پنجاه و ششمین کنفرانس ریاضی ایران
56TH ANNUAL IRANIAN MATHEMATICS CONFERENCE
دانشگاه ولی عصر

# Breast Cancer Diagnosis Via Twin Support Vector Machine Optimized by Metaheuristic Algorithms

Ayda Rahimi *†1, Omid SolaymaniFard[2]

[1]Master's student in Mathematical Biology, Ferdowsi University of Mashhad

[2]Associate Professor, Ferdowsi University of Mashhad

**Abstract**

Breast cancer is one of the most common diseases among women. Early and accurate diagnosis is essential to improve treatment outcomes and patient survival. In this study, we propose two hybrid classification models for breast cancer diagnosis by combining Twin Support Vector Machine (Twin SVM) with two metaheuristic optimization algorithms: Starfish Optimization Algorithm (SFOA) and Particle Swarm Optimization (PSO). These models, named SFOA-Twin SVM and PSO-Twin SVM, use a Radial Basis Function (RBF) kernel and aim to optimize the dual variables in the Twin SVM framework. Both approaches are applied to improve classification accuracy. A 30-run cross-validation experiment is conducted on a breast cancer dataset to evaluate their performance. The results show that the PSO-Twin SVM model achieves higher accuracy and better overall performance compared to the SFOA-Twin SVM model.

**Keywords:** Breast Cancer, Twin Support Vector Machine, Optimization Algorithm, Metaheuristic Algorithm

**Mathematics Subject Classification (2020):** 68T05, 90C27, 92C50.

---

## 1 Introduction

Breast cancer remains a leading cause of cancer-related mortality among women worldwide, with millions of new cases diagnosed annually, underscoring the urgent need for effective early detection methods. The advent of machine learning has transformed medical diagnostics, enabling predictive models to support clinicians in achieving timely and accurate diagnoses. In particular, supervised learning approaches have shown significant promise in breast cancer detection tasks [1]. Twin Support Vector Machine (Twin SVM), an advanced variant of the traditional SVM, has emerged as a powerful tool for classification tasks. Unlike standard SVM, which solves a single large quadratic programming problem, Twin SVM addresses two smaller quadratic programming problems, resulting in faster computation and enhanced generalization, which are particularly valuable for processing complex medical datasets [3]. To optimize its parameters, we employ Starfish Optimization Algorithm (SFOA), a recent nature-inspired metaheuristic modeled after the movement

*Corresponding author: ayda.rahimi8880@gmail.com

†Speaker

and regeneration behavior of starfish [2], and Particle Swarm Optimization (PSO), a swarm-based metaheuristic. SFOA and PSO are applied separately to optimize Twin SVM's dual variables, utilizing SFOA's robust search capabilities and PSO's swarm-based strategies, respectively, to compare their effectiveness in the end. This study evaluates the SFOA-Twin SVM and PSO-Twin SVM frameworks for breast cancer classification, aiming to deliver high accuracy and robustness on a real-world medical dataset, with results indicating that PSO outperforms SFOA.

# 2 Brief Review of Two Metaheuristic Algorithms

## 2.1 Starfish Optimization Algorithm (SFOA)

Here, we briefly review the Starfish Optimization Algorithm (SFOA), which incorporates two key strategies: **exploitation** (local refinement using historical memory) and **exploration** (global search via random movement).

**Exploration Phase:** When the problem dimension $D > 5$, the candidate solution is updated as:

$$Y_{i,p}^T = \begin{cases} X_{i,p}^T + a_1 \left( X_{(\text{best,p})}^T - X_{i,p}^T \right) \cos\theta & \text{if } r \leq 0.5 \\ X_{i,p}^T - a_1 \left( X_{(\text{best,p})}^T - X_{i,p}^T \right) \sin\theta & \text{if } r > 0.5 \end{cases} \tag{1}$$

where $a_1 = (2r-1)\psi$, $\theta = \frac{\pi\psi}{2} \cdot \text{ctd}(T_{\text{t}}, T_{\max})$, $r$ is a random number, $T$ is the current iteration, and $\psi \in (0,1)$ is a random number.

When $D \leq 5$, the formula becomes:

$$Y_{i,p}^T = E_t \times X_{i,p}^T + A_1 \left( X_{k1,p}^T - X_{i,p}^T \right) + A_2 \left( X_{k2,p}^T - X_{i,p}^T \right) \tag{2}$$

where $E_t = \frac{T_t}{T_{\max}} \cos\theta$, and $A_1, A_2$ are random numbers in $(-1, 1)$.

**Exploitation Phase:** In the preying step, the solution is refined using:

$$Y_{i,p}^T = X_{i,p}^T + r_1 d_{m1} + r_2 d_{m2} \tag{3}$$

where $r_1, r_2$ are random numbers in $(0,1)$, and $d_{m1}, d_{m2}$ denote the distances to the best position found.

## 2.2 Particle Swarm Optimization (PSO)

Additionally, we utilize Particle Swarm Optimization (PSO), a metaheuristic inspired by the social behavior of bird flocks or fish schools. PSO updates particle positions based on both individual and global best solutions.

**Velocity Update:** The velocity $v_i^{t+1}$ of particle $i$ at iteration $t+1$ is given by:

$$v_i^{t+1} = wv_i^t + c_1 r_1 (p_{\text{best},i} - x_i^t) + c_2 r_2 (g_{\text{best}} - x_i^t) \tag{4}$$

where $v_i^t$ is the current velocity, $w$ is the inertia weight, $c_1$ and $c_2$ are the cognitive and social coefficients, $r_1, r_2 \in (0,1)$ are random numbers, $p_{\text{best},i}$ is the personal best position, and $g_{\text{best}}$ is the global best position.

**Position Update:** The position $x_i^{t+1}$ is updated using:

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{5}$$

# 3 Twin Support Vector Machine

**Definition 3.1.** *Twin Support Vector Machine (Twin SVM) constructs two nonparallel hyperplanes by solving two smaller-sized quadratic programming problems (QPPs), each aiming to be* *closer to one class and farther from the other. Given a dataset* $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ *with* $y_i \in \{+1, -1\}$, *the primal problems are:*

$$\min_{\mathbf{w}_1, b_1} \frac{1}{2}\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|^2 + c_1\|\boldsymbol{\xi}_1\|^2, \quad s.t. \ -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \boldsymbol{\xi}_1 \geq \mathbf{e}_2, \ \boldsymbol{\xi}_1 \geq 0, \tag{6}$$

*and*

$$\min_{\mathbf{w}_2, b_2} \frac{1}{2}\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|^2 + c_2\|\boldsymbol{\xi}_2\|^2, \quad s.t. \ (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \boldsymbol{\xi}_2 \geq \mathbf{e}_1, \ \boldsymbol{\xi}_2 \geq 0, \tag{7}$$

*where* $\mathbf{A}$ *and* $\mathbf{B}$ *are data matrices for the positive and negative classes, respectively,* $\mathbf{w}_1, \mathbf{w}_2$ *are weight vectors defining the hyperplanes,* $b_1, b_2$ *are bias terms,* $c_1, c_2 > 0$ *are regularization parameters controlling the penalty on slack variables,* $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ *are slack variable vectors and* $\mathbf{e}_1$, $\mathbf{e}_2$ *are vectors of ones.*

*The corresponding dual problems,which are solved in this study using a Metaheuristic Algorithm, are:*

$$\max_{\boldsymbol{\alpha}} \ \mathbf{e}_2^\top \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^\top \mathbf{H}_1(\mathbf{H}_1^\top \mathbf{H}_1 + c_1\mathbf{I})^{-1}\mathbf{H}_1^\top \boldsymbol{\alpha}, \quad 0 \leq \boldsymbol{\alpha} \leq c_1, \tag{8}$$

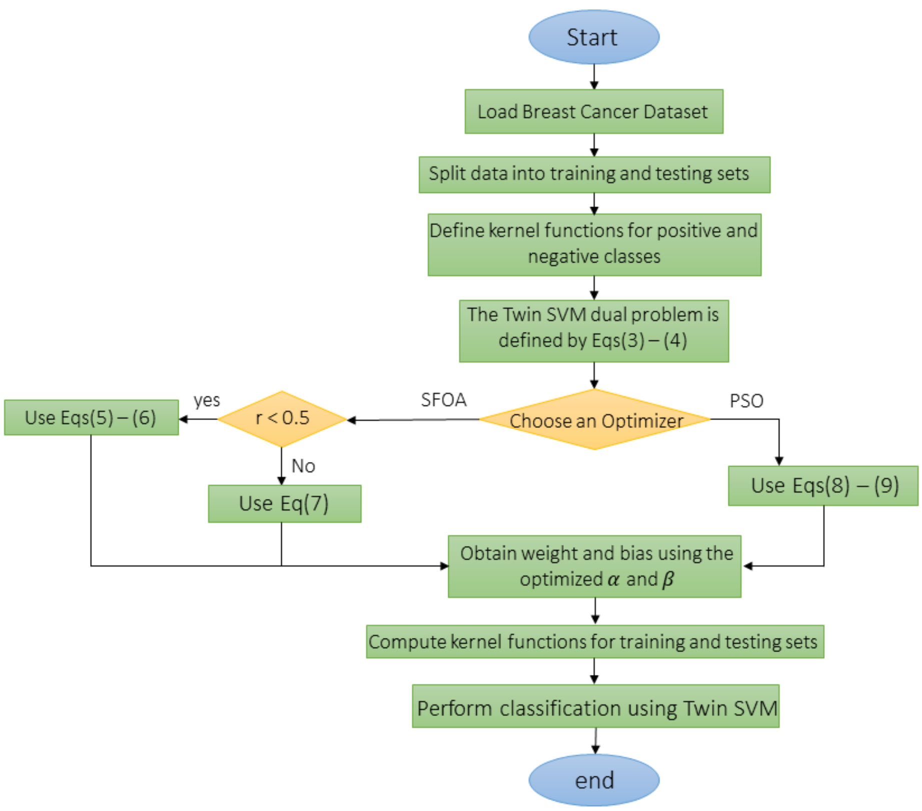
*and*

$$\max_{\boldsymbol{\beta}} \ \mathbf{e}_1^\top \boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}^\top \mathbf{H}_2(\mathbf{H}_2^\top \mathbf{H}_2 + c_2\mathbf{I})^{-1}\mathbf{H}_2^\top \boldsymbol{\beta}, \quad 0 \leq \boldsymbol{\beta} \leq c_2, \tag{9}$$

*where* $\mathbf{H}_1 = [\mathbf{B} \ \mathbf{e}_2]$, $\mathbf{H}_2 = [\mathbf{A} \ \mathbf{e}_1]$.

# 4 The Hybrid Twin SVM

The flowchart of the algorithm is as follows:



## 5 Numerical Results

In this work, we use a Gaussian kernel function as the radial basis function (RBF) in two proposed classification models. To evaluate the classification accuracy, we apply the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where:

- **TP (True Positive):** Number of correctly predicted positive samples.
- **TN (True Negative):** Number of correctly predicted negative samples.
- **FP (False Positive):** Number of negative samples incorrectly predicted as positive.
- **FN (False Negative):** Number of positive samples incorrectly predicted as negative.

The dataset used in this experiment is taken from the LIBSVM collection, available at: https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#breast-cancer

A summary of the results obtained using two approaches is presented in Table 1.

Table 1: Accuracy Metrics Over 30 Runs

| Metric | SFOA | PSO |
|---|---|---|
| Minimum Training Accuracy | 93.91 | **98.57** |
| Mean Training Accuracy | 94.37 | **98.86** |
| Minimum Testing Accuracy | 83.82 | **94.80** |
| Mean Testing Accuracy | 92.35 | **95.51** |

## 6 Conclusion

This paper proposed hybrid model for breast cancer classification using Twin Support Vector Machine (Twin SVM) combined with two optimization algorithms, SFOA and PSO. The results demonstrated that PSO outperformed SFOA in optimizing Twin SVM parameters, achieving higher accuracy in both training and testing data. These findings highlight the significant potential of PSO in medical diagnostic applications.

## References

[1] Yifeng Dou and Wentao Meng, *An Optimization Algorithm for Computer-Aided Diagnosis of Breast Cancer Based on Support Vector Machine*, Frontiers in Bioengineering and Biotechnology, 9(2021), No. 698390.

[2] Changting Zhong, Gang Li, Zeng Meng, Haijiang Li, Ali Riza yildiz and Seyedali Mirjalili, *Starfish optimization algorithm (SFOA): a bio-inspired metaheuristic algorithm for global optimization compared with 100 optimizers*, Neural Computing and Applications, 37(2025), 3641 - 3683.

[3] Divya Tomar and Sonali Agarwal. *Twin Support Vector Machines: A review from 2007 to 2014*,Egyptian Informatics Journal, 16(2015), 55 - 69.