Communications in Computational and Applied Mathematics, Vol. 1 No. 1 (2019) p. 12-17

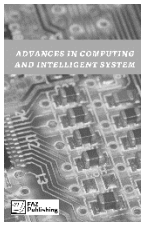
CCAM

FAZ Publishing

Communications in Computational and Applied Mathematics

Journal homepage : www.fazpublishing.com/ccam

e-ISSN : 2682-7468



Pixel Machine Learning with Clonal Selection Algorithm for Lung Nodules Visualization

Pang Yuen Yuen[1], Hang See Pheng[2*], Yan Soon Weei[3]

[1,2,3]Department of Mathematical Science, Universiti Teknologi Malaysia, Skudai, 81310, MALAYSIA

*Corresponding Author

Received 26 December 2018;
Accepted 03 March 2019;
Available online 07 March 2019

Abstract: The early detection of lung nodules is critical to provide a better chance of survival from lung cancer. Since benign/malignant lung cancer may be caused by the growth of lung nodules, the diagnosis of an early detection of lung nodules is important. With rapidly development of advanced technology, detection of lung nodules becomes efficient by utilizing computer-aided detection (CAD) systems that can automatically detect and localize the nodules from computed tomography (CT) scans. CAD is fundamentally based on pattern recognition by extensive use of machine learning approaches which is highly interrelated to mathematical algorithms. In this study, a pixel machine learning algorithm which is developed by artificial immune system (AIS) based algorithm – Clonal Section Algorithm (CSA) is proposed for lung nodules visualization. By using pixel machine learning algorithm, several pre-processing procedures can be avoided to prevent the loss of information from image intensities. It is found that the proposed classification algorithm using original intensity values from CT scans is able to provide reasonable visualization results for lung nodules detection.

Keywords: pixel machine learning, clonal selection algorithm, lung nodules detection

1. Introduction

Lung cancer is one of the top-ranking, or the highest causes of cancer deaths for men and women. Due to the dramatic statistics of lung cancer, the diagnoses of lung cancer always gain the universal attention from medical experts to improve the survival rate. Therefore, the early detection of lung nodules through Computed Tomography (CT) scans is essential to assist the radiologists in decision making and provide early diagnosis in medical treatment. The nodules are the small and rounded or oval shaped appear in the lungs. The size of lung nodule usually is smaller than 3 centimeters in diameter. The nodules can be considered as benign or malignant tumors when they are growing up to more than 3 centimeters. In the past one decade, the lung nodules can be discovered certainly by CT examinations. According to [1], CT scan is commonly used in the diagnostic machine to detect the lung nodules, it generates an extensive data set which includes hundreds pieces images. This requires the radiologists to analyses the data set manually which is time-consuming. Therefore, computerized system is helpful to automated the detection and visualization of lung nodules.

In the automated medical analysis, machine learning approaches are useful in the development of Computer Aided Detection (CAD) [2]. To avoid the loss of information from conventional medical image processing for lesions detection, the pixel machine learning algorithm is proposed in our study for lung nodules detection on CT scans. Learning the pattern of lesions directly though pixels of image could avoid the loss of information during the process of segmentation ad feature extraction. It is also found that the performance of pixel machine learning in visualization results of output image can potentially be better than the conventional classification algorithms. There is no segmentation and features calculation involve to avoid inaccuracy in classification of pixel on medical images. In this study, a pixel machine learning is integrated with an artificial immune system based algorithm-Clonal Selection Algorithm (CSA) to classify the pixels on CT scans for lung nodules visualization. We employed the data from the publicly accessed database which providing CT images with lung nodules and intensive nodules information. The performance of this algorithm will be seen from the output images after the classification of pixels.

**Corresponding author: sphang@utm.my*

2. Pixel Machine Learning

Pixel machine learning is one of the machine learning which can identify the medical images by using original pixel values in medical images. According to [2], pixel machine learning can directly process the original pixels from the medical images. The pixel machine learning can prevent the error such as the incorrect calculation of features in segmentation from the segmented regions as the input information. Fig. 1 show the workflow of pixel machine learning in classification and visualization of medical images.

The pixel machine learning algorithms have been applied in different classifiers such as artificial neural network, linear discriminant analysis, and support vector machine [2]. In the classification process, pixel machine learning machine employing pixels/voxels of medical image directly as the input information. In addition, pixel machine learning is time consuming compared to the conventional feature training. This is due to all the pixels in the training images are involved as features data. Some efforts have been done to reduce the dimension of redundant data in order to increase the efficiency of computation time. Nevertheless, the accuracy of classification results should be optimized while the efficiency can be improved. According to [3], the subdivisions of medical images are transformed into pixels form where each pixel can be classified or categorized as attributes for a specific class. In recent review, the popular pixel machine learning approach to classify medical images is massive-training artificial neural networks (MTANN).

2.1 Massive Training Artificial Neural Network (MTANN)

Artificial neural networks (ANN) is one type of machine learning algorithm. This algorithm mimics the process neuron in human brain. It consists of several non-linear computing elements operating in parallel and they are arranged in patterns alike to biological neural sets. Fig. 2 shows the basic architecture of ANN. The ANN is famous among the researchers for classification of data since it can be modified as different behavior in response to its environment, learn from experience, and extends from previous examples to new ones as stated by [4]. ANN consists of several advantages namely the users do not need the specific problem-solving algorithms, and ANN has inherent generalization ability [5]. In other words, ANN can recognize and react to the patterns that have the similarity but not exactly the patterns in which they have trained. ANN is normally trained to identify the data that is primarily given by experience and intuition of human experts, and involving a high degree of uncertainty [5].

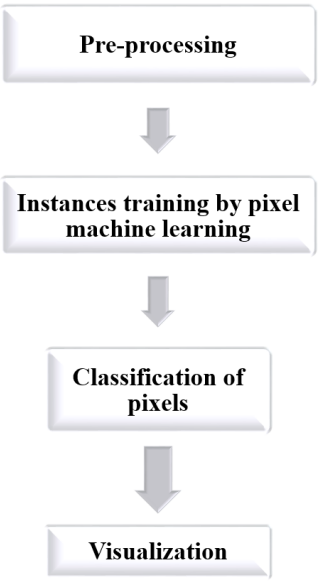


Fig. 1 – Basic work flow of pixel machine learning for medical images classification

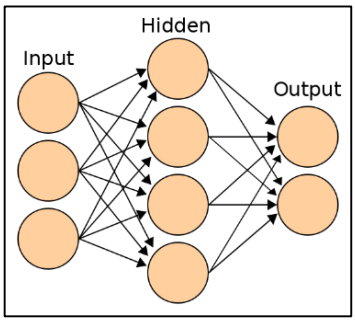


Fig. 2 – Architecture of ANN

By implementing the pixel machine learning, massive-training artificial neural network (MTANN) is built to classify the medical images in order to detect abnormality and lesions [6]. MTANN is able to classify tumors with related genes and classify tumors as benign and malignant [2]. MTANN can solve the high complexity problems and it is suitable to solve complex pattern-recognition tasks. The MTANN can be applied in 2-dimensional (2D) and 3-dimensional (3D) classification for differentiating a specified pattern of an object. For instances, the 2D MTANN has been applied for the detection of lung nodules from CT images [2]. This algorithm involves expensive computational time during the data training.

3. Artificial Immune System

The Artificial Immune System (AIS) is one of the computational intelligences, rule-based machine learning principle and an inspiration system for the actions of the vertebrate immune system [7]. AIS has been used or applied to solve problem in different research areas (such as the functional optimization, virus detection, and classification) These types of algorithms commonly explain the problem of simulating the learning and memory characteristics of the immune system. There are three typical types of intelligent computational algorithms in AIS such as clonal selection algorithm, dendritic cell algorithm, and negative selection algorithm [8].

AIS is having its advantages in solving classification problem due to its affluent metaphorical artificial complement is provided from AIS's learning, memory, feature extraction and recognize the features patterns [8]. Therefore, it is highly distributed, adaptive and self-organizing to solve convolution computational problems. According to [9], the immune system consists of many specialized cells and molecules as well as immune organs, providing a place for the maturation and function of immune cells. The relationship between immune cells and other cells in the body produces rich and complex immune behaviors that lead to the identification of pathogens and the arousal of appropriate pathogen elimination responses.

3.1 Clonal Selection Algorithm (CSA)

The clonal selection algorithm (CSA) is one of the computational algorithms from AIS. According to [8], this algorithm describes the idea for selecting cells which are T-cells and B-cells when recognizing the antigen proliferate. It is used in affinity maturation by implementing several elementary principles of clonal selection process. This can lead to provide reliable solution for pattern matching [12].

3.2 Design of Clonal Selection Algorithm

The CSA is built based on explanation of the basic reaction of the adaptive immune system to antigenic stimulation. In the human immune system, the B-cells come from the bone marrow which helping the body to grow. A B-cell (parent cell) will generate the daughter cell of B-cell, and every single daughter cell is different from the other daughter cells [10]. The parent cell has a unique receptor while each daughter cell will also have its own unique receptor. Moreover, these unique receptors will identify and react with a unique bacterium. In other words, there be one receptor will react to this unique bacterium. In the other hand, T-cell also has a similar process to B-cell that occur in the thymus. Fig. 3 and Fig. 4 illustrate the architecture of B-cell and T-cell respectively. The major different between B-cell and T-cell is T-cell does not become an antibody, but B-cell does.

According to [8], the theory states that the clonal expansion of the original lymphocyte happens when the primary lymphocyte is activated by binding to the antigen. The activated antigen will then generate the clone lymphocytes that express the same receptor as the primary lymphocytes, and this implies that the primary lymphocytes clonal expansion is occurring. This approach explains a population of adaptive information and procedure of cell selection in immune system It also illustrates the resultant duplication and variation ultimately improves the adaptive fit of the information units to their environment. The phenomena of the clonal process had been transformed into machine learning algorithm for classification purpose.

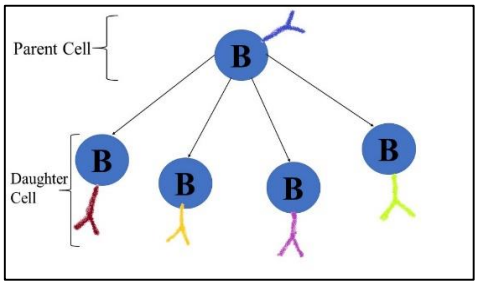


Fig. 3 – Architecture of B-Cell

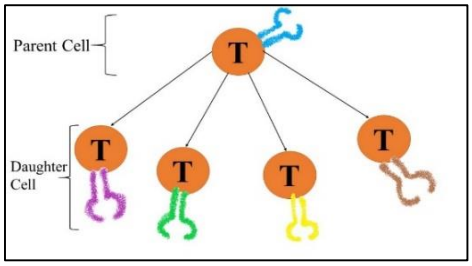


Fig. 4 – Architecture of T-Cell

4. Pixel Machine Learning in Clonal Selection Algorithm (CSA)

In this research, the CSA is proposed to be implemented in the lung nodules detection by using pixel machine learning approach. The framework of CSA is represented in Fig. 5. The algorithm aims to develop a memory pool of antibodies that provides a solution to pattern recognition problem where an antibody represents an element of a solution for a single solution to a problem, and antigens represent an element of the problem space. The algorithm needs to be initialized where a pile of antibodies is prepared by the fixed size, N. These antibodies are separating into two groups which are memory antibody m where represented by the solution of the algorithm and remaining antibody r where introduce additional diversity into the system [11]. Next, the algorithm will proceed into the looping process by calculating the iterations number of exposing the system to all known antigens. The number of iterations can be stopped through some specific condition. In this algorithm, the stopping benchmark is a predefined maximum number of generations [12].

4.1 The Clonal Selection Algorithm with Pixel Machine Learning in Pixels Classification

The training and testing images are collected from the Reference Image Database to Evaluate Therapy Response

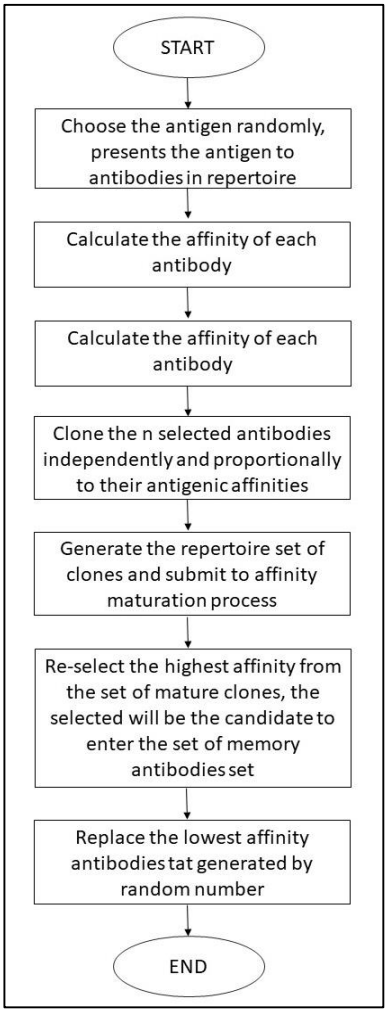


Fig. 5 – Framework of CSA

(RIDER) project. The RIDER dataset can be downloaded from the open source website - The Cancer Imaging Archive (TCIA) [13]. TCIA is an official and licensed website that can be made available to researchers. The database contains diagnostic data for 32 lung nodules patients which are generated by CT scans. To prepare the training, the CT images with presence of oval-shape nodules are chosen. These nodule images are recognized as region of interest (ROI) in the experiments. In the pre-processing, sub-regions of image with lung nodules are cropped and they are presented as features for the formation of the teaching image. The preparation for collect ROI data is done by MATLAB software. Next, the teaching image is created by the original intensities from CT images attributes and the training classes are generated by Gaussian function. The preparation of teaching image, $T(x, y)$ is formed by a piecewise function that including the Gaussian distribution represents nodule cases and otherwise, zero intensity for non-nodules cases (Equation 1). The matching process of extracted ROI with training class is shown in Fig. 6. In the procedures, the original intensity values are matched with the Gaussian values as training classes.

$$T(x,y)=\begin{cases}\dfrac{1}{\sqrt{2\pi\sigma_T}}\exp\left\{-\dfrac{x^2+y^2}{2\sigma_T^2}\right\}, & \text{for a nodule,}\\ 0, & \text{otherwise.}\end{cases} \quad (1)$$

The size of the two dimensional (x × y) teaching image can be adjusted by using the standard deviation, σ in the Gaussian distribution function. The adjustment of standard deviation in formulation of teaching image is based on the size of the ROI file. From the Gaussian distribution, the generated values are between 0 to 1 where 0 represents the dark color and 1 represents the bright white color. The teaching image is saved as the matrices with the size of 50×50 where are total of 2500 training instances are formed in each training file. This process is repeated to obtained testing images as an input of classification model.

The data training and testing are done by means of an open source software - Waikato Knowledge Analysis Environment (WEKA). WEKA is widely used for data training and data classification. It provides a collection of machine learning algorithms to perform data mining tasks. A built-in CSA classifier – CLONALG algorithm is chosen for training and testing purpose. The classification results are generated by the classifier in WEKA. Then, the prepared testing input files are used to test the classification of the CSA computational model.

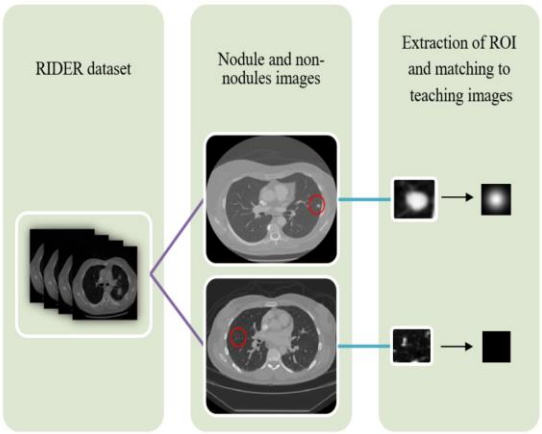


Fig. 6 – The extraction of ROI and training image preparation.

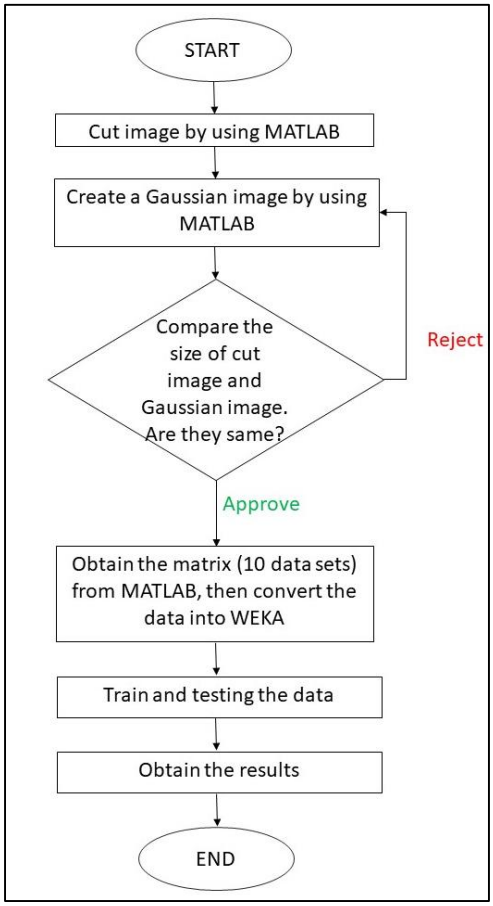


Fig. 7 – Flowchart for pixel machine learning with CSA framework

The parameter of the algorithm in the WEKA can be adjusted based on the types of training data. In the experiment, the trained model is built for the classification of pixels for lung nodule images. The testing data is also formed from the RIDER dataset using the same procedures in training image preparation except the training class matching The classification results can be obtained and the performance for the computational model and they are evaluated through the visualization of output images. Fig. 7 shows the procedure used in this research for the classification of lung nodules. The "cut image" in the framework is the ROI that consists of lung nodule. In the training data, the sub-region matrix with the size of 3×3 is formed as features for each training instances. A sample of matrix represent a sub-region, W which can be written as Equation 2. Each training region contains nine features and mapped with a class with the values in six decimal places. The sample of training file that consists of header and training instances with class is shown in Fig. 8. The class values are obtained from the Gaussian function. These data is viewed in as text file and it is saved as ".arff" according to the format required by WEKA software.

$$W=\begin{bmatrix}611 & 433 & 153\\ 696 & 542 & 259\\ 869 & 603 & 320\end{bmatrix} \quad (1)$$

```
@relation lungnoduledetection

@attribute F1 numeric
@attribute F2 numeric
@attribute F3 numeric
@attribute F4 numeric
@attribute F5 numeric
@attribute F6 numeric
@attribute F7 numeric
@attribute F8 numeric
@attribute F9 numeric
@attribute class {0.004006,0.005042,0.006282,...,0.000000}
@data

611,433,153,696,542,259,869,603,320,0.004006
433,153,203,542,259,144,603,320,197,0.005042
153,203,360,259,144,288,320,197,397,0.006282
203,360,301,144,288,252,197,397,385,0.007750
360,301,274,288,252,189,397,385,264,0.009466
.
.
.
0,0,0,0,0,0,0,0,0,0.000000
```

Fig. 8 – Sample of training file in WEKA

```
@relation lungnoduledetection

@attribute F1 numeric
@attribute F2 numeric
@attribute F3 numeric
@attribute F4 numeric
@attribute F5 numeric
@attribute F6 numeric
@attribute F7 numeric
@attribute F8 numeric
@attribute F9 numeric
@attribute class {0.004006,0.005042,...,0.000000}
@data

1165,1148,963,1131,1195,978,1097,1206,1065,?
1148,963,375,1195,978,212,1206,1065,398,?
963,375,190,978,212,31,1065,398,200,?
.
.
.
0,0,0,0,0,0,0,0,0,?
```

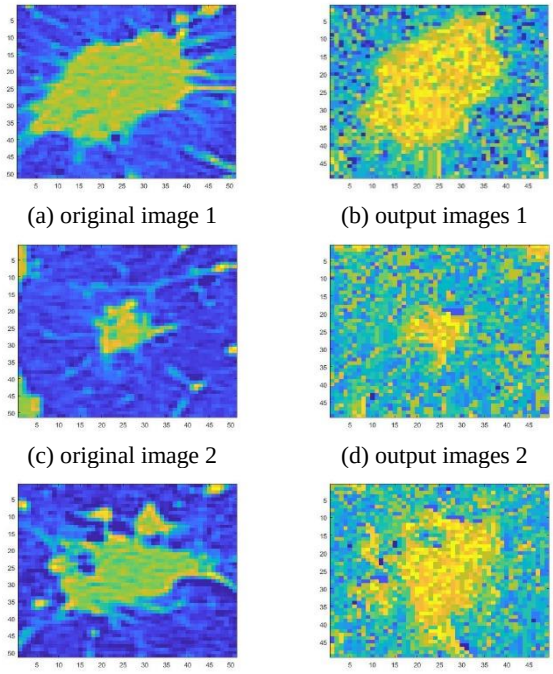
Fig. 9 – Sample of testing file in WEKA

For the testing data, all instances are classified by using the CLONALG algorithm in WEKA. This mean that each the small sub-regions will be given a class value after the classification. Fig. 9 shows the sample of input data to WEKA in order to obtain the classification results for each sub-region. The attributes in testing file are mapped "?". Every "?" symbol in testing file represents unknown class before classification result is obtained.

For the testing data, all instances are classified by using the CLONALG algorithm in WEKA. This mean that each the small sub-regions will be given a class value after the classification. Fig. 9 shows the sample of input data to WEKA in order to obtain the classification results for each sub-region. The attributes in testing file are mapped "?". Every "?" symbol in testing file represents unknown class before classification result is obtained.

5. Results and Discussion

From pixels training and testing processes by WEKA, the class for each of the attributes is generated by CSA algorithm. The class values are then transformed into an output image. Fig. 10 shows the sample of three set input and output images for lung nodules. Fig. 10 (a), (c) and (e) are three different size of original medical images extracted from CT scans while Fig. 10 (b), (d) and (f) are output images which are generated by class values obtained from CSA classification. From the visualization results that formed by 50×50 classes, it can be seen that the pixels of the output images are able to highlight the area of lung nodules. However, there exists some indistinct small false positives found at the surrounding nodules in the classified images. These false positives can be eliminated by improving the optimization of parameter selection in pixel machine learning and CSA algorithm.



(a) original image 1 (b) output images 1

(c) original image 2 (d) output images 2

(e) original image 3 (f) output images 3

Fig. 10 –Visualization of input and output images

6. Conclusion

As a conclusion, pixel machine learning with clonal selection algorithm could provide visualization of detected lung nodules from the testing images. The implementation of the proposed algorithm also found revealing better performance in classification compared to conventional method in term of visualization. Furthermore, the reliability of pixel machine learning algorithms is achieved by using appropriate parameters where the adjustment of parameters is sometimes important. The detection of lung nodules in the computation results are crucial to alert the CAD user to identify the lung nodules as abnormal tissue on CT scans.

Acknowledgement

The authors wish to thank the Malaysian Ministry of Education (MOE), Universiti Teknologi Malaysia (UTM) and Research Management Centre (RMC) for financial sponsorship to this work through grant funding number Q.J130000.2754.03K40.

References

[1] Zhao, B., G. Gamsu, M.S. Ginsberg, L. Jiang, & L.H. Schwartz (2003). Automatic detection of small

lung nodules on CT utilizing a local density maximum algorithm. journal of applied clinical medical physics, 4(3), 248-260.

[2] Suzuki, K. (2012). Pixel-based machine learning in medical imaging. Journal of Biomedical Imaging, 2012, 1.

[3] Arganda-Carreras, I., V. Kaynig, C. Rueden, J. Schindelin, K.W. Eliceiri, A. Cardona, et al. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics, 33(15), 2424-2426.

[4] Chaplot, S., L. Patnaik, & N. Jagannathan (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomedical signal processing and control, 1(1), 86-92.

[5] Benardos, P. & G.-C. Vosniakos (2007). Optimizing feedforward artificial neural network architecture. Engineering Applications of Artificial Intelligence, 20(3), 365-382.

[6] Khan, J., J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, et al. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature medicine, 7(6), 673.

[7] Xu, C., S. Xu, & W. Chen (2008). Artificial Immune System and its applications in gps single frequency precise point positioning, Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on. 180-183.

[8] Dasgupta, D., S. Yu, & F. Nino (2011). Recent advances in artificial immune systems: models and applications. Applied Soft Computing, 11(2), 1574-1587.

[9] Timmis, J., P. Andrews, N. Owens, & E. Clark (2008). An interdisciplinary perspective on artificial immune systems. Evolutionary Intelligence, 1(1), 5-26.

[10] Leandro, N. (2002). de Castro and Jonathan Timmis," Artificial Immune Systems: A New Computational Intelligence Approach", Springer Verlag, Berlin.

[11] Brownlee, J. (2005). Clonal selection theory & clonalg-the clonal selection classification algorithm (csca). Swinburne University of Technology,

[12] Ulutas, Berna Haktanirlar, & S. Kulturel-Konak (2011). A review of clonal selection algorithm and its applications. Artificial Intelligence Review, 36(2), 117-138.

[13] Clark, K., B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging, 26(6), 1045-1057.