

# Artificial Immune Systems Applications in Cancer Research

J Al-Enezi, M.F. Abbod, and S. Al-Sharhan\*

School of Engineering and Design, Brunel University, UK

\*Computer Science Department, Gulf University for Science and Technology

P.O. Box: 7207, Hawalli, 32093, Kuwait

Maysam.Abbod@brunel.ac.uk

**Abstract**—Artificial Immune System (AIS) is a branch of computational intelligence that has gained increasing interest among researchers in the development of immune-based models and techniques to solve diverse complex computational problems. This work focuses on the application of AIS techniques to cancer research and specifically for prediction of the recurrence of cancer in patients. The objective is to test AIS models and algorithms for cancer research by validation against actual cancer datasets.

**Keywords**- Artificial immune systems, artificial immune networks, clonal selection, negative selection

## I. INTRODUCTION

There is an increasing body of research on the application of novel artificial intelligence paradigms to the diagnosis and prognosis of various kinds of cancer. Typical examples are neural networks, fuzzy logic and evolutionary computing. Artificial intelligence techniques offer advantages - such as adaptation, fault tolerance, learning and human-like behaviour - over conventional computing techniques. The rapid increase in cancers, such as lung cancer, breast cancer, liver cancer, brain cancer and skin cancer resulting from modern life-styles, and the parallel exponential growth in affordable computing power has lent fresh urgency to researching, finding, developing and harnessing pattern recognition and machine intelligence based solutions in this arena.

The objective of this study was to test Artificial Immune Systems (AIS) applications, models and algorithms for cancer research by validation against actual cancer datasets. The development of intelligent techniques based on Artificial Immune Systems that are potentially useful to solve clinical cancer problems - in particular, for the prognosis of cancer recurrence in patients - is discussed, along with their underlying theory and the implementation of applications in cancer research. Three popular AIS algorithms are considered, and case studies based on actual cancer datasets are presented. The outcome, assumptions and limitations of the proposed methods with its implications for further research in this area draw this paper to its conclusion.

## A. Artificial Immune Systems

Artificial Immune Systems incorporate many properties of natural immune systems, including diversity, distributed computation, error tolerance, dynamic learning and adaptation, and self-monitoring. The biological immune system has an amazing pattern matching ability, used to distinguish between foreign cells entering the body (referred to as non-self, or antigen) and the cells belonging to the body (referred to as self). Immune systems have many characteristics such as uniqueness, autonomy, recognition of foreigners, distributed detection, and noise tolerance [1]. Artificial immune systems can be defined as a computational paradigm that is inspired by theoretical immunology, observed immune functions, principles and mechanisms [2].

Three immune mechanisms are primarily used in the development of AIS methods. These include the immune network theory, clonal selection principles and negative selection mechanisms.

Many AIS based algorithms have been introduced in the literature. A survey of the developed AIS algorithms can be found in [3] and [12].

## B. AIS and application in Health Sector - Cancer

ClonalG, V-Detector, and aiNet are algorithms that could, in principle, be applied to many domains. In this paper, they are applied to applications in cancer research.

The rest of this report is organized as follows: Section 2 presents an overview of other conventional classification models; Section 3 presents an introduction to three AIS models and techniques developed in the literature with particular reference made to their application in the Health Sector. It considers three of the most well-known, leading AIS algorithms and highlights their relative application areas, strengths and weaknesses; Section 4 presents case studies, using three datasets specific to cancer and briefly discusses the findings and outcome; Section 5 highlights the conclusion of this paper and the future work that it suggests.

## II. CLASSIFICATION MODELS IN CANCER RESEARCH

### A. Classification Models in Cancer Research - Conventional Methods

Classification is an important research area in cancer diagnosis. Reference is made to some outstanding work done in this area based on alternative approaches.

#### 1) Artificial Intelligence Techniques in Breast Cancer Diagnosis and Prognosis

Artificial Intelligence Techniques in Breast Cancer Diagnosis and Prognosis [4] describes the development of new statistical techniques and artificial intelligence methods that are key to recent advances in prognostic and predictive techniques in breast cancer research. These cover pathological, intelligent and statistical approaches to enable more precise definition of disease extent and prediction of tumor behavior and response to treatment. For example, neural networks can be applied to effectively recognize and represent the complicated dependence of the disease on a range of demographic and clinical factors and to distinguish this dependence from the noise. Another tool mentioned is MammoNet, a Bayesian networks-based tool for breast cancer. Yet another approach is the use of continuous and categorical biological variables as prognostic biomarkers for breast cancer patients. Mention is also made of the application of artificial intelligence paradigms to computer-assisted interpretation of mammograms and cytodiagnosis of breast cancer, including computer vision techniques and the current state of the ongoing research program in the USA on Xcyt, a new system for remote cytological diagnosis of breast cancer.

#### 2) Computational Intelligence Techniques

Artificial Intelligence Techniques in Breast Cancer Diagnosis and Investigating Human Cancer with Computational Intelligence Techniques [5] refers, inter alia, to Unsupervised breast cancer discovery: a comparative study on model-based and neural computing; Stratification of severity of illness indices and out-of-sample validation: a case study for breast cancer prognosis; Feature and model selection in single voxel spectra for cancer classification; and Statistical assessment of MSigDB gene sets in colon cancer.

#### 3) Fuzzy C-means

Simulated Annealing Fuzzy Clustering (SAFC) in Cancer Diagnosis [6] claims to improve and extend the ideas present in Variable String Length Simulated Annealing (VFC-SA) that was proposed by Bandyopadhyay. The data from seven oral cancer patients tissue samples, obtained through Fourier Transform Infrared Spectroscopy (FTIR), were clustered using Fuzzy C-means (FCM), VFC-SA and the proposed SAFC algorithm. Experimental results provided and comparisons made illustrate that the SAFC algorithm is able to find better clusters than the other two methods.

## III. AIS ALGORITHMS

Features of three AIS algorithms are described in brief below. Greater details may be found in the original papers cited.

### A. ClonalG Algorithm

The Clonal Selection principle is the whole process of antigen recognition, cell proliferation and differentiation into memory cell [7]. Several artificial immune algorithms have been proposed in the literature based on the clonal selection theory. Castro and Zuben proposed in 2002 a clonal selection algorithm named ClonalG for learning and optimization, ClonalG generates a population of  $N$  antibodies, each specifying a random solution for the optimization process. During each iteration, some of the best existing antibodies are selected, cloned and mutated in order to construct a new candidate population. New antibodies are then evaluated and certain percentage of the best antibodies is added to the original population. Finally a percentage of worst antibodies of previous generation are replaced with new randomly created ones [8].

### B. V-Detector Algorithm

Negative selection is a biological process by which the natural immune system generates non-self detectors that do not detect self structures. Similarly, the negative selection algorithm generates detector set by eliminating any detector candidate that match elements from a group of self samples. Negative selection based algorithms have been used in different applications areas, such as anomaly detection.

A Real-Valued Negative Selection Algorithm with Variable-Sized Detectors named V-Detector has been introduced by Ji and Dasgupta in [9]. It has many characteristics such as: simple generation strategy and detector scheme, variable sized detectors, coverage estimate and boundary-aware technique to interpret the training data set as a whole and not as independent points.

### C. aiNet Model

Several artificial immune network models have been proposed based on Jerne's theory [10] and applied to many application domains. Different from the clonal selection principle, the immune network theory assumes that the immune system maintains a regulated network of cells and molecules which maintain interactions between not only an antibody and an antigen, but also between antibodies.

Inspired by the immunological principles including the clonal selection, affinity maturation, and immune network theory, the artificial immune network model "aiNet" is presented by de Castro and Zuben [11]. The aiNet model has the capability of reducing redundancy and describing immune network structure, including data distribution and clustering.

#### IV. CASE STUDIES

Experiments were carried out to test three of the AIS algorithms and explore their capabilities. Three cancer data sets were used for testing the classification problem. This paper builds on earlier work done by the authors [12] [13]. The ClonalG, V-Detector and aiNet algorithms were chosen for these case studies and tested on three cancer data sets.

Initially, the datasets were normalised to unity before being fed to the algorithms. Once the data was normalised a five percentage of the samples was chosen at random and removed from the data set. This then became the testing samples and the rest used for the training phase. For each experiment, we have repeated the training and the testing phases ten times, in which we have used different samples chosen randomly.

##### A. Experiment # 1 : Cancer Data set

The first data set consisted of 693 instances and the number of attributes was 12. The predicted output of this dataset represents the recurrence status where the value one as an output indicates the possibility of the patient getting the cancer again in the future.

For this dataset, the results obtained for accuracy as shown in Table I indicating that ClonalG algorithm has achieved better average testing accuracy compared to the other two algorithms.

TABLE I

| Round #        | ClonalG       |               | V-Detector    |               | aiNet         |               |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                | Train.        | Test.         | Train.        | Test.         | Train.        | Test.         |
| 1              | 0.7178        | 0.7941        | 0.8300        | 0.6765        | 0.6753        | 0.7647        |
| 2              | 0.7204        | 0.7714        | 0.8313        | 0.6857        | 0.3237        | 0.2000        |
| 3              | 0.7143        | 0.7714        | 0.8374        | 0.5143        | 0.6824        | 0.6857        |
| 4              | 0.7015        | 0.6970        | 0.8227        | 0.8182        | 0.3197        | 0.3333        |
| 5              | 0.7193        | 0.7941        | 0.8392        | 0.5294        | 0.6768        | 0.7647        |
| 6              | 0.7167        | 0.7576        | 0.8288        | 0.6364        | 0.6727        | 0.8485        |
| 7              | 0.6980        | 0.7647        | 0.8346        | 0.6176        | 0.3232        | 0.2353        |
| 8              | 0.7238        | 0.5882        | 0.8498        | 0.3529        | 0.6889        | 0.5294        |
| 9              | 0.7056        | 0.7647        | 0.8316        | 0.6471        | 0.6798        | 0.7059        |
| 10             | 0.7102        | 0.7647        | 0.8346        | 0.5588        | 0.3187        | 0.3235        |
| <b>Average</b> | <b>0.7127</b> | <b>0.7468</b> | <b>0.8340</b> | <b>0.6037</b> | <b>0.5361</b> | <b>0.5391</b> |

##### B. Experiment # 2 : Wisconsin Breast Cancer Dataset

The Wisconsin breast cancer data set was obtained from the University of Wisconsin Hospitals and can be found in [14]. It has 699 samples with 9 attributes, where each instance has one of two possible classes: benign or malignant.

The result of this experiment presented in Table II has shown that the V-Detector algorithm has achieved a very promising accuracy results. Furthermore, the ClonalG

algorithm also achieved a very good accuracy results compared to the previous dataset. However, the results obtained from the aiNet algorithm still were unacceptably low and need more improvements.

TABLE II

| Round #        | ClonalG       |               | V-Detector    |               | aiNet         |               |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                | Train.        | Test.         | Train.        | Test.         | Train.        | Test.         |
| 1              | 0.8554        | 0.8571        | 0.9925        | 0.9429        | 0.4820        | 0.4545        |
| 2              | 0.8524        | 0.9429        | 0.9925        | 1.0000        | 0.4812        | 0.5294        |
| 3              | 0.8651        | 0.9063        | 0.9925        | 1.0000        | 0.6607        | 0.6970        |
| 4              | 0.8660        | 0.8571        | 0.9940        | 0.9429        | 0.5375        | 0.5455        |
| 5              | 0.8720        | 0.8000        | 0.9940        | 0.9143        | 0.3509        | 0.2000        |
| 6              | 0.8677        | 0.8235        | 0.9925        | 1.0000        | 0.5678        | 0.6286        |
| 7              | 0.8569        | 0.9429        | 0.9925        | 0.9714        | 0.6370        | 0.7143        |
| 8              | 0.8660        | 0.8286        | 0.9925        | 0.9429        | 0.7481        | 0.7500        |
| 9              | 0.8679        | 0.7576        | 0.9925        | 0.9394        | 0.4932        | 0.6176        |
| 10             | 0.8645        | 0.9143        | 0.9925        | 0.9714        | 0.5783        | 0.6286        |
| <b>Average</b> | <b>0.8634</b> | <b>0.8630</b> | <b>0.9928</b> | <b>0.9625</b> | <b>0.5537</b> | <b>0.5765</b> |

##### C. Experiment # 3 : Haberman's Survival Dataset

In this dataset, cases included from a study that was conducted at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer [14]. It contains 306 instances and 3 attributes, and the predicted output represents two classes. The first class represents the patient survived 5 years or longer, and the second class is for the patient died within 5 years.

In this experiment, we found also that the V-Detector algorithm achieving higher average accuracy rates compared to the other two algorithms as shown in Table III.

TABLE III

| Round #        | ClonalG       |               | V-Detector    |               | aiNet         |               |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                | Train.        | Test.         | Train.        | Test.         | Train.        | Test.         |
| 1              | 0.8419        | 0.9333        | 0.7629        | 0.9333        | 0.3540        | 0.3333        |
| 2              | 0.7663        | 0.5333        | 0.7869        | 0.4667        | 0.7491        | 0.4667        |
| 3              | 0.8041        | 0.8667        | 0.7629        | 0.8667        | 0.2715        | 0.1333        |
| 4              | 0.8007        | 0.6667        | 0.7766        | 0.6000        | 0.4914        | 0.7333        |
| 5              | 0.8316        | 0.8000        | 0.7663        | 0.8000        | 0.2749        | 0.1333        |
| 6              | 0.7904        | 0.6000        | 0.7698        | 0.7333        | 0.6426        | 0.6000        |
| 7              | 0.8110        | 0.6667        | 0.7732        | 0.6667        | 0.7388        | 0.6000        |
| 8              | 0.7904        | 0.8667        | 0.7663        | 0.8000        | 0.7354        | 0.8667        |
| 9              | 0.8328        | 0.6154        | 0.7713        | 0.7692        | 0.2662        | 0.2308        |
| 10             | 0.8110        | 0.5333        | 0.7801        | 0.6000        | 0.7491        | 0.6000        |
| <b>Average</b> | <b>0.8080</b> | <b>0.7082</b> | <b>0.7716</b> | <b>0.7236</b> | <b>0.5273</b> | <b>0.4697</b> |

#### D. Discussion

Against the background of current research being done, both by conventional approaches, as well as by the application of AIS algorithms, recent advances have prompted further investigation by experimentation of the more promising approaches in applying AIS techniques to cancer research, as an aid to prognosis, in particular.

The case studies carried out clearly demonstrate how the AIS approaches can be employed in dealing with real world problems in health and cancer research. The three experiments conducted to test the ClonalG, V-Detector and aiNet algorithms respectively against three cancer datasets yielded mixed results. In the first test, the results obtained for accuracy were unacceptably low and more improvements required for getting better outcomes. On the other hand, a better performance result was achieved in the second experiment especially with the V-Detector algorithm by detecting successfully the number and the clusters for the tested dataset.

This outcome leads to the conclusion that some of the AIS techniques are found to be more suitable for cancer research than other AIS approaches.

#### V. CONCLUSIONS & FUTURE WORK

The success rates achieved with AIS models till now indicates that there is further need for refinement and modifications of the approaches considered to gain greater accuracy and reliability to derive even greater benefit in cancer research. The developments of the artificial immune systems would benefit not only from the inspiration of biological immune principles and mechanisms, but also hybridization with other soft computing paradigms, such as neural networks, fuzzy logic, and genetic algorithms.

As a future work, we are aiming to build an ensemble model integrating different artificial immune systems techniques. Our main objective is to integrate different learning and adaptation techniques to overcome individual limitations and to achieve synergetic effects through the combination of these techniques.

#### REFERENCES

- [1] L de Castro and F Zuben, "Artificial Immune Systems: Part I – Basic Theory and Applications", TR – DCA 01/99, 1999.
- [2] L de Castro and J Timmis, "Artificial Immune Systems as a Novel Soft Computing Paradigm", *Soft Computing Journal*, vol. 7, Issue 7, July 2003.
- [3] D Dasgupta, Z Ji and F Gonzalez, "Artificial Immune System (AIS) Research in the Last Five Years", *The 2003 Congress on Evolutionary Computation (CEC '03)*, vol. 1, pp. 123-130, 8-12 Dec. 2003.
- [4] Ashlesha Jain, Ajita Jain, Sandhya Jain, Lakhmi Jain, "Artificial Intelligence Techniques In Breast Cancer Diagnosis & Prognosis", World Scientific Pub Co Inc (2000) ISBN-10: 981-02-4374-X.
- [5] Alfredo Vellido and Paulo J.G. Lisboa (Editors): "Investigating Human Cancer with Computational Intelligence Techniques", *Future Technology Press*, ISBN 978-0-9561516-0-5 2009
- [6] Xiao-Ying Wang, Jonathan M. Garibaldi, "Simulated Annealing Fuzzy Clustering in Cancer Diagnosis", *Automated Scheduling, Optimisation and Planning (ASAP) Research Group, Department of Computer Science & Information Technology, The University of Nottingham, Jubilee Campus, Wollaton Road, United Kingdom* {xyw, jmg}@cs.nott.ac.uk
- [7] F.M. Burnet, "The Clonal Selection Theory of Acquired Immunity", Cambridge University Press, 1959.
- [8] L. de Castro and F. Zuben, "Learning and Optimization Using the Clonal Selection Principle", *IEEE Transactions on Evolutionary Computation*, 2002, 6(3):239-251.
- [9] Zhou Ji and Dipankar Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In *LNCS 3102, Proceedings of GECCO 2004*, Seattle, Washington, June 2004.
- [10] N K Jerne, "Towards a network theory of the immune system", *Ann. Immunology*, vol. 125C, pp. 373–389, 1974.
- [11] Leandro Nunes de Castro & Fernando Jose Von Zuben. aiNet: An Artificial Immune Network for Data Analysis. in *Data Mining: A Heuristic Approach*. 2001
- [12] J.R. Al-Enezi, M.F. Abbod and Salah Al-Sharhan, "Advancement in Artificial Immune Systems: A perspective of Models, Algorithms and Applications", *The 5th IEEE-GCC Conference*, March 2009, Kuwait.
- [13] J.R. Al-Enezi, M.F. Abbod and S. Al-Sharhan, "Artificial Immune Systems – Models, Algorithms and Applications", *International Journal of Research and Reviews in Applied Sciences*, Volume 3, Issue 2, May 2010.
- [14] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.