Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 20, Iss. 2, pp. 21-33, 2022

BENIGN AND MALIGN BREAST CANCER CLASSIFICATION USING SUPPORT VECTOR MACHINES OPTIMIZED WITH PARTICLE SWARM AND GENETIC ALGORITHMS

Uriel Abe Contardi , Paulo Rogério Scalassara
Department of Electrical Engineering, Federal University of Technology - Paraná, Cornélio Procópio, Brazil
urielcontardi@hotmail.com, prscalassara@utfpr.edu.br

Douglas Vieira Thomaz
Faculty of Pharmacy, Federal University of Goiás, Goiânia, Brazil
douglasvthomaz@gmail.com

Resumo – Câncer de mama é uma doença neoplásica que pode ser diagnosticado como benigno ou maligno de acordo com a taxa de crescimento da lesão neoplásica. Devido à relevância de se obter melhores ferramentas de detecção, este trabalho descreve o desenvolvimento e otimização de máquinas de vetores de suporte para classificação dos tipos desse câncer. Foram realizados testes usando o banco de dados de câncer de mama dos hospitais da universidade de Wisconsin, EUA, disponibilizados pelo repositório de aprendizagem de máquina da Universidade da Califórnia, Irvine. Usou-se núcleo com função de base radial para o classificador e os hiperparâmetros foram refinados usando dois métodos: otimização por enxame de partículas e algoritmos genéticos. Os resultados para o primeiro método exibiram 97,71% de acurácia, 96,30% de sensibilidade e 98,65% de seletividade. Já usando o segundo método, a acurácia foi 95,78% com sensibilidade e seletividade de 96,73% e 95,25% respectivamente. Dessa forma, tem-se o indicativo que esses algoritmos de busca são ferramentas viáveis para otimizar modelos de aprendizagem de máquina para o propósito de classificação de câncer de mama.

Palavras-chave – algoritmo evolucionário, algoritmo de inteligência de enxame, modelos neurais, otimização, tumor neoplásico

Abstract – Breast cancer is a neoplastic disease that can be diagnosed either as benign or malign according to the growth-rate of the neoplastic lesion. Owing to the relevance of obtaining better detection tools, this work describes the development and optimization of support vector machines for the classification of the types of such cancer. Tests were performed using the breast cancer dataset of the University of Wisconsin Hospitals, USA, available at the Machine Learning Repository of the University of California Irvine. The radial basis function kernel was selected for the classifier and its hyperparameters were refined using two methods: particle swarm optimization and genetic algorithms. The results for the first method exhibited 97.71% accuracy, 96.30% sensitivity, and 98.65% of selectivity. On the other hand, using the second method, the accuracy was 95.78%, with sensitivity and selectivity of 96.73% and 95.25%, respectively. Therefore, there is an indication that these search algorithms are viable tools to optimize machine learning models for the purpose of breast cancer classification.

Keywords – evolutionary algorithm, swarm intelligence algorithm, neural models, optimization, neoplastic tumor.

1. INTRODUCTION

Breast cancer (BC) is a neoplastic disease characterized by the abnormal development and growth of cells in the breast tissue. This condition may be diagnosed as benign (B) or malign (M) according to the growth-rate of the neoplastic lesion, being malignancy widely considered a life-threatening prognostic. BC lesions may be classified by a trained physician according to their morphology, size and texture. In this regard, guidelines suggest the evaluation of the following characteristics of the lesions, among others: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension [1]. Nonetheless, most visual diagnoses are confirmed thereafter by laboratory tests such as biopsy; what highlights the importance of adequate and non-biased visual inspection of the lesion by trained health professionals [2].

Regarding BC progression, several authors emphasize the upmost importance of early-diagnosis for a good clinical outcome [3], thereby leading medical staff to promote informative outreaches and campaigns regarding self-assessment by patients in order to foster public awareness [4]. Nevertheless, BC still takes its toll in healthcare worldwide, as 2.3 million new cases and almost 685 thousand deaths were reported in 2020 [5]. Notwithstanding, the majority of the reported deaths are late-diagnosed malign BC, what henceforth demonstrates the relevance of early-identification of the neoplastic lesions and their correct differential diagnosis, that is distinguishing M from B cases [6].

Concerning the classification of neoplastic lesions, some authors have proposed the use of machine learning algorithms to accelerate diagnosis [7, 8]. Recent approaches, such as [9], use deep learning techniques, however they are still large models that have computationally demanding training processes. In this regard, support vector machines (SVM) have been favored due to the possibility of implementing kernel transformations of the input space into hyperplanes, thereby providing adequate nonlinear classification even when several attributes are considered simultaneously [10]. Despite SVM being a powerful tool in separating classes of attributes, the optimization of hyperparameters is highly relevant to ensure a robust and efficient classification method.

The use of polynomial, radial basis function (RBF), sigmoidal and Mahalanobis-based kernel functions has been reported to yield efficient SVM models for BC differential diagnosis [11]. However, the kernel selection and adjustment through the optimization of hyperparameters, such as C and Gamma values, was rarely explored in attempts to provide classification of BC into M and B. In this context, the use of optimization techniques could improve data classification, thereby aiding adequate diagnosis [12–14].

Recent tools applied to machine learning hyperparameters optimization are particle swarm optimization (PSO) and genetic algorithms (GA) [15], which are meta-heuristic methods of iterative search for the best parameters that optimize a function. PSO treats each candidate solution, in our case, hyperparameters to be tunned, as "particles" whose movements are traced on a performance landscape aiming the best response of a cost function. On the other hand, GA mimics natural selection and its effects on the evolution of a population over time. In this sense, GA generates random mutations on the candidate parameters upon each iteration, and selects them according to the best response up to obtaining optimal values. Nonetheless, these approaches are well-regarded methods for optimization of parameters of learning tools, having been reported by several authors [16].

A common approach to optimize hyperparameters of learning methods is to employ exhaustive search in the space of possible values, which can be computationally intensive [17]. A popular algorithm is grid search, where training data can be used to find the best classifier parameters in a brute-force search for the minimum error in the classification of test data [18]. But this strategy is time consuming and, recently, meta-heuristic techniques are being employed instead [19].

Due to the importance of providing tools for the differential diagnosis of BC, this work analyzes the use of PSO and GA to optimize SVM classifier hyperparameters, that is, the C and Gamma values, in order to develop a reliable BC classification model. After this brief introduction, this paper is divided in four more sections. In Theory, the SVM, PSO, and GA algorithms are detailed. In Materials and Methods, we present information about the Wisconsin BC dataset (WBCD). In Results and Discussion, the optimization

results are discussed considering the BC classification using SVM optimized by PSO and GA, along with a comparison with grid search, and, in Conclusion, we present the final comments of the work.

2 THEORY

In this section, we present some theoretical details of the SVM classifier along with the optimization methods.

2.1 SVM Classifier

SVM is a machine learning approach that is usually applied for classification purposes. This tool targets the best hyperplane in order to separate classes of samples. The use of manifold dimensions turns SVM into a powerful method for data classification, as nonlinear data can also be separated using nonlinear mappings, also known as the kernel trick [10, 20]. Figure 1 illustrates a linear SVM model with two classes (hexagons and stars) and only two input features represented by the two-dimensional space.

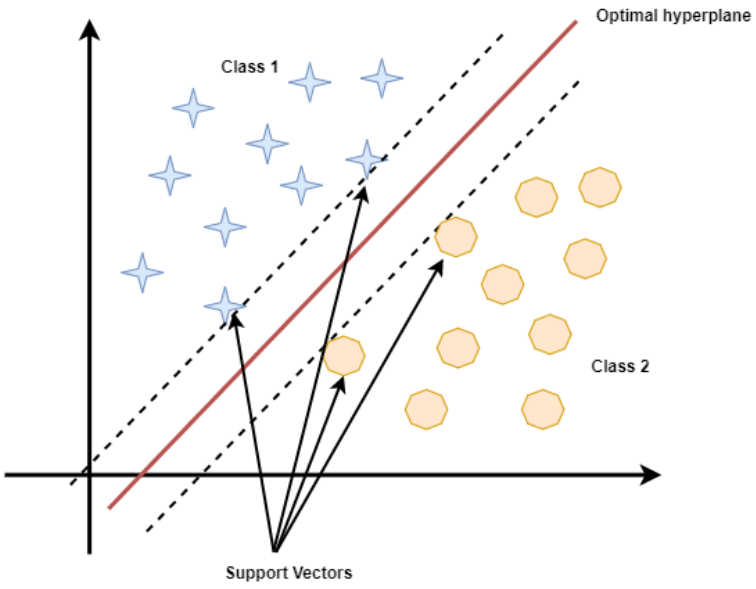


Figure 1: Illustration of a SVM model with linear hyperplane and two input features (the axis) discriminating between two classes (hexagons and stars). The support vectors of each class are also indicated.

Given N inputs $\mathbf{x}_i \in \mathbb{R}^d$, the separation hyperplane is defined as $\mathbf{w}^T\mathbf{x} + b = 0$, where $\mathbf{w} \in \mathbb{R}^d$ are the coefficients and b is a bias. The optimum solution is obtained by the optimization in Eq. 2, with $y_i \in [-1, 1]$ the classes targets and ξ_i the slack variables to allow for flexibility (outliers) by the classifier [21]. C is a regularization hyperparameter to control the amount of flexibility. High C values result in rigid margins whereas low values permit the restrictions to be easily ignored.

$$\min_{\mathbf{w}\in\mathbb{R}^d, \xi_i \in \mathbb{R}^+} ||\mathbf{w}||^2 + \mathrm{C}\sum_{i}^{N} \xi_i \tag{1}$$
$$\text{subjected to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i \qquad \text{for} \quad i = 1, \ldots, N$$

The objective of the nonlinear mapping used in SVM is to transform the data to higher dimensions and find a projection where the classes of samples are linearly separable. This strategy permits that classes which are not linearly separable to be correctly discriminated [21]. This approach maps the samples $\mathbf{x}_i$ of

dimension d to a feature (Hilbert) space of dimension $D > d$ according to a function $\Phi(\mathbf{x})$. The advantage of this technique, what caused it to be named kernel trick, is that using a symmetrical and continuous function $K(\mathbf{x}_i, \mathbf{x}_j)$ that satisfies the Mercer's theorem [20], it is possible to perform operations with the inputs in the Hilbert space without explicitly constructing the nonlinear mapping $\Phi(\mathbf{x}_i)$ that projects the data samples to that space.

In this work, we employed the RBF kernel as this nonlinear mapping since other authors reported satisfactory outcomes for the application of SVM on BC classification [11, 22]. Eq. (2) presents this kernel that depends on the gamma (γ) hyperparameter.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2\gamma^2), \quad \gamma > 0 \tag{2}$$

The adjustment of the SVM classifier is performed by tuning the C and gamma values. These hyperparameters are fundamental to obtain a optimum classifier [22, 23]. The C value regulates how flexible the SVM is regarding the separation margins, that is how much slack these margins can have [20]. On the other hand, gamma is a parameter of the RBF kernel, that quantifies the spread of the Gaussian function [23].

Among the most common methodologies to optimize hyperparameters in SVM is grid search, which consists of a systematic calculation of all possible hyperparameters and their performance in separating the data [13]. However, this approach is highly computationally demanding and may lead to oversearching. Therefore, specifically for the BC classification, we opted instead to optimize SVM-RBF kernel hyperparameters using less-demanding approaches, namely PSO and GA, similarly to [24, 25], and described in the next section.

2.2 PSO and GA Algorithms

Both algorithms are bio-inspired techniques, PSO is based on collective swarm intelligence, that is the behavior of interacting agents that follow simple rules [26]. On the other hand, GA is inspired in the evolution of living creatures.

In PSO, the method aims to search a feature space based on the behavior of social animals, like bird flocks trying to optimize a cost function, usually called fitness function in this application [27]. This approach is based on the cooperation and competition of populations of candidate solutions or particles. In this sense, individual particles socially evolve, adjusting their trajectory and position according to the behavior of the best candidate solution nearby. The optimal outcome is reached when candidate solutions reach the stop criterion at a specific iteration [16, 28].

The position in the search space of each particle is the optimized parameters. The best fitness obtained by the particle is compared to the best value obtained by any particle of the swarm. The search iteration is performed by updating each particles' position by means of velocity changes seeking the local or global best values. Considering $\mathbf{b}_p$ and $\mathbf{v}_p$, in $\mathbb{R}^P$, the position and velocity of the P particles in the swarm, the position update rule with respect to iteration n is given by Eq. (3) [27].

$$\mathbf{b}_p[n+1] = \mathbf{b}_p[n] + \mathbf{v}_p[n+1] \tag{3}$$

The velocity update is given by Eq. (4), where $\omega[n]$ is the inertia weight of the previous velocity, $\mathbf{U}_1$ and $\mathbf{U}_2$ are matrices of uniform random numbers, and $\mathbf{b}_p^B$ and $\mathbf{b}_g^B$ are, respectively, the best global and local positions. Finally, ϕ_1 and ϕ_2 are acceleration coefficients for the best particle and best global position, respectively.

$$\mathbf{v}_p[n+1] = \omega[n]\mathbf{v}_p[n] + \phi_1\mathbf{U}_{p1}[n](\mathbf{b}_p^B[n] - \mathbf{b}_p[n]) + \phi_2\mathbf{U}_{p2}[n](\mathbf{b}_g^B[n] - \mathbf{b}_p[n]) \tag{4}$$

In the case of the SVM optimization, the fitness is function is the BC classification accuracy for the training data set.

On the other hand, the GA algorithm is based on an evolutionary process, simulating Darwinian evolution by generating random populations and submitting them to natural selection according to the calculated

outcome. In this sense, each new iteration configures a selection step wherein only the best individual candidate solutions are selected, and their genes/behaviors will be mutated in the following iteration until the optimal outcome is reached [29].

The fittest individuals (similar to the particles in PSO) transmit their information to the next generation, whereas the less fit, do not [30]. The procedure is divided in five steps: population generation, fitness evaluation, selection, crossover, and mutation. The first phase is generating a random population composed of the individuals that are characterized by a set of parameters. These parameters, called genes in GA jargon, are the features to be searched, and their collection is called chromosome. The second step is evaluating the fitness function for all the individuals, likewise for PSO. The third step is the selection, where the fittest individuals pass to children some of their genes, which are crossed in the fourth step, crossover. The genes of a pair of selected individuals are combined to create the next generation, which tends to produce more fit individuals. Finally, the fifth step is mutation, where some of the genes are randomly changed in order to maintain population diversity, what could avert a premature algorithm convergence to a local minimum. The procedure is terminated when the maximum number of generations is achieved or if a optimum solution (to some criterion) is obtained.

3 MATERIALS AND METHODS

In this section, some details of the WBCD are presented and the methodology SVM optimization using the PSO and GA algorithms is discussed. All the code for this work was written in Python, using the following libraries: sklearn, optunity, and numpy, while seaborn was used in graphics rendering.

3.1 Wisconsin Breast Cancer Dataset

The data set employed in this study is the WBCD [31], which was constructed at the University of Wisconsin Hospitals, USA, and made available at the Machine Learning Repository of the University of California Irvine [32]. It is based on experimental information gathered from patients following fine needle aspirate diagnostic procedure. The cancer lesions of BC-afflicted patients were punctured and the collected cells were investigated through optic microscopy [33].

The optical micrographs were then submitted to image processing software, and the ten attributes of the cells from the lesions (mentioned in Introduction) were obtained [1], namely:

- radius (mean of distances from center to points on the perimeter of the cell),
- texture (standard deviation of gray-scale values),
- perimeter of the cell,
- area of the cell,
- smoothness (local variation in radius lengths),
- compactness (given by perimeter squared divided by the area),
- concavity (severity of concave portions of the cell contour),
- concave points (number of concave portions of the cell contour),
- symmetry, and
- fractal dimension (obtained using the “coastline approximation”).

Data regarding mean, standard deviation, and qualitative classification of outcomes (M or B) from each processed optical micrograph is provided in the WBCD, thereby leading to 30 attributes/features encompassing 569 samples. Benign diagnostics accounted for 357 samples, while malign cases accounted for 212 samples.

3.2 Methodology

In this work, PSO and GA were used to tune the optimal C and gamma (γ) hyperparameters of SVM using RBF kernel in order to develop an optimal classifier for the WBCD. Figure 2 illustrates each step of the proposed methodology.

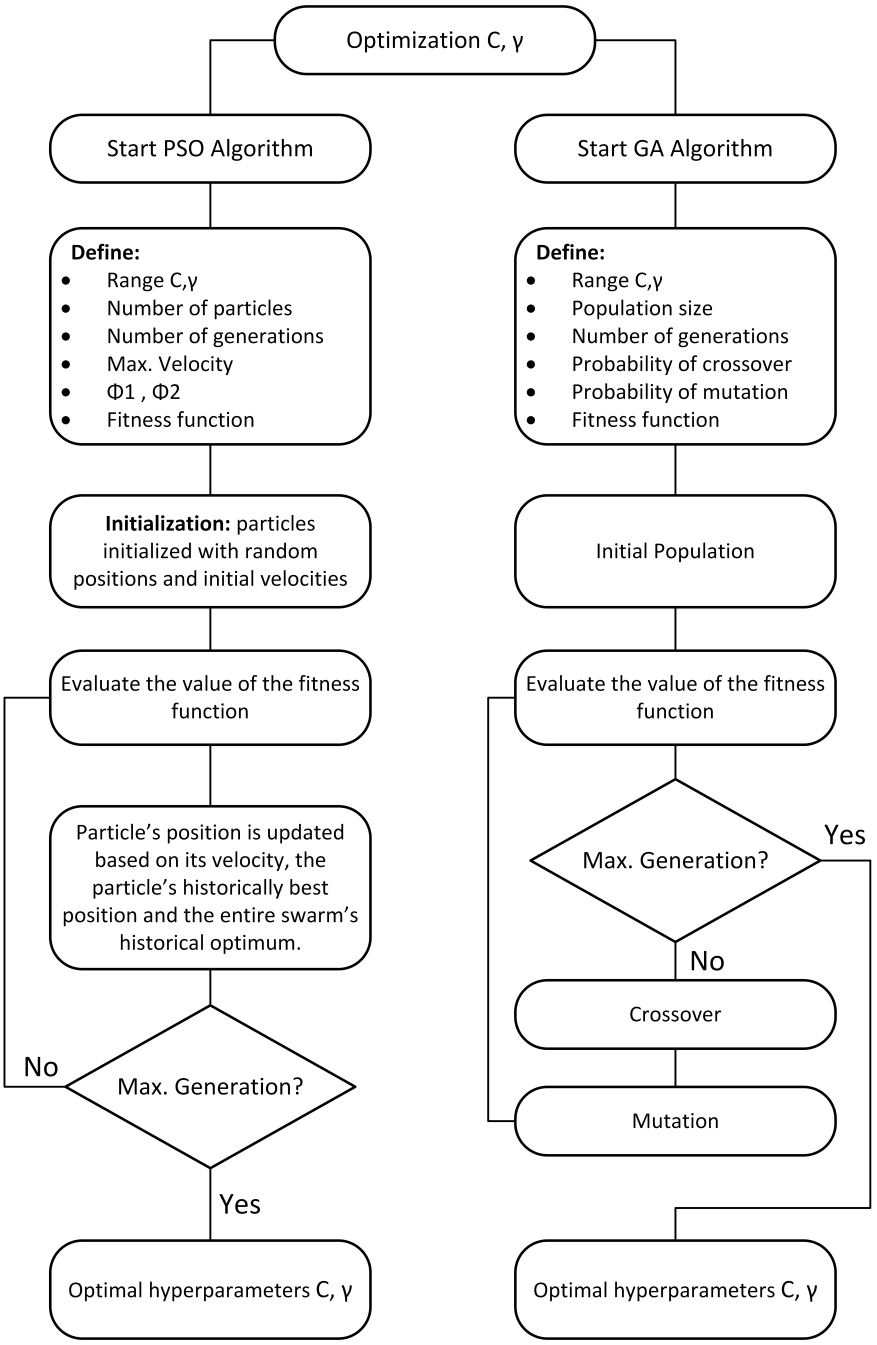


Figure 2: Proposed methodology to optimize the hyperparameters of the SVM-RBF kernel classifier using PSO and GA algorithms.

In order to normalize the input data, the z-score standard [34], Eq. (5), was applied to all features. Each feature input x was transformed to z, with zero mean and unit standard deviation. In this equation, μ and σ are, respectively, the mean and standard deviation of each attribute. After this procedure, the normalized data underwent classification.

$$z = \frac{x - \mu}{\sigma} \tag{5}$$

The optimization and classifier training procedure was performed using k-folds cross-validation with k equals to 10. In this regard, the training data is separated in ten groups, while the model is trained with nine groups and tested with the remaining one. This is then repeated until all groups were used either for training or testing. Through this method, the algorithm can maximize the accuracy of cross validation whilst minimizing error. Thus, the training group encompassing 90% of all data and the test group 10%. The training procedure was repeated ten times with a different test set each time. Figure 3 illustrated the training and testing of the two proposed methods: SVM with PSO and GA optimizations.

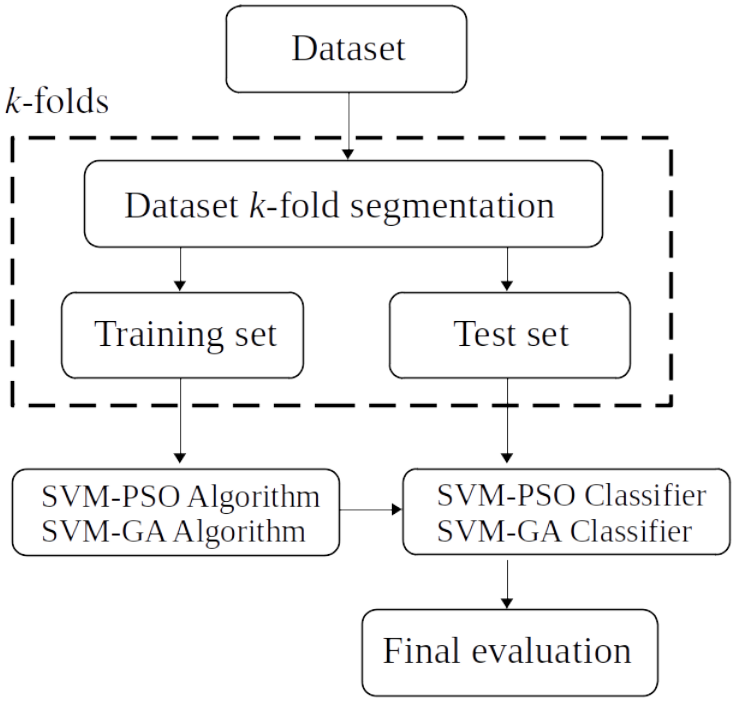


Figure 3: Procedure of training and testing the two proposed methods of optimization and classification: SVM-PSO and SVM-GA.

The GA parameters were: size of the generated population, maximum number of generations, crossover probability, and mutation probability. On the other hand, PSO parameters were: number of particles, number of generations, ϕ_1 that ponders the update particle position according to the local optimum value, and ϕ_2 that ponders the update particle position according to global optimum value. In both algorithms, the search range was delimited as $[1\,,\,10000]$ for C values and $[0.0001\,,\,150]$ for gamma values. This broad range was selected in order to cover as many values as possible. The configurations of the GA and PSO parameters are presented in Table 1. These parameters were selected after exhaustive tests based on the best obtained SVM accuracy rates for BC classification.

Table 1: Configurations of GA and PSO parameters.

GA		**PSO**	
Parameter	**Value**	**Parameter**	**Value**
Population size	60	Particles number	60
Max. generations	100	Max. generations	100
Probability of crossover	0.5	ϕ_1	1.5
Probability of mutation	0.15	ϕ_2	2.0

4 RESULTS AND DISCUSSION

In this section, we present and discuss the results of the SVM BC classification after hyperparameters optimization using PSO and GA algorithms.

After obtaining the best hyperparameter values with both PSO and GA, the classifier models were validated, with results presented in the form of confusion matrices. The figures-of-merit for this analysis were accuracy, sensitivity (also known as recall), and selectivity (specificity), calculated by Eqs. (6), (7), and (8), respectively, where TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative.

$$\text{Accuracy } = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Sensitivity } = \frac{TP}{TP + FN} \tag{7}$$

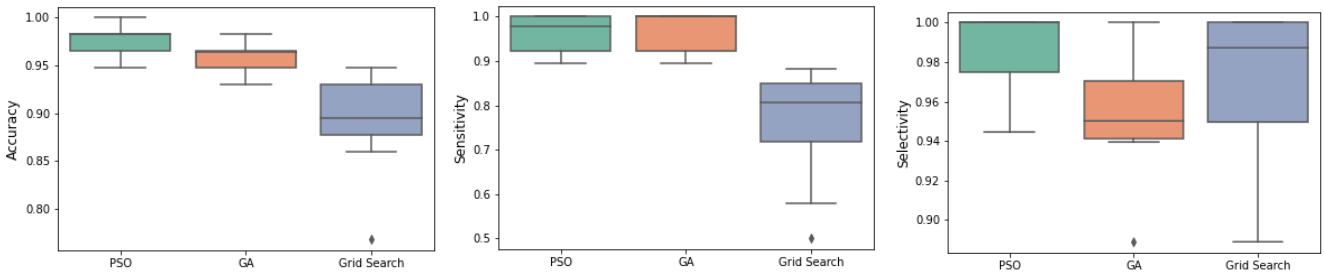
$$\text{Selectivity } = \frac{TN}{TN + FP} \tag{8}$$

Table 2 presents the mean and standard deviation results, over the k-folds validation, with $k = 10$, of accuracy, sensitivity, and selectivity for both optimization methods: SVM-PSO and SVM-GA, along with a comparison using grid search. For the latter, the SVM hyperparameters were delimited to the same range, values of C in $[1, 10000]$ with increments of 1 and Gamma in $[0.0001, 150]$ with increments of 0.015. Figure 4 presents the boxplots of these figures-of-merit.

Table 2: Results (mean and standard deviation) of SVM BC classification using PSO, GA, and grid search optimizations.

Performance	**PSO**		**GA**		**Grid Search**	
	mean	std	mean	std	mean	std
Accuracy	97.71%	1.58%	95.78%	1.79%	89.26%	5.15%
Sensitivity	96.30%	4.12%	96.73%	4.26%	76.08%	12.28%
Selectivity	98.65%	1.84%	95.25%	2.80%	97.19%	3.59%

All the classifiers presented accuracy rates higher than 89% in separating data between benign and malign BC. However, using grid search to optimize the SVM hyperparameters resulted in a lower accuracy and sensitivity, demonstrating that PSO and GA can obtain a better optimization. According to a Welch's t-test [35], the accuracy mean of the grid search approach is statistically different from those of the PSO and GA optimizers with p values of 0.0007 and 0.0041, respectively. Also, the lower sensitivity indicates the

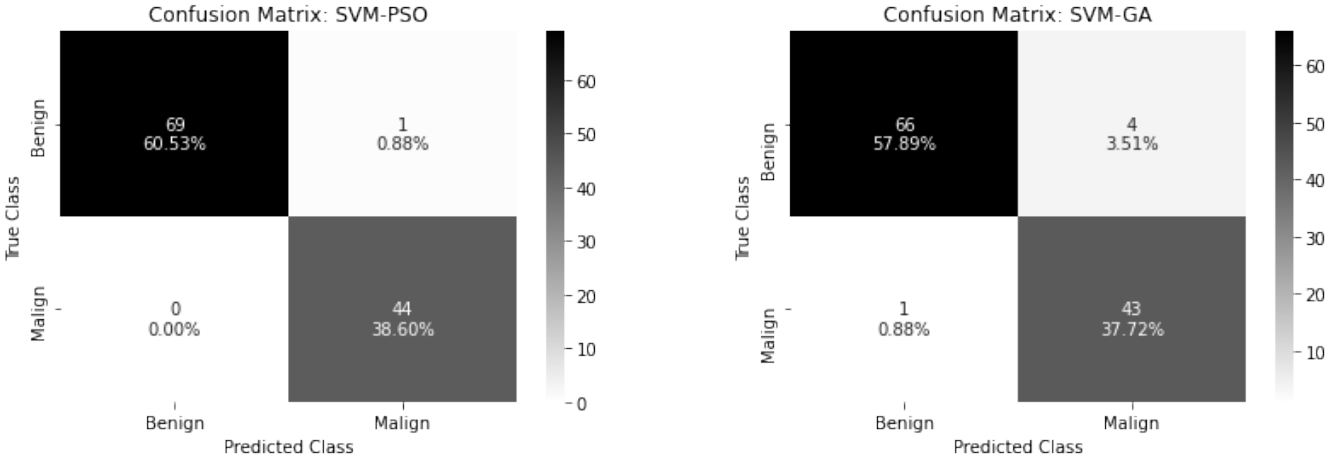


(a) (b) (c)

Figure 4: Boxplots of the result metrics presented in Table 2. (a) Accuracy, (b) sensitivity, and (c) selectivity.

resulting classifier is missing malign BC cases, labeling them as benign. Finally, as shown by the boxplots, the grid search method obtained a higher variation of the figures-of-merit, what could be explained by different initial conditions and obtaining a suboptimal result in the optimization. In order to improve this solution, a finer grid resolution would have to be used.

Considering only PSO and GA, the SVM classifier optimized by the former exhibited better mean accuracy and selectivity than the latter, with equal sensitivity. This may be attributed to the faster convergence of the PSO in comparison to GA [36]. Despite similar accuracy values, a Welch's t-test indicated their mean scores are significantly different with p value of 0.0262. Moreover, it is noteworthy that the same number of generations was used for both models. Figure 5 presents the confusion matrices of the best results obtained by the SVM-PSO and SVM-GA, respectively, during the validation process.



(a) (b)

Figure 5: Confusion matrices for the best result of the validation of: (a) SVM-PSO and (b) SVM-GA.

Other researchers analyzed the WBCD and discriminated BC cases using a plethora of classifiers. In [37], SVM, multilayer perceptron networks (MLP), and minimum distance classifiers using genetic programming generated features (GP-MDC) were compared. The average accuracy rates were 96.32%, 96.21%, and 97.47%, respectively. In [38], a proposed random forest classifier tuned by Bayesian optimization (RF-BO) is compared with SVM tuned by GA and by salp swarm optimization (SVM-SSO). The results using the Wisconsin BC Prognostic data set, that is similar to the WBCD, were 94.10%, 95.90%, and 96.40% for SVM-SSO, SVM-GA, and RF-BO, respectively. Finally, in [39], a review of WBCD classification is presented, focusing mainly in MLP, SVM, decision trees (DT), and k-nearest neighbors (k-NN). The average accuracy rates ranged from 94.36% for the simple DT classifier [40], passing to 97.20% and

98.70% for SVM [41] and k-NN [42], respectively, to 99.26% for MLP with genetic programming optimization (MLP-GPO) [43]. The best results were usually obtained for combinations of methods or based on computationally intensive optimizations or training procedures. Nevertheless, our approach resulted in similar accuracy rates as shown in Table 3.

Table 3: Comparison of WBCD classification accuracy rates.

Reference	Method	Accuracy
Guo [37]	SVM	96.32%
Guo [37]	MLP	96.21%
Guo [37]	GP-MDC	97.47%
Kumar et al. [38]	SVM-SSO	94.10%
Kumar et al. [38]	SVM-GA	95.90%
Kumar et al. [38]	RF-BO	96.40%
Sumbaly et al. [40]	DT	94.36%
Bennett & Blue [41]	SVM	97.20%
Medjahed et al. [42]	k-NN	98.70%
Bhardwaj & Tiwari [43]	MLP-GPO	99.26%
Proposed	SVM-PSO	97.37%
Proposed	SVM-GA	95.61%

5 CONCLUSIONS

In this work, we presented an approach to optimize hyperparameters of SVM classifier with RBF kernel using PSO and GA algorithms. The main objective was to obtain a more refined tool for the classification of benign and malign breast cancer. In order to test such application, the Wisconsin BC dataset was employed, which presents several features obtained from optical micrographs of the lesions.

The proposed methodology was developed using k-folds validation and the accuracy, sensitivity, and selectivity figures-of-merit were used to assess the quality of the results. Also, the confusion matrices of the best result of each approach, PSO and GA, where presented to illustrate behavior of the optimized SVM classifier. For comparison, we also considered grid search, an exhaustive search method that is more computationally intensive and presents no guarantees to achieve a better solution than meta-heuristic techniques.

The three search methods presented similar results, with accuracy rates ranging from 89.26% for grid search, passing from 95.78% for GA to 97.71% for PSO, which is in accordance to other results for the WBCD [25, 39]. However, in [11], where WBCD was used with SVM, the best results with 10-folds cross-validation in similar conditions to our study reached 97.08%, a slightly lower rate. Also, as presented in the previous section, for the WBCD, the range of classification rates are 94% to close to 100% [39], depending on the complexity and combination of methods. Finally, considering a Welch's t-test, we observed that the mean accuracy values of the approaches were statistically different, indicating an advantage of meta-heuristic methods over grid search, and a slight better mean performance of PSO over GA optimized SVM. Nonetheless, the meta-heuristic approaches are less computationally intensive than grid search.

The advantage of the proposed method lies in the fact that improvements to BC classification accuracy rates using simple methods help to design better automatic diagnostic systems. Such systems can be embedded in mobile devices and assist physicians make decisions in any place based on data from biopsy exams. For future works, we intend to analyze and optimize other classifiers in order to compare with the optimized SVM, but still aiming at simple tools instead of deep learning approaches. Also, compare modifications of PSO and other swarm methods. Finally, new breast cancer data sets can be considered to give a better perspective of the advantages of the proposed optimized classifier.

Acknowledgments

We would like to thank the Federal University of Technology - Paraná and the Araucária Foundation for the Support of the Scientific and Technological Development of the State of Paraná. Also, this study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] W. N. Street, W. H. Wolberg and O. L. Mangasarian. "Nuclear feature extraction for breast tumor diagnosis". In *Biomedical Image Processing and Biomedical Visualization*, edited by R. S. Acharya and D. B. Goldgof, volume 1905, pp. 861–870. International Society for Optics and Photonics, SPIE, 1993.

[2] M. J. Silverstein, M. D. Lagios, A. Recht *et al.*. "Image-Detected Breast Cancer: State of the Art Diagnosis and Treatment". *Journal of the American College of Surgeons*, vol. 201, no. 4, pp. 586–597, 2005.

[3] O. Ginsburg, C.-H. Yip, A. Brooks *et al.*. "Breast cancer early detection: A phased approach to implementation". *Cancer*, vol. 126, no. S10, pp. 2379–2393, 2020.

[4] N. Azamjah, Y. Soltan-Zadeh and F. Zayeri. "Global Trend of Breast Cancer Mortality Rate: A 25-Year Study". *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 7, pp. 2015–2020, 2019.

[5] W. H. Organization. "Breast cancer". `https://www.who.int/news-room/fact-sheets/detail/breast-cancer`. Accessed: 2022-06-14.

[6] K. Hu, P. Ding, Y. Wu, W. Tian, T. Pan and S. Zhang. "Global patterns and trends in the breast cancer incidence and mortality according to sociodemographic indices: an observational study based on the global burden of diseases". *BMJ Open*, vol. 9, no. 10, pp. e028461, 2019.

[7] I. Kononenko. "Machine learning for medical diagnosis: history, state of the art and perspective". *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001.

[8] M. Karabatak and M. C. Ince. "An expert system for detection of breast cancer based on association rules and neural network". *Expert Systems with Applications*, vol. 36, no. 2, pp. 3465–3469, 2009.

[9] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran and K. Ur Rehman. "A Brief Survey on Breast Cancer Diagnostic With Deep Learning Schemes Using Multi-Image Modalities". *IEEE Access*, vol. 8, pp. 165779–165809, 2020.

[10] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, second edition, 2010.

[11] K. Polat and S. Güneş. "Breast cancer diagnosis using least square support vector machine". *Digital Signal Processing*, vol. 17, no. 4, pp. 694–701, 2007.

[12] W. M. Czarnecki, S. Podlewska and A. J. Bojarski. "Robust optimization of SVM hyperparameters in the classification of bioactive compounds". *Journal of Cheminformatics*, vol. 7, no. 38, pp. 1–15, 2015.

[13] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl and A. C. P. L. F. Carvalho. "To tune or not to tune: Recommending when to adjust SVM hyper-parameters via meta-learning". In *Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Killarney, Ireland, 2015. IEEE.

[14] W. Jiang and S. Siddiqui. "Hyper-parameter optimization for support vector machines using stochastic gradient descent and dual coordinate descent". *EURO Journal on Computational Optimization*, vol. 8, no. 1, pp. 85–101, 2020.

[15] A. A. M. Lima, F. K. H. Barros, V. H. Yoshizumi, D. H. Spatti and M. E. Dajer. "Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm". *Journal of Control, Automation and Electrical Systems*, vol. 30, pp. 371–379, 2019.

[16] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh and S. Mirjalili. "Particle Swarm Optimization: A Comprehensive Survey". *IEEE Access*, vol. 10, pp. 10031–10061, 2022.

[17] T. Xiao, D. Ren, S. Lei, J. Zhang and X. Liu. "Based on grid-search and PSO parameter optimization for Support Vector Machine". In *Proc. 11th World Congress on Intelligent Control and Automation*, pp. 1529–1533, 2014.

[18] X. Yang, T. Wei, C. Qi and P. Yuan. "Research on the Method of Station Load Prediction Based on SVR Optimized by GS-PSO". In *Proc. 11th International Conference on Power and Energy Systems*, pp. 575–579, 2021.

[19] Z. Guo, M. Liu, Y. Wang and H. Qin. "A New Fault Diagnosis Classifier for Rolling Bearing United Multi-Scale Permutation Entropy Optimize VMD and Cuckoo Search SVM". *IEEE Access*, vol. 8, pp. 153610–153629, 2020.

[20] A. R. Rocha Neto. "Máquinas de vetores-suporte: uma revisão". *Learning & Nonlinear Models*, vol. 15, no. 1, pp. 16–44, 2017.

[21] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[22] M. Hussain, S. Wajid, A. Elzaart and M. Berbar. "A comparison of SVM kernel functions for breast cancer detection". In *Proceedings of the Eighth International Conference Computer Graphics, Imaging and Visualization*, pp. 145–150, 2011.

[23] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, 2018.

[24] J. Huang, X. Hu and F. Yang. "Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker". *Measurement*, vol. 44, no. 6, pp. 1018–1027, 2011.

[25] A. Rojas-Domínguez, L. C. Padierna, J. M. C. Valadez, H. J. Puga-Soberanes and H. J. Fraire. "Optimal Hyper-Parameter Tuning of SVM Classifiers With Application to Medical Diagnosis". *IEEE Access*, vol. 6, pp. 7164–7176, 2018.

[26] X. Yang. *Nature-Inspired Optimization Algorithms*. Elsevier, London, UK, first edition, 2014.

[27] L. R. R. Santos, F. R. Durand and T. AbrÃ£o. "Adaptive PID Scheme for OCDMA Next Generation PON Based on Heuristic Swarm Optimization". *IEEE Systems Journal*, vol. 13, no. 1, pp. 500–510, 2019.

[28] A. Arafi, R. Fajr and A. Bouroumi. "Breast cancer data analysis using support vector machines and particle swarm optimization". In *Proceedings of the Second World Conference on Complex Systems*, pp. 1–6, 2014.

[29] Y. Ren and G. Bai. "Determination of optimal SVM parameters by using GA/PSO". *Journal of Computers*, vol. 5, no. 8, pp. 1160–1168, 2010.

[30] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, first edition, 1996.

[31] W. H. Wolberg, W. N. Street and O. L. Mangasarian. "UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set", 1995.

[32] D. Dua and C. Graff. "UCI Machine Learning Repository", 2017.

[33] W. H. Wolberg, W. N. Street and O. L. Mangasarian. "Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates". *Cancer Letters*, vol. 77, no. 2-3, pp. 163–171, 1994.

[34] R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification*. Wiley-Interscience, second edition, 2000.

[35] J. H. Zar. *Biostatistical Analysis*. Prentice Hall, Upper Saddle River, NJ, fifth edition, 2010.

[36] V. Kachitvichyanukul. "Comparison of three evolutionary algorithms: GA, PSO, and DE". *Industrial Engineering and Management Systems*, vol. 11, no. 3, pp. 215–223, 2012.

[37] H. Guo and A. K. Nandi. "Breast cancer diagnosis using genetic programming generated feature". *Pattern Recognition*, vol. 39, pp. 980–987, 2006.

[38] P. Kumar P, M. A. Bai V and G. G. Nair. "An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization". *Biomedical Signal Processing and Control*, vol. 68, pp. 102682, 2021.

[39] W. Yue, Z. Wang, H. Chen, A. Payne and X. Liu. "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis". *Designs*, vol. 2, no. 2, pp. 1–17, 2018.

[40] R. Sumbaly, N. Vishnusri and S. Jeyalatha. "Diagnosis of Breast Cancer using Decision Tree Data Mining Technique". *International Journal of Computer Applications*, vol. 98, no. 10, pp. 16–24, 2014.

[41] K. P. Bennett and J. A. Blue. "A support vector machine approach to decision trees". In *Proc. IEEE International Joint Conference on Neural Networks*, volume 3, pp. 2396–2401, 1998.

[42] S. A. Medjahed, T. A. Saadi and A. Benyettou. "Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules". *International Journal of Computer Applications*, vol. 62, no. 1, pp. 1–5, 2013.

[43] A. Bhardwaj and A. Tiwari. "Breast cancer diagnosis using Genetically Optimized Neural Network model". *Expert Systems with Applications*, vol. 42, no. 10, pp. 4611–4620, 2015.