MASSIVE TRAINING ARTIFICIAL IMMUNE RECOGNITION SYSTEM FOR LUNG NODULES DETECTION

HANG SEE PHENG

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

OCTOBER 2014

To my beloved parent and husband for their support and encouragement during this graduate program.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my main supervisor, Professor Dr. Siti Mariyam Shamsuddin for her guidance and dedication towards the completion of my thesis. I would like to extend my sincere appreciation to my co-supervisor, Dr. Kenji Suzuki for his assistance in achieving the objectives of this thesis. Besides, I would like to thank Professor Anca L Ralescu for her advice and suggestions in my thesis.

In addition, I would like to thank the School of Graduate Studies (SPS) for the financial support in my studies. This study is also partially supported by The Ministry of Higher Education (MOHE) under Long Term Research Grant Scheme (LRGS/TD/2011/UTM/ICT/03, VOT 4L805). Furthermore, I would like to express my gratitude to Dr Jahari bin Mahbar in the Department of Radiology and Imaging, KPJ Johor Specialist Hospital for his advice in this study. I am also grateful to the Faculty of Computing and library of UTM for providing various facilities and references. Besides, I would also like to thank the developers of the utmthesis LaTeX project for making the thesis writing process a lot easier for me.

Last but at least, I would like to thank my beloved family for their great moral support and to my friends for their contributions in completing my thesis.

ABSTRACT

In the early detection and diagnosis of lung nodule, computer aided detection (CAD) has become crucial to assist radiologists in interpreting medical images and decision making. However, some limitations have been found in the existing CAD algorithms for detecting lung nodules, such as imprecision classification due to inaccurate segmentation and lengthy computation time. In this research, Massive Training Artificial Immune Recognition System (MTAIRS) is proposed to detect lung nodules on Computed Tomography (CT) scans. MTAIRS is developed based on the pixel machine learning and artificial immune-based system-Artificial Immune Recognition System (AIRS). Two versions of proposed algorithms have been investigated in the study: MTAIRS 1 and MTAIRS 2. Since segmentation and feature calculation are not implemented in the pixel-based machine learning, the loss of information can be avoided during the data training in MTAIRS 1 and MTAIRS 2. The experiment and analysis find that MTAIRS 1 and MTAIRS 2 have successfully reduced the computation time and accomplished good accuracy in the detection of lung nodules on CT scans compared to other well-known pixel-based classification algorithms. Furthermore, MTAIRS 1 and MTAIRS 2 are investigated to improve their performance in eliminating the false positives. A weighted non-linear affinity function is employed in the training of MTAIRS 1 and MTAIRS 2 to replace Euclidean distance in affinity measurement. The enhanced algorithms named, E-MTAIRS 1 and E-MTAIRS 2 are capable to reduce the false positives in the non-nodule classification while maintaining the accuracy in nodule detection. In order to further provide comparative analysis of pixel-based classification algorithms in lung nodules detection, a pixel-based evaluation method of Kullback Leibler (KL) divergence is proposed in this study. Based on the pixel-based quantitative analysis, MTAIRS 1 performs better in the elimination of false positives, while MTAIRS 2 in lung nodules detection. The average detection accuracy for both MTAIRS algorithms is 95%.

ABSTRAK

Dalam pengesanan awal dan diagnosis nodul peparu, pengesanan berbantukan komputer (CAD) telah menjadi penting untuk membantu ahli radiologi dalam mentafsirkan imej perubatan dan membuat keputusan. Walaubagaimanapun, terdapat keterbatasan yang telah ditemui dalam algoritma CAD sedia ada dalam mengesan nodul peparu seperti ketakpersisan pengelasan yang disebabkan oleh segmentasi yang tidak tepat dan masa penghitungan yang panjang. Dalam kajian ini, Sistem Pengecaman Imun Buatan Latihan Besar (MTAIRS) telah dicadangkan untuk mengesan nodul peparu pada imbasan Tomografi Berkomputer (CT). MTAIRS dibangunkan berdasarkan kepada pembelajaran mesin piksel dan sistem imun buatan - Sistem Pengecaman Imun Buatan (AIRS). Dua versi algoritma dicadanganakan telah diselidik dalam kajian ini: MTAIRS 1 dan MTAIRS 2. Oleh kerana segmentasi dan penghitungan fitur tidak dilaksanakan dalam pembelajaran mesin berasaskan piksel, kehilangan maklumat boleh dielakkan semasa latihan data pada MTAIRS 1 dan MTAIRS 2. Ujikaji dan analisis mendapati MTAIRS 1 dan MTAIRS 2 telah berjaya mengurangkan masa penghitungan dan mencapai ketepatan yang baik dalam mengesan nodul peparu pada imbasan CT berbanding dengan algoritma pengelasan bersandarkan piksel lain yang terkenal. Tambahan pula, MTAIRS 1 dan MTAIRS 2 telah dikaji untuk meningkatkan prestasi mereka dalam menghapuskan positif palsu. Fungsi afiniti tak linear berpemberat digunakan dalam melatih MTAIRS 1 dan MTAIRS 2 bagi menggantikan pengukuran afiniti jarak Euclidean. Algoritma yang dipertingkatkan dinamakan sebagai E-MTAIRS 1 dan E-MTAIRS 2 dan algoritma ini berupaya untuk mengurangkan positif palsu dalam pengelasan bukan nodul di samping mengekalkan ketepatan dalam mengesan nodul tersebut. Bagi mendapatkan analisis perbandingan lanjutan terhadap algoritma pengelasan berasaskan piksel bagi pengesanan nodul peparu, kaedah penilaian bersandarkan capahan *Kullback Leibler* (KL) telah dicadangkan. Daripada analisis kaedah penilaian berasaskan piksel, MTAIRS 1 menunjukkan prestasi yang lebih baik dalam menghapuskan positif palsu manakala MTAIRS 2 adalah lebih baik dalam pengesanan nodul peparu. Purata ketepatan pengesanan bagi kedua-dua MTAIRS algoritma adalah 95%.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	–	Artificial Intelligence
AIRS	–	Artificial Immune Recognition System
AIS	–	Artificial Immune System
ANN	–	Artificial Neural Network
ARB	–	Artificial Recognition Ball
CAD	–	Computer Aided Detection
CT	–	Computed Tomography
DICOM	–	Digital Imaging and Communications in Medicine
DVDM	–	Discretised Value Difference Metric
E-MTAIRS	–	Enhanced Artificial Immune Recognition System
FP	–	False Positive
GA	–	Genetic Algorithm
GLCM	–	Grey Level Co-occurrence Matrix
GPU	–	Graphics Processor Unit
HEOM	–	Heterogeneous Euclidean-overlap Metric
HU	–	Hounsfield values
HVDM	–	Heterogeneous Value Difference Metric
KL	–	Kullback Leibler
kNN	–	*k*-nearest-neighbor
LDA	–	Linear Discriminant Analysis
LIDC	–	Lung Image Database Consortium
MATLAB	–	Matrix Laboratory
MIPAV	–	Medical Image Processing, Analysis, and Visualisation
MRF	–	Markov Random Field
MRI	–	Magnetic Resonance Imaging
MTAIRS	–	Massive Training Artificial Immune Recognition System
MTANN	–	Massive Training Artificial Neural Network
NA	–	Not Available

NCI	–	National Cancer Institution
NIBIB	–	National Institute of Bioimaging and Bioengineering
NMR	–	Nuclear Magnetic Resonance
PET	–	Positron Emission Tomography
RIDER	–	Reference Image Data to Evaluate Response
ROC	–	Receiver-Operating Characteristic
ROI	–	Region of Interest
RSNA	–	Radiological Society of North America
s-MTANN	–	Standard Massive Training Artificial Neural Network
SOM	–	Self Organizing Maps
SVM	–	Support Vector Machine
USG	–	Ultrasonographic
VDM	–	Value Difference Metric
WHO	–	World Health Organisation
XML	–	Extensible Markup Language

LIST OF SYMBOLS

ag	–	Training antigen in AIRS
n	–	Number of training vector
mc_{match}	–	The best match memory cell
Mc	–	Memory cell
$MC_{ag.c}$	–	Memory pool with class c
d	–	Dimension of vector
X_t	–	Testing vector in AIRS algorithm
X_i	–	Training vector in AIRS algorithm
p_i	–	Probability distribution for image P
q_i	–	Probability distribution for image Q
T	–	Two-dimension teaching image
σ_T	–	Standard deviation for teaching image
ν	–	Training vector in MTAIRS
w	–	Weight in distance function
δ_j^{ti}	–	Distance between input and training points along j th feature
$Dist$	–	Distance in non-linear affinity function
D	–	Divergence function
Y_T	–	Predicted standard deviation in regression model
X	–	Size of nodules in regression model
Y	–	Actual standard deviation in regression model
σ_{est}	–	Standard error in estimation
R	–	Correlation coefficient
ω	–	Size of window for training sub-region
O	–	Output image
HO	–	Original testing image
HA	–	Computed image by MTAIRS
$h0_i$	–	Probability distribution of image HO
hA_i	–	Probability distribution of image HA

J_{max}	–	Argument of the maximum column in KL divergence
I_{max}	–	Argument of the maximum value in KL divergence
ϕ	–	Null set
z	–	Distance between concerned training point and each training instances

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

The recent statistics reported by World Health Organisation (WHO) shows that lung cancer continues to contribute the highest number of death among other types of cancer in the world (WHO, 2014). Lung cancer is also known as the most common cancer for both men and women in many Asia countries as well as the developing countries (Liam *et al.*, 2006). In Malaysia, the National Cancer Registry of Malaysia has shown that lung cancer is the third most common cancer in the country (Sachithanandan and Badmanaban, 2012). Due to the dramatic statistics of lung cancer, its diagnoses have always gained global attention from medical experts in an attempt to improve the patients' survival rate. The survival rate of lung cancer patients can possibly be reduced if the cancer is detected in the early pre-clinical stage (Jemal *et al.*, 2005). In modern medical diagnosis, screening is an effective manner to discover lung lesion in the early stage, and hence reduces the lung cancer mortality (Alex, 2005, WHO, 2014). Thus, to view the condition of human body noninvasively through screening, medical imaging plays an important role to detect the early symptoms of lung cancer.

Recently, the development of medical imaging by various technologies is rapidly emerging in the field of medical diagnosis. There are different types of medical modalities that have been actively used, such as, planar X-ray, computed tomography (CT) scans, magnetic resonance imaging (MRI), ultrasonographic diagnostics (USG), positron emission tomography (PET) and nuclear magnetic resonance (NMR) (Alex, 2005). Compared to conventional planar X-ray images and other modalities, CT images provide accurate visualisation of lung conditions especially in the lung nodules detection. Lung nodule is a pathological spot, measuring three centimeters or less in diameter on the lung. The growth of a nodule larger than three centimeters is considered as lung mass which is more likely to represent lung cancer. Therefore, early detection of lung nodules is crucial as they are likely to be curable lung cancer (Delogu *et al.*, 2005). With a CT scan, a series of cross-sections is obtained and analysed by a radiologist to detect lung nodules. The interpretation of CT images requires significant time from radiologists, as hundreds of multi-sections per patient is acquired from a CT scan. Consequently, an automated system is crucial in assisting the detection and classification of disease patterns in the lung medical images.

Computer aided detection (CAD) is an automated system used to enhance the performance of radiologists and thus, reduces their workload. As reported by Girvin and Ko (2008), performance of radiologists in the detection of lung nodules has been significantly increased by mean of CAD schemes when interpreting a large-scale dataset. In recent decade, research efforts have been focusing on the development of CAD system that could accurately recognise lesions in CT imagery, and consequently reduces false positive detection in the existing system. In the detection of lung nodules, computational intelligence techniques is constantly applied as these techniques provide promising results (Shiraishi *et al.*, 2011, Suzuki *et al.*, 2005b). It was reported that a variety of intelligence techniques were also successfully used in different types of medical images processing (BagcI *et al.*, 2012, Jiang *et al.*, 2010, Shi and He, 2010, Stoitsis *et al.*, 2006). In this chapter, the problem background of computational intelligent algorithms in CAD in detecting lung nodules will be discussed in the following section. It is followed by the discussion on the problem statements, the objectives of the study, the research contributions, the research methodology and the scope of the study. Lastly, the organisation of chapters is summarised in this thesis.

1.2 Background of Problem

Since the majority of imaging modalities provide a large number of datasets, the interpretation of medical imaging by automatically classifying medical images is crucial. As an automatic system, CAD has become one of the main research concerns in medical imaging and diagnostic radiology. In the current lung nodules assessment by CAD, the computerised systems are able to detect lung nodules and categorise the types of nodules, for example malignant and benign. Thus, the improvement of efficiency and accuracy in lung nodules identification has always gained interest of radiologists and researchers alike. In order to classify these abnormalities, pattern recognition techniques are usually adopted in the computerised detection system (BagcI *et al.*, 2012, Kilic *et al.*, 2009, Shiraishi *et al.*, 2011, Stoitsis *et al.*, 2006).

There are several pattern recognition techniques available in CAD scheme, such as, methods based on decision theory, structural methods and computational intelligence methods (Ogiela and Tadeusiewicz, 2008). As a subset of pattern recognition approaches, computational intelligence algorithms are popular since they provided promising results and more efficient diagnosis (Stoitsis *et al.*, 2006, Verma and Zakos, 2001). There are a number of typical computational intelligence approaches used to analyse medical imaging such as artificial neural network (ANN) (Alex, 2005, Suzuki, 2009, Suzuki *et al.*, 2003a, 2005a,b), genetic algorithm (GA) (Ghosh and Mitchell, 2006, Jennane *et al.*, 2010), support vector machine (SVM) (Abdullah *et al.*, 2011, Kakar and Olsen, 2009, Shao *et al.*, 2010, Xing-li and Xu, 2008) and artificial immune system (AIS) (Li *et al.*, 2009, Wang *et al.*, 2008, Zhou *et al.*, 2006). Nevertheless, these algorithms have their own limitations. For example, GA takes longer time to converge to optimum output, and the convergence time could not be estimated during the image analysis (Karkavitsas and Rangoussi, 2005). Based on the statistical learning theory, SVM also provide efficiency performance, but it is scale sensitive and time consuming in processing large scale of images (Kakar and Olsen, 2009, Zhang and Ma, 2008). In addition, ANN provides high performance results and achieves high identification rate in image classification, but it requires complex computation and large amount of data input for training (Shi and He, 2010).

Suzuki (2009), Suzuki *et al.* (2003a), Suzuki and Doi (2005), Suzuki *et al.* (2005b) employed massive training artificial neural network (MTANN) model for pattern analysis on lung CT scans. Based on their expertise in radiology, the computational model was successfully trained to differentiate almost all types of lung nodules with massive training instances. The architecture of MTANN is described in Figure 1.1 (Suzuki *et al.*, 2005a), whereby the sub-region with Hounsfield values(HU) was extracted from region of interest (ROI) to be the input of MTANN. A large number of overlapping sub-regions were extracted from the ROI to form training vectors in the learning process of the computational model. Then, the training vectors were matched with each pixel on teaching images to generate training instances. In the training of nodule case, the teaching images were produced by the Gaussian function with standard deviation as key parameter. The standard deviation determined the distribution at the center of teaching images which represented the size of lung nodules. Since the MTANN was able to produce regular-sized distribution for different diameters of nodules, only single value for standard deviation had been used during the training process. Back propagation algorithm performed the massive training for all the extracted sub-region. This well-trained model gave promising results and low false positive rate in the classification without segmentation. However, the computation time for training took extremely long 30 hours.

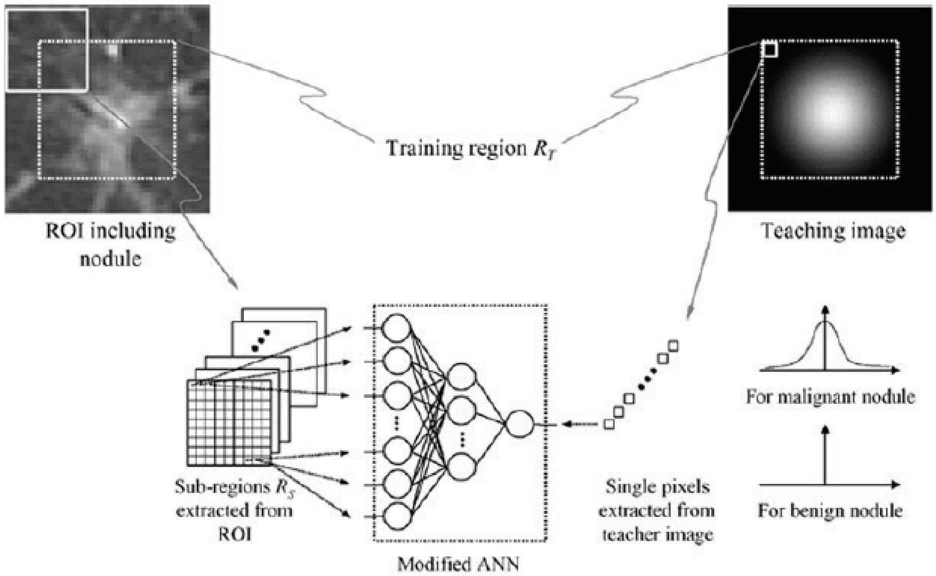

Figure 1.1: Architecture of MTANN (Suzuki *et al.*, 2005a)

Furthermore, inaccurate segmentation is a common limitation in most of the existing CAD algorithms when classifying lung nodules. The segmentation procedures in CAD algorithm development are difficult as they can affect the accuracy of features extraction and classification results. Meanwhile, many research have been conducted to improve the segmentation of lung nodules as this provides significant impacts in the subsequent process. It was reported that CAD system missed some nodules due to the inaccurate lung segmentation (Eva *et al.*, 2009). This shows that inappropriate segmentation algorithm may directly cause a decrease in accuracy in lung nodules detection. For instance, segmentation process may exclude low contrast lung nodules adjacent to blood vessels and airways, and nodules with Hounsfield values similar to blood vessels and airways, and cases where both densities as well as shapes of the lungs are altered by abnormalities (Delogu *et al.*, 2005, Sluimer *et al.*, 2006, van Rikxoort *et al.*, 2009). Therefore, there is a challenge in handling the segmented algorithm which will guarantee satisfying results in detection of lung nodule. In addition, most of the feature training in CAD development is based on the calculated features, but these statistical features will cause the loss of original information in medical images. Therefore, appropriate features selection in computerised system is still a main concern for researchers. To avoid the loss of information from the original intensities, the pixel machine learning was proposed by Suzuki *et al.* (2005a) in the training of CAD algorithm. Nevertheless, this approach required huge computational power to handle massive training of instances which were formed by the original intensities extracted directly from medical images. Due to the huge number of sub-volume in the training of MTANN, several dimension reduction approaches such as Laplacian-eigenfunction-based, support vector regression and principal component analysis were incorporated with the algorithm to increase the efficiency of training process (Suzuki *et al.*, 2010a,b, Xu and Suzuki, 2011a). However, the use of these dimension reduction approaches increased the complexity of pre-processing steps in the development of algorithm. Thus, selection of appropriate and intelligent computational algorithm is crucial to ensure that the computational cost be reduced efficiently for the massive training approach.

In this research, Artificial Immune Recognition System (AIRS) is investigated to be applied in the pixel machine learning. AIRS is one of the intelligent algorithms that can reduce the redundant features in training, as well as provide promising classification results. This machine learning algorithm mimics the several processes in biological immune system, such as, mutation, immune memory and recognition of antigen. In the algorithm, the Euclidean distance is used as affinity measure representing the antibody-antigen interaction. Based on previous works by Watkins *et al.* (2004), there were two versions of AIRS being developed to solve the classification problem. AIRS was named as AIRS1 after AIRS2 was developed. AIRS2 is the updated versions of AIRS1 that was claimed to have slightly better improvement as compared to the original version algorithm. However, the fundamental of AIRS2 algorithm are still similar to AIRS1. According to several experiments in literature, there was no significant difference in classification accuracy between AIRS1 and AIRS2 in most of the application area (Brownlee, 2005). Furthermore, AIRS has been applied in the classification in many research fields including medical data, and a good accuracy has been shown (Polat and Gunes, 2007a, 2008a,b, Polat *et al.*, 2006a,b, 2007a,b, Zhao and Davis, 2011). Nevertheless, there is no significant improvement shown in the application of AIRS algorithms due to the satisfying results obtained from the application of algorithm. Recently, Jenhani and Elouedi (2012) have improved the AIRS algorithm and introduced AIRS3, which could further enhance the classification accuracy on UCI machine learning repository data set. Therefore, due to the high accuracy found in many applications using AIRS algorithm, AIRS is intended to be implemented in the detection of lesions by CAD on medical imaging. Besides, the advantage of AIRS in reducing the redundant data intelligently in the massive training of algorithm is worth evaluated. Furthermore, to test the performance of AIRS algorithm in medical images processing, the classification accuracy is normally measured quantitatively as well as qualitatively.

Usually, the qualitative analysis on the performance of CAD is visualisation evaluation based on human prejudices on how accurate the abnormality is revealed by

CAD. Through this manner, several physicians concluded the usefulness of system and the results were subjective. Therefore, quantitative analysis is also crucial, especially for the researchers to measure the performance of CAD algorithms. For common quantitative analysis, effectiveness of CAD in selecting parameters in algorithm is commonly evaluated by Receiver-Operating Characteristic (ROC) curve. ROC curve is also used as a probability-based approach to assess whether the performance of a physician can be increased by implementing CAD (Doi, 2007). In addition, ROC curve is implemented by researchers to present accuracy of classification results that has been performed by CAD algorithms. In the research of medical image processing, golden standard is always required to evaluate the performance of algorithms when using ROC analysis (Zou *et al.*, 2007). In the detection of lung nodules using pixel-based machine learning approaches, output images are revealed by classes of instances, where the pixel values of results are different from the original intensity values.

1.3 Problem Statement

In this study, there are five problem statements and potential solution have been addressed as following:

Related issue 1: In the preprocessing stage of classification algorithms, segmentation procedures would sometimes exclude some small and low contrast lung nodules, as well as affect the accuracy of feature extraction. The inaccurate segmentation has then affected the overall accuracy of classification results. Besides, calculated features used in common features training of classification algorithms can also cause the loss of important information in medical imaging.

Hypothesis 1: Pixel machine learning is implemented in the classification algorithm, whereby no segmentation and statistical calculation of features are required. The errors caused by both, segmentation and feature calculation can be avoided since the training of sub-region with original intensity values is extracted directly from the original medical images. Although the required time in pixel by pixel-based training is longer, the dramatic growth of computational power has solved this problem.

Related issue 2: In MTANN training, the teaching images are formed by single value of the standard deviation to represent all sizes of lung nodules. The MTANN is claimed to be able to train the model for diverse sizes of lung nodules using average-sized distributions in teaching images due to its particular mechanism. This has been proven by using ROC curve where the fixed value of standard deviation provides the highest performance measure. However, the training of model using only one teaching image may not be suitable to present all sizes of trained lung nodules as the nodules can differ by more than 20mm, especially in other computational models.

Hypothesis 2: To ensure the generality of performance of algorithms in pixel machine learning using Gaussian teaching images as likelihood of lung nodules, the size of standard deviation can be adjusted according to the size of lung nodules for each of the training cases. Since the standard deviation is proportional to the size of the lung nodules, statistical model can be used to analyse the association and correlation of parameters.

Related issue 3: In previous research, Massive Training Artificial Neural Network (MTANN) was developed with pixel-based machine training to detect lesions from CT scans (Suzuki, 2009, 2012, Suzuki *et al.*, 2005b). MTANN has achieved high accuracy in classification problems but the long training time of this classifier is still required to be reduced by using appropriate dimension reduction algorithm such as Laplacian-eigenfunction-based, support vector regression and principal-component-analysis (PCA) (Suzuki *et al.*, 2010a,a,b, Xu and Suzuki, 2011a). Besides, comprehensive adjustment of weights and parameters in MTANN based on the training cases is needed to ensure the accuracy of classification results. Thus, the computational time for massive training processes became very expensive.

Hypothesis 3: In this study, Artificial Immune System (AIS) based algorithm – Artificial Immune Recognition System (AIRS) is employed with the pixel machine learning for lung CT scans to classify nodule and non-nodule cases. AIRS is a classification algorithm that can intelligently reduce the redundancy of training instances and evolving best memory instances for each class in training process, while maintain the classification performance. Therefore, the proposed algorithm – Massive Training Artificial Immune Recognition System (MTAIRS) is implemented to perform the classification of lung nodules on medical imaging. As AIRS1 and AIRS2 algorithms have shown different performance measure in diverse of application area. Thus, they are intended to be investigated in this massive training manner to assess their performance in medical imaging classification.

Related issue 4: In the AIRS algorithm, a linear affinity measure namely Euclidean distance function is used to compute the similarity between two vectors. Although

Euclidean distance has the advantages of simplicity and efficiency in calculation, the formulation is sometimes inadequate to measure the similarity between two cases especially in image processing (Bhattacharya *et al.*, 2012, Wu *et al.*, 2005). Fundamentally, distance function satisfyingly measures of distance between two objects but not the similarity of them. In implementation of MTAIRS for processing lung nodule images, the application of Euclidean distance in affinity function may not suitable calculate the proximity between two training sub-regions from CT images.

Hypothesis 4: The appropriate function ought to be chosen to measure the divergence between two sub-regions in the proposed approach. A weighted affinity function can be employed to measure the similarity between two instances, instead of using Euclidean distance. In the MTAIRS algorithm training, the weighted affinity function is expected to be able to take into account the impact of vicinity level of other training points during the training of particular features. The classification results in both MTAIRS and modified MTAIRS are compared in terms of false positive rate and quantitative analysis from their output images.

Related issue 5: The images and results produced by the CAD are normally analysed by qualitative analysis. Although qualitative evaluation is a quick judgment on how good the CAD performs, this method is subjective for the scientists and experts in different fields. Besides obtaining the ROC curve of CAD performance, the image quality produced by the algorithm should also be analysed by quantitative manner to prevent the argument of image quality. In the implementation of MTAIRS, the outcome images have different scale of intensity values in comparison to the original intensity values. Therefore, statistical analysis such as significant tests for mean comparison are not suitable to be applied to assess whether there is a significant difference in intensity between computed and original images.

Hypothesis 5: A measure based on the frequency occurrence of intensities should be implemented to evaluate two images with different scales of intensity that represent the same illustration. An expectation of the logarithm of probability density relation – Kullback Leibler (KL) divergence is implemented in the quantitative evaluation procedure to evaluate the discrepancy of two images Kullback and Leibler (1951). In this analysis, KL divergence is able to provide discrepancy measurement from probability distribution between columns of two images. Subsequently, pixel-based evaluation is performed to calculate the accuracy of classification based on the number of false positives obtained from the range of column that has the highest inconsistency measurement.

Based on the statement above, the research questions that need to be addressed are:

(a) How can the inaccuracy of segmentation in the detection of lung nodules be avoided?

(b) What is the appropriate method suitable to improve the efficiency of classifier for large datasets?

(c) How can the MTAIRS models provide the optimal classification results for detection of lung nodules on CT scans?

(d) Can enhanced affinity function in MTAIRS reduce false positives in the classification of pixels on CT scans?

(e) How can the performance of MTAIRS in detection of lung nodule images be evaluated by quantitative method?

1.4 Objectives of Study

The objectives of the study are as follows:

1. To propose the adaptive parameters in the pixel machine learning;
2. To propose Massive Training Artificial Immune Recognition System (MTAIRS) for detection of lung nodules on CT scans;
3. To propose a pixel-based quantitative analysis based on Kullback Leibler divergence; and
4. To enhance Massive Training Artificial Immune Recognition System (MTAIRS) for false positives reduction.

1.5 Contribution of Study

In this research, the contribution of study will focus on the proposed MTAIRS algorithms for lung nodules detection. There are several roles in the contribution of research:

(a) Proposed of MTAIRS as pixel-based classification algorithms to improve the lung nodules detection on CT scans.

(b) Proposed of pixel-based evaluation method based on KL divergence for image quantitative analysis.

(c) Enhancement of MTAIRS to reduce false positives of lung nodules detection.

1.6 Research Overview

The research methodology of this study focuses on the proposed algorithm - MTAIRS, which is developed based on the pixel by pixel-based training in the AIRS algorithms. The aim of the proposed MTAIRS is to detect lung nodules on CT scans. Besides, the MTAIRS algorithms will be further enhanced based on the implementation of appropriate affinity function in the study domain. Figure 1.2 shows the overview of the conducted research.

Firstly, samples of nodules and non-nodules are selected from Lung Image Database Consortium (LIDC) database for pixel-based training whereby the locations and diameter of the nodules are extracted from the documentary of database. The estimation of parameter is applied on teaching images which are formed by Gaussian function. A massive training data will be formed based on the pixel machine learning approach.

Secondly, the prepared massive training data are going through the mechanism of MTAIRS which mimics the process of several processes in mammalian immune system. There are several term relations between immune system and component of AIRS such as antibody - features of vector, antigen - training vector, artificial recognition ball (ARB) - combination of feature vectors and classes, clonal expansion - production of ARBs that match with antigen, and affinity maturation - removes the redundant recognition balls from the mutation and immune memory - remained memory attributes for classification. The MTAIRS consists of four main stages: initialisation, memory instances identification and ARB generation, competition for limited resources, and memory instances introduction. There are two versions of MTAIRS: MTAIRS 1 and MTAIRS 2 implemented in the experiments. The differences between both MTAIRS algorithms are in the mechanism of competition for limited resources. MTAIRS 2 is a simplified version compared to MTAIRS 1. Thirdly, the modification of affinity function will be done in the stage of memory instances

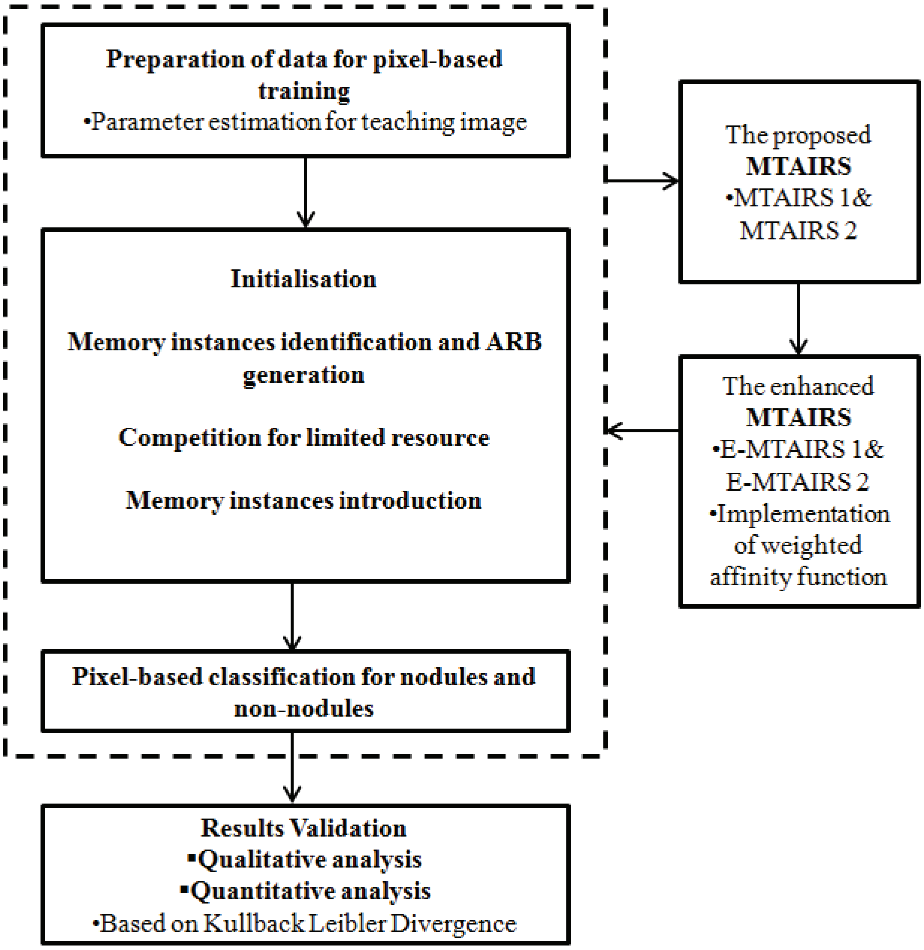

Figure 1.2: The overview of proposed research flow

identification and ARB generation. Furthermore, in the pixels classification, the outcome is indicated by a set of class for each instance and the images will be built based on the class to reveal the pattern of nodules and non-nodules. It is expected that a nodule cases will be highlighted as light region, while non-nodule cases will be revealed as dark regions which represent normal tissues. In addition, the appropriateness of affinity function in MTAIRS algorithms is further examined in the application domain. Based on the potential solution suggested in literature review, a weighted affinity function will be implemented in the enhanced MTAIRS algorithms (E-MTAIRS 1 and E-MTAIRS 2). The impact of the use of weighted affinity function in E-MTAIRS 1 and E-MTAIRS 2 will be tested in the detection of lung nodules as well as the elimination of false positives.

Lastly, qualitative and quantitative analysis are performed in the image evaluation. It is crucial to examine the quality output as the outcome images will reveal the illustration of nodule and non-nodule cases. Moreover, a quantitative analysis on the basis of KL divergence is proposed to evaluate the dissimilarity of original images and classified images. The pixel-based evaluation based on the KL divergence measurement is performed to calculate the accuracy of pixels in the classification results for lung nodules detection.

1.7 Scope of Study

The scope of the study is limited to the following:

(a) Due to the confidential stipulation, only publicly available dataset is used in the experiments and method validation. The publicly available dataset: Lung Image Database Consortium (LIDC) and the Reference Image Data to Evaluate Response (RIDER) are used as study cases in this research.

(b) The medical images used in the experiment are CT scans for lung in Digital Imaging and Communications in Medicine (DICOM) format.

(c) The sizes of nodules in study cases are limited to measure between 5mm to 18mm.

(d) Features selection is done based on Hounsfield values (HU) only on CT scans and pixel by pixel-based extraction for lung region characterisation.

(e) There are two versions of classification algorithms are developed, which are MTAIRS 1 and MTAIRS 2.

(f) To avoid the inaccuracy of classification, the segmentation of lesions in lung images and statistical calculation of selected features are not included in the work flow of research.

(g) Several specific image viewer computation tools are used. They are MIPAV (Medical Image Processing, Analysis, and Visualisation) version 7.0.1, ImageJ and MATLAB.

(h) JAVA programming language has been used to develop the proposed algorithms.

1.8 Thesis Organisation

Chapter 1 presents the objective of the studies by reviewing the research area and the problem background. The scope, brief research methodology and contribution of research are also highlighted.

Chapter 2 presents the literature review on the application of various techniques in CAD for medical imaging. The role and recent review of CAD in lung nodules detection by CT scans are discussed. As the focus of research is on computational intelligent approach, various types of artificial intelligence classifiers are discussed on the application of CAD. In addition, the pixel machine learning classifier – MTANN is investigated and its limitation is identified. Further, the overview of AIRS is presented and the exploration of its potential in classification of lesions on medical imaging is studied. Besides, the study of affinity function that applicable in AIRS is included. Furthermore, the evaluation methods for testing the performance of pixel machine learning classifiers models are reviewed.

Chapter 3 reveals the methodology applied in this study. Firstly, this chapter introduces research framework of the study. Subsequently, the publicly available datasets: Reference Image Data to Evaluate Response (RIDER) and Lung Image Database Consortium (LIDC) are described since they are the input data in development of algorithms for testing and training purposes. The proposed algorithm MTAIRS which is developed based on the concept of pixel machine learning and Artificial Immune System (AIS) is introduced. There are three major phases in the research design of MTAIRS: input of pixel-based training data, massive data training, pixel classification and reconstruction of pixels for nodule and non-nodule images. The methods and processes involved are thoroughly discussed here. Consequence steps and mathematical formulas are also included in the chapter. This chapter briefly introduces the proposed pixel-based evaluation method which is based on the concept of KL divergence in the results validation of classification algorithms.

Chapter 4 explains the details of fundamental of pixel machine learning and proposed algorithm Massive Training Artificial Immune System (MTAIRS). There are two versions of MTAIRS: MTAIRS 1 and MTAIRS 2 are implemented in the experiments of lung nodules detection. The differences of mechanisms in both proposed methods are also revealed in this chapter. The experiments are conducted to examine the performance of MTAIRS algorithms in the classification of pixels for nodule and non-nodule cases on CT images. Besides, the performances of these algorithms are also compared with the standard pixel machine learning algorithm – Massive Training Neural Network (MTANN) in term of efficiency and accuracy. Furthermore, the experimental results produced by MTAIRS algorithms are assessed qualitatively and quantitatively.

Chapter 5 describes the enhancement of affinity function in the training process of MTAIRS. The enhanced algorithms are named as E-MTAIRS 1 and E-MTAIRS 2. In this chapter, non-linear affinity function is proposed in both enhanced algorithms. The results of modification of affinity function in pixel classification by E-MTAIRS 1 and E-MTAIRS 2 are compared with the MTAIRS 1 and MTAIRS 2 for both, nodule and non-nodule cases. The experimental results are examined by both, qualitative and proposed quantitative methods. Lastly, the results produced by E-MTAIRS 1 and E-MTAIRS 2 in the classification of lung nodules are concluded.

Chapter 6 concludes the results and discussion of the research contribution and findings in lung nodules detection by using both, MTAIRS and enhanced affinity function in MTAIRS algorithms. The chapter ends with suggestions of future study of this research.

REFERENCES

Abdullah, N., Ngah, U. K. and Aziz, S. A. (2011). Image classification of brain MRI using support vector machine. In *IEEE International Conference on Imaging Systems and Techniques*. 242–247.

Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., Gareen, I. F., Gatsonis, C., Marcus, P. M. and Sicks, J. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England Journal of Medicine*. 362(5), 395–409.

Alex, E. (2005). Medical imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 546(1-2), 1–13.

Arimura, H., Katsuragawa, S., Suzuki, K., Li, F., Shiraishi, J., Sone, S. and Doi, K. (2004). Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening. *Academic Radiology*. 11(6), 617–629.

Armato, S. G., Giger, M. L., Moran, C. J., Blackburn, J. T., Doi, K. and MacMahon, H. (1999). Computerized Detection of Pulmonary Nodules on CT Scans. *Radiographics*. 19(5), 1303–1311.

Armato III, S. G., Giger, M. L. and MacMahon, H. (2001). Automated detection of lung nodules in CT scans: preliminary results. *Medical Physics*. 28(8), 1552–1561.

Awai, K., Murao, K., Ozawa, A., Komi, M., Hayakawa, H., Hori, S. and Nishimura, Y. (2004). Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on Radiologists Detection Performance 1. *Radiology*. 230(2), 347–352.

BagcI, U., Bray, M., Caban, J., Yao, J. and Mollura, D. J. (2012). Computer-assisted detection of infectious lung diseases: A review. *Computerized Medical Imaging and Graphics*. 36(1).

Basseville, M. (2013). Divergence measures for statistical data processing– annotated bibliography. *Signal Processing*. 93(4), 621–633.

Bateni, M., Baraani, A. and Ghorbani, A. (2012). Alert correlation using artificial immune recognition system. *International Journal of Bio-Inspired Computation*.

4(3), 181–195.

Bhattacharya, G., Ghosh, K. and Chowdhury, A. S. (2012). An affinity-based new local distance function and similarity measure for kNN algorithm. *Pattern Recognition Letters*. 33(3), 356–363.

Brown, M. S., McNitt-Gray, M. F., Goldin, J. G., Suh, R. D., Sayre, J. W. and Aberle, D. R. (2001). Patient-specific models for lung nodule detection and surveillance in CT images. *IEEE Transactions on Medical Imaging*. 20(12), 1242–1250.

Brownlee, J. (2005). *Artificial Immune Recognition System (AIRS): A Review and Analysis*. Technical report.

Casro, L. D. and Timmis, J. (2002). *Artificial Immune Systems: A New Computational Intelligence Approach*. Spinger.

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *City*. 1(2), 1.

Chang, S., Emoto, H., Metaxas, D., Axel, L., Barillot, C., Haynor, D. and Hellier, P. (2004). *Pulmonary Micronodule Detection from 3D Chest CT Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, Springer Berlin / Heidelberg, *Lecture Notes in Computer Science*, vol. 3217. 821–828.

Chen, C., Xu, C., Bie, R. and Gao, X. (2008). Artificial immune recognition system for DNA microarray data analysis. In *Fourth International Conference on Natural Computation*, vol. 6. IEEE, 633–637.

Chikh, M. A., Saidi, M. and Settouti, N. (2012). Diagnosis of diabetes diseases using an artificial immune recognition system 2 (AIRS 2) with fuzzy k-nearest neighbor. *Journal of medical systems*. 36(5), 2721–2729.

Choi, W.-J. and Choi, T.-S. (2012). Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images. *Information Sciences*. 212, 57–78.

Chung, A. C., Wells III, W. M., Norbash, A. and Grimson, W. E. L. (2002). *Multi-modal image registration by minimising kullback-leibler distance*, Springer. 525–532.

De Vito, S., Martinelli, E., Di Fuccio, R., Tortorella, F., Di Francia, G., D'Amico, A. and Di Natale, C. (2010). Artificial immune systems for Artificial Olfaction data analysis: Comparison between AIRS and ANN models. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*. 1–7.

Delibasis, K. K., Asvestas, P. A. and Matsopoulos, G. K. (2010). Automatic point correspondence using an artificial immune system optimization technique for medical image registration. *Computerized Medical Imaging and Graphics*. 35(1), 31–41.

Delogu, P., Cheran, S. C., De Mitri, I., De Nunzio, G., Fantacci, M. E., Fauci, F., Gargano, G., Lopez Torres, E., Massafra, R., Oliva, P., Preite Martinez, A., Raso, G., Retico, A., Stumbo, S. and Tata, A. (2005). Preprocessing methods for nodule detection in lung CT. *International Congress Series*. 1281, 1099–1103.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*. 31(4-5), 198–211.

Duda, R. O., Hart, P. E. and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A. and Abdollahi, B. (2013). Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *International Journal of Biomedical Imaging*. 2013.

Eva, M. v. R., Bartjan de, H., Max, A. V., Mathias, P. and Bram van, G. (2009). Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*. 36(7), 2934–2947.

Ezoe, T., Takizawa, H., Yamamoto, S., Shimizu, A., Matsumoto, T., Tateno, Y., Iimura, T. and Matsumoto, M. (2002). Automatic detection method of lung cancers including ground-glass opacities from chest X-ray CT images. In *Medical Imaging*. International Society for Optics and Photonics, 1672–1680.

Farag, A. A., El-Baz, A., Gimelfarb, G., El-Ghar, M. A. and Eldiasty, T. (2005). Quantitative nodule detection in low dose chest CT scans: new template modeling and evaluation for CAD system design. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*. (pp. 720–728). Springer.

Farag, A. A., El-Baz, A., Gimelfarb, G. G., Falk, R. and Hushek, S. G. (2004). Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*. (pp. 856–864). Springer.

Farmer, J. D., Packard, N. H. and Perelson, A. S. (1986). The immune system, adaptation, and machine learning. *Physica D: Nonlinear Phenomena*. 22(1-3), 187–204.

Findik, O., Babaouglu, I. and Ulker, E. (2011). A color image watermarking scheme based on artificial immune recognition system. *Expert Systems with Applications*. 38(3), 1942–1946.

Forouzideh, N., Mahmoudi, M. T. and Badie, K. (2011). Organizational texts classification using artificial immune recognition systems. In *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2011 IEEE Symposium on*. 1–8.

Fraioli, F., Serra, G. and Passariello, R. (2010). CAD (computed-aided detection) and CADx (computer aided diagnosis) systems in identifying and characterising lung nodules on chest CT: overview of research, developments and new prospects. *Radiologia Medica*. 115, 385 – 402.

Freitas, A. A. and Timmis, J. (2007). Revisiting the foundations of artificial immune systems for data mining. *IEEE Transactions on Evolutionary Computation*. 11(4), 521–540.

Frome, A., Singer, Y., Sha, F. and Malik, J. (2007). Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007*. IEEE, 1–8.

Gan, R., Wu, J., Chung, A. C., Simon, C. and Wells III, W. M. (2004). *Multiresolution image registration based on Kullback-Leibler distance*, Springer. 599–606.

Ge, Z., Sahiner, B., Chan, H.-P., Hadjiiski, L. M., Cascade, P. N., Bogot, N., Kazerooni, E. A., Wei, J. and Zhou, C. (2005). Computer-aided detection of lung nodules: false positive reduction using a 3D gradient field method and 3D ellipsoid fitting. *Medical Physics*. 32(8), 2443–2454.

Ghosh, P. and Mitchell, M. (2006). Segmentation of medical images using a genetic algorithm. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*. ACM, 1171–1178.

Giger, M. L., Bae, K. T. and MacMahon, H. (1994). Computerized detection of pulmonary nodules in computed tomography images. *Investigative Radiology*. 29(4), 459–465.

Girvin, F. and Ko, J. P. (2008). Pulmonary Nodules: Detection, Assessment, and CAD. *American Journal of Roentgenology*. 191(4), 1057–1069.

Golosio, B., Masala, G. L., Piccioli, A., Oliva, P., Carpinelli, M., Cataldo, R., Cerello, P., De Carlo, F., Falaschi, F. and Fantacci, M. E. (2009). A novel multithreshold method for nodule detection in lung CT. *Medical Physics*. 36(8), 3607–3618.

Gomathi, M. and Thangaraj, P. (2010). Lung Nodule Detection using a Neural Classifier. *International Journal of Engineering and Technology*. 2(3), 291–295.

Guetter, C., Xu, C., Sauer, F. and Hornegger, J. (2005). *Learning based non-rigid* *multi-modal image registration using Kullback-Leibler divergence*, Springer. 255–262.

Guo, X.-H., Sun, T., Wang, H. and Liang, Z. (2011). *Prediction Models for Malignant Pulmonary Nodules Based-on Texture Features of CT Image, Theory and Applications of CT Imaging and Analysis*, InTech.

Gurcan, M. N., Sahiner, B., Petrick, N., Chan, H.-P., Kazerooni, E. A., Cascade, P. N. and Hadjiiski, L. (2002). Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. *Medical Physics*. 29(11), 2552–2558.

Hamaker, J. S. and Boggess, L. (2004). Non-Euclidean distance measures in AIRS, an artificial immune classification system. In *Congress on Evolutionary Computation*, vol. 1. 1067–1073.

Hu, S., Hoffman, E. A. and Reinhardt, J. M. (2001). Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Transactions on Medical Imaging*. 20(6), 490–498.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011). Global cancer statistics. *CA: A Cancer Journal for Clinicians*. 61(2), 69–90.

Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R. C., Ghafoor, A., Feuer, E. J. and Thun, M. J. (2005). Cancer statistics, 2005. *A Cancer Journal for Clinicians*. 55(1).

Jenhani, I. and Elouedi, Z. (2012). Re-visiting the artificial immune recognition system: a survey and an improved version. *Artificial Intelligence Review*, 1–13.

Jennane, R., Almhdie-Imjabber, A., Hambli, R., Ucan, O. N. and Benhamou, C. L. (2010). Genetic algorithm and image processing for osteoporosis diagnosis. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. 5597–5600.

Jerne, N. K. (1974). Towards a network theory of the immune system. In *Annales d'immunologie*, vol. 125. 373–389.

Jiang, J., Trundle, P. and Ren, J. (2010). Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics*. 34(8), 617–631.

Kakar, M. and Olsen, D. R. (2009). Automatic segmentation and recognition of lungs and lesion from CT scans of thorax. *Computerized Medical Imaging and Graphics*. 33(1), 72–82.

Kanazawa, K., Kawata, Y., Niki, N., Satoh, H., Ohmatsu, H., Kakinuma, R., Kaneko, M., Moriyama, N. and Eguchi, K. (1998). Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Computerized Medical Imaging and Graphics*. 22(2), 157–167.

Kara, S., Aksebzeci, B. H., Kodaz, H., Gunes, S., S. and Kaya, H., EsmaGunesOzbilge (2009). Medical application of information gain-based artificial immune recognition system (IG-AIRS): Classification of microorganism species. *Expert Systems with Applications*. 36(3), 5168–5172.

Karkavitsas, G. and Rangoussi, M. (2005). Object localization in medical images using genetic algorithms. *International Journal of Information and Communication Engineering*. 1(i), 204–208.

Kerridge, D. (1961). Inaccuracy and Inference. *Journal of the Royal Statistical Society, 23 (Ser. B)*, 184–94.

Kilic, N., Ucan, O. N. and Osman, O. (2009). Colon segmentation and colonic polyp detection using cellular neural networks and three-dimensional template matching. *Expert Systems*. 26(5), 378–390.

Ko, J. P. and Betke, M. (2001). Chest CT: Automated Nodule Detection and Assessment of Change over Time–Preliminary Experience. *Radiology*. 218(1), 267–273.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.

Latifoglu, F., Kodaz, H., Kara, S. and Gunes, S. (2007). Medical application of Artificial Immune Recognition System (AIRS): Diagnosis of atherosclerosis from carotid artery Doppler signals. *Computers in Biology and Medicine*. 37(8), 1092–1099.

Lee, S., Kouzani, A. and Hu, E. (2009). Automated detection of lung nodules in computed tomography images: a review. *Machine Vision and Applications*, 1–13.

Lee, Y., Hara, T., Fujita, H., Itoh, S. and Ishigaki, T. (2001). Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Transactions on Medical Imaging*. 20(7), 595–604.

Li, H., Wang, C. and Wang, A. (2009). Medical Image Registration Based on More Features and Artificial Immune Algorithm. In *2009 International Joint Conference on Artificial Intelligence*.

Li, Q. and Doi, K. (2004). New selective nodule enhancement filter and its application for significant improvement of nodule detection on computed tomography. In *Medical Imaging*. International Society for Optics and Photonics, 1–9.

Liam, C.-K., Pang, Y.-K., Leow, C.-H., Poosparajah, S. and Menon, A. (2006).

Changes in the distribution of lung cancer cell types and patient demography in a developing multiracial Asian country: Experience of a university teaching hospital. *Lung cancer (Amsterdam, Netherlands)*. 53(1), 23–30.

Lo, S. C. B., Li, H., Yue, W., Kinnard, L. and Freedman, M. T. (2002). A multiple circular path convolution neural network system for detection of mammographic masses. *IEEE Transactions on Medical Imaging*. 21(2), 150–158.

Ma, W., Jiao, L., Shang, R. and Zhao, F. (2009). Medical image segmentation based on immune clonal optimization. In *IEEE International Conference on Intelligent Computing and Intelligent Systems 2009*, vol. 1. IEEE, 377–381.

Matsumoto, S., Kundel, H. L., Gee, J. C., Gefter, W. B. and Hatabu, H. (2006). Pulmonary nodule detection in CT images with quantized convergence index filter. *Medical Image Analysis*. 10(3), 343–352.

Mekada, Y., Kusanagi, T., Hayase, Y., Mori, K., Hasegawa, J.-i., Toriwaki, J.-i., Mori, M. and Natori, H. (2003). Detection of small nodules from 3D chest X-ray CT images based on shape features. In *International Congress Series*, vol. 1256. Elsevier, 971–976.

Mendonca, P. R., Bhotika, R., Sirohey, S. A., Turner, W. D., Miller, J. V. and Avila, R. S. (2005). Model-based analysis of local shape for lesion detection in CT scans. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*. (pp. 688–695). Springer.

Messay, T., Hardie, R. C. and Rogers, S. K. (2010). A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Medical Image Analysis*. 14(3), 390–406.

Mishra, V., Dwivedi, R., Das, R. *et al.* (2013). Gases/Odors Identification With Artificial Immune Recognition System Using Thick Film Gas Sensor Array Responses. *Sensors Journal*. 13(8), 3039–3045.

Murphy, K., van Ginneken, B., Schilham, A. M., De Hoop, B., Gietema, H. and Prokop, M. (2009). A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Medical Image Analysis*. 13(5), 757–770.

Oda, S., Awai, K., Suzuki, K., Yanaga, Y., Funama, Y., MacMahon, H. and Yamashita, Y. (2009). Performance of Radiologists in Detection of Small Pulmonary Nodules on Chest Radiographs: Effect of Rib Suppression With a Massive-Training Artificial Neural Network. *American Journal of Roentgenology*. 193(5), W397–W402.

Ogiela, M. R. and Tadeusiewicz, R. (2008). *Modern Computational Intelligence Methods for the Interpretation of Medical Images*. Studies in Computational

Intelligence 84. Spinger.

Ondimu, S. N. and Murase, H. (2008). Effect of probability-distance based Markovian texture extraction on discrimination in biological imaging. *Computers and Electronics in Agriculture*. 63(1), 2–12.

Paik, D. S., Beaulieu, C. F., Rubin, G. D., Acar, B., Jeffrey Jr, B., Yee, J., Dey, J. and Napel, S. (2004). Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. *IEEE Transactions on Medical Imaging*. 23(6), 661–675.

Park, S., Kim, B., Lee, J., Goo, J. and Shin, Y. (2011). GGO Nodule Volume-Preserving Nonrigid Lung Registration using GLCM Texture Analysis. *IEEE Transactions on Biomedical Engineering*. PP(99), 1–1.

Perelson, A. S. and Oster, G. F. (1979). Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology*. 81(4), 645–670.

Pokrajac, D., Megalooikonomou, V., Lazarevic, A., Kontos, D. and Obradovic, Z. (2005). Applying spatial distribution analysis techniques to classification of 3D medical images. *Artificial Intelligence in Medicine*. 33(3), 261–280.

Polat, K. and Gunes, S. (2006). Automated identification of diseases related to lymph system from lymphography data using artificial immune recognition system with fuzzy resource allocation mechanism (fuzzy-AIRS). *Biomedical Signal Processing and Control*. 1(4), 253–260.

Polat, K. and Gunes, S. (2007a). Medical decision support system based on artificial immune recognition immune system (AIRS), fuzzy weighted pre-processing and feature selection. *Expert Systems with Applications*. 33(2), 484–490.

Polat, K. and Gunes, S. (2007b). Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system. *Applied Mathematics and Computation*. 189(2), 1282–1291.

Polat, K. and Gunes, S. (2008a). Computer aided medical diagnosis system based on principal component analysis and artificial immune recognition system classifier algorithm. *Expert Systems with Applications*. 34(1), 773–779.

Polat, K. and Gunes, S. (2008b). Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. *Expert Systems with Applications*. 34(1), 214–221.

Polat, K., Gunes, S. and Tosun, S. (2006a). Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing. *Pattern*

Recognition. 39(11), 2186–2193.

Polat, K., Sahan, S. and Gunes, S. (2006b). A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Systems with Applications*. 31(2), 264–269.

Polat, K., Sahan, S. and Gunes, S. (2007a). Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing. *Expert Systems with Applications*. 32(2), 625–631.

Polat, K., Sahan, S. and Gunes, S. (2007b). A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis. *Expert Systems with Applications*. 32(4), 1141–1147.

Pu, J., Zheng, B., Leader, J. K., Wang, X.-H. and Gur, D. (2008). An automated CT based lung nodule detection scheme using geometric analysis of signed distance field. *Medical Physics*. 35(8), 3453–3461.

Puspitasari, F., Volkau, I., Ambrosius, W. and Nowinski, W. L. (2009). Robust calculation of the midsagittal plane in CT scans using the KullbackLeiblers measure. *International Journal of Computer Assisted Radiology and Surgery*. 4(6), 535–547.

Renyi, A. (1961). On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. 547–561.

Retico, A., Delogu, P., Fantacci, M. E., Gori, I. and Preite Martinez, A. (2008). Lung nodule detection in low-dose and thin-slice computed tomography. *Computers in Biology and Medicine*. 38(4), 525–534.

Riccardi, A., Petkov, T. S., Ferri, G., Masotti, M. and Campanini, R. (2011). Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and Zernike MIP classification. *Medical Physics*. 38(4), 1962–1971.

Roy, A. S., Armato III, S. G., Wilson, A. and Drukker, K. (2006). Automated detection of lung nodules in CT scans: False-positive reduction with the radial-gradient index. *Medical Physics*. 33(4), 1133–1140.

Sachithanandan, A. and Badmanaban, B. (2012). Screening for Lung cancer in Malaysia: Are we there yet? *Medical Journal of Malaysia*. 67(1), 3–6.

Sakthivel, N., Nair, B. B., Sugumaran, V. and Rai, R. S. (2011). Decision support system using artificial immune recognition system for fault classification of centrifugal pump. *International Journal of Data Analysis Techniques and Strategies*. 3(1), 66–84.

Seeker, A. and Freitas, A. A. (2007). WAIRS: improving classification accuracy by weighting attributes in the AIRS classifier. In *IEEE Congress on Evolutionary Computation*. 3759–3765.

Shao, H., Ni, T.-y., Kang, Y. and Zhao, H. (2010). Chest DR Image Classification Based on Support Vector Machine. In *Second International Workshop on Education Technology and Computer Science (ETCS), 2010*, vol. 1. 170–173.

Shi, Z. and He, L. (2010). Application of Neural Network for Medical Image Processing. In *Proceedings of the Second International Symposium on Networking and Network Security*. 023–026.

Shiraishi, J., Li, Q., Appelbaum, D. and Doi, K. (2011). Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging. *Seminars in Nuclear Medicine*. 41(6), 449–462.

Sluimer, I., Schilham, A., Prokop, M. and van Ginneken, B. (2006). Computer analysis of computed tomography scans of the lung: A survey. *IEEE Transactions on Medical Imaging*. 25(4), 385–405.

Soltani Halvaiee, N. and Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. *Applied Soft Computing*. 24(0), 40–49.

Stoitsis, J., Valavanis, I., Mougiakakou, S. G., Golemati, S., Nikita, A. and Nikita, K. S. (2006). Computer aided diagnosis based on medical image processing and artificial intelligence methods. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 569(2), 591–595.

Studler, U., Gluecker, T., Bongartz, G., Roth, J. and Steinbrich, W. (2005). Image Quality from High-Resolution CT of the Lung: Comparison of Axial Scans and of Sections Reconstructed from Volumetric Data Acquired Using MDCT. *American Journal of Roentgenology*. 185(3), 602–607.

Suarez-Cuenca, J. J., Tahoces, P. G., Souto, M., Lado, M. J., Remy-Jardin, M., Remy, J. and Jos Vidal, J. (2009). Application of the iris filter for automatic detection of pulmonary nodules on computed tomography images. *Computers in Biology and Medicine*. 39(10), 921–933.

Summers, R. M. (2003). Road Maps for Advancement of Radiologic Computer-aided Detection in the 21st Century. *Radiology*. 229(1), 11–13.

Suzuki, K. (2009). A supervised ’lesion-enhancement’ filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). *Physics in Medicine and Biology*. 54(18).

Suzuki, K. (2012). Pixel-Based Machine Learning in Medical Imaging. *International Journal of Biomedical Imaging*. 2012(Article ID 792079), 18.

Suzuki, K., Abe, H., MacMahon, H. and Doi, K. (2006a). Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Transactions on Medical Imaging*. 25(4), 406–416.

Suzuki, K., Armato III, S. G., Feng, L., Sone, S. and Do, K. (2003a). Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose CT. *Medical Physics*. 30, 1602–1617.

Suzuki, K. and Doi, K. (2005). How Can a Massive Training Artificial Neural Network (MTANN) Be Trained With a Small Number of Cases in the Distinction Between Nodules and Vessels in Thoracic CT?1. *Academic Radiology*. 12(10), 1333–1341.

Suzuki, K., Horiba, I. and Sugie, N. (2002a). Efficient approximation of neural filters for removing quantum noise from images. *IEEE Transactions on Signal Processing*. 50(7), 1787–1799.

Suzuki, K., Horiba, I. and Sugie, N. (2003b). Neural edge enhancer for supervised edge enhancement from noisy images. *Pattern Analysis and Machine Intelligence*. 25(12), 1582–1596.

Suzuki, K., Horiba, I., Sugie, N. and Nanki, M. (2002b). Neural filter with selection of input features and its application to image quality improvement of medical image sequences. *IEICE TRANSACTIONS on Information and Systems*. E85-D(10), 1710–1718.

Suzuki, K., Horiba, I., Sugie, N. and Nanki, M. (2004). Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector. *IEEE Transactions on Medical Imaging*. 23(3), 330–339.

Suzuki, K., Jun, Z. and Jianwu, X. (2010a). Massive-Training Artificial Neural Network Coupled With Laplacian-Eigenfunction-Based Dimensionality Reduction for Computer-Aided Detection of Polyps in CT Colonography. *IEEE Transactions on Medical Imaging*. 29(11), 1907–1917.

Suzuki, K., Li, F., Sone, S. and Doi, K. (2005a). Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*. 24(9), 1138 – 1150.

Suzuki, K., Shiraishi, J., Abe, H., MacMahon, H. and Doi, K. (2005b). False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. *Academic*

Radiology. 12(2), 191–201.

Suzuki, K., Xu, J., Zhang, J. and Sheu, I. (2010b). *Principal-Component Massive-Training Machine-Learning Regression for False-Positive Reduction in Computer-Aided Detection of Polyps in CT Colonography*, Springer Berlin Heidelberg, *Lecture Notes in Computer Science*, vol. 6357, chap. 23. 182–189.

Suzuki, K., Yoshida, H., Nappi, J., Armato, S. G. and Dachman, I. A. H. (2008). Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography. *Medical Physics*. 35(2), 694–703.

Suzuki, K., Yoshida, H., Nappi, J. and Dachman, A. H. (2006b). Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: Suppression of rectal tubes. *Medical Physics*. 33(10), 3814–3824.

Takizawa, H. and Yamamoto, S. (2002). Recognition of lung nodules from X-ray CT images using 3D Markov random field models. In *Proceedings. 16th International Conference on Pattern Recognition, 2002.*, vol. 1. 99–102 vol.1.

Tan, M., Deklerck, R., Jansen, B., Bister, M. and Cornelis, J. (2011). A novel computer-aided lung nodule detection system for CT images. *Medical Physics*. 38(10), 5630–5645.

Tanino, M., Takizawa, H., Yamamoto, S., Matsumoto, T., Tateno, Y. and Iinuma, T. (2003). A detection method of ground glass opacities in chest X-ray CT images using automatic clustering techniques. In *Medical Imaging*. International Society for Optics and Photonics, 1728–1737.

Tao, Y., Lu, L., Dewan, M., Chen, A. Y., Corso, J., Xuan, J., Salganicoff, M. and Krishnan, A. (2009). *Multi-level ground glass nodule detection and segmentation in CT lung images*, Springer. 715–723.

Tartar, A., Kilic, N. and Akan, A. (2013). Bagging support vector machine approaches for pulmonary nodule detection. In *International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 047–050.

Timmis, J. and Neal, M. (2001). A resource limited artificial immune system for data analysis. *Knowledge-Based Systems*. 14(3), 121–130.

Timmis, J., Neal, M. and Hunt, J. (2000). An artificial immune system for data analysis. *Biosystems*. 55(13), 143–150.

van Rikxoort, E. M., de Hoop, B., Viergever, M. A., Prokop, M. and van Ginneken, B. (2009). Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*. 36(7), 2934–2947.

Verma, B. and Zakos, J. (2001). A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Transactions on Information Technology in Biomedicine*. 5(1), 46–54.

Volkau, I., Bhanu Prakash, K., Ananthasubramaniam, A., Aziz, A. and Nowinski, W. L. (2006). Extraction of the midsagittal plane from morphological neuroimages using the Kullback Leibler measure. *Medical Image Analysis*. 10(6), 863–874.

Wang, C.-M., Kuo, C.-T., Lin, C.-Y. and Chang, G.-H. (2008). Application of Artificial Immune System Approach in MRI Classification. *EURASIP Journal on Advances in Signal Processing*. 2008(Article ID 547684), 8.

Wang, X.-L., Cheng, J.-H., Yin, Z.-J. and Guo, M.-J. (2011). A new approach of obtaining reservoir operation rules: Artificial immune recognition system. *Expert Systems with Applications*. 38(9), 11701–11707.

Watkins, A., Timmis, J. and Boggess, L. (2004). Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm. *Genetic Programming and Evolvable Machines*. 5(3), 291–317.

Watkins, A. B. and Boggess, L. C. (2002). A resource limited artificial immune classifier. In *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 1. 926–931.

Wei, G.-Q., Fan, L. and Qian, J. (2002). *Automatic detection of nodules attached to vessels in lung CT by volume projection analysis*, Springer. 746–752.

Weisenfeld, N. I. and Warfteld, S. K. (2004). Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*. 101–104 Vol. 1.

WHO, W. H. O. (2014). *Cancer*.

Wiemker, R., Rogalla, P., Zwartkruis, A. and Blaffert, T. (2002). Computer-aided lung nodule detection on high-resolution CT data. In *Medical Imaging*. International Society for Optics and Photonics, 677–688.

Worthy, S. (1995). High resolution computed tomography of the lungs. *BMJ*. 310(6980), 616.

Wu, G., Chang, E. Y. and Panda, N. (2005). Formulating distance functions via the kernel trick. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 703709.

Xing-li, B. and Xu, Q. (2008). Medical Image Classification Based on Fuzzy Support Vector Machines. In *International Conference on Intelligent Computation*

Technology and Automation, vol. 2. 145–149.

Xu, J.-W. and Suzuki, K. (2011a). *False-Positive Reduction in Computer-Aided Detection of Polyps in CT Colonography: A Massive-Training Support Vector Regression Approach*, Springer Berlin Heidelberg, *Lecture Notes in Computer Science*, vol. 6668, chap. 7. 47–52.

Xu, J.-W. and Suzuki, K. (2011b). Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography. *Medical Physics*. 38(4), 1888–1902.

Yuan, R., Vos, P. M. and Cooperberg, P. L. (2006). Computer-aided detection in screening CT for pulmonary nodules. *American Journal of Roentgenology*. 186(5), 1280–1287.

Zezula, P., Amato, G., Dohnal, V. and Batko, M. (2006). *Similarity search: the metric space approach*. vol. Similarity search: the metric space approach. Springer.

Zhang, L., Fang, M., Naidich, D. P. and Novak, C. L. (2004). Consistent interactive segmentation of pulmonary ground glass nodules identified in CT studies. In *Medical Imaging 2004*. International Society for Optics and Photonics, 1709–1719.

Zhang, R. and Ma, J. (2008). An improved SVM method P-SVM for classification of remotely sensed data. *International Journal of Remote Sensing*. 29(20), 6029–6036.

Zhao, W. and Davis, C. E. (2011). A modified artificial immune system based pattern recognition approach-An application to clinical diagnostics. *Artificial Intelligence in Medicine*. 52(1), 1–9.

Zhou, J., Dasgupta, D., Zhiling, Y. and Hongmei, T. (2006). Analysis of Dental Images using Artificial Immune Systems. In *IEEE Congress on Evolutionary Computation*. 528–535.

Zou, K. H., Oalley, A. J. and Mauri, L. (2007). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation*. 115(5), 654–657.

APPENDIX A

LIST OF PUBLICATION

1 Hang, S.P., Shamsuddin, S.M. and Kenji, S. Application of Intelligent Computational Models on Computed Tomography Lung Images. International Journal of Advances in Soft Computing and its Application, 3, 2 (2011), 1-15. (Scopus Indexed)

2 Hang, S. P. and Shamsuddin, S.M. Texture Classification of Lung Computed Tomography Images. In proceeding of the 4th International Conference on Graphic and Image Processing (ICGIP 2012). 5-6 October 2012, Singapore. Vol. 8768 87683Z-2. (Ei Compendex and Thomson ISI Indexed)

3 Hang, S. P., Shamsuddin, S.M, and Ralescu, A. Massive Training in Artificial Immune Recognition Algorithm for Enhancement of Lung CT Scans. To be appear in proceeding of Fifth International Conference of Soft Computing and Pattern Recognition (SoCPaR 2013). 15-18 December 2013, Hanoi, Vietnam. (Scopus Indexed)

Submitted Paper:

1 Hang, S. P. and Shamsuddin, S.M. Massive Training Artificial Immune Recognition System for Lung Nodules Detection and False Positives Reduction. Manuscript submitted to Computerized Medical Imaging and Graphics. (IF:1.496)