

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/>).

¹Corresponding author

E-mail addresses: (Ayda Rahimi) ayda.rahimi@mail.um.ac.ir, (Omid SolaymaniFard) soleimani@um.ac.ir

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)

^(۱) گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران

^(۲) گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران

تاریخ انتشار: ۱۴۰۵/۰۷/۰۱

تاریخ پذیرش: ۱۴۰۵/۰۶/۰۳

تاریخ دریافت: ۱۴۰۵/۰۱/۲۵

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 فرااکتشافی الهام‌گرفته از اکوسیستم دریایی — SSA، WOA، PSO و MPA — را به عنوان 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 به صورت 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 قابل توجهی 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)

^۲ نویسنده مسئول مقاله

و بدخیم (Malignant) تقسیم می شود، نقشی حیاتی در افزایش نرخ بقا و بهبود کیفیت زندگی بیماران ایفا می کند [۱]. در دهه های اخیر، با پیشرفت فناوری های تصویر برداری پزشکی و ظهور کلان داده ها، استفاده از روش های هوش مصنوعی و یادگیری ماشین به ابزاری قدرتمند برای تحلیل داده های پزشکی و پشتیبانی از تصمیم گیری بالینی مبدل شده است [۲]. این الگوریتم ها قادرند الگوهای پیچیده و نامشهود در داده ها را شناسایی کرده و مدل های پیش بینی کننده ای با دقت بالا ارائه دهند [۳-۶].

در میان انبوه روش های طبقه بندی، 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 محدودیت هایی همچون حساسیت بالا به پارامترها و عدم توانایی در مدیریت بهینه ی داده های نامتوازن با چالش هایی نظیر پیچیدگی محاسباتی قابل توجه در مجموعه داده های بزرگ که ناشی از حل یک مسئله برنامه ریزی درجه دوم (QPP) است، نیز روبروست. برای رفع این چالش ها ماشین بردار پشتیبان دوقلو (Twin SVM) معرفی شد که با حل دو مسئله ی بهینه سازی کوچک تر به جای یک مسئله بزرگ، سرعت محاسبات را به طور چشم گیری افزایش می دهد و دقت طبقه بندی را بهبود می بخشد [۸].

با وجود برتری محاسباتی، کارایی ماشین بردار پشتیبان دوقلو نیز همانند نسخه کلاسیک آن، به شدت به تنظیم بهینه های پارامترهای 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 روش های سنتی مانند جستجوی شبکه ای (Grid Search) ناکارآمد و زمان بر است. اینجاست که الگوریتم های فرااکتشافی (Metaheuristic Algorithms) به عنوان ابزارهایی قدرتمند برای جستجو در فضاهای پیچیده و یافتن تنظیمات بهینه برای پارامترها، راهگشا بوده اند.

الگوریتم های 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 بهینه سازی نهنگ ها (WOA) [۱۲] در سال های اخیر کاربردهای موفقی در حل مسائل بهینه سازی پیچیده داشته اند.

با این حال، قضیه No Free Lunch Theorem بیان می دارد که هیچ الگوریتم فرااکتشافی واحدی وجود ندارد که برای تمام مسائل بهینه سازی بهترین عملکرد را داشته باشد. هر الگوریتم نقاط قوت و ضعف خود را در توازن میان دو فاز اصلی جستجو، یعنی کاوش (Exploration) و بهره برداری (Exploitation) داراست. به بیان دیگر هر یک از این الگوریتم ها دارای نقاط قوت و ضعف خاصی هستند؛ برای مثال، PSO در کاوش سریع فضای جستجو توانمند است اما ممکن است در بهره برداری محلی ضعیف عمل کند، در حالی که WOA در بهره برداری دقیق تر 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 می شود که از قابلیت چهار الگوریتم الهام گرفته از طبیعت یعنی PSO، SSA، MPA و WOA به صورت همزمان بهره می برد. برای مدیریت هوشمندانه این اکوسیستم و ایجاد توازن پویا میان کاوش و بهره برداری، از یادگیری تقویتی چندعاملی (Multi Agent Reinforcement Learning) مبتنی بر الگوریتم Q-Learning استفاده می شود. این عامل هوشمند در هر مرحله از فرآیند بهینه سازی، بر اساس وضعیت فعلی جستجو، تصمیم می گیرد 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 بهینه های محلی جلوگیری می نماید. نوآوری اصلی این کار در سه جنبه خلاصه می شود: (۱) طراحی یک اکوسیستم فرااکتشافی متشکل از الگوریتم های 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 بهره گیری از روش ماشین بردار پشتیبان دوقلو و استفاده از مفهوم یادگیری تقویتی چندعاملی (MARL) برای ایجاد یک اکوسیستم بر پایه الگوریتم های فرااکتشافی منتخب پایه ریزی شده است. برای این منظور، این بخش اختصاص می یابد 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 i ام 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

$$f_1(x) = K(x, D^T)w_1 + b_1 = 0 \quad (1.2)$$

و

$$f_2(x) = K(x, D^T)w_2 + b_2 = 0 \quad (2.2)$$

بیان می‌شوند که در آن $K(\cdot, \cdot)$ تابع هسته و w_i, b_i پارامترهای ابرصفحه i ام 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

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} \|K(A, D^T)w_1 + e_1 b_1\|^2 + C_1 e_1^T \xi \\ \text{s.t.} \quad & -(K(B, D^T)w_1 + e_2 b_1) + \xi \geq e_2, \quad \xi \geq 0 \end{aligned} \quad (3.2)$$

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

$$\begin{aligned} \min_{w_2, b_2, \eta} \quad & \frac{1}{2} \|K(B, D^T)w_2 + e_2 b_2\|^2 + C_2 e_2^T \eta \\ \text{s.t.} \quad & (K(A, D^T)w_2 + e_1 b_2) + \eta \geq e_1, \quad \eta \geq 0 \end{aligned} \quad (4.2)$$

حاصل می شوند.

فرم دوآل

برای استخراج فرم دوآل، ماتریس های

$$H = [K(A, D^T) \ e_1], \quad G = [K(B, D^T) \ e_2] \quad (5.2)$$

و بردارهای پارامتر

$$u_1 = \begin{bmatrix} w_1 \\ b_1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} w_2 \\ b_2 \end{bmatrix} \quad (6.2)$$

در نظر گرفته می شوند.

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

$$\begin{aligned} \max_{\alpha} \quad & e_1^T \alpha - \frac{1}{2} \alpha^T G (H^T H + \delta I)^{-1} G^T \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq C_1 \end{aligned} \quad (7.2)$$

به دست می آید و پارامترهای ابرصفحه اول از رابطه

$$u_1 = -(H^T H + \delta I)^{-1} G^T \alpha \quad (8.2)$$

محاسبه می شوند.

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

$$\begin{aligned} \max_{\beta} \quad & e_2^T \beta - \frac{1}{2} \beta^T H (G^T G + \delta I)^{-1} H^T \beta \\ \text{s.t.} \quad & 0 \leq \beta \leq C_2 \end{aligned} \quad (9.2)$$

و پارامترهای ابرصفحه دوم از رابطه

$$u_2 = (G^T G + \delta I)^{-1} H^T \beta \quad (10.2)$$

به دست می آیند.

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

$$\text{Class}(i) = \arg \min_{i=1,2} |K(x^T, D^T) u_i + b_i|, \quad \text{for } i = 1, 2 \quad (11.2)$$

تعریف می شود (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 به صورت خطی تفکیک پذیر 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 عبارت‌اند از:

(۱) هسته خطی

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

$$K(x_i, x_j) = x_i \cdot x_j \quad (۱۲.۲)$$

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

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (۱۳.۲)$$

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

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \quad (۱۴.۲)$$

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

$$K(x, x_i) = \tanh(v(x \cdot x_i) + c) \quad (۱۵.۲)$$

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 شاخه های 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 می گیرد و باید از طریق آزمون و خطا، بهترین تصمیم ها را بگیرد [۱۵، ۱۶]. در اینجا عامل با انجام هر اقدام، یک بازخورد از محیط دریافت می کند. اگر اقدامش درست باشد، پاداش می گیرد و اگر اشتباه باشد، جریمه می شود (شکل ۱). نکته کلیدی در این روش، آینده نگری است؛ یعنی سیستم فقط به دنبال سود سریع و لحظه ای نیست بلکه یاد می گیرد تصمیماتی بگیرد که شاید در ابتدا پاداش کمی داشته باشند اما در نهایت منجر به بهترین نتیجه ممکن می شوند. با این حال چارچوب فوق بیانگر حالت پایه و تک عاملی یادگیری تقویتی است در حالی که در بسیاری از مسائل پیچیده - از جمله مسئله مورد بررسی در این پژوهش - تصمیم گیری به صورت هم زمان توسط چندین عامل انجام می شود که رفتار هر یک می تواند بر عملکرد سایرین و بر وضعیت کلی محیط اثرگذار باشد؛ از این رو، در این مطالعه از رویکرد یادگیری تقویتی چندعاملی استفاده می شود که در آن مجموعه ای از عامل ها در یک محیط مشترک تعامل کرده و هر عامل بر اساس سیاست تصمیم گیری مستقل خود عمل می کند در حالی که نتیجه نهایی حاصل برهم کنش جمعی تصمیم های تمامی عامل ها خواهد بود (شکل ۲).

۱.۲.۲ اجزای اصلی مسئله یادگیری تقویتی

در مدل سازی یادگیری تقویتی، معمولاً اجزای زیر تعریف می شوند [۱۶]:

(۱) محیط (Environment)

محیط سامانه ای است که عامل یا مجموعه ای از عامل ها در آن قرار دارند و با آن تعامل می کنند. این سامانه در واکنش به تصمیم های اتخاذ شده، پیامد متناظر با آن تصمیم را به صورت پاداش یا جریمه بازمی گرداند. در واقع، محیط چارچوب پویای مسئله را تعیین می کند و مشخص می سازد که هر تصمیم در یک وضعیت مشخص، به چه نتیجه ای منجر خواهد شد.

(۲) عمل (Action) عمل تصمیمی است که عامل در هر لحظه بر اساس وضعیت جاری اتخاذ می کند. مجموعه تمام عمل ها را با نماد A و عمل انتخاب شده در زمان t را با A_t نمایش می دهیم. فضای عمل ها بسته به ماهیت مسئله می تواند گسسته یا پیوسته تعریف شوند.

(۳) حالت (State)

حالت مجموعه ای از اطلاعات است که وضعیت فعلی محیط را توصیف می کند و اطلاعاتی را در اختیار عامل قرار می دهد تا بر اساس آن تصمیم گیری کند. متغیر S مجموعه حالات ممکن را نشان می دهد و در زمان t ، وضعیت سیستم با S_t نمایش داده می شود. حالت می تواند شامل هرگونه اطلاعات مرتبط با وضعیت جاری سیستم باشد و بسته به نوع مسئله، به صورت گسسته یا پیوسته تعریف شود. تغییر وضعیت سیستم به حالت فعلی و عمل انتخاب شده وابسته است و احتمال انتقال به حالت بعدی با توزیع شرطی

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

توصیف می شود.

(۴) پاداش (Reward) پاداش (R_t) ، بازخوردی عددی و دارای مقدار مثبت یا منفی است که در زمان t و در پاسخ به عمل اتخاذ شده، از طرف محیط، تولید و به عامل ارائه می شود. این کمیت میزان مطلوبیت پیامد حاصل از تصمیم اتخاذ شده را بیان می کند. به صورت کلی، پاداش می تواند تابعی از حالت فعلی، عمل انجام شده و حالت بعدی باشد و به شکل

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

تعریف می شود.

(۵) سیاست (Policy)

سیاست قاعده‌ای است که تعیین می‌کند عامل در هر وضعیت چه عملی را انتخاب کند. اگر سیاست به صورت 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

$$\pi(a | s) = P(A_t = a | S_t = s), \quad (۱۸.۲)$$

و در حالت قطعی، سیاست به صورت تابعی از وضعیت

$$A_t = \pi(S_t). \quad (۱۹.۲)$$

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. به عبارت دیگر، تابع ارزش معیار سنجش عملکرد یک سیاست در افق زمانی آینده است. نسخه بهینه این تابع از طریق بیشینه‌سازی روی تمامی سیاست‌های ممکن به دست می‌آید. تابع ارزش بهینه به شکل

$$V_t^*(s_t) = \max_{\pi \in \Pi} V_t^\pi(s_t) = \max_{\pi \in \Pi} \mathbb{E}_\pi \left[\sum_{k=0}^{T-t} \gamma^k R_{t+k} \mid S_t = s_t \right] \quad (۲۰.۲)$$

تعریف می‌شود که در آن، $\gamma \in [0, 1]$ 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 دورمدت در آینده بهای بیشتری می‌دهد (دید بلندمدت).

۲.۲.۲ الگوریتم Q-Learning

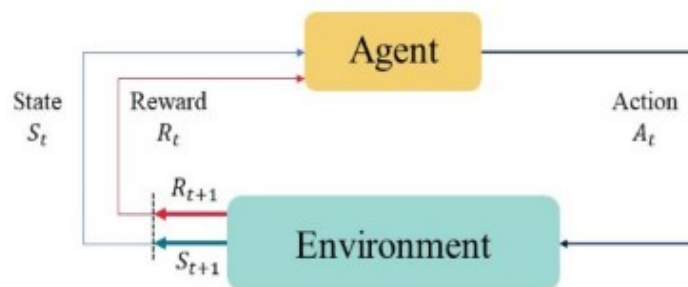
هرچند تابع ارزش رکن اصلی یادگیری تقویتی است، اما حتی با وجود در اختیار داشتن مدلی دقیق از محیط، یافتن بهترین استراتژی از طریق محاسبات مستقیم معمولاً بسیار پیچیده است. به همین دلیل، از روش یادگیری تفاضل زمانی (TD) استفاده می‌شود. در این روش، عامل منتظر نمی‌ماند تا تمام مراحل به پایان برسد، بلکه در هر 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

$$Q_{t+1}(S_t, A_t) = Q_t(S_t, A_t) + \alpha [R_t + \gamma \max_a Q_t(S_{t+1}, a) - Q_t(S_t, A_t)] \quad (۲۱.۲)$$

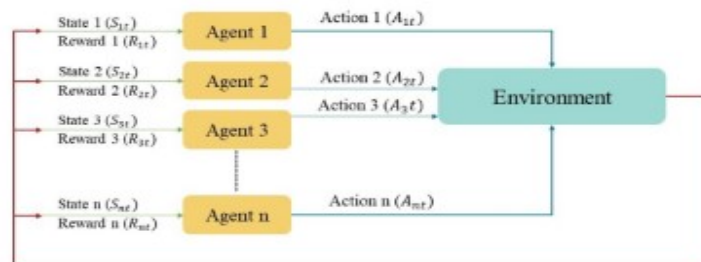
در واقع، با این روش کافی است وضعیت فعلی را بدانیم تا بتوانیم سیاست بهینه را به کمک رابطه‌ی

$$\pi^*(s) = \arg \max_a Q(s, a) \quad (۲۲.۲)$$

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 به‌روزرسانی سرعت ذره v در گام زمانی $t + 1$ به صورت

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (pbest_i - x_i^t) + c_2 \cdot r_2 \cdot (gbest - x_i^t) \quad (23.2)$$

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 i -ام 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_2 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 کشف شده 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_1 و r_2 با توزیع یکنواخت در بازه $[0, 1]$ استفاده می‌شود. سپس موقعیت جدید هر ذره x_i با استفاده از سرعت به‌روز رسانی شده به صورت

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (24.2)$$

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_1 و اعداد تصادفی c_2 و c_3 در بازه $[0, 1]$ به شکل

$$x_j^1 = \begin{cases} F_j + c_1((ub_j - lb_j)c_2 + lb_j) & c_3 \geq 0.5 \\ F_j - c_1((ub_j - lb_j)c_2 + lb_j) & c_3 < 0.5 \end{cases} \quad (25.2)$$

به‌روز رسانی می‌شود.

c_1 به عنوان مهم‌ترین پارامتر در این الگوریتم عمل کرده زیرا تعادل بین فاز اکتشاف و استخراج را حفظ می‌کند و به صورت

$$c_1 = 2e^{-\left(\frac{t}{L}\right)^2} \quad (26.2)$$

تعریف می‌شود که در آن l تکرار فعلی و L حداکثر تعداد تکرارها است. همچنین معادله بروزرسانی موقعیت سالپ پیرو i -ام در بعد j -ام

$$x_j^i = \frac{1}{\gamma}(x_j^i + x_j^{i-1}), \quad i \geq 2 \quad (27.2)$$

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 $\frac{1}{3}$ ابتدایی تکرارها انجام می شود. در این شرایط، بهترین استراتژی برای شکارچی این است که ثابت بماند و موقعیت طعمه با حرکت برآونی به صورت

$$\overrightarrow{stepsize}_i = \overrightarrow{R}_B \otimes (\overrightarrow{Elite}_i - \overrightarrow{R}_B \otimes \overrightarrow{Prey}_i), \quad i = 1, \dots, n \quad (28.2)$$

$$\overrightarrow{Prey}_i = \overrightarrow{Prey}_i + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \quad (29.2)$$

به روز رسانی شود.

فاز ۲: زمانی که شکارچی و طعمه با سرعت تقریباً یکسان حرکت می کنند
این فاز برای $\frac{1}{3}$ میانی تکرارها انجام می شود. در این بخش اکتشاف و استخراج هر دو مهم هستند بنابراین نیمی از جمعیت به کاوش و نیمی دیگر به بهره برداری می پردازد. بدین منظور، طعمه به صورت لوی و شکارچی به صورت برآونی حرکت می کند.
برای نیمه اولیه جمعیت، طعمه به صورت لوی

$$\overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{Elite}_i - \overrightarrow{R}_L \otimes \overrightarrow{Prey}_i), \quad i = 1, \dots, \frac{n}{2} \quad (30.2)$$

$$\overrightarrow{Prey}_i = \overrightarrow{Prey}_i + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \quad (31.2)$$

و برای نیمه ثانویه، طعمه به صورت برآونی

$$\overrightarrow{stepsize}_i = \overrightarrow{R}_B \otimes (\overrightarrow{R}_B \otimes \overrightarrow{Elite}_i - \overrightarrow{Prey}_i), \quad i = \frac{n}{2} + 1, \dots, n \quad (32.2)$$

$$\overrightarrow{Prey}_i = \overrightarrow{Elite}_i + P \cdot CF \otimes \overrightarrow{stepsize}_i \quad (33.2)$$

حرکت می کند که در آن ضریب همگرایی CF به شکل

$$CF = \left(1 - \frac{\text{Iter}}{\text{Max_Iter}}\right)^{\frac{\text{Iter}}{\text{Max_Iter}}} \quad (34.2)$$

تعریف می شود.

فاز ۳: زمانی که شکارچی سریع تر از طعمه حرکت می کند
این فاز برای $\frac{1}{3}$ انتهایی تکرارها انجام می شود. در این فاز بهترین استراتژی برای شکارچی این است که به صورت لوی حرکت کند. بروزرسانی موقعیت طعمه به صورت

$$\overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{R}_L \otimes \overrightarrow{Elite}_i - \overrightarrow{Prey}_i), \quad i = 1, \dots, n \quad (35.2)$$

$$\overrightarrow{Prey}_i = \overrightarrow{Elite}_i + P \cdot CF \otimes \overrightarrow{stepsize}_i \quad (36.2)$$

خواهد بود.

WOA

الگوریتم بهینه سازی 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 نهنگ ها با ایجاد حباب های ماریچی، طعمه های خود را محاصره و به سمت خود می کشند.

در طبیعت نهنگ‌های گوزپشت 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

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

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (۳۸.۲)$$

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

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (۳۹.۲)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (۴۰.۲)$$

هستند به طوری که $\vec{a}$ به طور خطی از ۲ تا ۰ کاهش می‌یابد و $\vec{r}$ نیز یک بردار تصادفی در بازه $[0, 1]$ 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

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (۴۱.۲)$$

استفاده می‌کند.

در مرحله (بهربرداری)، 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 $\vec{A}$ 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

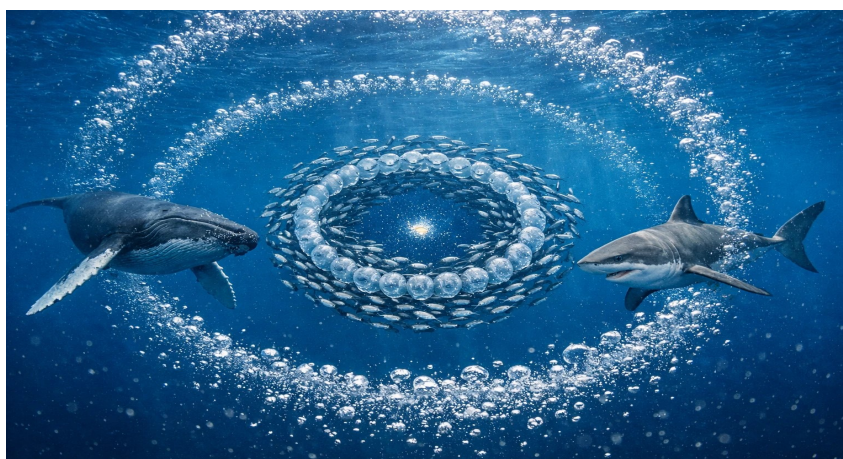
$$\vec{D} = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}| \quad (۴۲.۲)$$

$$\vec{X}(t+1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D} \quad (۴۳.۲)$$

محاسبه شود.

۳ اکوسیستم دریایی

در اکوسیستم دریایی (Marine Ecosystem) مبتنی بر یادگیری تقویتی چندعاملی، فرآیند بهینه سازی به صورت تکراری انجام می شود و در هر تکرار چند الگوریتم فرااکتشافی به طور هم زمان در فضای جستجو فعالیت می کنند. در این چارچوب، چهار الگوریتم الهام گرفته از موجودات دریایی یعنی (PSO)، (SSA)، (MPA) و (WOA) مورد استفاده قرار می گیرند. این الگوریتم ها به دلیل تفاوت در ساز و کار جستجو و نحوه هدایت عامل ها به سمت نواحی امیدبخش فضای جواب انتخاب شده اند. هر یک از این الگوریتم ها فرآیند جستجوی فضای جواب را بر اساس موقعیت های کلیدی متفاوتی هدایت می کنند. در الگوریتم PSO فرآیند جستجو بر اساس بهترین تجربه فردی ذرات و بهترین تجربه سراسری جمعیت انجام می شود؛ به این صورت که هر ذره موقعیت خود را با توجه به این دو مرجع به روزرسانی می کند. در الگوریتم MPA موقعیت شکارچیان و طعمه ها در تعامل با یکدیگر تعیین کننده مسیر جستجو است و عامل ها با استفاده از الگوهای حرکتی برآونی و لوی در مراحل مختلف به سمت بهترین جواب ها هدایت می شوند. در الگوریتم SSA نیز فرآیند جستجو با استفاده از موقعیت رهبر (Leader) که نشان دهنده بهترین جواب فعلی است هدایت می شود و سایر اعضای جمعیت با پیروی از آن موقعیت های خود را تنظیم می کنند. در الگوریتم WOA نیز جستجو عمدتاً بر اساس بهترین جواب سراسری انجام می شود و عامل ها با شبیه سازی رفتار محاصره و حرکت مارپیچی نهنگ ها به سمت این موقعیت حرکت می کنند، در حالی که در برخی مراحل نیز حرکت های تصادفی برای افزایش تنوع جستجو به کار گرفته می شود (شکل ۳).



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 می شود. فرآیند تولید این جمعیت پایه به صورت تصادفی و بر اساس توزیع یکنواخت در محدوده کران های بالا و پایین مجاز فضا انجام می گیرد. بر این اساس، موقعیت ابتدایی ذره i -ام در بعد j -ام $(x_{i,j}^0)$ برای تمامی الگوریتم ها به صورت یکسان و مطابق با رابطه

$$x_{i,j}^0 = LB_j + \text{rand}(0, 1) \times (UB_j - LB_j), \quad \forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, D\} \quad (1.3)$$

تعیین می گردد که در آن، UB_j و LB_j به ترتیب بیانگر حد پایین و حد بالای تعریف شده برای بعد j -ام فضای جستجوی مسئله هستند و $\text{rand}(0, 1)$ یک عدد تصادفی با توزیع یکنواخت در بازه $[0, 1]$ است. این مکانیزم تضمین می کند که در تکرار اول، تمام الگوریتم ها 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 می شوند و هر یک تلاش می کند بر اساس مکانیزم اختصاصی خود، بهترین جواب ممکن را بیابد. پس از پایان اجرای الگوریتم ها در تکرار جاری، عملکرد آن ها بر اساس میزان شاخص برازش با یکدیگر مقایسه شده و الگوریتمی که بهینه ترین مقدار تابع هدف را تولید کرده باشد، به عنوان الگوریتم برتر آن تکرار انتخاب می شود. سپس، کل ماتریس جمعیتی که

توسط این الگوریتم برتر ارتقا یافته و به روز رسانی شده است، به عنوان جمعیت مشترک برای تکرار بعدی در نظر گرفته می‌شود. به عبارت دیگر، در تکرار بعدی، هر چهار الگوریتم مجدداً 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 (مثلاً 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 برآزش فعلی و برآزش 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 i -ام در تکرارهای گذشته از طریق اعمال استراتژی m -ام در حالت s_i به دست آورده است. مکانیزم تصمیم‌گیری و منطق بهینه‌سازی ارزش در این چارچوب به صورت زیر تبیین می‌گردد:

۱. **بیشینه بودن $Q(s_i, a_{i,1})$ (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 $Q(s_i, a_{i,2})$ (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 $Q(s_i, a_{i,3})$ (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 به دست آمده، 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_1	Sphere	$f(x) = \sum_{i=1}^n x_i^2$	۳۰	$[-100, 100]$	۰
F_2	Rosenbrock	$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i)^2 + (1 - x_i)^2]$	۳۰	$[-100, 100]$	۰
F_3	Rastrigin	$f(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	۳۰	$[-5.12, 5.12]$	۰
F_4	Ackley	$f(x) = -20 \exp\left(-0.2\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + \exp(1)$	۳۰	$[-32, 32]$	۰

جدول ۲: مقایسه عملکرد الگوریتم ها روی توابع معیار

Func.	Metric	PSO	SSA	MPA	WOA	Marine Ecosystem
F_1	Mean	۱.۱۷e-۲۳	۵.۴۰e-۰۹	۳.۶۲e-۲۱	۱.۱۸e-۲۷	۱.۰۴e-۵۲
F_2	Mean	۳۱.۴۲	۶۸.۲۵	۴۸.۴۲	۲۷.۱۸	۲۵.۰۹
F_3	Mean	۲۸.۲۶	۱۸.۹۰	۰.۰	۲۹.۸۱	۰.۰
F_4	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 مجموعه ای 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 صحت سنجی انجام گرفته است. هدف از این بخش، تحلیل رفتاری الگوریتم، اثبات توانایی ریاضی آن در حفظ تعادل میان کاوش و بهره برداری، و اطمینان از عدم گرفتار شدن در بهینه های محلی در مواجهه با فضاهاى جستجوی متنوع است.

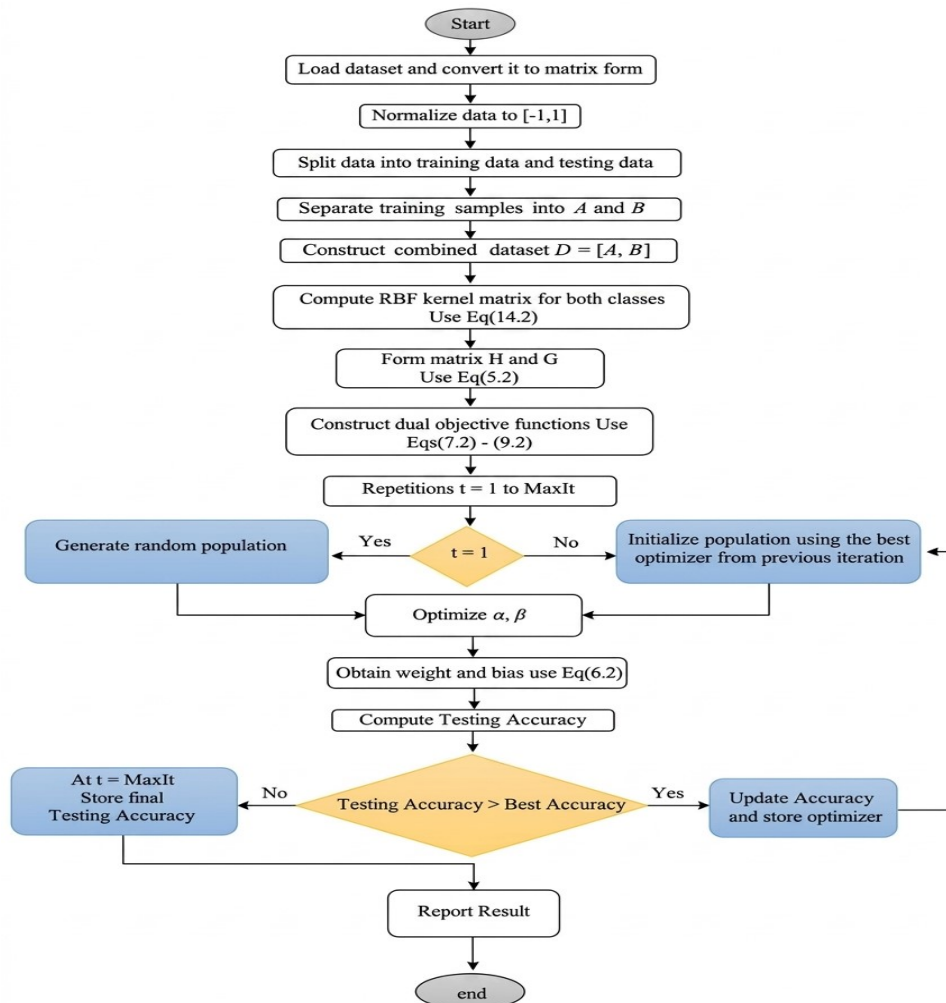
نمای کلی فرآیند الگوریتم پیشنهادی در شکل ۴ ارائه می شود. این الگوریتم ترکیبی، پارامترهای کلیدی (α, β) در ماشین بردار پشتیبان دوقلو را با استفاده از اکوسیستم دریایی به صورت همزمان بهینه می سازد تا بهترین دقت طبقه بند حاصل شود.

۴ ارزیابی تجربی

۱.۴ مجموعه داده

به منظور ارزیابی دقت و پایداری عملکرد الگوریتم پیشنهادی، از دو مجموعه داده ی شناخته شده در حوزه ی تشخیص سرطان پستان استفاده شده است. نخست، مجموعه داده ی سرطان پستان ویسکانسین متعلق به دکتر ویلیام اچ. وولبرگ از دانشکده پزشکی دانشگاه ویسکانسین مورد استفاده قرار گرفت (لینک مجموعه داده). این مجموعه داده شامل ۱۰ ویژگی برای هر نمونه است که عبارت اند از: شماره شناسه ی نمونه، ضخامت تومور، یکنواختی اندازه سلول، یکنواختی شکل سلول، چسبندگی لبه، اندازه ی سلول اپی تلیال منفرد، هسته ی برهنه، کروموزوم بورینگ (chromosome boring)، هسته ی طبیعی و تعداد میتوز. به جز شماره شناسه ی نمونه، مقادیر سایر ویژگی ها در بازه ی ۱ تا ۱۰ تعریف شده اند. در این مسئله، متغیر هدف یک متغیر دودویی تشخیصی است که نمونه ها را در دو گروه بدخیم و خوش خیم طبقه بندی می کند. برای حفظ اصالت داده ها، نسخه ی اصلی فایل مورد استفاده قرار گرفت و در مرحله ی پیش پردازش، ویژگی های زائد حذف شدند و ۱۶ نمونه ی ناسازگار یا ناقص کنار گذاشته شد؛ در نتیجه، تعداد نهایی داده ها به ۶۸۳ نمونه معتبر رسید.

علاوه بر این، از مجموعه داده ی دیگری از سرطان پستان که در مخزن داده های Kaggle در دسترس است نیز برای انجام آزمایش ها بهره گرفته شد (لینک مجموعه داده). این مجموعه داده شامل ۵۶۹ نمونه است که ویژگی های استخراج شده از تصاویر دیجیتال هسته ی سلول های بافت پستان را در بر می گیرد. در این داده ها برای هر نمونه ۳۲ ویژگی در نظر گرفته شده است که شامل یک شناسه، یک متغیر تشخیصی و مجموعه ای از ویژگی های عددی مرتبط با ساختار هندسی و بافتی هسته ی سلول ها می باشد. این ویژگی ها شامل مشخصه هایی مانند شعاع، بافت، محیط، مساحت، همواری، فشردگی، تقعر، نقاط تقعر، تقارن و بعد فراکتالی هستند که به صورت مقادیر میانگین، خطای استاندارد و بیشینه برای هر نمونه گزارش شده اند. متغیر هدف در این مجموعه داده نیز یک متغیر طبقه بندی 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 مجموعه داده‌ها نسبت به سایر ویژگی‌ها بزرگ‌تر است و این موضوع می‌تواند موجب غالب شدن 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 به کار 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 به درستی 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 درهم ریختگی.

بنا به ماتریس درهم ریختگی، هرچه مجموع نمونه های صحیح پیش بینی شده $(TP + TN)$ بزرگ تر و تعداد خطاها $(FP + FN)$ کوچک تر باشد، مقدار دقت کلی افزایش می یابد که نشان دهنده عملکرد بهتر مدل در تشخیص صحیح نمونه ها است. این معیار مطابق رابطه ی زیر محاسبه می شود

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% . \quad (1.4)$$

معیار صحت نشان دهنده ی نسبت نمونه هایی است که طبقه بند به درستی به عنوان یک گروه خاص تشخیص داده و کل نمونه هایی که باید به عنوان آن گروه تشخیص داده شوند. همان طور که در رابطه زیر مشاهده می شود، 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 بزرگ تر خواهد شد و احتمال خطای طبقه بند در تشخیص آن کلاس کاهش می یابد

$$\text{Precision} = \frac{TP_i}{TP_i + FP_i} \times 100\% . \quad (2.4)$$

معیار بازخوانی بیانگر نسبت نمونه هایی است که توسط طبقه بند به درستی تشخیص داده شده اند به کل نمونه های موجود در کلاس مورد نظر. همان طور که در رابطه ی زیر دیده می شود، بازخوانی نشان دهنده ی کامل بودن نتایج مدل در شناسایی نمونه های یک کلاس خاص است؛ به این معنا که هر چه TP_i بزرگ تر و FN_i کوچک تر باشد، مقدار بازخوانی بزرگ تر شده و نشان می دهد تعداد نمونه هایی که باید توسط مدل شناسایی شوند اما از قلم افتاده اند، کمتر است

$$\text{Recall} = \frac{TP_i}{TP_i + FN_i} \times 100\% . \quad (3.4)$$

معیار F-Score میانگین همساز میان دو معیار صحت و بازخوانی است و به عنوان یک شاخص ترکیبی برای ارزیابی همزمان دقت و کامل بودن مدل عمل می کند. همان طور که در رابطه زیر مشاهده می شود، این معیار تعادلی ساختاری میان نرخ مثبت های کاذب FP و منفی های کاذب FN برقرار می سازد؛ به این معنا که بالا بودن مقدار F-Score نشان دهنده عملکرد متوازن و پایدار طبقه بند است.

$$\text{F-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% . \quad (4.4)$$

۵ نتایج

به منظور افزایش دقت طبقه بند ماشین بردار پشتیبان دوقلو، پارامترهای α و β باید بهینه سازی شوند. بدین منظور، از مجموعه ای از الگوریتم های فرااکتشافی استفاده شده است که در قالب یک «اکوسیستم دریایی» با یکدیگر تعامل و همکاری می کنند تا بهترین جواب (مقادیر بهینه برای این دو پارامتر) را بیابند.

برای اعتبارسنجی اثربخشی روش پیشنهادی، نسخه های مختلفی مورد مقایسه تجربی قرار گرفتند: نسخه پایه ماشین بردار پشتیبان دوقلو (Twin SVM) و نسخه های ترکیب شده با الگوریتم های 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 سیستم عامل ویندوز اجرا گردیده‌اند. علاوه بر این، جهت حفظ عدالت در مقایسه تجربی میان الگوریتم‌های فرااکتشافی مختلف، تنظیمات پارامترهای اولیه درون هر الگوریتم دقیقاً مطابق با مقادیر استاندارد و پیش فرض 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 ارائه شده با استفاده از معیارهای، دقت کلی صحت و بازخوانی برای دو مجموعه داده «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 نتیجه گیری

در این پژوهش یک چارچوب نوآورانه به نام الگوریتم اکوسیستم دریایی ارائه شد که در آن چهار الگوریتم قدرتمند فرااکتشافی، یعنی بهینه سازی ازدحام ذرات، الگوریتم ازدحام سالپ‌ها، الگوریتم شکارچیان دریایی و الگوریتم بهینه سازی نهنگ‌ها به صورت هم زمان و در قالب یک سازوکار هوشمند، به منظور بهینه سازی پارامترهای ماشین بردار پشتیبان دوقلو با یکدیگر ترکیب شدند. در روش پیشنهادی، در هر تکرار فرایند بهینه سازی، عملکرد الگوریتم‌ها به صورت پویا ارزیابی شده و مؤثرترین الگوریتم جستجو به عنوان «برنده» انتخاب می شود تا مسئولیت تولید جمعیت مرحله بعد را بر عهده بگیرد. این سازوکار انتخاب پویا به گونه ای طراحی شده است که درون هر الگوریتم، از یادگیری تقویتی چندعاملی برای تنظیم هوشمندانه تعادل بین کاوش و بهره برداری استفاده می شود؛ در نتیجه، فضای جستجو به شکل هدفمند و کارآمدتری پیمایش می گردد و خطر گیر افتادن در بهینه های محلی کاهش می یابد. فرایند ارزیابی مدل پیشنهادی، نشان دهنده دو محدودیت ساختاری کلیدی است که نیازمند توجه در تحقیقات آتی هستند. نخست، افزایش ابعاد فضای حالت-عمل در جدول پایه ای $Q - table$ که مستقیماً تحت تأثیر بزرگ شدن حجم جمعیت اولیه قرار دارد؛ این پدیده منجر به تشدید پیچیدگی زمانی و حافظه ای الگوریتم شده و در نتیجه، سرعت همگرایی بهینه را کاهش می دهد. دومین محدودیت، وابستگی شدید کارایی الگوریتم Q-Learning به حجم داده های ورودی در مدل ماشین بردار پشتیبان دوقلو است. از آنجا که حل مسائل الگوریتم های مرتبه دوم در ماشین بردار پشتیبان دوقلو با چالش داده های مقیاس بزرگ مواجه است، ترکیب آن با فرایند یادگیری تقویتی می تواند به عنوان یک گلوگاه محاسباتی عمل کرده و مقیاس پذیری روش پیشنهادی را در مواجهه با پایگاه داده های وسیع با محدودیت روبه رو سازد. نتایج تجربی به دست آمده بر روی انواع مجموعه داده های سرطان پستان نشان داد که چارچوب پیشنهادی، هم از نظر دقت و هم از نظر پایداری، عملکردی برتر نسبت به روش های کلاسیک و الگوریتم های تک عاملی دارد. در مسئله طبقه بندی دودویی، مقادیر شاخص هایی نظیر دقت کلی، صحت و بازخوانی به سطوح بسیار بالا رسیده است که این امر نشان دهنده توانایی بالای مدل در تشخیص صحیح و تفکیک دقیق تومورهای خوش خیم و بدخیم است. این نتایج، علاوه بر تأیید کارایی اکوسیستم دریایی پیشنهادی در بهینه سازی ماشین بردار پشتیبان دوقلو، بیانگر آن است که ترکیب چند الگوریتم فرااکتشافی در قالب یک چارچوب انتخاب گر پویا، می تواند به طور محسوسی کیفیت تصمیم گیری طبقه بند را بهبود دهد.

از منظر بالینی، یافته های این پژوهش نشان داد که چارچوب ارائه شده قابلیت تبدیل شدن به یک سامانه قدرتمند تشخیص کمک پزشکی را داراست. این سامانه می تواند با کاهش خطای انسانی، افزایش سرعت تحلیل داده های پزشکی و فراهم کردن یک ابزار تحلیلی دقیق برای پزشکان، از فرایند تصمیم گیری بالینی در تشخیص زودهنگام سرطان پستان پشتیبانی کند. به کارگیری چنین سامانه ای در محیط های واقعی درمانی، می تواند به بهبود کیفیت غربالگری، کاهش تشخیص های 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.