

Swarm intelligence-based neural network applied in predicting breast cancer

Le Anh Uyen Vu

Ho Chi Minh University of Transport

Email: vu.le@ut.edu.vn

Abstract:

This paper investigates the effectiveness of a swarm intelligence algorithm namely the time-varying binary particle swarm algorithm in finding an optimal subset of the breast cancer dataset's features. After the feature selection phase, an artificial neural network was used for predicting the presence of malignant lesions in female breasts. Empirical results show that our approach achieved significant performance in the breast cancer prediction task: Accuracy of 0.9807 (± 0.01), Precision of 0.9766 (± 0.02), Recall of 0.9728 (± 0.02), and F1-score of 0.9742 (± 0.01).

Keywords: Evolutionary algorithms; feature selection; machine learning; disease prediction.

1. Introduction

Breast cancer (BC) prediction is specific research attracting vast attention. The earlier the radiologist successfully diagnoses BC the larger the survival chance of the patient [1][2]. Recently, the role of machine learning in any disease diagnostic program is significant such as in [3][4]. However, those applications are data-driven and require high speed and accuracy which leads to several challenges: First is the curse of dimension: The exponential increase of computational resources required to analyze data with high dimensions. The second is that they are discrete problems that are hard to solve using the convex optimization method. Machine learning-based techniques such as neural network (NN), decision tree (DT), and support vector machine (SVM) faced difficulty to find the global minimum of these problems.

In this paper, we propose a feature selection method namely Time-Varying Binary Particle Swarm Optimization (TV-BPSO) based on the wrapper approach to extract the optimal subset of the input feature set. The rest of the paper is organized as follows: Section 2 provides background information and related works, section 3 describes our proposed approach, and

section 4 shows quantitative results before presenting conclusions in section 5.

2. Literature Review

2.1. Feature Selection

Feature selection (FS) is a procedure to extract the most relevant part of the feature map while also maintaining good performance. This method, therefore, reduces the effect of the curse of dimension. FS can be divided into three groups:

- Filter method: Features were selected based on their relation to the output. The relationship can be calculated using statistics and probabilistic data [5][6].

- Wrapper method: Evaluating a group of potential subsets by training them using a machine learning classifier [7][8].

- Hybrid method: Combining both Filter and Wrapper methods that aim to provide an efficient and effective solution [9][10].

2.2. Binary Particle Swarm Optimization

The usage of SI algorithms such as bee colony optimization (BCO) [11], cuckoo search (CS) [12], firefly (FF) algorithm [13], Jaya algorithm [14], and particle swarm optimization (PSO) [15] has been applied in various fields due to its

generalization, easy implementation, and capability of finding the global minimum solution. The PSO algorithm encodes a group of candidate solutions as a biological swarm such as an ant swarm, or a flock of birds. Each particle searches for the food (optimal solution) in the search space over an iterative process. During the process, they change their position by modifying the velocity based on the capability of sharing positions between individuals. The mathematical equations for updating position and velocity are as follows:

$$x_{i,t+1}^d = x_{i,t}^d + v_{i,t+1}^d \quad (1)$$

$$V_{i,t+1}^d = wv_{i,t}^d + c_1r_{1i}(p_{i,t}^d - x_{i,t}^d) + c_2r_{2i}(p_{g,t}^d - x_{i,t}^d) \quad (2)$$

Where $x_{i,t+1}^d$ and $v_{i,t+1}^d$ are the updating position and velocity of the i^{th} individual in d^{th} dimension search space, respectively. w is inertia weight which controls the exploitation and the

exploration of the swarm, c_1 and c_2 are acceleration coefficients. r_{1i} and r_{2i} are two random variables in the range $[0, 1]$, $p_{i,t}^d$ denotes the local best position and $p_{g,t}^d$ denotes the global best position.

However, the PSO algorithm is unable to apply directly to discrete optimization problems such as feature selection. Thus, a binary version of PSO called BPSO [16] was introduced. The difference between BPSO and PSO is the position updating term. BPSO using equation (3) to update the individual position:

$$x_{i,t+1}^d = \begin{cases} 0 & \text{if } rand() \geq S(v_{i,t+1}^d) \\ 1 & \text{if } rand() < S(v_{i,t+1}^d) \end{cases} \quad (3)$$

Where $S(.)$ represents the sigmoid function to transform the velocity into probability. $rand()$ is a random number in the range $[0,1]$.

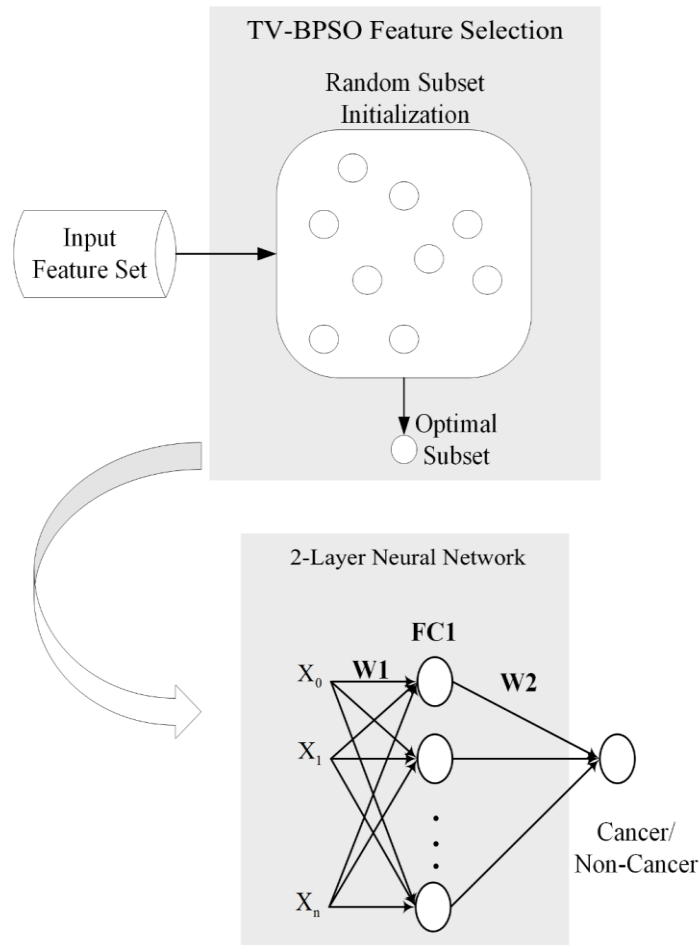


Figure 1. Proposed SI-based Neural Network.

2.3. Related Work

Various SI-based methods have been employed to feature selection. Quy et al [17] proposed an adaptive PSO-based neural network to tackle the problem of the traditional backpropagation (BP) algorithm. In [18], the authors proposed a PSO-based feature selection with multiple classifiers to reduce the computational complexity. Haghghi and Hoseini [7] used the Fruit-Fly algorithm for both feature selection and training of the multi-layer perceptron (MLP) network.

Besides, Mafarja et al [19] investigated time-varying updating strategies to hyperparameters in BPSO such as inertia weight, and acceleration coefficients such as linear decreasing, non-linear decreasing oscillating, and logarithmic decreasing. The results show significant improvements in the feature selection task. Another approach in [20] modified the sigmoid transfer function.

The authors proposed a dynamic transfer function that has high flipping probabilities at the early stage of the searching process and low flipping probabilities at the late stage. Authors in [21] also research the effectiveness of the dynamic transfer function in the SI-based feature selection method.

3. Proposed Approach

The proposed approach is demonstrated in figure (fig.) 1, it contains two main phases: the feature selection phase and the classification phase. The former used proposed TV-BPSO algorithms to select an optimal subset of input feature maps, the latter utilizing a typical two-layer neural network to predict breast cancer cases.

3.1. Time-varying Binary PSO for feature selection

Using SI-based algorithms for feature selection belongs to the wrapper method. First, the algorithm randomly initializes 40 particles, each of which represents a potential subset of the input feature set and has its position and velocity. Second, a classification algorithm is applied to form a multi-objective cost function J as follows:

$$J = \beta \times E(x_{i,t}^d) + (1 - \beta) \times \frac{\text{sum}(x_{i,t}^d)}{N} \quad (4)$$

Where β denotes is a pre-defined weight for the relative importance of error rate (i.e., $E(\cdot)$), N is the number of original feature set, $\text{sum}(x_{i,t}^d)$ is the number of selected features of the i^{th} particle. The algorithm defined the global best solution (i.e., $p_{g,t}^d$) and the local best solution (i.e., $p_{i,t}^d$) based on the value of the cost function.

Using the above multi-objective cost function helps the algorithm optimize both the classification accuracy and reduce redundant features effectively. Research on the impact of hyperparameters w [19][22] pointed out that in the early stages of searching, a large inertia weight encourages exploitation of the swarm while a small one result in strong exploration. Thus, in this paper, we proposed a linear increase inertia weight strategy demonstrated in equation (eq.) (5). Where w_{init} represents for the initialized weight, w_{end} represents the final weight at the end of searching. Table 1 lists the hyperparameter settings of the proposed TV-BPSO specifically:

$$w_t = w_{init} + \left(\frac{\text{iteration} - i \times (w_{end} - w_{init})}{w_{end} \times \text{maxiter}} \right) \quad (5)$$

Table 1. The setting of TV-BPSO.

Parameter	Notation	Values
Number of iterations	$Maxiter$	50
Swarm size	$size$	40
Initialized inertia weight	w_{init}	0.4

Parameter	Notation	Values
Final inertia weight	w_{end}	0.9
Personal factor	c_1	2
Social factor	c_2	2

Algorithm 1 TV-BPSO

Input
Initialize the swarm
Initialize hyperparameters
Begin
Define $p_{i,t}^d$ and $p_{g,t}^d$
While $t = 1$ to the maximum iteration **do**:
 For $i = 0$ to the swarm size **do**:
 Calculate $J(x_{i,t}^d)$ using eq. (4)
 If $J(x_{i,t}^d) < J(p_{i,t}^d)$ **then**
 $p_{i,t}^d \leftarrow x_{i,t}^d$
 $J(p_{i,t}^d) \leftarrow J(x_{i,t}^d)$
 end
 If $J(x_{i,t}^d) < J(p_{g,t}^d)$ **then**
 $p_{g,t}^d \leftarrow x_{i,t}^d$
 $J(p_{g,t}^d) \leftarrow J(x_{i,t}^d)$
 end
 Updating $V_{i,t+1}^d$ using eq. (2)
 Updating $x_{i,t+1}^d$ using eq. (3)
 Updating w_i using eq. (5)
 Checking terminate criterion
 end
end

3.2. Experiments

Database and Settings: The BC database in this study [23] was collected at the University of Wisconsin. Fig. 2 illustrated the class distribution in the database. Specifically, it has 569 instances of which 357 are normal and 212 are biopsy-proven cancer cases. Radiologists manually extracted the location of the cell nucleus of breast

mass, each has 10 features: Area, radius, perimeter, symmetry, concavities, fractal dimension, compactness, smoothness, and texture (variance of grey levels inside the boundary). Three statistical values of each feature namely mean, worst (mean of three largest values), and standard deviations were calculated. Thus, there are a total of 30 features for each instance.

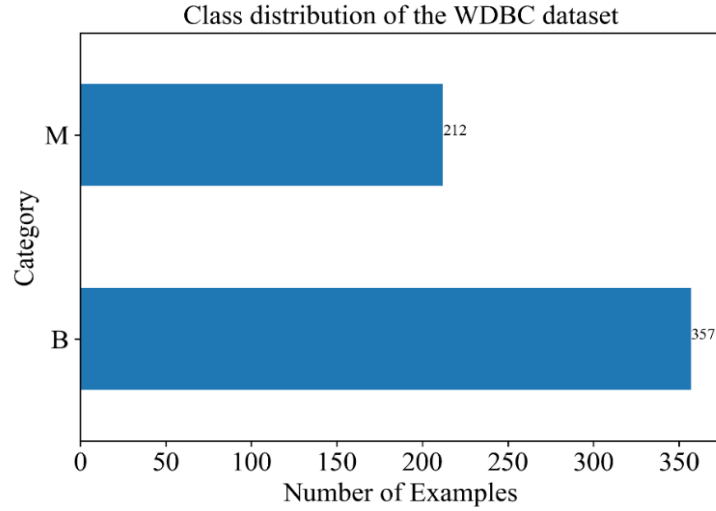


Figure 2. The distribution of two categories in the BC database. M is malignant, B is benign.

For the BC prediction task, a typical 2-layer neural network classifier was utilized to distinguish the malignant from the normal instance. As shown in fig. 2., we can see that the dataset is unbalanced which led to the risk of having bias prediction of the major category. Thus, we applied weight to the loss function in eq. (6) so that the learning algorithm focuses more on the minority category during the training phase. where w_p is the weight of the positive class, w_n is the weight of the negative class, y is the true label of output, and \hat{y} is the model output.

$$J = -[w_p y \log(\hat{y}) + w_n (1 - y) \log(1 - \hat{y})] \quad (6)$$

The performance of the classifier is evaluated by a 5-Fold cross-validation with the following metrics: an average classification, recall (true positive rate), precision, and the F1 score. They can be calculated by using eq. (7)-(10).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (10)$$

Accuracy is the ratio of correct prediction to the total cases. Precision measures the ratio of correctly predicted cancer cases to the total of cancer prediction cases, Recall is the ratio of correctly cancer predicted cases to all positive cases. Meanwhile, the F1 score measures the harmony between precision and recall.

4. Results and Discussion

Figure 3 illustrated the convergence of cost function J in two cases: training classifier with TV-BPSO-based feature selection and without feature selection. We trained each experiment for 200 iterations, with a batch size of 32, the learning rate is set as 0.001. The loss curve of the case without FS terminated after 114 iterations due to the early stopping condition had been met. On the other hand, we observed a linear decrease in the case with the TV-BPSO-based FS. This proved that our proposed method could avoid the local minimum and keep searching for the global minimum. On the other hand, table 2 shows the significant increase in performance of the proposed TV-BPSO-based neural network compared to the conventional neural network.

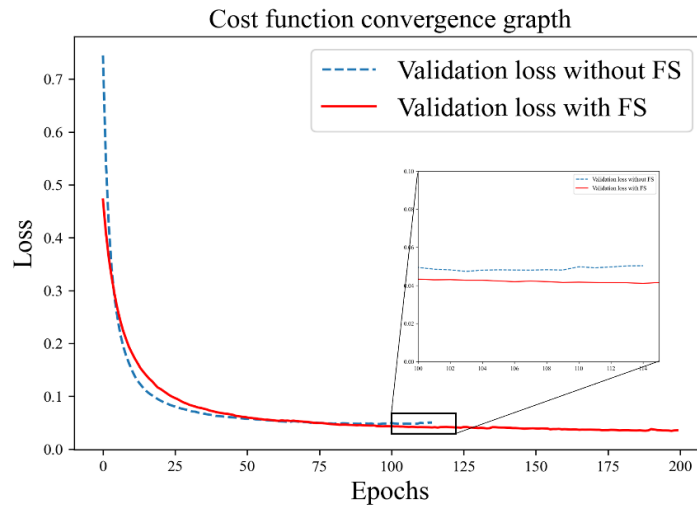


Figure 2. Loss convergence graph of with FS and without FS.

Table 2. Comparison of the results.

Model	Evaluation Metrics			
	Accuracy	Recall	Precision	F1-score
ANN without FS	0.9719 (± 0.01)	0.9638 (± 0.03)	0.9625 (± 0.04)	0.9621 (± 0.01)
TV-BPSO ANN	0.9807 (± 0.01)	0.9766 (± 0.02)	0.9728 (± 0.02)	0.9742 (± 0.01)

5. Conclusion

In this paper, we proposed an SI-based feature selection method namely TV-BPSO. We evaluated the performance of the NN classifier using our method in the breast prediction task. The results show that our model surpassed the performance of the NN classifier alone. In the future, we intend to analyze our method with other advanced feature selection algorithms. Another direction is to apply our method to the deep learning area where the data is very high dimensional.

Acknowledgment

We acknowledge the support of time and facilities from Ho Chi Minh City University of Transport (UTH) for this study.

References

- [1] M. Milosevic, D. Jankovic, A. Milenkovic and D. Stojanov; "Early diagnosis and detection of breast cancer". *Technology and health care*. 2018; 26(4):729–759. DOI:10.3233/THC-181277.
- [2] O. Ginsburg et al.; "Breast cancer early detection: A phased approach to implementation". *Cancer*. 2020; 126 (10):2379–2393. DOI: 10.1002/cncr.32887.
- [3] M. Tanveer et al.; "Machine learning techniques for the diagnosis of Alzheimer’s disease: A review". *ACM Transactions on Multimedia Computing, Communications, and Applications*. 2020; 16(1s):1–35. DOI: 10.1145/3344998.
- [4] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor; "Heart disease identification method using machine learning classification in e-healthcare". *IEEE Access*. 2020; 8:107562–107582. DOI: 10.1109/ACCESS.2020.3001149.
- [5] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili; "A novel multivariate filter method for feature selection in text classification problems". *Engineering Applications of Artificial Intelligence*. 2018; 70:25–37. DOI: 10.1016/j.engappai.2017.12.014.
- [6] M. Rostami, S. Forouzandeh, K. Berahmand, and M. Soltani; "Integration of multi-objective PSO based feature selection and node centrality for medical datasets". *Genomics*. 2020;

- 112(6):4370–4384. DOI: 10.1016/j.ygeno.2020.07.027.
- [7] M. S. Haghighi and M. J. M. Hoseini; “Improving machine learning accuracy in diagnosing diseases using feature selection based on the fruit-fly algorithm”; in Proc. 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS); 23-24 December 2020; Shahrood University of Technology, Semnan, Iran. IEEE. 2020; pp. 1–6. DOI: 10.1109/ICSPIS51611.2020.9349593.
- [8] S. Sayed, M. Nassef, A. Badr, and I. Farag; “A nested genetic algorithm for feature selection in high-dimensional cancer microarray datasets”. *Expert Systems with Applications*. 2019; 121:233–243. DOI: 10.1016/j.eswa.2018.12.022.
- [9] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elsoud; “A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier”. *Knowledge-based systems*. 2020; 205:106270. DOI: 10.1016/j.knosys.2020.106270.
- [10] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Z. Gao; “A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data”. *IEEE Transactions on Cybernetics*. 2021; pp.1-14. DOI: 10.1109/TCYB.2021.3061152.
- [11] D. Teodorovic, P. Lucic, G. Markovic, and M. Dell’Orco; “Bee colony optimization: principles and applications”; in Proc. 2006 8th Seminar on Neural Network Applications in Electrical Engineering; 25-27 September 2006; Belgrade, Serbia. IEEE. 2006; pp. 151–156. DOI: 10.1109/NEUREL.2006.341200.
- [12] X.-S. Yang and S. Deb; “Cuckoo search via Lévy flights”; in Proc. 2009 World congress on nature & biologically inspired computing (NaBIC); 09-11 December 2009; Coimbatore, India. 2009; pp. 210–214. DOI: 10.1109/NABIC.2009.5393690.
- [13] X.-S. Yang and A. Slowik; “Firefly algorithm”; *Swarm Intelligence Algorithms*. Florida, USA: CRC Press. 2020; pp. 163–174.
- [14] R. Rao; “Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems”. *International Journal of Industrial Engineering Computations*. 2016; 7(1):19-34. DOI: 10.5267/j.ijiec.2015.8.004.
- [15] J. Kennedy and R. Eberhart; “Particle swarm optimization”; in Proc. ICNN’95-International Conference on Neural Networks; 27 November - 01 December 1995; Perth, WA, Australia. IEEE. 1995; pp. 1942–1948. DOI: 10.1109/ICNN.1995.488968.
- [16] J. Kennedy and R. C. Eberhart; “A discrete binary version of the particle swarm algorithm”; in Proc. 1997 IEEE International conference on systems, man, and cybernetics. *Computational cybernetics and simulation*; 12-15 October 1997; Orlando, FL, USA. IEEE. 1997; pp. 4104–4108. DOI: 10.1109/ICSMC.1997.637339.
- [17] H. D. Quy, C. Van Kien, H. P. H. Anh, and N. N. Son; “Machine Learning-Based Evolutionary Neural Network Approach Applied in Breast Cancer Tumor Classification”; in Proc. International Conference on Green Technology and Sustainable Development; 27-28 November 2020; Da Nang City, Vietnam. Cham, Denmark: Springer. 2020; pp. 72–83. DOI: 10.1007/978-3-030-62324-1_7.
- [18] E.-S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein; “Novel Feature Selection and Voting Classifier Algorithms for COVID-19 Classification in CT Images”. *IEEE Access*. 2020; 8:179317–179335. DOI:10.1109/ACCESS.2020.3028012.
- [19] M. Mafarja, R. Jarrar, S. Ahmad, and A. A. Abusnaina; “Feature selection using binary particle swarm optimization with time-varying inertia weight strategies”; in Proc. The 2nd International Conference on Future Networks and Distributed Systems; 26 – 27 June 2018; Amman, Jordan. NY, USA: Association for Computing Machinery. 2018; pp. 1–9. DOI:10.1145/3231053.3231071.
- [20] M. J. Islam, X. Li, and Y. Mei; “A time-varying transfer function for balancing the exploration and exploitation ability of a binary PSO”. *Applied Soft Computing*. 2017; 59:182–196. DOI:10.1016/j.asoc.2017.04.050.
- [21] M. Mafarja et al.; “Binary dragonfly optimization for feature selection using time-varying transfer functions”. *Knowledge-Based*

- Systems. 2018; 161:185–204. DOI:10.1016/j.knosys.2018.08.003.
- [22] A. Jeyasothy, S. Ramasamy, and S. Sundaram; “Meta-neuron learning-based spiking neural classifier with time-varying weight model for credit scoring problem”. *Expert Systems with Applications*. 2021; vol. 178. DOI: 10.1016/j.eswa.2021.114985.
- [23] W. H. Wolberg, W. N. Street, D. M. Heisey, and O. L. Mangasarian; “Computer-derived nuclear features distinguish malignant from benign breastcytology”. *Human pathology*. 1995; 26(7):792–796. DOI:10.1016/0046-8177(95)90229-5.