**Sanford
Inter Science Press**

Artificial Immune Systems for Identifying Malignant Breast Cancer

Ramin Javadzadeh[*]

Department of Computer, Bojnourd Branch, Islamic Azad University, Bojnourd, Iran

Abstract— Cancer is a gene disruption based disease in which, body's cells behave abnormally. These cancerous cells start uncontrolled reproduction and usually offend organisms, and a mass of cancerous cells is called tumor. Epidemiological studies have revealed that breast cancer is the leading cause of mortality among female population. Owing to the essence of natural lipid masses in breast, identification of tumor type from mammographic images is challenging. In this sense, almost patients are supposed to have biopsy operation which might lead wide variety of side effects. On this basis, a novel expert system based on negative selection and fuzzy sets is proposed to save assistantship functionality for specialists and physicians. In some sense, multiple pattern repositories are injected into the infrastructure of artificial immune systems. To verify and validate the proposed approach, several computational simulations have been performed. Simulation results prove that proposed method provide more accurate results in identifying malignant tumors in comparison to traditional pattern recognition techniques as such artificial neural networks and neuro-fuzzy systems.

Keywords— Artificial Immune Systems; Breast Cancer; Pattern Recognition; Classification; E-Health; E-Medicine; Data-Mining.

I. INTRODUCTION

Cancer is a genetic disruption based disease in which, biological cells lose their natural reproduction behavior, and natural organisms are attacked by a mass of cancerous cells which is called tumor. Regarding to World Health Organization (WHO) annually reports, breast cancer is leading cause of morbidity and mortality among women. From the perspective of physicians, anatomy of natural breast consists of lipid masses, so identifying malignant and benign tumors based on mammographic images is performed with no desirable accuracy and efficacy. For instance, some natural lipid masses might be identified as malignant tumors. In this sense, patients are referred to biopsy which increases the healthcare expenditures remarkably. In addition, these types of false positive detections could also provide psychological and emotional side effects. In spite benign masses are more likely to be identified as malignant tumors, false negative would increase risk of late detection in which, patients would have less chance to survive [1].

Admittedly, it can be concluded that identifying the type of tumors through mammographic images is difficult enough to make issue [2]. To deal with this challenge, almost patients with malignant or benign tumors are referred to biopsy surgery to take more special tests [3]. On this basis, various efforts have been done to enhance the accuracy and reliability of medical identification devices [4,5], and develop physician assistantship artifacts. For instance, wide variety of applications based on digital images processing methods has been proposed to serve assistantship functionality for specialists [6,7].

Data-mining includes various applications – including but not limited to pattern recognition and classification, clustering, regression and size reduction. In some sense, pattern recognition applications can be considered as. Recently, cellular automata are used for many purposes ranging from cryptography [8], random number generation [8,9] to digital image processing [10].

In this research, Artificial Immune Systems (AIS) have been applied to develop a novel expert system for identifying malignant breast cancer though mammographic images. Accordingly, not only

---

[*] Corresponding author can be contacted via the journal website.

artificial immune systems are employed for optimization purposes, but they also have been used for pattern recognition applications. To this end, cancer as a disease is discussed in second section. Afterwards, concepts of artificial immune systems, and proposed approach for identifying malignant breast cancers are expressed in section three and four respectively. In this way, simulation results for verifying and validating proposed method is provided in section five, and conclusion is discussed in last section.

II. CANCER

Cells are elementary components in emerging life which look like a bobble that contains proteins, lipoid acids, DNA, etc. From the biological viewpoints, alive cells have reproduction and growing capabilities. In fact, behaviors of cells (i.e. time for mitosis, growing, metabolism, etc.) are defined based on their genetic structure that is recorded on chromosomes. In natural organisms, elderly cells or substituted with new born cells, and follow a regular predefined plan up. From another attitude, there is a kind of equilibrium in reproducing new cells and mortality of elderly cells [11].

Cancer is a special type of genetic disease in which, the equilibrium between mitosis and mortality rate is manipulated. In this sense, cancerous cells perform more mitosis and consequently reproduce more new cells. On the other hand, cancerous cells pay less or no attention to death signals. Undoubtedly, this process will lead a mass of abnormal cells which called tumor. Benign tumors only utilize resources of natural organisms; however, malignant tumors offend living being's organism. Thus, malignant tumors lead death, and identifying these types of tumors from benign masses will dramatically reduce risk of treatment [12]

When cancer was detected in a patient as disease, an efficient care plan must be followed up not only for preventing prevalence of disease but for radical treatment purposes. To this end, cancer in each patient case is sorted out through leveling methods. Afterwards, health care delivery services are applied for treatment purposes –including but not limited to chemotherapy, radiotherapy, surgery, gene therapy, etc.

III. ARTIFICIAL IMMUNE SYSTEMS

Immune system consists of cells, molecules and mechanisms which prevent external agents such as pathogens from harming the host body [13]. Antigen is part of pathogen recognized by immune system. Immune cells named lymphocytes detect and kill pathogens and are composed of two groups of cells each with different structure and functionality: B-cells and T-cells. B-cells produce antibody and by attaching themselves to antigens they cause pathogens to be destroyed. On the other hand part of T-cells stimulates B-cells to produce antibody and another part of T-cells

collaborate with rest of immune cells to eliminate the detected pathogens [7]. Upon recognition of antigens, B-cells begin to produce antibody. Within the produced receptor cells some are picked to be memory cells which yield to an enhanced response from immune system to secondary encounters with the same specific antigens or similar structure[14].

All cells produced in the immune system are identical to their parents because the only reproduction method for these cells is cell division and no crossover takes place. However, each cell is affected by mutation operator according to its affinity with antigen; lower the affinity with antigen, higher the transformation of the cell. The other factor that depends on the affinity with antigen is the number of cells that each cell can reproduce. Parent cell reproduces more cells when the affinity is higher. Selection and mutation process is called Affinity maturation [16–19].

For the sake of simplicity B-cells and T-cells are considered to be a unique set in artificial immune system. Samples of immune system algorithms customized for optimization problems are ClonalG and aiNet [15]. Furthermore aiNet algorithm can be placed within the clustering algorithms. Castro and Timmis (2002) classified artificial immune system algorithms into two population-based and network-based categories and thereby put the negative and clonal selection in first category and immune network model, subcategorized to continuous network and discrete network, in the second category (Fig. 1).
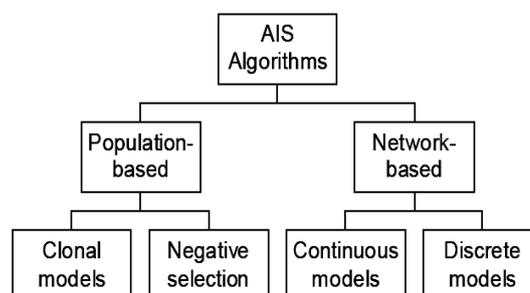

Figure 1. Taxonomy for AIS algorithms.

IV. AIS FOR BREAST CANCER IDENTIFICATION

Nowadays, various medical devices are employed to identify breast cancer; however, mammograms are mostly used by almost health care delivery service providers. In this sense, X-Ray based image of beasts are captured. Afterward, these images are applied by physicians to consider breast tissues and identify the type of mass (malignant or benign). Owing to the lower quality/resolution of mammograms, this identification process is more likely to provide false result. From another perspective, regarding to the existence of natural lipoid masses in breast's tissue, not only some cancerous tumors may not be detected, but benign masses might also be identified as malignant. It is

obvious that referring almost patients to biopsy in identifying breast cancer is plausible. In this section, a novel structure of artificial immune systems is proposed to serve decision support functionality for physicians and specialist.

A. Structure of the System

In Fig. 2, architecture of proposed intelligent system for identifying malignant breast cancer is illustrated. In this infrastructure, digital mammogram images are system inputs.
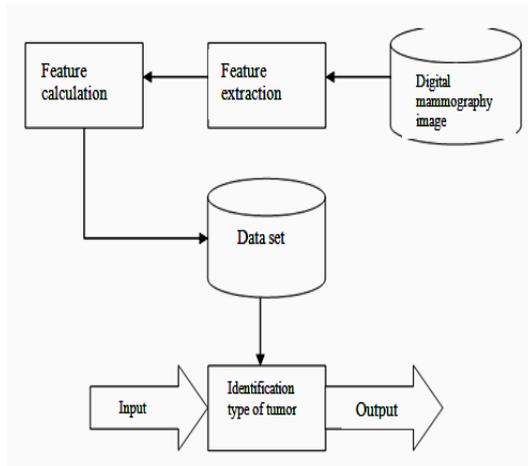

Figure 2. Artificial Immune System Based structure for identifying malignant breast cancer

The intelligent system receives digital mammograms and perform computational preprocesses on digital images. In this step not only resolution of mammograms are enhanced but some features are also extracted to be considered for decision making.

Apparently, wide variety of features can be extracted from these digital images; however, employing unnecessary features will increase system resource requirements (time and memory). In addition, large number of features will dramatically increase the complexity of decision making problem, so accuracy on identifying tumor type will be face remarkable reductions.

To deal with this issue, both feature selection and feature extraction approaches are applicable. It undoubtedly, these data-mining processes are performed offline. From another attitude, proposed intelligent system can be trained through standard data set of mammograms in which type of tumor for each instance is available. These identifications must be reliable; thus, specialist should be employed for creating such a standard data sets. Despite the fact that feature selection algorithms require remarkable system resources, they performed offline and will not provide any issue in real world applications.

In this study, a standard data set of mammographic images are utilized, therefore, feature selection/feature extracting module is eliminated in implementations. On this basis, input vector for each instance is processed by pattern recognition system to identify the type of tumor. Apparently, this intelligent system needs a training process to adjust system parameters. From the perspective of artificial intelligence, enough number of training instances must be considered by system to tune the variables up for further decision makings. Moreover, feedback signal has been provided to report false negatives and false positives (wrong identifications). This report can be applied for readjusting system parameters. Consequently, accuracy of further decisions will be enhanced through retraining processes.

B. AIS Based Pattern Recognition

As mentioned eelier, feature selection and feature extraction algorithms are performed offline. It can be concluded that in real world applications, it is obvious that which parameters are supposed to be measured. Thus, preprocesses on digital mammographic images are performed only for extracting desired features. Afterwards, these measured features are sent to pattern recognition module for ultimate division making purposes. In fact, the pattern recognition module serves intelligent functionality for decision support applications [16,17].

In this study case, tumors are sorted out into two categories: benign masses (C1) and malignant tumors (C2). Thus, pattern recognition module is supposed to be able to classify input mammograms into one of these benign or malignant classes. To this end, two separate antibody repositories are applied in this modified artificial immune system. First repository contains those antibodies that identify familiar antigens. On the other hand, second repository serves a library functionality that has complete set of foreign antigens. All those antigens that have a matching instance in second repository are assumed offensive, so immune system is supposed to attack to them. It is also worth to noting that, if there is no absolute match for an antigen in both repositories, antigen will be classified through a maximum likelihood evaluation.

To provide a computational model of above model, each type of cancerous tumors (benign masses and malignant tumors) are represented through an antibody repository. In this sense, each library consists of fixed number of instances (antibodies). When a new mammogram is entered as input, extracted features vector (antigen) is considered by both library members. Afterwards, among all antibodies in each repository, the antibody that has maximum match with antigen is selected. From another perspective, each library introduces a candidate to analyze the antigen. To make the ultimate decision, likelihood of antigen with both

antibodies are calculated. Finally, the antigen will be assign to the class that has more fit with specified antibody. For example, if the likelihood of an instance mammogram to C1 library is more than malignant repository, the patient are more likely to has benign mass (instead on malignant cancerous tumor)

In this study, antigens are represented in discrete state space. Thus, using distance as a benchmark to mature the likelihood parameter should be plausible. In this sense, mammographic image will be assigned to the class (benign or malignant) that has less distance in state space formulation. It is also worth to noting that these antibody repositories are created in training process. To this end, clonal selection algorithms are applied in this study. For this purpose, initially both repositories are generated with random antibodies. Afterwards, both libraries are trained based on training data set. In this sense, objective of training first repository (C1) is to be more likely for identifying benign masses. On the other hand, objective of training second repository (C2) is increasing odds of identifying malignant tumors.

Actually, these libraries compete to achieve maximum fitness, so the winner will put its class label (benign or malignant) on mammographic image. For misdiagnosis situations (false positive and false negative), feedback signal is used to retain both libraries. Apparently, this system serves decision support functionality, so physicians and specialists are required for fatal decision making.

V. EXPERIMENTAL RESULTS

As mentioned the main propose of this research is designing an intelligent expert system to determine tumor type in breast cancer of mammography image based on artificial immune system algorithm of people who has breast cancer. In this evaluation that was based on computer simulation operation of artificial immune system in feature selection and pattern recognition based on standard data set were analyzed.

In this study it was show that the proposed artificial immune system algorithm in feature selection better than genetic algorithm and in pattern recognition is better than artificial neural network. In collecting data set for using in intelligent pattern recognition system the main problem in addition of image standardizing was determining the type of cancer tumor of associated with each mammography image. To overcome this problem used a standard data set of mammography images to evaluate performance of expert system to deal with real world pattern recognition [15–17,20,21].

This data set is digital mammography image of 961 patients that realized and extract five features as useful feature. The sixth feature of this data set show type of

tumor. Useful properties of a data set of digital mammogram images are, respectively [12]:

BI-RADS: integer number one to four

Age: An integer number that represents the year of the age of the patient.

Mass shape: circular=1, elliptical (oval) = 2, labial (lobular) = 3 and irregular=4

Mass edge: clear = 1, indent (Tiny Dalbor) = 2, Obscure (Blur) = 3, none clear = 4 and Needle=5

Density: number between 1 to 4, number will decreases if density increase

Tumor types: benign =0 and malignant = 1

As mentioned above its clear that the range of feature vector value is a real number. Data set used in two ways for evaluate of proposed system in determining tumor type of breast cancer. In case one data set directly gives to classifier tools and in case two feature vectors convert in normalized feature vector.

As discussed for evaluation of designed expert system the efficiency and accuracy in classification with multilayer perceptron neural network has been compared. The simulation results of artificial neural network on mammography image's data set are presented in table 1 and table 2. Table 1 and table 2 show the training error and test error respectively.

According to the nature of statistical tests in simulation results may be different in different runs, each simulation was run 10 times consecutively, statistical feature such as average, best of solutions and variance feature are calculated. It is clear mean feature show performance and variance feature reflects the stability of expert system. If the mean feature is lower the expert system performance is better and if the variance feature is lower the expert system stability is better.

Table 1. Training error for determine tumor type (ANN)

Variance	Mean	Minimum	Hidden layer
7/3564*10-1	18/77	14/91	5-5
9/4509*10-1	19/31	15/42	5-10-5

Table 2. Test error for determine tumor type (ANN)

Variance	Mean	Minimum	Hidden layer

| 8/3481*10-1 | 19/53 | 16/89 | 5-5 |
| 7/9065*10-1 | 20/72 | 17/11 | 5-10-5 |

In simulation experiment size of database library for each class of tumor was same. The simulation results of purposed expert system to recognize tumor type are presented in table3 and table 4. Table 3 and table 4 show the training error and test error respectively. To investigate the dependence of algorithms to size of antibody library, simulations were performed for different sizes of database library. Similar to artificial neural network experiments, each simulation was run 10 times and statistical indicators such as best; mean and variance of training and testing error are calculated.

Table 3. Training error for determine tumor type (AIS)

Variance	Mean	Minimum	Library size
2/2337*10-2	15/93	14/20	10
4/3426*10-2	15/17	14/09	20
3/2337*10-2	14/83	13/20	40
7/4683*10-3	12/92	11/90	100

Table 4. Test error for determine tumor type (AIS)

Variance	Mean	Minimum	Library size
2/45678*10-2	18/36	16/29	10
6/9945*10-2	17/42	15/90	20
1/9335*10-3	17/39	14/10	40
3/5477*10-3	15/19	13/48	100

With investigate the numerical simulation result on recognize tumor type of mammography images it is clear that the performance of purposed expert system is introduced in this paper has better rate. A prominent feature of this algorithm is that with expanding antibody library size has a direct impact on the efficiency and stability of the algorithm. But increasing number of neuron and middle layer in artificial neural network after a step has negative impact on system operation.

VI. CONCLUSIONS

Breast cancer is the leading cause of morbidity and mortality among female population and almost mammographic images have no enough quality for accurate decision makings. In this paper artificial immune systems are employed to propose an intelligent system to serve decision support functionality. To this end, some modifications were performed on standard artificial immune systems. For instance, two antibody repositories are applied to identify the type of breast cancer through mammograms. In fact, antibodies in firs repository are responsible to identify benign masses; however, antibodies in second library are supposed to identify malignant tumors. These antibodies are tainted through clonal selection algorithm. Ultimate decision will be making based on the maximum likelihood parameter. In other words, the mammographic image will be assigned to a class that has less distance with it.

To evaluate the functionality of this method, computational simulations are implemented. In addition, a standard data set of mammographic images was used to measure the efficacy of proposed method. Experimental results revealed that artificial immune system based decision support perform more desirable functionality in comparison to artificial neural networks and neuro-fuzzy systems. Statistical analysis also demonstrated that proposed system behaves more stable than other methods. It is also worth to noting that this approach is a decision support system, so ultimate decision must be made by physicians and specialists.

REFERENCES

[1] Schottenfeld, David, and Joseph F. Fraumeni Jr. *Cancer epidemiology and prevention*. Eastbourne, UK; WB Saunders Co, 1982.

[2] Tabar, Lazio, A. Gad, L. H. Holmberg, U. Ljungquist, Kopparberg County Project Group, C. J. G. Fagerberg, L. Baldetorp et al. "Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare." *The Lancet* 325, no. 8433 (1985): 829-832.

[3] Veronesi, Umberto, Giovanni Paganelli, Viviana Galimberti, Giuseppe Viale, Stefano Zurrida, Marilia Bedoni, Alberto Costa et al. "Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes." *The Lancet* 349, no. 9069 (1997): 1864-1867.

[4] Boyd, N. F., J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and Martin J. Yaffe. "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study." *Journal of the National Cancer Institute* 87, no. 9 (1995): 670-675.

[5] Boyd, Norman F., Gillian S. Dite, Jennifer Stone, Anoma Gunasekara, Dallas R. English, Margaret RE McCredie, Graham G. Giles et al. "Heritability of mammographic density, a risk factor for breast cancer." *New England Journal of Medicine* 347, no. 12 (2002): 886-894.

[6] Karabatak, Murat, and M. Cevdet Ince. "An expert system for detection of breast cancer based on association rules and neural network." *Expert Systems with Applications* 36, no. 2 (2009): 3465-3469.

[7] Rueckert, Daniel, Luke I. Sonoda, Carmel Hayes, Derek LG Hill, Martin O. Leach, and David J. Hawkes. "Nonrigid registration using free-form deformations: application to breast MR images." *Medical Imaging, IEEE Transactions on* 18, no. 8 (1999): 712-721.

[8] Jaberi, Alireza, Ramin Ayanzadeh, Elaheh Raisi, and Aidin Sadighi. "Central Limit Theorem based Cellular Automata for Generating Normal Random Numbers." *Information Technology Journal* 12, no. 12 (2013).

[9] Ayanzadeh, Ramin, K. Hassani, Y. Moghaddas, H. Gheiby, and S. Setayeshi. "Multi-layer cellular automata for generating normal random numbers." In Electrical Engineering (ICEE), 2010 18th Iranian Conference on, pp. 495-500. IEEE, 2010.

[10] Dyer, Charles R., and Azriel Rosenfeld. "Parallel image processing by memory-augmented cellular automata." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3, no. 1 (1981): 29-41.

[11] Gutierrez, Andres A., Nick R. Lemoine, and Karol Sikora. "Gene therapy for cancer." *The Lancet* 339, no. 8795 (1992): 715-721.

[12] Farazi, Thalia A., Jelle J. Ten Hoeve, Miguel Brown, Aleksandra Mihailovic, Hugo M. Horlings, Marc J. van de Vijver, Thomas Tuschl, and Lodewyk FA Wessels. "Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets." *Genome Biol* 15 (2014): R9.

[13] Engelbrecht, Andries P. *Computational intelligence: an introduction*. John Wiley & Sons, 2007.

[14] Javadzadeh, Ramin, Zahra Afsahi, and MohammadReza Meybodi. "Improved Artificial Immune System Algorithm with Local Search." *World Academy of Science, Engineering and Technology* (2009): 654-657.s

[15] Aickelin, Uwe, Dipankar Dasgupta, and Feng Gu. "Artificial immune systems." In *Search Methodologies*, pp. 187-211. Springer US, 2014.

[16] De Castro, Leandro Nunes, and Jonathan Timmis. *Artificial immune systems: a new computational intelligence approach*. Springer Science & Business Media, 2002.

[17] Webb, Andrew R. *Statistical pattern recognition*. John Wiley & Sons, 2003.

[18] Ayanzadeh, Ramin, E. Shahamatnia, and S. Setayeshi. "Determining Optimum Queue Length in Computer Networks by Using Memetic Algorithms." *Journal of Applied Sciences* 9, no. 15 (2009): 2847-2851.

[19] Shahamatnia, Ehsan, Ramin Ayanzadeh, Rita A. Ribeiro, and Saeid Setayeshi. "Adaptive imitation scheme for memetic algorithms." In Technological Innovation for Sustainability, pp. 109-116. Springer Berlin Heidelberg, 2011.

[20] Marzban, Forough, Ramin Ayanzadeh, and Pouria Marzban. "Discrete Time Dynamic Neural Networks for Predicting Chaotic Time Series." Journal of Artificial Intelligence 7, no. 1 (2014).

[21] Ayanzadeh, R., K. Hassani, Y. Moghaddas, H. Gheiby, and S. Setayeshi. "Innovative Approach to Generate Uniform Random Numbers Based on a Novel Cellular Automata." Journal of Applied Sciences 9, no. 22 (2009).