعیت طعمه را بر اساس بهترین جواب یافت شده تا آن لحظه به‌روز رسانی می‌نماید[۱۱]. بروزرسانی موقعیت‌ها در سه فاز متمایز صورت می‌گیرد که هر کدام بر اساس نسبت سرعت شکارچی به طعمه تعریف می‌شوند. این رویکرد دینامیکی، الگوریتم را قادر می‌سازد تا از گیر افتادن در بهینه‌های محلی اجتناب کرده و قابلیت همگرایی به سمت بهینه‌های سراسری را افزایش دهد.

الگوریتم MPA در سه فاز اصلی عمل می‌کند:

فاز ۱ : زمانی که طعمه سریع‌تر از شکارچی حرکت می‌کند

این فاز برای $\frac{۱}{۳}$ ابتدایی تکرارها انجام می‌شود. در این شرایط، بهترین استراتژی برای شکارچی این است که ثابت بماند و موقعیت طعمه با حرکت برآونی به صورت

$$\overrightarrow{stepsize_i} = \overrightarrow{R}_B \otimes \left(\overrightarrow{Elite_i} - \overrightarrow{R}_B \otimes \overrightarrow{Prey_i}\right), \quad i = ۱, \ldots, n \tag{۲۸.۲}$$

$$\overrightarrow{Prey_i} = \overrightarrow{Prey_i} + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize_i} \tag{۲۹.۲}$$

به‌روز رسانی شود.

فاز ۲ : زمانی که شکارچی و طعمه با سرعت تقریباً یکسان حرکت می‌کنند

این فاز برای $\frac{۱}{۳}$ میانی تکرارها انجام می‌شود. در این بخش اکتشاف و استخراج هر دو مهم هستند بنابراین نیمی از جمعیت به کاوش و نیمی دیگر به بهره‌برداری می‌پردازد. بدین منظور، طعمه به صورت لوی و شکارچی به صورت برآونی حرکت می‌کند.

برای نیمه اولیه جمعیت، طعمه به صورت لوی

$$\overrightarrow{stepsize_i} = \overrightarrow{R}_L \otimes \left(\overrightarrow{Elite_i} - \overrightarrow{R}_L \otimes \overrightarrow{Prey_i}\right), \quad i = ۱, \ldots, \frac{n}{۲} \tag{۳۰.۲}$$

$$\overrightarrow{Prey_i} = \overrightarrow{Prey_i} + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize_i} \tag{۳۱.۲}$$

و برای نیمه ثانویه، طعمه به صورت برآونی

$$\overrightarrow{stepsize_i} = \overrightarrow{R}_B \otimes \left(\overrightarrow{R}_B \otimes \overrightarrow{Elite_i} - \overrightarrow{Prey_i}\right), \quad i = \frac{n}{۲} + ۱, \ldots, n \tag{۳۲.۲}$$

$$\overrightarrow{Prey_i} = \overrightarrow{Elite_i} + P \cdot CF \otimes \overrightarrow{stepsize_i} \tag{۳۳.۲}$$

حرکت می‌کند که در آن ضریب همگرایی CF به شکل

$$CF = \left(۱ - \frac{\text{Iter}}{\text{Max_Iter}}\right)^{۲\frac{\text{Iter}}{\text{MaxIter}}} \tag{۳۴.۲}$$

تعریف می‌شود.

فاز ۳ : زمانی که شکارچی سریع‌تر از طعمه حرکت می‌کند

این فاز برای $\frac{۱}{۳}$ انتهایی تکرارها انجام می‌شود. در این فاز بهترین استراتژی برای شکارچی این است که به صورت لوی حرکت کند. بروزرسانی موقعیت طعمه به صورت

$$\overrightarrow{stepsize_i} = \overrightarrow{R}_L \otimes \left(\overrightarrow{R}_L \otimes \overrightarrow{Elite_i} - \overrightarrow{Prey_i}\right), \quad i = ۱, \ldots, n \tag{۳۵.۲}$$

$$\overrightarrow{Prey_i} = \overrightarrow{Elite_i} + P \cdot CF \otimes \overrightarrow{stepsize_i} \tag{۳۶.۲}$$

خواهد بود.

WOA

الگوریتم بهینه‌سازی وال، یک روش فرااکتشافی مبتنی بر جمعیت است که از رفتار شکار گروهی نهنگ های گوژپشت الهام گرفته است[۱۲] . ایده اصلی این الگوریتم، تقلید از استراتژی شکار (Bubble-Net) است که در آن نهنگ‌ها با ایجاد حباب‌های مارپیچی، طعمه‌های خود را محاصره و به سمت خود می‌کشند.

در طبیعت نهنگ‌های گوژپشت ابتدا موقعیت طعمه را شناسایی نموده و سپس اقدام به محاصره آن می‌نمایند. به دلیل اینکه موقعیت بهترین طراحی در فضای جستجو از قبل مشخص نیست، الگوریتم WOA فرض می‌کند که بهترین جواب فعلی همان طعمه هدف است یا به آن نزدیک است. بعد از اینکه بهترین عامل جستجو مشخص شد، سایر عامل‌ها تلاش می‌کنند تا موقعیت خود را به سمت بهترین عامل به‌روز کنند. این رفتار با معادلات

$$\overrightarrow{D} = |\overrightarrow{C} \cdot \overrightarrow{X^*}(t) - \overrightarrow{X}(t)| \tag{۳۷.۲}$$

$$\overrightarrow{X}(t+۱) = \overrightarrow{X^*}(t) - \overrightarrow{A} \cdot \overrightarrow{D} \tag{۳۸.۲}$$

نمایش داده می‌شود که در آن بردارهای $\overrightarrow{A}$ و $\overrightarrow{C}$ به صورت

$$\overrightarrow{A} = ۲\overrightarrow{a} \cdot \overrightarrow{r} - \overrightarrow{a} \tag{۳۹.۲}$$

$$\overrightarrow{C} = ۲ \cdot \overrightarrow{r} \tag{۴۰.۲}$$

هستند به طوری که $\overrightarrow{a}$ به طور خطی از ۲ تا ۰ کاهش می‌یابد و $\overrightarrow{r}$ نیز یک بردار تصادفی در بازه $[۰, ۱]$ است.
نهنگ‌ها با استفاده از پارامتر a که به تدریج کوچک می‌شود شعاع حرکت خود را به دور طعمه کاهش می‌دهند. این کار باعث می‌شود که عامل‌ها به طور فزاینده‌ای به بهترین جواب شناخته شده نزدیک شوند و آن را بهتر کاوش کنند. در واقع، موقعیت جدیدشان بین جای فعلی خودشان و بهترین موقعیت موجود قرار می‌گیرد.
هم‌چنین، نهنگ‌ها حرکتی مارپیچی نیز از خود نشان می‌دهند. در این حالت، با محاسبه فاصله تا بهترین موقعیت (طعمه)، مسیری مارپیچ (مانند حرکت حلزونی) بین نهنگ و طعمه ایجاد می‌شود. این مسیر، حرکت واقعی نهنگ‌ها را که به دور طعمه می‌چرخند و همزمان به سمت آن حرکت می‌کنند، شبیه‌سازی می‌کند.
برای اینکه این رفتار پیچیده نهنگ‌ها به بهترین شکل مدل شود، الگوریتم به صورت تصادفی (با احتمال ۵۰٪) یکی از دو روش بالا را برای به‌روزرسانی موقعیت هر نهنگ

$$\overrightarrow{X}(t+۱) = \begin{cases} \overrightarrow{X^*}(t) - \overrightarrow{A} \cdot \overrightarrow{D} & \text{if } p < ۰/۵ \\ \overrightarrow{D} \cdot e^{bl} \cdot \cos(۲\pi l) + \overrightarrow{X^*}(t) & \text{if } p \geq ۰/۵ \end{cases} \tag{۴۱.۲}$$

استفاده می‌کند.
در مرحله(بهره‌برداری)، عامل‌ها به سمت بهترین موقعیت شناخته شده حرکت می‌کردند اما در مرحله(کاوش) آن‌ها به صورت تصادفی به اطراف نگاه می‌کنند و با استفاده بردار $\overrightarrow{A}$ که مقادیر مثبت و منفی دارد سعی می‌کنند از یک نهنگ خاص فاصله بگیرند. این کار باعث می‌شود که آن‌ها مناطق جدیدی را که شاید قبلاً بررسی نکرده‌اند، کشف کنند و موقعیت جدیدشان به صورت

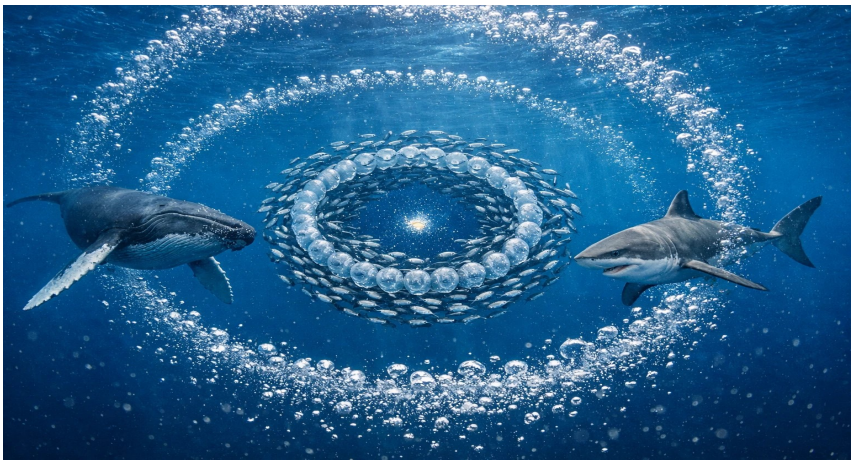
$$\overrightarrow{D} = \left|\overrightarrow{C} \cdot \overrightarrow{X_{\text{rand}}} - \overrightarrow{X}\right| \tag{۴۲.۲}$$

$$\overrightarrow{X}(t+۱) = \overrightarrow{X_{\text{rand}}} - \overrightarrow{A} \cdot \overrightarrow{D} \tag{۴۳.۲}$$

محاسبه شود.

۳ اکوسیستم دریایی

در اکوسیستم دریایی (Marine Ecosystem) مبتنی بر یادگیری تقویتی چندعاملی، فرآیند بهینه‌سازی به صورت تکراری انجام می‌شود و در هر تکرار چند الگوریتم فرااکتشافی به طور هم‌زمان در فضای جستجو فعالیت می‌کنند. در این چارچوب، چهار الگوریتم الهام‌گرفته از موجودات دریایی یعنی (PSO)، (SSA)، (MPA) و (WOA) مورد استفاده قرار می‌گیرند. این الگوریتم‌ها به دلیل تفاوت در ساز و کار جستجو و نحوه هدایت عامل‌ها به سمت نواحی امیدبخش فضای جواب انتخاب شده‌اند. هر یک از این الگوریتم‌ها فرآیند جستجوی فضای جواب را بر اساس موقعیت‌های کلیدی متفاوتی هدایت می‌کنند. در الگوریتم PSO فرآیند جستجو بر اساس بهترین تجربه فردی ذرات و بهترین تجربه سراسری جمعیت انجام می‌شود؛ به این صورت که هر ذره موقعیت خود را با توجه به این دو مرجع به‌روزرسانی می‌کند. در الگوریتم MPA موقعیت شکارچیان و طعمه‌ها در تعامل با یکدیگر تعیین‌کننده مسیر جستجو است و عامل‌ها با استفاده از الگوهای حرکتی برآونی و لوی در مراحل مختلف به سمت بهترین جواب‌ها هدایت می‌شوند. در الگوریتم SSA نیز فرآیند جستجو با استفاده از موقعیت رهبر (Leader) که نشان‌دهنده بهترین جواب فعلی است هدایت می‌شود و سایر اعضای جمعیت با پیروی از آن موقعیت‌های خود را تنظیم می‌کنند. در الگوریتم WOA نیز جستجو عمدتاً بر اساس بهترین جواب سراسری انجام می‌شود و عامل‌ها با شبیه‌سازی رفتار محاصره و حرکت مارپیچی نهنگ‌ها به سمت این موقعیت حرکت می‌کنند، در حالی که در برخی مراحل نیز حرکت‌های تصادفی برای افزایش تنوع جستجو به کار گرفته می‌شود(شکل۳).



شکل ۳: نمایی از اکوسیستم دریایی .

به منظور اجرای این اکوسیستم، در گام نخست یک جمعیت اولیه یکسان شامل N عامل در فضای جستجوی D بعدی مسئله تولید می‌شود. فرآیند تولید این جمعیتِ پایه به صورت تصادفی و بر اساس توزیع یکنواخت در محدوده کران‌های بالا و پایین مجاز فضا انجام می‌گیرد. بر این اساس، موقعیت ابتدایی ذره iـام در بعد jـام ($x^{\circ}_{i,j}$) برای تمامی الگوریتم‌ها به صورت یکسان و مطابق با رابطه

$$x^{\circ}_{i,j} = LB_j + \text{rand}(\circ, ۱) \times (UB_j - LB_j), \quad \forall i \in \{۱, \dots, N\}, \forall j \in \{۱, \dots, D\} \qquad (۱.۳)$$

تعیین می‌گردد که در آن، LB_j و UB_j به ترتیب بیانگر حد پایین و حد بالای تعریف شده برای بعد jـام فضای جستجوی مسئله هستند و $\text{rand}(\circ, ۱)$ یک عدد تصادفی با توزیع یکنواخت در بازه $[\circ, ۱]$ است. این مکانیزم تضمین می‌کند که در تکرار اول، تمام الگوریتم‌ها فرآیند رقابت و جستجوی خود را از شرایط کاملاً همسان و متوازن آغاز کنند. در تکرار اول، هر چهار الگوریتم یاد شده به طور مستقل بر روی این جمعیت اولیه یکسان اجرا می‌شوند و هر یک تلاش می‌کند بر اساس مکانیزم اختصاصی خود، بهترین جواب ممکن را بیابد. پس از پایان اجرای الگوریتم‌ها در تکرار جاری، عملکرد آن‌ها بر اساس میزان شاخص برازش با یکدیگر مقایسه شده و الگوریتمی که بهینه‌ترین مقدار تابع هدف را تولید کرده باشد، به عنوان الگوریتم برتر آن تکرار انتخاب می‌شود. سپس، کل ماتریس جمعیتی که

توسط این الگوریتم برتر ارتقا یافته و به‌روز رسانی شده است، به عنوان جمعیت مشترک برای تکرار بعدی در نظر گرفته می‌شود. به عبارت دیگر، در تکرار بعدی، هر چهار الگوریتم مجددا کار خود را به طور هم‌زمان از همین موقعیت‌های بهینه شده توسط الگوریتم پیروز آغاز می‌کنند. این مکانیزم انتقال و اشتراک‌گذاری سراسری اطلاعات، نوعی رقابت سازنده میان الگوریتم‌ها ایجاد می‌کند؛ چرا که نقاط قوت کشف شده توسط یک استراتژی جستجو (مثلا جهش‌های تصادفی موثر در WOA) در گام بعدی در اختیار سایر مکانیزم‌ها مانند (جستجوی محلی دقیق در PSO) قرار می‌گیرد و فرآیند بهینه‌سازی را به سمت اهداف مطلوب هدایت می‌کند.

از سوی دیگر، یکی از چالش‌های اصلی در الگوریتم‌های فرااکتشافی، ایجاد تعادل پویا میان دو فرآیند کاوش و بهره‌برداری است. در بسیاری از الگوریتم‌های استاندارد، این تعادل به صورت تصادفی تعیین می‌شود که انعطاف‌پذیری لازم را در مواجهه با فضاهای جستجوی پیچیده ندارند. در روش پیشنهادی، برای مدیریت هوشمندانه این تعادل در ساختار هر الگوریتم، از چارچوب یادگیری تقویتی چندعاملی استفاده شده است. فضای حالت برای الگوریتم یادگیری تقویتی چند عاملی یکی از سه حالت کاوش، بهره‌برداری و تعادل بین این دو است و عمل انتخاب یکی از این سه حالت است. در هر تکرار تجربه‌های شخصی هر عامل ارزیابی می‌شود و پاداش بر اساس تفاوت مثبت بین برازش فعلی و برازش قبل محاسبه می‌شود و به $Q-table$ داده می‌شود و الگوریتم Q-Learning سیاست جدید یعنی انتخاب یکی از فرآیندهای کاوش، بهره‌برداری و یا تعادل بین کاوش و بهره‌برداری را برای تکرار بعد تعیین می‌کند. این رویکرد باعث می‌شود الگوریتم‌های فرااکتشافی هوشمندانه و سریع‌تر به نتیجه مطلوب برسند. به عبارت دیگر، در هر یک از الگوریتم‌های PSO، SSA، MPA و WOA، یک عامل یادگیری مبتنی بر Q-learning تعبیه شده است که وظیفه دارد در هر تکرار، بهینه‌ترین استراتژی حرکتی را از میان سه عمل اصلی شامل «کاوش»، «تعادل میان کاوش و بهره‌برداری» و «بهره‌برداری» بر اساس ماتریس ارزش ($Q-table$) خود انتخاب نماید. توابع ارزش $Q(s_i, a_{i,m})$ بازتاب‌دهنده میزان موفقیتی هستند که عامل iـ‌ام در تکرارهای گذشته از طریق اعمال استراتژی mـ‌ام در حالت s_i به دست آورده است. مکانیزم تصمیم‌گیری و منطق بیشینه‌سازی ارزش در این چارچوب به صورت زیر تبیین می‌گردد:

۱. **بیشینه بودن $Q(s_i, a_{i,۱})$ (انتخاب کاوش):** این وضعیت نشان می‌دهد که الگوریتم در تکرارهای اخیر در یک بهینه محلی گرفتار شده و تنوع جواب‌ها کاهش یافته است. بنابراین، عامل با انتخاب این گزینه، دستور استفاده از پارامترهایی با قابلیت کاوش قوی را صادر می‌کند تا ذرات با گام‌های بلندتر از تله‌های محلی خارج شده و نواحی ناشناخته فضا را جستجو کنند.

۲. **بیشینه بودن $Q(s_i, a_{i,۲})$ (انتخاب حالت متعادل):** این حالت بیانگر آن است که الگوریتم توانسته توازن پایداری میان کشف نواحی جدید و بررسی دقت جواب‌های گذشته برقرار سازد. در این شرایط، پارامترهای متعادل انتخاب می‌شوند تا ضمن حفظ جهت‌گیری به سمت بهترین جواب‌ها، پویایی فرآیند جستجو برای یافتن مسیرهای موازی نیز حفظ شود.

۳. **بیشینه بودن $Q(s_i, a_{i,۳})$ (انتخاب بهره‌برداری):** بیشینه شدن این مقدار بدان معناست که الگوریتم در نزدیکی جواب بهینه سراسری قرار گرفته است و جستجوی وسیع مزیتی ندارد. در نتیجه، عامل با انتخاب بهره‌برداری و تنظیم پارامترها به سمت گام‌های کوچک و محلی، تمرکز الگوریتم را بر بهبود دقت جواب فعلی معطوف می‌سازد..

در این سازوکار، عامل‌های یادگیرنده با دریافت بازخورد از کیفیت پاسخ‌های به‌دست‌آمده، به تدریج می‌آموزند که در شرایط مختلف فضای جستجو کدام رفتار می‌تواند عملکرد بهتری ایجاد کند. به این ترتیب، به جای استفاده از انتخاب‌های تصادفی یا تنظیمات ثابت پارامترها، فرآیند تنظیم رفتار الگوریتم‌ها به صورت تطبیقی و مبتنی بر یادگیری انجام می‌پذیرد. این رویکرد باعث می‌شود هر الگوریتم بتواند به طور پویا استراتژی جستجوی خود را تنظیم کرده و تعادل مؤثرتری میان کاوش نواحی جدید و بهره‌برداری از نواحی امیدبخش فضای جواب برقرار کند.

جدول ۱: توابع معیار

Func.	Benchmark Function	Formula	Dim	Range	Optimal
$F_۱$	Sphere	$f(x)=\sum_{i=۱}^{n} x_i^۲$	۳۰	$[-۱۰۰, ۱۰۰]$	۰
$F_۲$	Rosenbrock	$f(x)=\sum_{i=۱}^{n-۱}[۱۰۰(x_{i+۱}-x_i^۲)^۲ + (۱-x_i)^۲]$	۳۰	$[-۱۰۰, ۱۰۰]$	۰
$F_۳$	Rastrigin	$f(x)=\sum_{i=۱}^{n}[x_i^۲ - ۱۰\cos(۲\pi x_i) + ۱۰]$	۳۰	$[-۵/۱۲, ۵/۱۲]$	۰
$F_۴$	Ackley	$f(x)=-۲۰\exp\left(-۰/۲\sqrt{\frac{۱}{n}\sum_{i=۱}^{n} x_i^۲}\right) - \exp\left(\frac{۱}{n}\sum_{i=۱}^{n}\cos(۲\pi x_i)\right) + ۲۰ + \exp(۱)$	۳۰	$[-۳۲, ۳۲]$	۰

جدول ۲: مقایسه عملکرد الگوریتم‌ها روی توابع معیار

Func.	Metric	PSO	SSA	MPA	WOA	Marine Ecosystem
$F_۱$	Mean	۱.۱۷e-۲۳	۵.۴۰e-۰۹	۳.۶۲e-۲۱	۱.۱۸e-۲۷	۱.۰۴e-۵۲
$F_۲$	Mean	۳۱.۴۲	۶۸.۲۵	۴۸.۴۲	۲۷.۱۸	۲۵.۰۹
$F_۳$	Mean	۲۸.۲۶	۱۸.۹۰	۰.۰	۲۹.۸۱	۰.۰
$F_۴$	Mean	۷.۴۱e-۰۶	۶.۱۴e-۰۶	۱.۹۲e-۱۱	۷.۹۹e-۱۴	۲.۶۶e-۱۵

به منظور ارزیابی الگوریتم اکوسیستم دریایی، این روش بر روی مجموعه‌ای از توابع معیار در جدول ۱ اعمال شده و نتایج حاصله در جدول ۲ آورده شده است. شایان ذکر است که ارزیابی اولیه الگوریتم بر روی این توابع تست استاندارد ریاضی، به عنوان یک گام صحت‌سنجی انجام گرفته است. هدف از این بخش، تحلیل رفتاری الگوریتم، اثبات توانایی ریاضی آن در حفظ تعادل میان کاوش و بهره‌برداری، و اطمینان از عدم گرفتار شدن در بهینه‌های محلی در مواجهه با فضاهای جستجوی متنوع است.

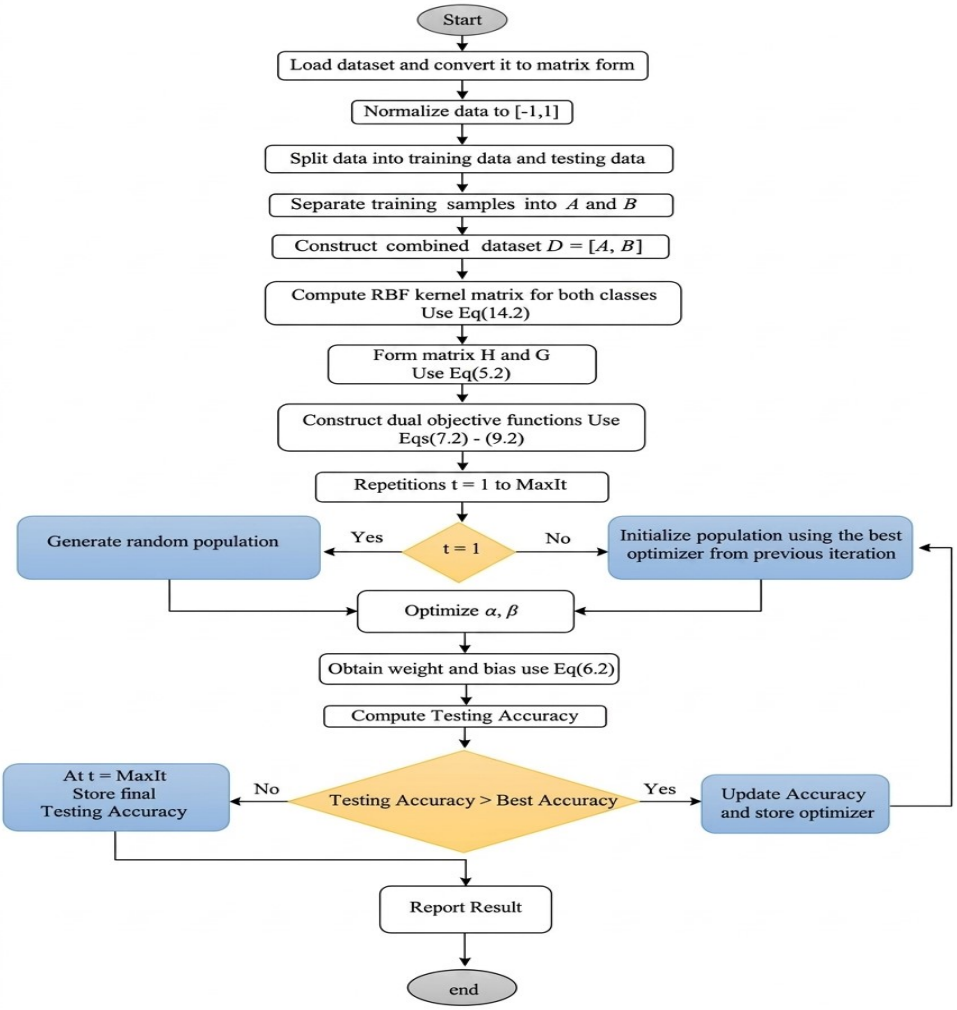
نمای کلی فرآیند الگوریتم پیشنهادی در شکل ۴ ارائه می‌شود. این الگوریتم ترکیبی، پارامترهای کلیدی (α, β) در ماشین بردار پشتیبان دوقلو را با استفاده از اکوسیستم دریایی به صورت همزمان بهینه می‌سازد تا بهترین دقت طبقه‌بند حاصل شود.

۴ ارزیابی تجربی

۱.۴ مجموعه داده

به‌منظور ارزیابی دقت و پایداری عملکرد الگوریتم پیشنهادی، از دو مجموعه‌داده‌ی شناخته‌شده در حوزه‌ی تشخیص سرطان پستان استفاده شده است. نخست، مجموعه‌داده‌ی سرطان پستان ویسکانسین متعلق به دکتر ویلیام اچ. وولبرگ از دانشکده پزشکی دانشگاه ویسکانسین مورد استفاده قرار گرفت(لینک مجموعه‌داده). این مجموعه‌داده شامل ۱۰ ویژگی برای هر نمونه است که عبارت‌اند از: شماره شناسه‌ی نمونه، ضخامت تومور، یکنواختی اندازه سلول، یکنواختی شکل سلول، چسبندگی لبه، اندازه‌ی سلول اپی‌تلیال منفرد، هسته‌ی برهنه، کروموزوم بورینگ (chromosome boring)، هسته‌ی طبیعی و تعداد میتوز. به‌جز شماره شناسه‌ی نمونه، مقادیر سایر ویژگی‌ها در بازه‌ی ۱ تا ۱۰ تعریف شده‌اند. در این مسئله، متغیر هدف یک متغیر دودویی تشخیصی است که نمونه‌ها را در دو گروه بدخیم و خوش‌خیم طبقه‌بندی می‌کند. برای حفظ اصالت داده‌ها، نسخه‌ی اصلی فایل مورد استفاده قرار گرفت و در مرحله‌ی پیش‌پردازش، ویژگی‌های زائد حذف شدند و ۱۶ نمونه‌ی ناسازگار یا ناقص کنار گذاشته شد؛ در نتیجه، تعداد نهایی داده‌ها به ۶۸۳ نمونه معتبر رسید.

علاوه بر این، از مجموعه‌داده‌ی دیگری از سرطان پستان که در مخزن داده‌های Kaggle در دسترس است نیز برای انجام آزمایش‌ها بهره گرفته شد(لینک مجموعه‌داده). این مجموعه‌داده شامل ۵۶۹ نمونه است که ویژگی‌های استخراج‌شده از تصاویر دیجیتال هسته‌ی سلول‌های بافت پستان را در بر می‌گیرد. در این داده‌ها برای هر نمونه ۳۲ ویژگی در نظر گرفته شده است که شامل یک شناسه، یک متغیر تشخیصی و مجموعه‌ای از ویژگی‌های عددی مرتبط با ساختار هندسی و بافتی هسته‌ی سلول‌ها می‌باشد. این ویژگی‌ها شامل مشخصه‌هایی مانند شعاع، بافت، محیط، مساحت، همواری، فشردگی، تقعر، نقاط تقعر، تقارن و بعد فراکتالی هستند که به‌صورت مقادیر میانگین، خطای استاندارد و بیشینه برای هر نمونه گزارش شده‌اند. متغیر هدف در این مجموعه‌داده نیز یک متغیر طبقه‌بندی دودویی



شکل ۴: فلوچارت الگوریتم ترکیبی ماشین بردار پشتیبان دوقلو و اکوسیستم دریایی.

است که نمونه‌ها را در دو گروه بدخیم و خوش‌خیم قرار می‌دهد و تمامی ۵۶۹ نمونه موجود در فایل داده دارای مقادیر معتبر بوده و مستقیماً در آزمایش‌ها مورد استفاده قرار گرفته‌اند.

با توجه به اینکه دامنه‌ی برخی ویژگی‌ها در این مجموعه‌داده‌ها نسبت به سایر ویژگی‌ها بزرگ‌تر است و این موضوع می‌تواند موجب غالب‌شدن مقادیر بزرگ‌تر در فرآیند یادگیری و کاهش حساسیت مدل نسبت به ویژگی‌های کوچک‌تر شود، پیش از مرحله‌ی آموزش، تمامی ویژگی‌های داده‌های آموزش و آزمون نرمال‌سازی شدند. در این مرحله، مقادیر ویژگی‌ها با استفاده از روش Min-Max به بازه‌ی $[-1, 1]$ منتقل گردیدند تا توزیع متعادلی برای فرآیند یادگیری ایجاد شده و از بروز مشکلات محاسباتی در مراحل آموزش مدل جلوگیری شود.

۲.۴ معیارهای ارزیابی

در این پژوهش برای ارزیابی عملکرد طبقه‌بند از سه معیار اصلی شامل دقت کلی (Accuracy)، صحت (Precision) و بازخوانی (Recall) استفاده شده است.

برای تبیین بهتر شاخص‌های ارزیابی به‌کار رفته در این مقاله، ماتریس درهم ریختگی مربوط به مسئله طبقه‌بندی دودویی در شکل ۵ معرفی می‌شود.

معیار دقت کلی یکی از رایج‌ترین شاخص‌ها در مسائل طبقه‌بندی است و بیان می‌کند چه نسبتی از کل نمونه‌ها توسط مدل به‌درستی شناسایی شده‌اند. این معیار از مجموع پیش‌بینی‌های صحیح در کلاس‌های مثبت و منفی به کل تعداد نمونه‌ها به دست می‌آید. مقدار دقت کلی به‌طور مستقیم نشان می‌دهد عملکرد کلی مدل تا چه اندازه رضایت‌بخش است.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

شکل ۵: ماتریس درهم‌ریختگی .

بنا به ماتریس درهم‌ریختگی، هرچه مجموع نمونه‌های صحیح پیش‌بینی‌شده $(TP + TN)$ بزرگ‌تر و تعداد خطاها $(FP + FN)$ کوچک‌تر باشد، مقدار دقت کلی افزایش می‌یابد که نشان‌دهنده‌ی عملکرد بهتر مدل در تشخیص صحیح نمونه‌ها است. این معیار مطابق رابطه‌ی زیر محاسبه می‌شود

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times ۱۰۰ \,\%\,. \tag{۱.۴}$$

معیار صحت نشان‌دهنده‌ی نسبت نمونه‌هایی است که طبقه‌بند به‌درستی به عنوان یک گروه خاص تشخیص داده و کل نمونه‌هایی که باید به عنوان آن گروه تشخیص داده شوند. همان‌طور که در رابطه زیر مشاهده می‌شود، مقدار صحت بیانگر صحت مدل در شناسایی یک کلاس مشخص است؛ به این معنا که اگر تعداد TP_i زیاد و تعداد FP_i کم باشد، مقدار صحت بزرگ‌تر خواهد شد و احتمال خطای طبقه‌بند در تشخیص آن کلاس کاهش می‌یابد

$$\text{Precision} = \frac{TP_i}{TP_i + FP_i} \times ۱۰۰ \,\%\,. \tag{۲.۴}$$

معیار بازخوانی بیانگر نسبت نمونه‌هایی است که توسط طبقه‌بند به‌درستی تشخیص داده شده‌اند به کل نمونه‌های موجود در کلاس مورد نظر. همان‌طور که در رابطه‌ی زیر دیده می‌شود، بازخوانی نشان‌دهنده‌ی کامل بودن نتایج مدل در شناسایی نمونه‌های یک کلاس خاص است؛ به این معنا که هر چه TP_i بزرگ‌تر و FN_i کوچک‌تر باشد، مقدار بازخوانی بزرگ‌تر شده و نشان می‌دهد تعداد نمونه‌هایی که باید توسط مدل شناسایی شوند اما از قلم افتاده‌اند، کمتر است

$$\text{Recall} = \frac{TP_i}{TP_i + FN_i} \times ۱۰۰ \,\%\,. \tag{۳.۴}$$

معیار F-Score بیانگر میانگین همساز میان دو معیار صحت و بازخوانی است و به عنوان یک شاخص ترکیبی برای ارزیابی هم‌زمان دقت و کامل بودن مدل عمل می‌کند. همان‌طور که در رابطه زیر مشاهده می‌شود، این معیار تعادلی ساختاری میان نرخ مثبت‌های کاذب FP و منفی‌های کاذب FN برقرار می‌سازد؛ به این معنا که بالا بودن مقدار F-Score نشان‌دهنده عملکرد متوازن و پایدار طبقه‌بند است.

$$\text{F-Score} = ۲ \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times ۱۰۰ \,\%\,. \tag{۴.۴}$$

۵ نتایج

به منظور افزایش دقت طبقه‌بند ماشین بردار پشتیبان دوقلو، پارامترهای α و β باید بهینه‌سازی شوند. بدین منظور، از مجموعه‌ای از الگوریتم‌های فرااکتشافی استفاده شده است که در قالب یک «اکوسیستم دریایی» با یکدیگر تعامل و همکاری می‌کنند تا بهترین جواب (مقادیر بهینه برای این دو پارامتر) را بیابند.

برای اعتبارسنجی اثربخشی روش پیشنهادی، نسخه‌های مختلفی مورد مقایسه تجربی قرار گرفتند: نسخه پایه ماشین بردار پشتیبان دوقلو (Twin SVM) و نسخه‌های ترکیب‌شده با الگوریتم‌های ازدحام ذرات (PSO)، ازدحام

سالپ (SSA)، بهینه‌سازی نهنگ (WOA)، شکارچیان دریایی (MPA) و در نهایت الگوریتم ترکیبی اکوسیستم دریایی (روش پیشنهادی).

تمامی شبیه‌سازی‌ها و پیاده‌سازی‌های عددی این پژوهش در محیط نرم‌افزار MATLAB R2021b انجام شده است که تمام داده‌ها و کدهای مربوطه در آدرس github.com/aydarahimi8880/Marine-Ecosystem قابل دسترسی است. به منظور تضمین یکپارچگی و یکسانی شرایط ارزیابی، کلیه آزمایش‌ها بر روی سیستمی با پردازنده 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz و ۸ گیگابایت حافظه رم با فرکانس 3200 MHz در محیط سیستم‌عامل ویندوز اجرا گردیده‌اند. علاوه بر این، جهت حفظ عدالت در مقایسه تجربی میان الگوریتم‌های فرااکتشافی مختلف، تنظیمات پارامترهای اولیه درون هر الگوریتم دقیقا مطابق با مقادیر استاندارد و پیش‌فرض پیشنهاد شده در مقالات مرجع و کدهای اصلی آن‌ها در نظر گرفته شده است. معیار توقف بر اساس تعداد تکرار مشخصی تعیین شده است؛ بدین ترتیب که فرآیند جستجو پس از طی ۱۰ تکرار متوالی (Max Iteration = 10) متوقف شده و در نهایت، میانگین نتایج به دست آمده گزارش می‌گردند. جهت ارزیابی دقیق‌تر، داده‌ها به صورت تصادفی با نسبت‌های ۵۰%، ۶۰%، ۷۰% ، ۸۰% و ۹۰% به دو بخش آموزش و آزمون تقسیم شدند.

در این بخش، عملکرد مدل‌های ارائه‌شده با استفاده از معیارهای ،دقت کلی صحت و بازخوانی برای دو مجموعه‌داده «ویسکانسین» و «Kaggle» مورد ارزیابی قرار گرفت. نتایج حاصل از این ارزیابی که در جداول ۳ و ۴ ارائه شده‌اند، بیانگر توانمندی مدل‌های بهینه‌سازی شده در مقایسه با مدل پایه ماشین بردار پشتیبان دوقلو هستند.

نتایج جدول ۳ نشان می‌دهد که مدل‌های بهینه‌سازی شده در تمامی نسبت‌های داده‌های آموزشی، عملکرد بهتری را نسبت به مدل پایه ماشین بردار پشتیبان دوقلو به ثبت رسانده‌اند. در این میان، مدل « Marine Ecosystem » در اکثر سناریوها بالاترین مقادیر را در هر سه معیار کسب کرده و دقت پیش‌بینی بالاتری را نشان می‌دهد. به‌طور مشخص، در نسبت داده آموزشی ۹۰%، الگوریتم اکوسیستم دریایی با دستیابی به Accuracy برابر با ۰.۹۸۵۳ و Precision برابر با ۱.۰۰۰۰ ، بهترین عملکرد را در مجموعه داده ویسکانسین ارائه داده است که بیانگر توان بالای این مدل در تفکیک دقیق کلاس‌ها و کاهش نرخ خطای طبقه‌بندی است.

بررسی نتایج جدول ۴ نیز این یافته‌ها را تأیید می‌کند. مطابق این جدول، الگوریتم اکوسیستم دریایی در تمامی نسبت‌های داده‌های آموزشی، کمترین اختلاف را با دقت ایده‌آل داشته و در معیار Accuracy نسبت به سایر مدل‌ها پیشتاز است. به‌عنوان نمونه، در نسبت داده آموزشی ۹۰%، این مدل با مقدار Accuracy برابر با ۰.۹۹۲۸ عملکرد قابل‌توجهی از خود نشان داده است. همچنین، مقایسه بین نسبت‌های مختلف داده‌های آموزشی حاکی از آن است که با افزایش حجم داده‌های آموزشی (از ۵۰% به ۹۰%)، تمامی مدل‌ها روند صعودی در دقت را تجربه کرده‌اند که نشان‌دهنده پایداری و یادگیری مؤثر الگوریتم‌ها با افزایش اطلاعات ورودی است.

نتایج مربوط به معیارهای Precision و Recall که در جداول ۳ و ۴ ارائه شده‌اند نیز با تحلیل‌های پیشین هم‌راستا بوده و برتری مدل‌های بهینه‌سازی شده را بیش از پیش تأیید می‌کنند. بر اساس این جداول، الگوریتم اکوسیستم دریایی به‌طور مداوم مقادیر بالاتری را در این دو معیار ثبت کرده است که نشان‌دهنده توان بالای این مدل در حفظ تعادل میان Precision و Recall است. علاوه بر این، مشاهده می‌شود که مدل‌های مبتنی بر بهینه‌سازی در این دو مجموعه داده، استحکام ساختار پیشنهادی و تأثیر مثبت استفاده از الگوریتم‌های فراابتکاری در بهبود عملکرد نهایی مدل کلاسیک Twin SVM را به‌خوبی نشان داده‌اند. در نهایت، برتری معنادار الگوریتم اکوسیستم دریایی در تمامی معیارها، مؤید کارایی این الگوریتم در استخراج الگوهای مؤثر و افزایش دقت طبقه‌بندی در مسائل پزشکی است.

جدول ۳: نتایج ارزیابی شاخص‌های چندگانه مربوط به مجموعه داده ویسکانسین

Evaluating indicator	Proportion of training data	Twin SVM	PSO	SSA	MPA	WOA	Marine Ecosystem
Accuracy score	۵۰%	۰.۹۱۴۶	**۰.۹۷۱۵**	۰.۹۵۴۸	۰.۹۵۲۵	۰.۹۵۴۸	۰.۹۶۱۹
	۶۰%	۰.۹۲۱۵	۰.۹۴۰۷	**۰.۹۵۹۷**	۰.۹۵۳۸	۰.۹۵۸۲	**۰.۹۵۹۷**
	۷۰%	۰.۹۳۸۸	۰.۹۵۶۹	۰.۹۶۰۶	۰.۹۵۴۹	۰.۹۴۶۲	**۰.۹۷۱۰**
	۸۰%	۰.۹۴۰۲	۰.۹۷۴۱	۰.۹۶۳۲	۰.۹۴۸۵	۰.۹۵۵۹	**۰.۹۸۸۴**
	۹۰%	۰.۹۴۵۹	۰.۹۶۲۵	۰.۹۵۳۵	۰.۹۴۲۹	۰.۹۲۷۱	**۰.۹۸۵۳**
Precision score	۵۰%	۰.۸۹۴۴	۰.۹۶۴۲	**۰.۹۶۵۷**	۰.۹۴۳۶	۰.۹۵۴۰	۰.۹۵۱۲
	۶۰%	۰.۹۰۰۷	**۰.۹۸۰۲**	۰.۹۵۸۳	۰.۹۵۹۸	۰.۹۴۱۲	۰.۹۷۰۰
	۷۰%	۰.۹۲۶۴	۰.۹۴۴۹	۰.۹۳۸۹	۰.۹۵۱۸	۰.۹۴۷۸	**۱.۰۰۰۰**
	۸۰%	۰.۹۴۰۵	۰.۹۵۱۴	۰.۹۱۱۹	۰.۹۲۱۸	۰.۹۳۶۱	**۱.۰۰۰۰**
	۹۰%	۰.۹۲۵۴	۰.۹۵۵۶	۰.۹۴۱۷	۰.۸۹۶۱	۰.۸۷۴۵	**۱.۰۰۰۰**
Recall score	۵۰%	۰.۹۰۱۰	۰.۸۹۷۸	۰.۹۰۹۰	۰.۹۲۴۱	۰.۹۱۴۲	**۰.۹۴۰۰**
	۶۰%	۰.۹۰۹۶	۰.۹۱۹۹	۰.۹۲۷۶	۰.۹۰۶۳	۰.۹۴۰۵	**۰.۹۵۱۲**
	۷۰%	۰.۹۱۵۸	۰.۹۴۴۵	**۰.۹۷۱۱**	۰.۹۱۳۱	۰.۸۶۸۲	۰.۹۶۰۰
	۸۰%	۰.۹۲۵۴	**۰.۹۷۱۸**	۰.۹۶۹۸	۰.۹۲۹۰	۰.۹۴۹۴	۰.۸۵۰۰
	۹۰%	۰.۹۴۶۱	۰.۹۶۰۸	**۰.۹۸۳۳**	۰.۹۷۶۲	۰.۹۴۶۶	۰.۹۶۰۰
F-Score	۵۰%	۰.۸۹۷۷	۰.۹۲۹۸	۰.۹۳۶۵	۰.۹۳۳۷	۰.۹۳۳۷	**۰.۹۴۵۶**
	۶۰%	۰.۹۰۵۱	۰.۹۴۹۱	۰.۹۴۲۷	۰.۹۳۲۳	۰.۹۴۰۸	**۰.۹۶۰۵**
	۷۰%	۰.۹۲۱۱	۰.۹۴۴۷	۰.۹۵۴۷	۰.۹۳۲۰	۰.۹۰۶۳	**۰.۹۷۹۶**
	۸۰%	۰.۹۳۲۹	**۰.۹۶۱۵**	۰.۹۴۰۰	۰.۹۲۵۴	۰.۹۴۲۷	۰.۹۱۸۹
	۹۰%	۰.۹۳۵۶	۰.۹۵۸۲	۰.۹۶۲۱	۰.۹۳۴۴	۰.۹۰۹۱	**۰.۹۷۹۶**

جدول ۴: نتایج ارزیابی شاخص‌های چندگانه مربوط به مجموعه داده Kaggle

Evaluating indicator	Proportion of training data	Twin SVM	PSO	SSA	MPA	WOA	Marine Ecosystem
Accuracy score	۵۰%	۰.۹۲۸۶	**۰.۹۶۸۳**	۰.۹۴۶۰	۰.۹۴۹۶	۰.۹۵۱۴	۰.۹۵۱۰
	۶۰%	۰.۹۰۶۵	۰.۹۵۵۱	۰.۹۱۷۲	۰.۹۵۲۴	۰.۹۲۲۵	**۰.۹۷۳۶**
	۷۰%	۰.۹۰۹۱	۰.۹۵۵۳	۰.۹۱۷۶	۰.۹۵۲۹	۰.۹۲۴۷	**۰.۹۷۸۹**
	۸۰%	۰.۹۱۹۴	۰.۹۵۵۸	۰.۹۵۰۴	۰.۹۳۶۲	۰.۹۴۲۴	**۰.۹۸۸۲**
	۹۰%	۰.۹۲۵۶	۰.۹۶۷۹	۰.۹۷۵۰	۰.۹۲۸۶	۰.۹۲۱۴	**۰.۹۹۲۸**
Precision score	۵۰%	۰.۹۲۴۲	**۰.۹۷۰۲**	۰.۹۰۹۳	۰.۹۴۸۹	۰.۹۵۱۲	۰.۹۷۰۰
	۶۰%	۰.۹۰۴۲	**۰.۹۵۲۸**	۰.۹۰۴۶	۰.۹۳۹۰	۰.۹۰۴۸	۰.۹۲۴۸
	۷۰%	۰.۹۱۹۵	۰.۹۳۹۴	۰.۹۱۸۱	۰.۹۵۴۲	۰.۹۱۴۶	**۰.۹۸۰۰**
	۸۰%	۰.۹۲۶۴	۰.۹۱۴۱	۰.۹۱۴۰	۰.۹۲۶۴	۰.۹۱۹۱	**۰.۹۸۰۰**
	۹۰%	۰.۹۳۴۵	۰.۹۵۰۲	۰.۹۲۵۶	۰.۹۴۸۶	۰.۹۴۷۴	**۰.۹۸۱۰**
Recall score	۵۰%	۰.۹۰۵۹	۰.۹۱۸۹	**۰.۹۴۷۷**	۰.۹۰۲۴	۰.۹۲۱۰	۰.۹۱۰۰
	۶۰%	۰.۹۰۱۲	۰.۹۲۹۸	۰.۹۴۳۲	۰.۹۲۱۲	۰.۹۰۵۹	**۰.۹۵۵۲**
	۷۰%	۰.۹۰۱۸	۰.۹۲۸۱	۰.۸۷۹۶	۰.۹۲۶۲	۰.۹۴۵۸	**۰.۹۷۰۶**
	۸۰%	۰.۹۰۰۲	۰.۹۲۴۵	۰.۹۵۴۹	۰.۹۰۹۴	۰.۹۲۷۸	**۰.۹۸۰۰**
	۹۰%	۰.۹۱۱۱	۰.۹۸۸۹	۰.۹۵۲۲	۰.۹۰۸۹	۰.۹۵۱۴	**۱.۰۰۰۰**
F-Score	۵۰%	۰.۹۱۴۹	**۰.۹۴۳۹**	۰.۹۲۸۱	۰.۹۲۵۱	۰.۹۳۵۹	۰.۹۳۹۱
	۶۰%	۰.۹۰۲۷	۰.۹۴۱۲	۰.۹۲۳۵	۰.۹۳۰۰	۰.۹۰۵۳	**۰.۹۳۹۸**
	۷۰%	۰.۹۱۰۶	۰.۹۳۳۷	۰.۸۹۸۴	۰.۹۴۰۰	۰.۹۳۰۰	**۰.۹۷۵۳**
	۸۰%	۰.۹۱۳۱	۰.۹۱۹۳	۰.۹۳۴۰	۰.۹۱۷۸	۰.۹۲۳۴	**۰.۹۸۰۰**
	۹۰%	۰.۹۲۲۷	۰.۹۶۹۲	۰.۹۳۸۷	۰.۹۲۸۳	۰.۹۴۹۴	**۰.۹۹۰۴**

۶ نتیجه‌گیری

در این پژوهش یک چارچوب نوآورانه به نام الگوریتم اکوسیستم دریایی ارائه شد که در آن چهار الگوریتم قدرتمند فرااکتشافی، یعنی بهینه‌سازی ازدحام ذرات، الگوریتم ازدحام سالپ‌ها، الگوریتم شکارچیان دریایی و الگوریتم بهینه‌سازی نهنگ‌ها به‌صورت هم‌زمان و در قالب یک سازوکار هوشمند، به‌منظور بهینه‌سازی پارامترهای ماشین بردار پشتیبان دوقلو با یکدیگر ترکیب شدند. در روش پیشنهادی، در هر تکرار فرایند بهینه‌سازی، عملکرد الگوریتم‌ها به‌صورت پویا ارزیابی شده و مؤثرترین الگوریتم جستجو به‌عنوان «برنده» انتخاب می‌شود تا مسئولیت تولید جمعیت مرحله بعد را بر عهده بگیرد. این سازوکار انتخاب پویا به‌گونه‌ای طراحی شده است که درون هر الگوریتم، از یادگیری تقویتی چندعاملی برای تنظیم هوشمندانه تعادل بین کاوش و بهره‌برداری استفاده می‌شود؛ در نتیجه، فضای جستجو به شکل هدفمند و کارآمدتری پیمایش می‌گردد و خطر گیر افتادن در بهینه‌های محلی کاهش می‌یابد. فرآیند ارزیابی مدل پیشنهادی، نشان‌دهنده دو محدودیت ساختاری کلیدی است که نیازمند توجه در تحقیقات آتی هستند. نخست، افزایش ابعاد فضای حالت‌ـعمل در جدول پایه‌ای $Q-table$ که مستقیماً تحت تأثیر بزرگ شدن حجم جمعیت اولیه قرار دارد؛ این پدیده منجر به تشدید پیچیدگی زمانی و حافظه‌ای الگوریتم شده و در نتیجه، سرعت همگرایی بهینه را کاهش می‌دهد. دومین محدودیت، وابستگی شدید کارایی الگوریتم Q-Learning به حجم داده‌های ورودی در مدل ماشین بردار پشتیبان دوقلو است. از آنجا که حل مسائل الگوریتم‌های مرتبه دوم در ماشین بردار پشتیبان دوقلو با چالش داده‌های مقیاس‌بزرگ مواجه است، ترکیب آن با فرآیند یادگیری تقویتی می‌تواند به عنوان یک گلوگاه محاسباتی عمل کرده و مقیاس‌پذیری روش پیشنهادی را در مواجهه با پایگاه‌داده‌های وسیع با محدودیت روبه‌رو سازد. نتایج تجربی به‌دست‌آمده بر روی انواع مجموعه‌داده‌های سرطان پستان نشان داد که چارچوب پیشنهادی، هم از نظر دقت و هم از نظر پایداری، عملکردی برتر نسبت به روش‌های کلاسیک و الگوریتم‌های تک‌عاملی دارد. در مسئله طبقه‌بندی دودویی، مقادیر شاخص‌هایی نظیر دقت کلی، صحت و بازخوانی به سطوح بسیار بالا رسیده است که این امر نشان‌دهنده توانایی بالای مدل در تشخیص صحیح و تفکیک دقیق تومورهای خوش‌خیم و بدخیم است. این نتایج، علاوه بر تأیید کارایی اکوسیستم دریایی پیشنهادی در بهینه‌سازی ماشین بردار پشتیبان دوقلو، بیانگر آن است که ترکیب چند الگوریتم فرااکتشافی در قالب یک چارچوب انتخاب‌گر پویا، می‌تواند به‌طور محسوسی کیفیت تصمیم‌گیری طبقه‌بند را بهبود دهد.

از منظر بالینی، یافته‌های این پژوهش نشان داد که چارچوب ارائه‌شده قابلیت تبدیل شدن به یک سامانه قدرتمند تشخیص کمک‌پزشکی را داراست. این سامانه می‌تواند با کاهش خطای انسانی، افزایش سرعت تحلیل داده‌های پزشکی و فراهم کردن یک ابزار تحلیلی دقیق برای پزشکان، از فرآیند تصمیم‌گیری بالینی در تشخیص زودهنگام سرطان پستان پشتیبانی کند. به‌کارگیری چنین سامانه‌ای در محیط‌های واقعی درمانی، می‌تواند به بهبود کیفیت غربالگری، کاهش تشخیص‌های اشتباه و در نهایت ارتقای سطح سلامت بیماران منجر شود.

مراجع

[1] Siegel, R.L., Miller, K.D. and Jemal, A., 2016. Cancer statistics. CA: A Cancer Journal for Clinicians, 66(1), pp.7–30.

[2] Dou, Y., and Meng, W., 2021. An optimization algorithm for computer-aided diagnosis of breast cancer based on support vector machine. *Frontiers in Bioengineering and Biotechnology, 9*, pp.1-12.

[3] Rahimi, A. and SolaymaniFard, O., 2025. Breast cancer diagnosis via twin support vector machine optimized by metaheuristic algorithms. *56th Annual Iranian Mathematics Conference*, pp.1-4.

[4] Rahimi, A. and SolaymaniFard, O., 2025. A novel metaheuristic-optimized fuzzy twin SVM with dissimilarity measure for breast cancer diagnosis. *18th International Conference of Iranian Operations Research Society*, pp.1-5.

[5] Rahimi, A. and SolaymaniFard, O., 2026. BMI-driven multi-task TSVM with PSO/GWO-based optimization for breast cancer diagnosis. *8th National Seminar on Control and Optimization*, pp.1-7.

[6] Rahimi, A. and SolaymaniFard, O., 2026. A bio-inspired optimization framework for BMI-guided multi-task TSVM in breast cancer diagnosis. *5th National and 1st International Conference on Soft Computing of Engineering Sciences in Industry and Society*, pp.1-6.

[7] Akay, M. F., 2009. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications, 36(2)*, pp.3240-3247.

[8] Tomar, D. and Agarwal, S., 2015. Twin support vector machines: A review from 2007 to 2014. *Egyptian Informatics Journal, 16*, pp.55-69.

[9] Kennedy, J. and Eberhart, R., 1995. Particle swarm optimization. *Proceedings of the IEEE International Conference on Neural Networks, 4*, pp.1942-1948.

[10] Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H. and Mirjalili, S.M., 2017. Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software, 114*, pp.163-191.

[11] Faramarzi, A., Heidarinejad, M., Mirjalili, S. and Gandomi, A.H., 2020. Marine predators algorithm: A nature-inspired metaheuristic. *Expert Systems with Applications, 152*, pp.113377.

[12] Mirjalili, S. and Lewis, A., 2016. The whale optimization algorithm. *Advances in Engineering Software, 95*, pp.51-67.

[13] Qiu, J., Xie, J., Zhang, D., Zhang, R. and Lin, M., 2024. A fuzzy twin support vector machine based on dissimilarity measure and its biomedical applications. *International Journal of Fuzzy Systems, 26*(8), pp.2750–2766.

[14] Bartlett, P., 2008. Reproducing Kernel Hilbert Spaces . Lecture notes of CS281B/Stat241B Statistical Learning Theory, University of California at Berkeley.

[15] Zhao, Y., Kosorok, M.R. and Zeng, D., 2009. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine, 28*(26), pp.3294-3315.

[16] Sutton, RS., Barto, AG., 1998. Reinforcement Learning: An Introduction. MIT Press; Cambridge, MA.