BioPublisher©
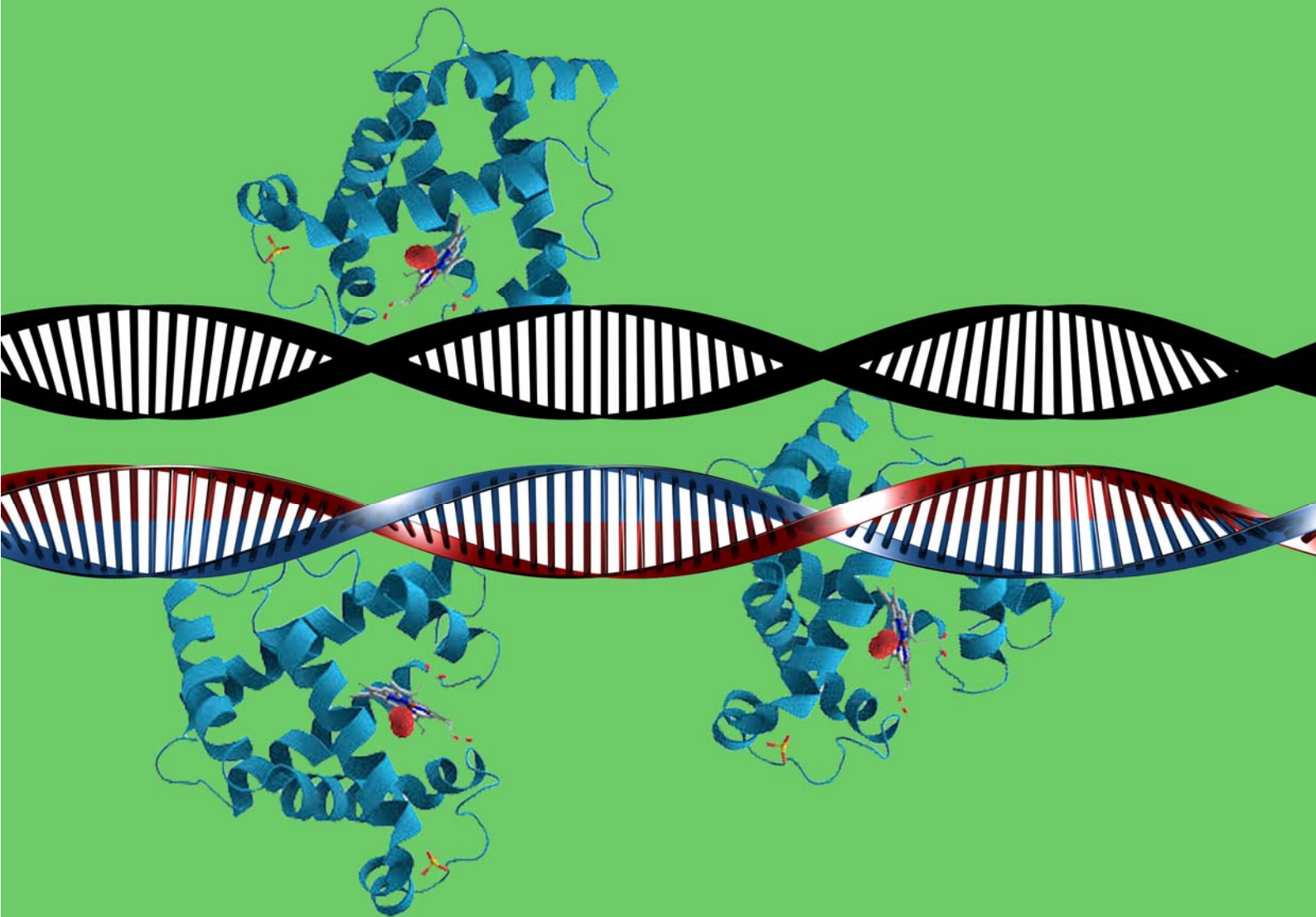
Genomics and Applied Biology

An international journal of applied biology for genomics era

2014
Vol. 5

Publisher

Edited by

Genomics and Applied Biology (GAB) (ISSN1925-1602, online) is an open access and peer review journal, publishing original research papers involving in the structural genomics, functional genomics, comparative genomics and basic theory and application technology of applied biology based on genomics.

All papers chosen for publishing should be innovative research work in field of genomics and applied biology, particular in the research on genomic structure and function, evolutionary and comparative genomics, genomics and bioinformatics, gene expression and its function identification, etc. The primary criteria for publication is new insights that are not only of broad interest to biologists, geneticist and molecular biologist, but also to satisfy the populations of this journals who are engaging in the development and marketing of biotechnology.

BioPublisher, operated by Sophia Publishing Group (SPG), is an international Open Access publishing platform that publishes scientific journals in the field of life science. Sophia Publishing Group (SPG), founded in British Columbia of Canada, is a multilingual publisher.

BioPublisher uses CrossCheck service to identify academic plagiarism through the world's leading plagiarism prevention tool, iParadigms, and to protect the original authors' copyrights.

Latest Content

Inhibition of microRNA-let-7a Increases the Specific Productivity of Antibody-Producing CHO Cell Lines

Lydia Greenlees, Robert Georgantas, III, Jie Zhu, Huifang Dong, Gargi Roy, Jonathan Jacobs, Lori Clarke, Christina Stracener, Hui Feng, Yihong Yao, Michael Bowen, Koustubh Ranade, Katie Streicher

Genomics and Applied Biology, 2014, Vol. 5, No. 1

Expression Analysis of Arabidopsis thaliana Genes *ARGAH1* 、 *ARGAH2* and the Response to Sodium Chloride Stress During Seed Germination

Xiaoxu Zhang, Xinxin Zhang, Tetsuo Takano, Shenkui Liu, Yuanyuan Bu

Genomics and Applied Biology, 2014, Vol. 5, No. 2

Classification of Prostate Cancer with the Use of Artificial Immune System and ANN

Hasibe Cingilli Vural, Seral ÖZŞEN

Genomics and Applied Biology, 2014, Vol. 5, No. 3

Classification of Prostate Cancer with the Use of Artificial Immune System and ANN

Hasibe Cingilli VURAL[1,] ✉ and Seral ÖZŞEN[2]

1. Selcuk University, Department of Biology, Molecular Biology, 42079 Selçuklu, Konya, Turkey
2. Selcuk University,Department of Electrical and Electronics Engineering, 42079 Selçuklu, Konya, Turkey
✉ Corresponding Author email: hcingilli@selcuk.edu.tr

Abstract Before analyzing cells in Laboratory in prostate cancer detection, a classification system can give valuable information about the cancer. The purpose of this paper is to assess the value of Artificial Immune System (AIS) and Artificial Neural Networks (ANN) for classification of prostate cancer cases. Paraffine-embedded prostate cancer tissue specimens of 50 prostate cancer subjects were used in this study. Age range was 35-72 years and all subjects were males. 10 subjects had family history of cancer and 40 patients were non family. An Artificial Immune System (AIS) which is based on clonal selection theory was used to classify these 50 subjects as healthy and patient. With the correct arrangement in system parameters, AIS has reached a classification accuracy of 93.33%. This ratio in 50 data means that in test phase, only one data was misclassified as healthy whereas indeed that data was belonging to a patient. The classification procedure was also done with another method which is a well-known effective classification method for biomedical data: Artificial Neural Networks. The result for this application was 100% with ANN method. While it seems that there is a big difference in the performances of AIS and ANN in the classification accuracy, this difference was only because of 1 data. Thus, it can be said that, AIS is also a good performing classification algorithm as well as ANN for this application.

Keywords Prostate cancer classification; Artificial immune system; Artificial neural networks

Introduction

Prostate cancer is the most common malignancy in men and majority leading cause of cancer deaths in the Eastern world. The genetic predisposition to prostate cancer is well established, as genomic instability is a common feature of many human cancers. Epidemiological studies have suggested that several risk factors (Li, et al., 1997; Steck et al., 1997; Magnusson et al., 1998; Marshall, 1991). Knowing about the genetic markers of prostate cancer in men with prostate cancer diagnosis could help. Inactivation or deletion a large number of tumor-related genes, which otherwise regulate normal cellular growth and suppression of abnormal cell proliferation, is recognized to be one of the major mechanisms of tumorigenesis (Goddard and Solomon, 1993; Waite and Protean, 2002). The successful treatment of prostate cancer relies on detection of the disease at its earliest stages. Although prostate-specific antigen (PSA)-based screening has been a significant advance in the early diagnosis of prostate cancer, identifying specific genetic alterations in a given family or patient

will allow more appropriate screening for early disease. Mapping and identification of specific prostate cancer susceptibility genes is slowly becoming a reality.

Immune System (IS) can be regarded as a defence mechanism of the body. It seeks the condition of body and explores if there is any dangerous situation. If so, the related units are put into effect and necessary processes are held. Artificial Immune System (AIS) is an artificial intelligence method whose roots lie to simple mathematical models developed for the understanding of natural immune system [(L. N. de Castro, and Timmis, 2002; L.N. de Castro, and Von Zuben, 1999; Dasgupta, 1998). Later, with properties like learning, memory, distributed and organized working, etc. in these models, researchers had begun to scrutinize the natural immune system as an inspiration source of AIS. Since that time, many methods modelling or inspiring from some metaphors in natural immune system have been developed and

applied to various problems (Chaudhuri et al., 2007). Medical classification problems are among them. In this study we applied an Artificial Immune System (AIS) to the problem of prostate cancer classification. The used AIS models the clonal selection theory in IS as many other AISs. 17 data from healthy people and 33 data from patients were used for this classification process. The features of data are age, gleason score, personal history of cancer and family history of cancer. The 70% of data was taken for training and the remaining 30% was used for the test process. After analysing AIS with different parameters, best classification accuracy on test data was recorded as 93.33%. In other words, only one data was incorrectly classified.

The other type of classification with prostate cancer data was conducted with artificial neural networks (ANN) (Principe, et al., 1999). When this widely used effective classifier was applied the performance has increased to 100% by correctly classifying the whole test data.

1 Materials and Methods

1.1 Dataset formation

Paraffine blocks of prostate pathologies were derived from the archives of the Department of Pathology in Faculty of Medicine at the University of Selcuk, Turkey. Namely, Paraffine-embedded prostate cancer tissue specimens of 50 subjects were used in this study. Age range was 35-72 years, all subjects were males. 10 subjects had family history of cancer and 40 subjects were non family. These subjects went to physicians to demonstrate a variety of serious symptoms of prostate cancer, e.g., difficulty in voiding, urodynia, urgent and frequent urination, and hematuria. Their prostates were examined by one or more of the following means: rectal ultrasound detection, digital rectal examination, computed tomography, and magnetic resonance imaging. Biopsy was performed for the subjects who were suspected to have prostate cancer, and all specimens were from archived paraffin blocks.

Diagnosis of prostate cancer requires the tissue and cell specimens. These specimens are screened and analyzed by a pathologist using a microscope. Optimum medical treatment is decided according to this information gathered by the pathologist. In some cases, correct diagnosis is very hard and there can be 30-40% difference between pathologists' decisions (Schenck, and Planding, 1998). Dramatic results about wrong diagnosis of cancer cases from biopsy slides can be found in (Kopec et al., 2003). Prostate cancer is evaluated using two staging systems: the Jewett-Whitmore system and the TNM (tumor, node, metastases) system. In TNM system, T refers to the size of the primary tumor, N will describe the extent of lymph node involvement, and M refers to the presence or absence of metastases.

In this study, 50 data with 4 features were used. The features are:

1. Age
2. Gleason score (PSA*)
3. Personal History of Cancer
4. Family History of Cancer

Here, PSA is a protein produced by the prostate gland that can be detected in the blood. Levels rise with age and when the prostate is enlarged. Significantly increased levels of PSA in the blood can indicate prostate cancer. PSA levels are also known to rise in other prostate conditions such as prostatitis (inflammation of the prostate). Normal values of PSA are as the following:

AGE / PSA Value
Under 50 years / < 2.5
50 – 59 years / < 3.5
60 – 69 years / < 4.5
70 years and over / < 6.5

The personal and family history of cancer in 3rd and 4th features are taken as 0 if there is no presence of cancer and 1 if there has been a cancer in the subject or his family. The 17 data belong to the healthy subjects while the remaining 33 data were of subjects with prostate cancer. The 70% of the whole data was taken for training, and 30% was taken for test. The

division of dataset for training and test is shown in Table 1.

Table 1 The number of healthy and patient data in training and test sets

	Training	Test	Total
Healthy	12	5	17
Patient	23	10	33
Total	35	15	50

The AIS system was trained using training data and its performance in the test set was recorded. Also, to have an insight about the result of this study, a widely used method- Artificial Neural Networks- was also applied to the same data.

1.2 Used AIS system

The applied method is a simple AIS algorithm that mimics the clonal selection theory in natural immune system as many other clonal selection-based AIS methods (Ada, and Nossal, 1987; Chen, and Mahfouf, 2006; Cutello et al., 2005; L. N. de Castro, and Von Zuben, 2000; Garrett, 2003; Ong et al., 2005; Perelson, and Oster, 1979; Timmis, and Neal, 2001). The biological base for this theory can be found in (Abbas, 1994). The block diagram of used AIS method is shown in Figure 1.

Here, the system units are named as Antibodies (*Ab*) and inputs that are presented to the system are regarded as Antigens (*Ag*). A random population is formed in the beginning and the affinity of each member of this population to a presented input (*Ag*) is calculated using Euclidean distance criterion given below in Equation 1:

Affinity=1-D ;where

$$D = \frac{1}{L}\sqrt{\sum_{k=1}^{L}\left(Ab_k - Ag_k\right)^2}.$$

Here, Ab_k is the kth feature of *Ab* vector and Ag_k is the kth feature of *Ag* vector.

If the affinity of an *Ab* in the population exceeds a threshold value named as *supp*, that *Ab* is selected for cloning. Cloning is done by simply copying the vector of *Ab* with a number proportional to the *Ab*'s affinity and after that, a clone population is formed. After cloning, a mutation procedure is applied to the some *Ab* clones to have diversity in the population. The mutation is done through changing some values of *Ab* vectors by a new value which is determined randomly. That is, besides of random determination of which *Ab*s will be mutated, the selection of features that will be changed and the determination of new feature values are also conducted in a random fashion. After these processes, a number of best *Ab*s in mutated population is taken for the next iteration (the best *Ab*s are the *Ab*s whose affinities are highest). Also some randomly generated *Ab*s are added to these *Ab*s for the use in next iteration as a beginning population. The iterations are conducted a number of times a memory population is formed using best *Ab*s produced after the iterations conducted for each *Ag*. The class information of that *Ab* is also taken with the same class of presented *Ag*. As a last step, the memory *Ab*s are deleted from the memory population, if the affinity of them is higher that *supp* value with any other memory *Ab* (Figure 1).

Here a memory population including memory *Ab*s and their class information is formed in training. The classification procedure on the other hand is done through finding the nearest memory *Ab* to the presented *Ag* and class decision about that *Ag* is given by looking the class information of that nearest *Ab*.

In our application, we conducted 100 iterations for each presented input data (*Ag*). The 50 *Ab*s were used in the beginning population in each iteration. The determination of correct *supp* value was done in experimentally. That is the *supp* value was changed between 0.9-0.01 and for each experienced *supp* value, a training-test process was conducted to see the resulted test classification accuracy in percentage. Here this accuracy is calculated as:

$$accuracy(T) = \frac{\sum_{i=1}^{|T|} assess(t_i)}{|T|}, t_i \in T$$

$$assess(t) = \begin{cases} 1, & \text{if } classify(t) = t.c \\ 0, & \text{otherwise} \end{cases}$$

where *T* is the set of data items to be classified (the test set), *t∈T*, *t.c* is the class of the item *t*, and classify(*t*) returns the classification of *t* by method. The best *supp* value is determined as the *supp* value for which the test classification accuracy is the highest.

After conducting the classification of prostate cancer data with AIS, an ANN structure was also trained and tested for the same data. One-layer ANN architecture was used and gradient descent learning rule was utilized in the training. The optimum value of learning rate parameter, the number of hidden nodes and the momentum constant are also determined in the experiments to give the highest test classification accuracy. The results of ANN were compared with AIS.
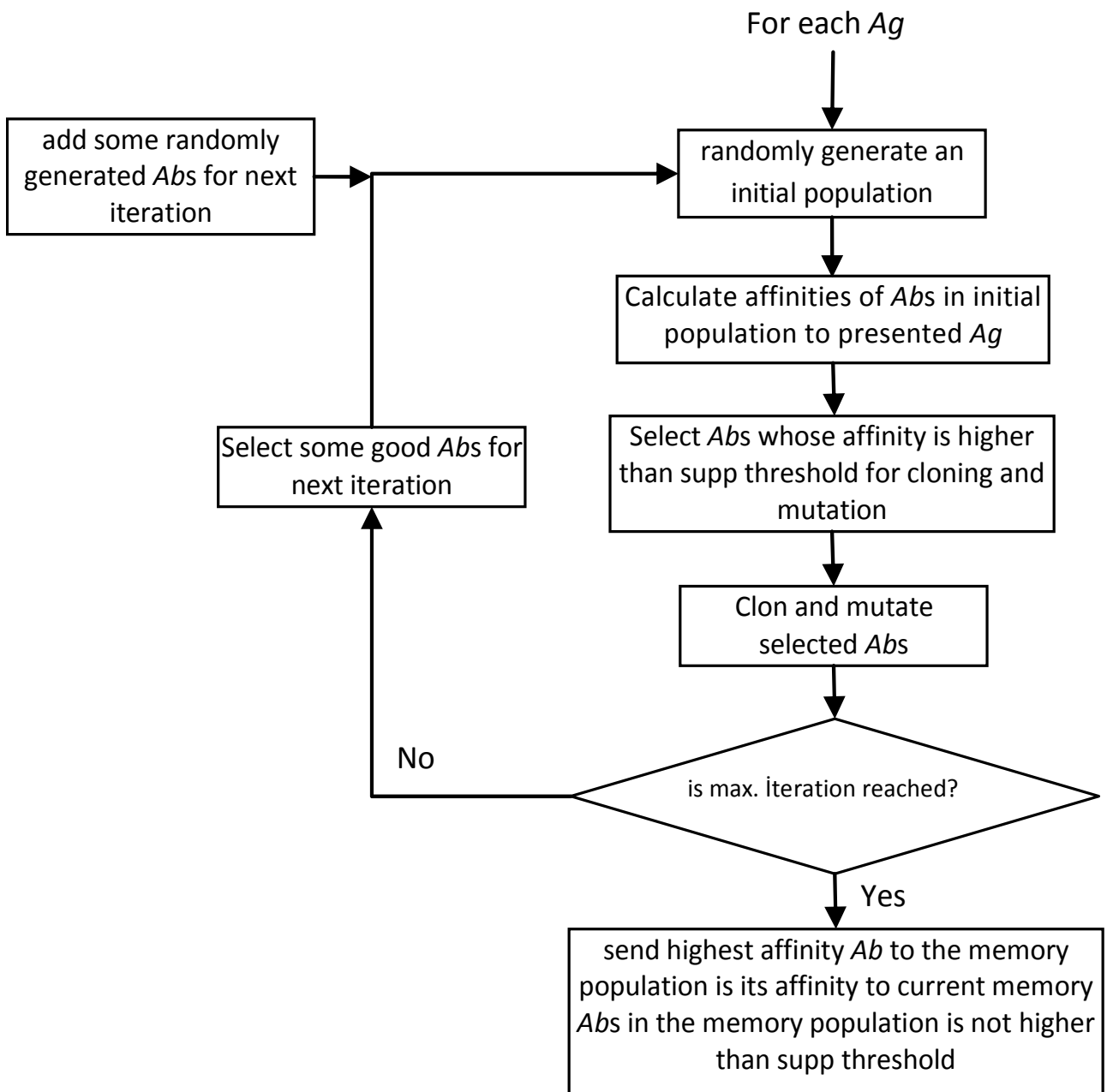

Figure 1 The Flow chart of clonal based AIS system

2 Results

As stated in the previous section, the first application study was conducted with AIS and the optimum value of *supp* value was searched through changing its value in 0.01-0.9 interval to have maximum test classification accuracy. Normally, used data and system units are normalized vectors and the affinity value lie only in the [0-1] interval. So, *supp* value should be selected in that interval and we begun with 0.9 value and decrease it with 0.1. Then, around some values which give high classification accuracies we decreased or increased *supp* value more tenderly. That is, we first had a rough insight about *supp* values and scrutinized some good values in a more detailed way. The test classification accuracies for searched *supp* values are given in Figure 2.
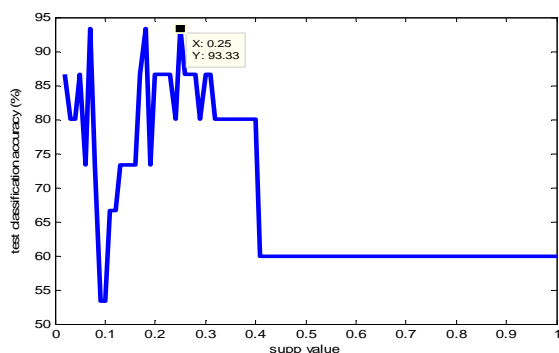

Figure 2 The change of test classification accuracy with regard to the change in *supp* value

As shown from the figure, the maximum test classification accuracy was detected as 93.33% for 0.07, 0.18 and 0.25 values for the *supp* value. Because 15 data in the test procedure is used, this accuracy means that only one data is misclassified by the system. Figure 2 shows that there isn't a specific way of determining *supp* value. That is, we cannot say that the accuracy decreases as *supp* increases or otherwise. The only way of finding best *supp* value is to search for *supp* value through experimentation.

The other application of prostate cancer classification was with ANN method. As stated in Section 2, gradient descent learning algorithm was taken for a one-hidden layered ANN. The searched parameters in this application are optimum number of hidden nodes

(*hn*), learning rate (*lr*) and momentum constant (*mc*). In this search procedure, firstly by fixing *lr* and *mc* to values of 2 and 0.8, *hn* is changed between 1 and 50 by steps of 1and for each experimented *hn* values, the test classification accuracy was recorded. The change of test classification accuracy, according to the *hn* values is shown in Figure 3.

As can be seen from the figure, 100% test classification accuracy was reached for some hidden node numbers like 16, 35, 40 and 45. Thus we took *hn* as 16. Searches for best *lr* and *mc* parameters were also done but, because the 100% was reached, these are not presented here.
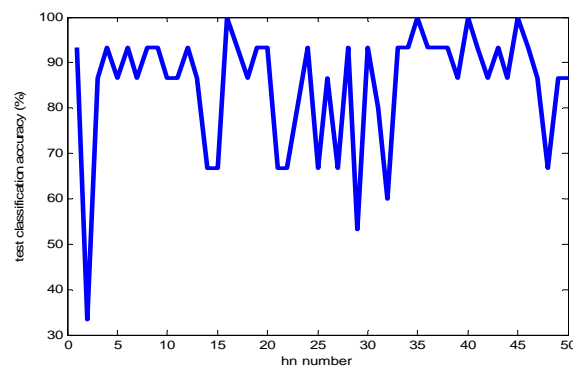

Figure 3 Change of test classification accuracy according to the changing *hn* number

In summary, ANN has reached a higher result than AIS but two points should be emphasized here:

- The difference was only for one test data. That is ANN has correctly classified one more data
- The number of dataset is not satisfactory to have a confidential comparison between two systems.
- Anymore, both methods have well performed for this classification task and encouraged us to use these systems with much more data in real applications of daily life.

3 Discussion

Application of some artificial intelligence methods to the classification of some disease has increasing day by day. While this kind of systems cannot be used in

their own in detecting abnormalities, at least they can help physician in selecting suspicious situations or in deciding his resultant decision.

In our study, we classified prostate cancer data of 50 subjects by AIS and ANN. By taking the clonal selection model in immune system, a code for AIS was written and applied to the dataset for classification. With best system parameters, AIS has reached a test classification ratio of 93.33% by misclassifying only 1 test data. On the other hand ANN has reached 100% accuracy. Whereas it is not possible to do a confidential comparison between AIS and ANN methods for 50 data, we can say that ANN can be searched for more crowded datasets for a real-life application to help doctors.

Acknowledgement

References

Abbas A.K., Lichtman A.H., and Pober J.S., 1994, Celluar and Molecular Immunology, United States of America, W. B. Sounders Com.

Ada G.L. and Nossal G.J.V, 1987, The Clonal Selection Theory, Scientific American, 257: (2), 50-57.
http://dx.doi.org/10.1038/scientificamerican0887-62

Chaudhuri K., Saha S., Azeem R., Balachandran S., Yu S., Majumdar N., and Nino F., 2007, Artificial Immune Systems: A Bibliography, Compiled by Dasgupta, D., CS Technical Report, No. CS-07-004.

Chen J., and Mahfouf, M., 2006, A Populastion Adaptive Based Immune Algorithm for Solving Multi-objective Optimization Problems, Lecture Notes in Computer Science, 4163, 280-293.
http://dx.doi.org/10.1007/11823940_22

Cutello V., Narzisi G., Nicosia G., Pavone M., "Clonal Selection Algorithms: A Comparative Case Study Using Effective Mutation Potentials", Lecture Notes in Computer Scienvce, Vol. 3627, pp. 13-28, 2005.
http://dx.doi.org/10.1007/11536444_2

Dasgupta D., 1998, Artificial Immune Systems and Their Applications, Berlin, Springer-Verlag.

Garrett S.M., 2003, A Paratope is not an Epitope: Implications for Immune Network Models and Clonal Selection",

Lecture Notes in Computer Scienvce, 2787: 217-228.
http://dx.doi.org/10.1007/978-3-540-45192-1_21

Goddard A.D., Solomon B., 1993, "Genetic aspects of cancer", Advances in Human Genetics, 21: 321-376.

Kopec D., Kabir, M.H., Reinharth D., Rothschild O., and Castiglione J.A., 2003, Human Errors in Medical Practice: Systematic Classification and Reduction with Automated Information Systems, J. of Medical Systems, 27: (4) 297-313.
http://dx.doi.org/10.1023/A:1023796918654

Li J., Yen C., Liaw D., Podsypanina K., Bose S.,Wang, S.I., 1997, PTEN a putative protein, Am. J. Hum. Genet. 1943-1947.

L. N. de Castro, Timmis J., 2002, Artificial Immune Systems: A New Compoutational Intelligence Approach, U.K., Springer Ed.

L. N. de Castro and Von Zuben F.J., 1999, Artificial Immune Systems: Part I- Basic Theory and Applications, Technical Report - DCA-RT 02/00.

L. N. de Castro and Von Zuben F.J., 2000, An Evolutionary Immune Network for Data Clustering", Proc. of the IEEE Brazilian Symposium on Artificial Neural Networks, 84-89.
http://dx.doi.org/10.1109/SBRN.2000.889718

Magnusson C., Baron, J., Persson, I., Wolk, A., Bergstrom R., Trichopoulos D., Adami H.O., 1998, "Body size in different periods of life and breast cancer risk in postmenopausal women", Int J Cancer, 76: 29-34.
http://dx.doi.org/10.1002/(SICI)1097-0215(19980330)76:1%3C29::AID-IJC6%3E3.0.CO;2-#

Marshall, C.J., 1991, Tumor suppressor genes, Cell, 64: 313-326.
http://dx.doi.org/10.1016/0092-8674(91)90641-B

Ong, Z.X., Tay J.C., Kwoh C.K., 2005, Applying the Clonal Selection Principle to Find Flexible Job-Shop Schedules, Lecture Notes in Computer Science, 3627: 442-455.
http://dx.doi.org/10.1007/11536444_34

Perelson, A.S. and Oster, G.F., 1979, "Theoretical Studies of Clonal Selection: Minimal Antibody Repertuarie Size and Reliability of Self-Nonself Discrimination", J. Theor. Biol., 81: 645-670.
http://dx.doi.org/10.1016/0022-5193(79)90275-3

Principe, J.C., Euliano, N.R., Lefebvre, W.C., 1999, Neural and Adaptive Systems (Fundemantals Through Simulations), USA, John WILEY & SONS, Inc.

Schenck U., and Planding W., 1998, Quantitation of visual screening technique in cytology, Proc. Image Analysis in

Medicine, II. National Symposium, 7-14.

Steck, P.A., Pershouse, M.A., Jasser, S.A., Yung, W.K., Lin H., Ligon, A.H., 1997, Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers, Nat. Genet., 15: 356-362.

http://dx.doi.org/10.1038/ng0497-356

Timmis J., and Neal M., 2001, A Resource Limited Artificial Immune System", Knowledge Based Systems, 14: 121-130.

http://dx.doi.org/10.1016/S0950-7051(01)00088-0

Waite K.A., and Protean C., 2002, PTEN: form and function, Am J Hum Genet, 70: 829-844.

http://dx.doi.org/10.1086/340026

Reasons to publish in BioPublisher

A BioScience Publishing Platform

★ Peer review quickly and professionally
☆ Publish online immediately upon acceptance
★ Deposit permanently and track easily
☆ Access free and open around the world
★ Disseminate multilingual available

Submit your manuscript at: http://bio.sophiapublisher.com