



ISSN: 0067-2904

## Smart Doctor: Performance of Supervised ART-I Artificial Neural Network for Breast Cancer Diagnoses

Kamal R. AL-Rawi<sup>1\*</sup>, Saifaldeen K. AL-Rawi<sup>2</sup>

<sup>1</sup>Computer Science Department, Faculty of Computer Science and Informatics, Amman Arab University, Amman, Jordan

<sup>2</sup>Department of Computer Engineering, School Electrical Engineering and Information Technology, German Jordanian University, Amman, Jordan

Received: 9/9/2019

Accepted: 21/1/2020

### Abstract

Wisconsin Breast Cancer Dataset (WBCD) was employed to show the performance of the Adaptive Resonance Theory (ART), specifically the supervised ART-I Artificial Neural Network (ANN), to build a breast cancer diagnosis smart system. It was fed with different learning parameters and sets. The best result was achieved when the model was trained with 50% of the data and tested with the remaining 50%. Classification accuracy was compared to other artificial intelligence algorithms, which included fuzzy classifier, MLP-ANN, and SVM. We achieved the highest accuracy with such low learning/testing ratio.

**Keywords:** Adaptive Resonance Theory, Artificial Neural Network, Breast Cancer Diagnoses, Machine Learning.

## الطبيب الذكي: أداء الشبكة العصبونية الإصطناعية "نظرية الرنين التكيفي الموجهة" في تشخيص سرطان الثدي

كمال الراوي<sup>1\*</sup>, سيف الدين الراوي<sup>2</sup>

<sup>1</sup>قسم علم الحاسوب، كلية العلوم الحاسوبية و المعلوماتية، جامعة عمان العربية، عمان، الأردن

<sup>2</sup>قسم هندسة الحاسوب، كلية هندسة الكهرباء و تكنولوجيا المعلومات، الجامعة الألمانية الأردنية، عمان، الأردن

### الخلاصة

تم توظيف بيانات ويسكونسن لسرطان الثدي لإظهار أداء الشبكة العصبونية الإصطناعية "نظرية الرنين التكيفي الموجهة" لبناء نظام ذكي لتشخيص سرطان الثدي. تم تعليم النظام بقيم متغيرات و مجاميع تعليم مختلفة. أفضل نتيجة حصلت عند تعليم النظام باستعمال 50% من البيانات و اختباره باستخدام البيانات الـ 50% المتبقية. تمت مقارنة دقة التشخيص مع خوارزميات الذكاء الاصطناعي الأخرى (fuzzy classifier, MLP-ANN, SVM). حصلنا على أعلى دقة عند هذه النسبة المنخفضة للتعليم/الفحص.

### 1. Introduction

Breast cancer is one of the most common types of cancers among women nowadays. It manifests in the form of a non-uniform tumour, but not all tumours are cancerous. Non-cancerous tumours are called benign while cancerous tumours are called malignant. Diagnosing tumours in the breast tissue as benign or malignant as early as possible is very important as it highly increases the chance of

\*Email: k.rawi@aau.edu.jo

surviving. A general physician will have to conduct the diagnoses when a specialist is not available. Hence, it is better to employ an artificial intelligent system, which is learned with the knowledge of experts, for the aid in diagnoses.

The main objective of this paper is to investigate the performance of the ART-ANN in the whole space of the learning parameters. To be specific, we will measure the performance of Supervised ART-I. We also aim at building an artificial intelligence system for Breast Cancer Diagnoses (BCD) based on the original WBC dataset.

The rest of this paper is organized as follows: section two provides a background on the classification of breast cancer. The description of the WBCD dataset is shown in section three. The Supervised ART-I algorithm is listed in section four. The performance of the supervised ART-I is shown in section five. The conclusions and discussion are presented in section six.

## 2. Background

Much research has been conducted on the diagnoses of WBCD [1] using different artificial intelligence algorithms. The Fuzzy classifier was used with 95.06%, 97.51%, 98.57%, and 98.2% accuracy [2-5]. Moreover, the Artificial Immune Recognition System (AIRS) classifier was implemented [6, 7] with 97.2% and 98.51% accuracy. Nearest Neighbour classifier was employed by Sahan *et al.* [8] with 99.14% and Sáez *et al.* [9] with 96.04% accuracy. Furthermore, an accuracy of 97.13% was achieved by Kumar *et al.* [10] using a voting classifier utilizing Naïve Bayes, Support Vector Machine (SVM), and J48 Decision Tree. In another work [11], the fusion of J48 and Multi Layered Perceptron Artificial Neural Network (MLP-ANN) with Principle Component Analysis (PCA) showed an accuracy of 97.57%. In addition, SVM was implemented with different approaches [12-17]. The best reported accuracy was 99.74% with 10-fold cross validation [15]. ANN was implemented in other studies [18-22]. The best accuracy was 100%, with 80% of the data used for learning and the remaining 20% for testing [20]. Furthermore, a spiking neural network approach was utilized, which achieved an accuracy of 99.26% [23]. In another paper, the XCS learning classifier system (LCS) was implemented, which achieved an average accuracy of 96.4% [24].

## 3. Description of WBCD

The WBC dataset contains 699 instances with 9 features and a class label (benign or malignant). Furthermore, the score of each feature is an integer value between 1 and 10. These features, according to their order in the dataset, are clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses. These features were driven from breast tissue using Fine Needle Aspirates (FNA) from women with breast tumours in order to diagnose the tumour as benign or malignant [1].

The dataset contains 16 instances with missing data. The remaining 683 instances are the scope of this paper. Specifically, it contains 444 benign and 239 malignant instances [1].

## 4. Supervised ART-I ANN

Any ANN should go through two phases, the learning phase and the testing phase. For this, the data is divided into two parts. The first part is used for learning while the second part is used for testing. At the end of the learning phase the weights that connect the input nodes with the category nodes must be determined. Then these weights are used during the testing phase in order to classify the second part of the data. The features and the class code, in a supervised form, are introduced to the supervised ANN during the learning phase. However, only the features are introduced to the ANN during the testing phase. It is the ANN's task to assign the class.

For all ART ANNs, the features are introduced to the ART-ANN in a normalized form between zero and one. Such approach has two advantages. First, the initial values for the weights are set to one. Second, the complements ( $1 - \text{normalized feature value}$ ) can be introduced together with their normalized features as well. ART ANNs has two parameters that need to be optimized for a specific problem, which are the vigilance parameter  $\rho$  and the learning parameter  $\beta$ .

There are many supervision approaches for ART-ANN, including the Mapfield approach as in ARTMAP [25] and Fuzzy ARTMAP [26], the Tagging approach as in Supervised ART-I [27], and the Bagging approach as in Supervised ART-II [28]. All these supervisions have the same classification accuracy from the theoretical aspect. However, Tagging and Bagging approaches are better than Mapfield approach from a memory requirement and execution time points of view. In addition to that, their architectures are simpler.

ART ANNs do not fall in local minima, as other approaches do, because they always converge. The weights are strictly decreasing during the learning phase. This can be clearly seen from the weight learning formula:

$$w_{i,j}^{new} = (1 - \beta) w_{i,j}^{old} + \beta \cdot \min(a_i, w_{i,j}^{old}) \quad (1)$$

Where  $w_{i,j}$  is the weight between the input node  $i$  and the winning category node  $J$ , and  $a_i$  is the  $i$ th feature in the feature vector. The weight decreases when  $a_i < w_{i,j}^{old}$ , otherwise the weight does not change.

From the computational aspect, the Supervised ART-I requires less learning and testing time than the Supervised ART-II when there are less than 1000 Committed Category Nodes (CCNs) during learning [29]. The number of CCNs is bounded by the learning size. However, it is normally much less than that depending on the vigilance parameter  $\rho = (0, 1]$  and learning parameter  $\beta = (0, 1]$ . Since the total number of WBCD instances is less than 1000, the Supervised ART-I will be employed for this task. Its architecture is shown in Figure- 1. The left array of the nodes represents the input features  $a_i$ ;  $i = 1, \dots, 2M$ , of a single instance with the class code "b". The right array of nodes represents the category nodes  $N$ . Only CCNs are assigned tags and weights that connect them with the input node [27].

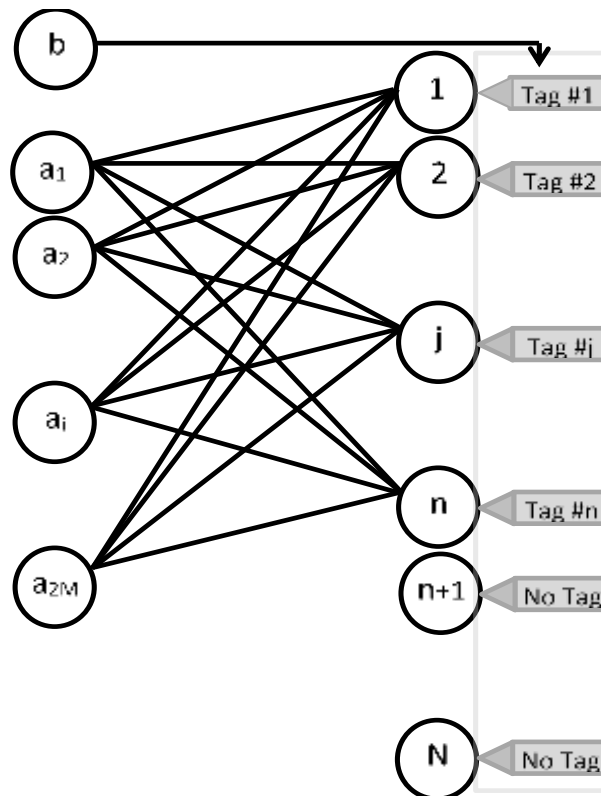


Figure 1- Architecture of the Supervised ART-I ANN

#### 4.1. Supervised ART-I Learning Algorithm

The learning algorithm of the Supervised ART-I is as follows:

1. Normalize the features for learning set to  $[0, 1]$ .
2. Input the size of the learning dataset, the number of features  $M$ , the initial vigilance parameter  $\rho_0$ , and the learning parameter  $\beta$ .
3. Take the first normalized instance  $a_i$ ;  $i = \{1, \dots, M\}$ , the complement of  $a_i$ ;  $i = \{M+1, \dots, 2M\}$ , and the class code "b" to create the first committed category node (CCN),  $n=1$  ( $n$  corresponds to the current CCN), then do the following:

a- Compute the weights that connect each input node with the CCN ( $n=1$ ):

For normal learning mode use Eq. 2:

$$w_{i,1} = (1 - \beta) + \beta \cdot a_i; i = 1, \dots, 2M \quad (2)$$

For fast learning mode use Eq. 3:

$$w_{i,1} = a_i; i = 1, \dots, 2M \quad (3)$$

b- Assign the class code (b) of the current input features to the new CCN ( $n=1$ ),  $\text{Tag}(1) = b$ .

4. For the next input of the learning dataset, introduce the normalized instance and its complement  $a_i; i = \{1, \dots, 2M\}$ , and the class code "b".

Reset the value of the vigilance parameter  $\rho$  to its initial value,  $\rho = \rho_0$ .

5. Compute the score  $T_j$ , where  $j$  corresponds to a CCN, using Eq. 4 for each CCN:

$$T_j = \frac{\sum_{i=1}^{2M} \min(a_i, w_{i,j})}{\sum_{i=1}^{2M} w_{i,j}}; j = 1, \dots, n \quad (4)$$

6. Find the max score  $T_j$  among all CNN scores using Eq. 5:

$$T_j = \max(T_j, j = 1, \dots, n) \quad (5)$$

7. If  $T_j = -1$  then a new category node must be committed and  $n$  is incremented. The weights of the new nodes are set according to the current input feature vector as follows:

For normal learning mode use Eq. 6:

$$w_{i,n} = (1 - \beta) + \beta \cdot a_i; i = 1, \dots, 2M \quad (6)$$

While for fast learning mode use Eq. 7:

$$w_{i,n} = a_i; i = 1, \dots, 2M \quad (7)$$

go to step 11.

8. Test the winning CCN "J" against the vigilance parameter using Eq. 8:

$$\text{If } \frac{\sum_{i=1}^{2M} \min(a_i, w_{i,j})}{M} < \rho \quad (8)$$

Put the CCN "J" in shut off mode  $T_j = -1$ .

go to step 6.

9. if  $\text{Tag}(J) \neq b$  then put CCN "J" in shut off mode,  $T_j = -1$ , and assign the new value for  $\rho$  using Eq. 9:

$$\rho = \frac{\sum_{i=1}^{2M} \min(a_i, w_{i,j})}{M} \quad (9)$$

go to step 6.

10. Learn the winning CCN "J" using Eq. 1:

$$w_{i,j}^{\text{new}} = (1 - \beta) w_{i,j}^{\text{old}} + \beta \cdot \min(a_i, w_{i,j}^{\text{old}}) \quad (1)$$

11. If there are more input features go to step 4.

12. Save all CCNs with their Tags and weights.

13. End of the learning phase.

#### 4.2. Supervised ART-I Testing Algorithm

The testing algorithm of the Supervised ART-I is as follows:

1. Normalize the features of the testing set to  $[0, 1]$ .

2. Input the size of the testing dataset, the number of features  $M$ , and the total number of CCNs with their tags and weights that are obtained from the learning phase.

3. For the next feature vector in the testing dataset, introduce the features with their complements to the learned ANN.

4. Compute the score for each CCN using Eq. 4:

$$T_j = \frac{\sum_{i=1}^{2M} \min(a_i, w_{i,j})}{\sum_{i=1}^{2M} w_{i,j}}; j = 1, \dots, n \quad (4)$$

5. Find the max score using Eq. 5:

$$T_j = \max(T_j, j = 1, \dots, n) \quad (5)$$

6. Assign  $\text{Tag}(J)$  as the class of the current features,  $b = \text{Tag}(J)$ .

7. If there is more data to be classified go to step 3.

8. End of the testing phase.

#### 5. Performance of the Supervised ART-I

The examination of the performance of the Supervised ART-I ANN is presented using the well-known WBC dataset benchmark. Specifically, the aim of the classification is to distinguish between benign and malignant patients according to the nine input features of WBC. We introduce  $a_1, a_2, \dots, a_9$  normalized to  $[0, 1]$  together with their complements  $1-a_1, 1-a_2, \dots, 1-a_9$  and class code (benign=1, malignant=2) to the Supervised ART-I ANN during the learning phase.

The performance is measured with different vigilance  $\rho$  and learning  $\beta$  parameters. Furthermore, the size of the learning set is 341, which represents about 50% of the total WBC dataset, while the remaining data is used for testing.

Learning is performed using every combination of the values of the vigilance parameter  $\rho = [0.01, 1]$  and the learning parameter  $\beta = [0.01, 1]$  with a step size of 0.01. Furthermore, the performance of both normal and fast modes is investigated. This represents a total of 20,000 different runs, which aims to show the performance of the Supervised ART-I ANN across the whole space of the  $\rho$  and  $\beta$  parameters.

The classification performance for learning in normal mode with any combination of  $\rho$  and  $\beta$  is more than 98%. Furthermore, the number of CCNs is shown in Figure-2 as a contour plot across the entire space of  $\rho$  and  $\beta$ . We achieved 99.71% accuracy with  $\rho = 0.82$  and  $\beta = 0.85$ . Furthermore, the accuracy for the malignant class is 100%, while the accuracy for the benign class is 99.62%, with only one miss-classified feature vector of the 262 benign feature vectors, as shown in Table- 1. Moreover, the learning time for this run was 17 ms, while the testing time was 10 ms.

**Table 1-**The system performance when learning with 341 instances (183 benign and 158 malignant) and testing the remaining 342 instances (261 benign and 81 malignant)

Learning mode		Normal	Fast
Learning parameters with max accuracy		$\rho=0.82 \beta=0.85$	$\rho=0.65 \beta=0.51$
Benign	Accuracy%	<b>99.62</b>	99.23
	Missed	1	2
	CCNs	15	6
Malignant	Accuracy%	<b>100.00</b>	100.00
	Missed	0	0
	CCNs	61	27
Overall	Accuracy%	<b>99.71</b>	99.42
	Missed	1	2
	CCNs	76	33

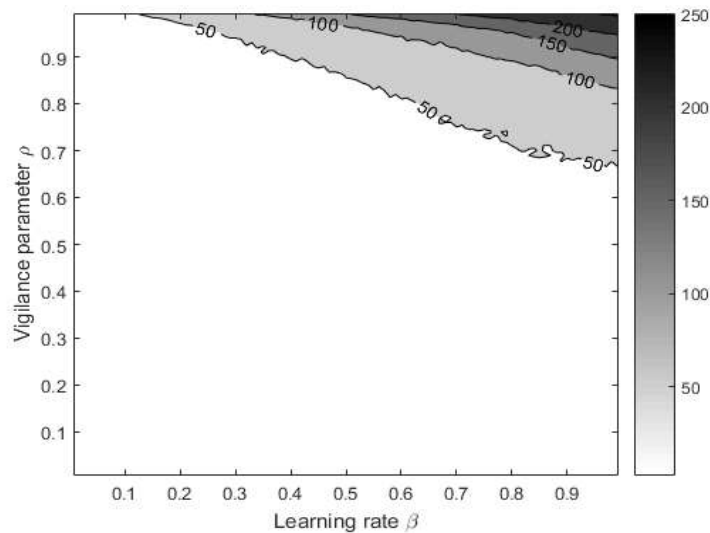
The classification performance for the fast learning mode is slightly less than that for the normal mode in terms of accuracy achieved. Specifically, it is more than 97% for the whole space. Moreover, the total number of CCNs is shown in Figure- 3. It can be observed in the normal mode that the number of CCN is higher when the values of  $\beta$  or  $\rho$  are high. The best overall accuracy in the fast mode is 99.42% at  $\rho = 0.65$  and  $\beta = 0.51$  with 33 CCNs. Specifically, the accuracy for the benign and malignant classes is 99.23% and 100%, respectively, as shown in Table-1. Furthermore, the learning time for this run was 0.011 sec, while the testing time was 0.006 sec.

Learning with just 25 instances, an accuracy of more than 90% is achieved with any combination of  $\rho$  and  $\beta$ . In particular, the highest accuracy for normal mode is 96.50%. Furthermore, learning in the normal mode with five epochs increased the overall accuracy to 96.81%. However, learning in the fast mode increased the accuracy to 97.26%, as shown in Table- 2. Moreover, the resulting network consists of only three category nodes. Figure- 4 shows the complete Supervised ART-I ANN with the values of the weights for each of the three category nodes and the 18 input nodes

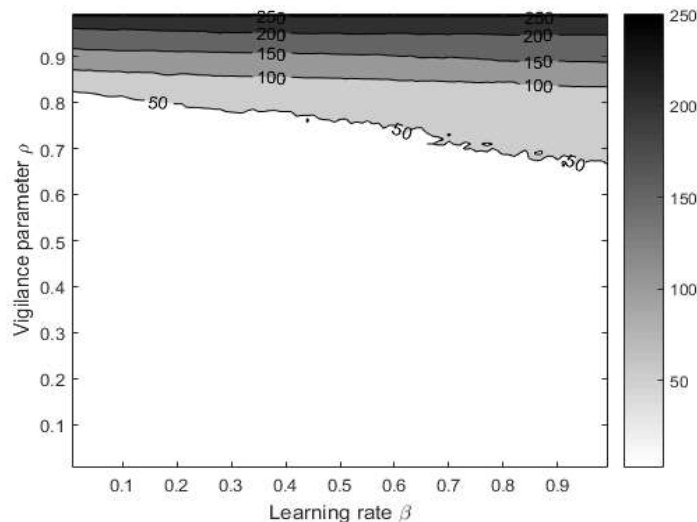
**Table 2-**The system performance when learning with 25 instances (17 benign and 8 malignant) and testing the remaining 658 instances (427 benign and 231 malignant)

Learning mode		Normal	Normal 5 epochs	Fast
Learning parameters with max accuracy		$\rho=0.72 \beta=0.61$	$\rho=0.50 \beta=0.45$	$\rho=0.01 \beta=0.06$
Benign	Accuracy%	97.89	97.66	97.42
	Missed	9	10	11

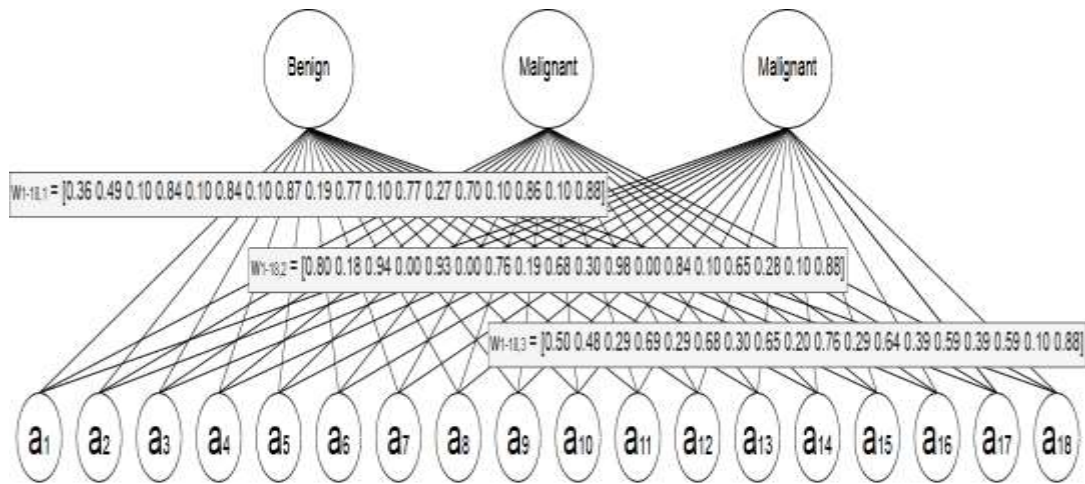
	CCNs	2	2	1
Malignant	Accuracy%	93.94	95.24	96.97
	Missed	14	11	7
	CCNs	2	3	2
Overall	Accuracy%	96.50	96.81	97.26
	Missed	23	21	2
	CCNs	4	5	3



**Figure 2-** Number of CCNs for the whole domain of the vigilance parameter  $\rho$  and learning parameter  $\beta$ . Learning with 341 instances in normal mode. Number of CCN is increased as  $\rho$  increases because matching criteria is larger.



**Figure 3-** Number of CCNs for the whole domain of the vigilance parameter  $\rho$  and learning parameter  $\beta$ . Learning with 341 instances in fast mode. Number of CCN is increased as  $\rho$  increases



**Figure 4-** The complete Supervised ART-I artificial neural network with the weight values trained with 25 input features

**Table 3-** The performance of different data classification techniques for classifying the original WBCD. Accuracies  $\geq 99.00\%$  are shown in bold

Author	classification technique	Acc %	Learning/testing
[2]	NEFCLASS	95.06	10-fold cross validation
[3]	fuzzy-genetic	97.36 97.51	50% / 50% 75% / 25%
[4]	Supervised Gath–Geva clustering algorithm	98.57	10-fold cross validation
[5]	Fuzzy Rules	98.2 96.5	7.3% / 92.7% 49.8% / 50.2%
[6]	AIRS	97.2	10-fold cross validation
[7]	Fuzzy AIRS	98.51	10-fold cross validation
[8]	<b>Fuzzy-AIS-kNN</b>	<b>99.14</b>	<b>10-fold cross validation</b>
[9]	FW-KNNI	96.04	10-fold cross validation
[10]	Voting classifier (Naïve Bayes + SVM + J48)	97.13	10-fold cross validation
[11]	J48 and MLP with PCA	97.57	10-fold cross validation
[12]	Least square SVM	98.53 97.08 96.59 95.89	10-fold cross validation 80% / 20% 70% / 30% 50% / 50%
[13]	<b>SVM + GA with feature chromosome</b>	<b>99.00</b>	<b>10-fold cross validation</b>
[14]	SVM + Hill Climber + Pedagogical	97.16	67% / 33%
[15]	<b>PTVPSO-SVM</b>	<b>99.74</b>	<b>10-fold cross validation</b>
[16]	SVM MLP BPN	96.85 95.71	10-fold cross validation
[17]	Logistic Regression C 4.5 Random Forest Bayes Net ANN (MLP) Radial Basis Function Networks SVM Rotation Forest	96.78 96.05 96.34 97.22 96.05 95.75 96.78 96.78	10-fold cross validation
[18]	MLP-NN	97.4	3-fold cross validation
[19]	<b>Artificial Metaplasticity Multilayer Perceptron</b>	<b>99.26</b>	<b>60% / 40%</b>
[20]	<b>RS + ELM</b>	<b>100</b>	<b>80% / 20%</b>

[21]	RS-BPNN	98.60	80% / 20%
[22]	Deep Belief Networks	99.68	54.9% / 45.1%
[23]	Spike Response Model with Spiking neural Network	99.26	60%/40%
This work	Supervised ART-I	99.71	49.9% / 50.1%

## 6. Discussion and Conclusions

All previous works trained their systems with more than or equal to 50% of the data. In particular, only two works [3, 12] trained their systems with 50% of the data. In a previous study [3] an accuracy of 97.36% was achieved using fuzzy-generic classifier, while another work [12] achieved 95.89% accuracy using least square (LS) SVM classifier. A better accuracy was achieved in this work with the same learning size. However, two works achieved better accuracy [15, 20] but with a learning size of more than 50%. In particular, an accuracy of 99.74% was achieved [15] with 10-fold cross validation using parallel time variant particle swarm optimization (PTVPSO) for parameter optimization and feature selection for SVM classifier. While an accuracy of 100% with 80% learning size was achieved using Rough Set (RS) and Extreme Learning Machine (ELM) classifier [20]. A comparison of this work with previous ones is shown in Table- 3.

It is important to note how fast the Supervised ART-I can execute the learning and testing phases. In particular, the learning and testing time is not longer than 200 ms for any single run. For this, it was able to perform the learning and testing for all the 20,000 different combinations of the learning parameters  $\rho$  and  $\beta$ .

The classification accuracy for the majority of the runs for the whole space of  $\rho$  and  $\beta$ , and different learning size, was higher than 90%, which indicates the high performance of the Supervised ART-I classifier. Furthermore, the best accuracy was 99.71% at  $\rho=0.82$  and  $\beta=0.85$  learning in the normal mode. In this run, only one benign instance was miss-classified as malignant by the system while correctly classifying all the other instances. Moreover, the time for the best accuracy run is 17 ms for the learning phase and 10 ms for the testing phase.

It can be seen that as  $\rho$  increases, more CCNs are generated. This happens because the matching criteria for the CCNs are higher. Moreover, CCNs is proportional with  $\beta$  as well. Furthermore, the Supervised ART-I can perform very well, even with a very small learning size, achieving 96.50% accuracy in the normal mode. The performance is improved further using the fast learning mode, achieving 97.26% accuracy. The increase in accuracy for the fast mode, learning with such small size, is due to the assignment of the feature values to the weights when a node is committed. However, for the normal mode, the weight value is between the features values and 1 when a node is committed. When  $\beta$  is higher, the weights are closer to the values of the input features. Thus, there are not enough training instances to decrement the weights to their optimal values.

## References

1. W. Wolberg and O. Mangasarian, O. 1990. "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", *Proceedings of the National Academy of Sciences*, **87**(23): 9193-9196.
2. Nauck, D. and Kruse, R. 1999. "Obtaining interpretable fuzzy classification rules from medical data", *Artificial Intelligence in Medicine*, **16**(2): 149-169.
3. Peña-Reyes, C. and Sipper, M. 1999. "A fuzzy-genetic approach to breast cancer diagnosis", *Artificial Intelligence in Medicine*, **17**(2): 131-155.
4. Abonyi, J. and Szeifert, F. 2003. "Supervised fuzzy clustering for the identification of fuzzy classifiers", *Pattern Recognition Letters*, **24**(14): 2195-2207.
5. Gadaras and L. Mikhailov, L. 2009. "An interpretable fuzzy rule-based classification methodology for medical diagnosis", *Artificial Intelligence in Medicine*, **47**(1): 25-41, 2009.
6. Goodman, D.E., Boggess, L. and Watkins, A. 2002. "Artificial immune system classification of multiple-class problems", *Proceedings of the artificial neural networks in engineering*, **2**: 179-183.



7. Polat, K., Şahan, S., Kodaz, H. and Güneş, S. **2007**. "Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism", *Expert Systems with Applications*, **32**(1): 172-183.
8. Şahan, S., Polat, K., Kodaz, H. and S. Güneş, S. **2007**. "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis", *Computers in Biology and Medicine*, **37**(3): 415-423.
9. Sáez, J., Derrac, J., Luengo, J. and Herrera, F. **2014**. "Statistical computation of feature weighting schemes through data estimation for nearest neighbor classifiers", *Pattern Recognition*, **47**(12): 3941-3948, 2014.
10. Kumar, U.K., Nikhil, M.B.S. and Sumangali, K. **2017**. "Prediction of breast cancer using voting classifier technique," *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Chennai, pp. 108-114.
11. Salama, G.I., Abdelhalim, M.B. and Zeid, M.A. **2012**. "Breast cancer diagnosis on three different datasets using multi-classifiers", *International Journal of Computer and Information Technology*, **1**(1): 36–43.
12. Polat, K. and Güneş, S. **2007**. "Breast cancer diagnosis using least square support vector machine", *Digital Signal Processing*, **17**(4): 694-701.
13. Zhao, M., Fu, C., Ji, L., Tang, K. and Zhou, M. **2011**. "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes", *Expert Systems with Applications*, **38**(5): 5197-5204, 2011.
14. Stoean, R. and Stoean, C. **2013**. "Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection", *Expert Systems with Applications*, **40**(7): 2677-2686.
15. Chen, H., Yang, B., Wang, S., Wang, G., Liu, D., Li, H. and Liu, W. **2014**. "Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy", *Applied Mathematics and Computation*, **239**: 180-197.
16. Ghosh, S., Mondal, S. and Ghosh, B. **2014**. "A comparative study of breast cancer detection based on SVM and MLP BPN classifier," In *IEEE First International Conference on Automation, Control, Energy and Systems*, pp. 1-4.
17. Aličković, E. and Subasi, A. **2017**. "Breast cancer diagnosis using GA feature selection and Rotation Forest", *Neural Computing and Applications*, **28**(4): 753-763.
18. Karabatak, M. and Ince, M. **2009**. "An expert system for detection of breast cancer based on association rules and neural network", *Expert Systems with Applications*, **36**(2): 3465-3469.
- A. Marcano-Cedeño, J., Quintanilla-Domínguez and Andina, D. **2011**. "WBCD breast cancer database classification applying artificial metaplasticity neural network", *Expert Systems with Applications*, **38**(8): 9573-9579, 2011.
19. Kaya, Y. **2013**. "A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: RS + ELM", *Turkish Journal of Electrical Engineering & Computer Sciences*, **21**: 2079-2091.
20. Nahato, K.B., Harichandran, K.N. and Arputharaj, K. **2015**. "Knowledge mining from clinical datasets using rough sets and backpropagation neural network", *Computational and mathematical methods in medicine*, vol. 2015.
21. Abdel-Zaher and Eldeib, A. **2016**. "Breast cancer classification using deep belief networks", *Expert Systems with Applications*, **46**: 139-144.
22. Ourdighi and A. Benyettou. **2016**. "An efficient spiking neural network approach based on spike response model for breast cancer diagnostic." *International Arab Journal of Information Technology*, **13**(6B): 1032-1038, 2016.

23. Kharbat, F., Odeh, M. and Bull, L. **2014**. "A new hybrid architecture for the discovery and compaction of knowledge from breast cancer datasets." *The International Arab Journal of Information Technology*, **11**(3): 116-123.
24. Carpenter, G., Grossberg, S. and Reynolds, J. **1991**. "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network", *Neural Networks*, **4**(5): 565-588.
25. Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H. and Rosen, D.B. **1992**. "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," in *IEEE Transactions on Neural Networks*, **3**(5): 698-713, Sep .
26. AL-Rawi, K.R. **1999**. "*Supervised ART-I: A new neural network architecture for learning and classifying multi-valued input patterns*", Lecture Notes in Computer Science, 1606, pp. 756-765.
27. AL-Rawi, K., Gonzalo, C. and Arquero, A. **1999**. "Supervised ART-II: A new neural network architecture, with quicker learning algorithm, for classifying multi-valued input patterns," in *Proceeding of the European Symposium on Artificial Neural Network*, Bruges, Belgium, pp. 289-294.
28. AL-Rawi, K.R. **2001**. "Design new supervised ART-type artificial neural networks, and their performance for classification Landsat TM images", Ph.D. Thesis, Faculty of Computer Science, University Polytechnic of Madrid, Spain, 2001. <http://oa.upm.es/42680/>