An Overview of Artificial Immune System in Pattern Recognition
Azah Kamila Muda & Siti Maryam Shamsuddin
FSKSM, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor.
{azahkm@gmail.com}

ABSTRACT
A miracle of our body, an Immune System (IS) prevents the enemy cells, such as bacteria or viruses, from entering the body. This biological IS has intelligent capabilities of detecting or recognizing self/non-self antigen in the body. Artificial Immune Systems (AIS) is one of the recent biologically inspired approaches to emerge from computer science, possess nonlinear classification properties along with biological properties such as self/non-self identification, Negative Selection and Clonal Selection. AIS has been proved as one of the mechanism that has a good potential in solving complex problem such as pattern recognition. Meanwhile, handwriting is one of the most popular areas of research in pattern recognition due to its immense potential on commercial perspective. There are many techniques and approaches have been done in order to recognize the handwriting pattern for completing daily activities. Thus, this paper discusses the potential of AIS in handwritten recognition area. It reviews basic immune system recognition with an emphasis on negative selection algorithm.

KEYWORD
Biological Immune System, Artificial Immune System, Negative Selection, Handwritten recognition

1. Introduction
The field of Artificial Immune Systems (AIS) is one of the recent biologically inspired approaches to emerge from computer science. The natural immune system is an adaptive learning system that employs many parallel and complementary mechanisms for defense against foreign pathogens. It is a distributed system, capable of learning to identify previously unseen invaders and remembering what it has learnt. Numerous immune algorithms now exist, based on processes identified within human immune systems. AIS computational technique has led to the development of useful computational tools for the solution of complex problems such as in pattern recognition, fault detection, classifications, computer security, and optimization [1], [2], [3], [4].

In the work of pattern recognition, Immune System (IS) is performed by using 3 mechanisms which are Negative Selection (T-cells that recognize self-antigens are excluded from the population of T-cells during the maturation process), Clonal Selection (if B-cell encounters a non-self antigen with a sufficient affinity, it will proliferates and differentiates into memory cells) and Immune Network (if B-cell recognizes a self-antigen, it might result in suppression). This study is to explore and apply the capabilities of AIS in pattern recognition, specifically in handwritten domain.

Area of handwritten recognition is firstly obtained by Tausheck [25] in 1929, which patent OCR in Germany [26]. After that, it becomes more popular in research community to explore this area. There are so many papers and patents advertising recognition rates as high as 99.99% and this gives the impression that automation problems seem to have been solved. However, the failure of some real application shows that performance problems exist on composite and degraded documents. So there is still room for progress and still needed to explore [20],[22],[37], while there is still a gap between human and machine reading capabilities[]. Furthermore, Allessandro in [24] mentioned that although many people assume that handwriting recognition is already a mature field, this is not entirely true. Handwriting recognition technology is still far from broad applications.

This paper is organized into several sections. Basic immune system recognition is described in section 2. After that, AIS in pattern recognition is described section 3 which is focus on Negative Selection Algorithm. Then pattern recognition –
related work is reviewed in section 4. Finally, some conclusion is drawn in the last section.

2. Basic Immune System Recognition

Biological immune system has intelligent capabilities of detecting or recognizing self / non self antigen in the body. The primary immune defense, also referred to as innate immunity is the immune mechanism our bodies are born with [8]. If the innate immune system cannot remove the pathogen, then the adaptive (secondary immune response) immune system will take over. This vertebrate immune system exhibits some remarkable properties that mentioned in [5], [6], including:

- Feature extraction to determine unique signature of the antigen.
- Learn to recognize new patterns / antigens.
- Work as distributed pattern recognizers.
- Use content-addressable memory to retrieve known pattern / antigens.
- Use of selective proliferation and self-replication for quick recognition and response.
- Eliminate / neutralize the effect of antigens in a systematic fashion.

Figure 2.1 adopted from [7], [2] illustrates a simplified view of the main immune recognition and activation mechanism. Phase (I) from the figure shows that whenever any non self antigen entered the body, antigen-presenting cell (APC) circulate throughout the body ingesting and digesting the antigen. If this antigen still cannot fully recognized at this stage, T cell will need Major Histocompatibility Complex (MHC) cell to bind the molecule and present in the APC cell surface as an MHC complex (II). This MHC/peptide will presented to T cell and allow them to recognize different MHC complex (III). After that, T cell will activated, divide and secrete a chemical signal (lymphokines) to stimulate the further action (IV).

Meanwhile, B cell has a receptor with ability to recognize antigen without MHC (V). Whenever B cell receive any signal, it will be activated and will perform proliferate and differentiation process to produce plasma cells in high volumes (VI). It contains antibody that be used to neutralized the panthogen (VII). Some of these T cell and B cell will differentiate into memory cells. Taken as a whole of this process, it is the miracle of immune system and being the soldiers to defense our body from any virus entered.

3. AIS in Pattern Recognition

As mentioned before, IS has three mechanisms in recognizing pattern but, this study will only focus on NSA. NSA in AIS has been applied to detect computer viruses [10], toll breakage detection and time-series anomaly detection [11] and network intrusion detection [12], [13] color image classification [14] and creative design classification [15].

The original work by [10] which negative selection algorithm was proposed has been inspirational to almost all the research in the AIS related to computer security [16]. In pattern recognition, most of the researches are also influenced by NSA besides Clonal Selection Algorithm (CSA). In [18], author tried to summarize the works on existing models and development of AIS research in the year of 1999 to 2003. It shows that there are still no applications or models are developed for handwritten classification or recognition using NSA in AIS.

3.1 Negative Selection Algorithm

The NSA is proposed by Forrest in [10] for various anomaly detection problems. It is inspired by the maturation of T-cells in the thymus gland. NSA uses the property of self/non self discrimination to detect foreign antigens. In the biological system, this is achieved in part by T-cells which have receptors on their surface that can detect foreign antigens. [17] has mentioned that a pseudo random genetic process makes the receptors during the generation of the T-cells. Then they undergo a process of filtering in the
thymus where T cells that react against self cells are destroyed and only those do not bind to self cells are allowed to leave thymus. These matured T cells circulate through out the body to protect against foreign antigens. NSA works on similar principles, generating detectors randomly and eliminating the ones that detect self, so that the remaining T cells can detect any non self.

The NSA as summarized by Dasgupta in [17] is as follows:
1) Define self as a collection of strings $S$ of finite length $L$ that needs to be monitored.
2) Generate a set of detectors $R$ each of which fails to match any self string in $S$ (Figure 3.2).
3) Monitor $S$ for changes by continually matching the detectors in $R$ against $S$. If any detector matches, then a change is known to have occurred, because the detectors are designed not to match any of the original strings in $S$.

Figure 3.1 and 3.2 shows the matching of detector set with new antigens based on certain matching rule. Non-self is detected if there is a match between the antigen and any of the detectors.

According to Xiao in [15], the original NSA is not suitable for pattern recognition task. The NSA is originally used to solve change detection problems with only two states defined, either normal or abnormal. In pattern recognition task, negative selection must be extended to multi-state, equivalent to number of classified groups. Xiao in [15] modified the original NSA to be used for pattern recognition in engineering creative design. The modified algorithm is discuss in section 4.1 below. Since this study aimed at pattern matching of handwritten, the modified NSA by Xiao in [15] is adopted. However, the algorithm is customized in hope to increase accuracy of classification.

4. Pattern Recognition – Related Work

4.1 AIS Related Work

One such natural system, AIS can be seen as a parallel and distributed adaptive system with many strength capability in recognition, feature extraction, diversity, learning, memory, distributed detection and self regulation. Nasaroui in [40] mentioned that, based on the complete capabilities of detection and recognition antigens in the body an intelligent computational technique has been developed for pattern recognition and data analysis.

AIS has been applied to detect computer viruses [10], tool breakage detection and time-series anomaly detection [11] and network intrusion detection [12],[13], color image classification [14], creative design classification [15], cancer classification[41], learning to detect texture objects[42], character recognition by using CSA[43] and ect.

Sathyanath et al. in [14], proposed a computational implementation of negative selection of immune system along with genetic algorithm (GA) to perform a color image classification task. Detectors are generated by using NSA but, this procedure had the major limitation of generating valid detectors due to its random generated detectors. Hence, GA was explored to improve the random generation detectors.

Xiao et al. in [15], modified the original NSA to be used for pattern recognition in engineering creative design. According to him the original NSA is not suitable for pattern recognition task. The NSA is originally used to solve change detection problems with only two states defined, either normal or abnormal. In pattern recognition task, negative selection must be extended to multi-state, equivalent to number of classified
groups. The modified algorithm by Xiao is as follow:

1) Define self as N sets of strings $S_i$ (the suffix $i$ starts from 1 to N) of limit length
2) Generate N sets of detectors and each should fail to match with members of the pre-classified groups.
3) Present the input pattern to the N sets of detectors respectively. If any match occurs at the $i^{th}$ set of detectors, the input pattern belongs to the corresponding groups and the pattern matching task is finished.

Shin et al. in [41], presented a method for cancer type classification based on micro-array data with the implementation of AIS for ALL (lymphoid precursor) / AML (myeloid precursor). Important aspects in this problem are selection of informative genes (feature selection) and optimization of strength (weight) of each gene. Shin solves the problems by group the ALL training data set as antigen and AML training data set as self-proteins. Classification rules represent antibody, which captures training samples (antigen/self-proteins) when its profile satisfies the condition of the rules. A population of antibody goes through a cycle of invasion by antigen and selective reproduction. Then, a successful antibody will converted into memory cell as ALL/AML class is learned. NSA and CSA was combined to improve the classification capability in this problem.

Zheng et al. in [42], proposed an AIS approach to automatically generate segmentation thresholds and texture filters. It will automatically detect texture object from panchromatic satellite images. CSA is used to learn the texture filters and segmentation thresholds because of it’s ability to search with optimal solution. Texture objects are regarded as antigens and texture object filters and segmentation thresholds are regarded as antibodies in this approach.

Jennifer et al. in [43], examined the CSA CLONALG over a series of binary character recognition, compared to basic binary matching algorithms. They have made few enhancements to CLONALG algorithms. CLONALG failed to capitalize on the information generated by each clone population. New algorithm of CSA called Clonal Classification (CLONCLAS) has been proposed to improve its performance of classification.

### 4.2 Handwritten Recognition

Pattern could be a fingerprint image, trademark, a handwritten word, digit character, a human face or a speech signal. There are many techniques and approaches have been done in order to recognize the pattern for completing daily activities. However, handwritten recognition still poses a challenge because human capability is beyond to computerized system.

Two characteristic of human recognition mentioned in [9] are robustness and accuracy and these are the goals one tries to reach when designing a recognition system. It still ignited researchers to explore this field in order to find the best technique to recognize the pattern. Furthermore, whenever it comes to real implementation, the recognition rate in these studies are still low, and there is still need to improve [23], especially in selecting better feature vectors as mentioned in [22], [23].

In pattern recognition community, it is a well known that feature extraction and classification task are important role to achieved a good performance in recognizing patterns. Cheng [19] and Liu [38] mentioned the performance of pattern recognition largely depends on the feature extraction approach and classification/learning scheme. Therefore, these two stages will be focus in this study.

#### 4.2.1 Feature Extraction

Extracting and selecting the meaningful features is a crucial step in the process of pattern recognition because it will be used in the next step of pattern recognition. It is a process of converts the input image into feature vectors and removes redundancy from the data. A good technique in extracting and selecting features is very important and being a focus of research for a long time as mentioned in [22], [30], [31], [32], [33].

There are many feature extraction technique have been explored such as Global Moment Function[32],[36]; Hough Transform and Fuzzy logic by Shamik in [35]; Brijesh in [34] proposed an approach of extracting structural features based on character’s contour code and stroke direction. Furthermore, Xuewen [37] present a method of extracting features directly from gray-scale image by Gabor filter to achieve better
performance on low quality images; Liu in [39] was proposed some improved normalization functions and direction feature extraction strategies with extract three types of direction features: chaincode, NCFE (normalization-cooperated feature extraction) and gradient features.

4.2.2 Classification

The task of classification is to partition the feature space into regions corresponding to source classes or assign class confidences to each location in the feature space [19]. It assigns the digitized image into its symbolic class and classifies the image into categories based on the image data samples that the classifier has learned during its training stage.

They are many methods have been explored where usually they are classified in three different ways [20] [44]:

1) Statistical pattern classification.
   Usually use of formal language-theoretic models to represent patterns. It is flexible and general method for modeling handwritten image and entails the use of probabilistic models with uncertain or incomplete information [24]. Example of this category is Hidden Markov Model (HMM) and Support Vector Machine (SVM).

2) Syntactic / structural / linguistic pattern classification.
   It can handle the structural information about the interconnections in complex patterns. The image is represented as unions of structural primitive [39].

3) Neural network-based pattern classification.
   It was widely used in classification due to the implementation efficiency [19]. These networks can be used as a combined feature extractor and classifier, where the inputs are scaled or sub-sampled input image, or as a “pure” classifier where the inputs are extracted features [24].

A few classifiers have been used and proposed to solve problem in pattern recognition by researchers. Lazzerini et al. in [20], presented EYE, which is a fuzzy logic-based classifier for off-line isolated character recognition. It is a linguistic fuzzy classification of the shape of character. Meanwhile, Sarfraz in [36] et al., proposed a neural classifier which is Radial Basis Function (RBF) was implemented to recognize the Arabic printed text. Soltanzadeh et al. in [46], proposed a method by using the support vector machines (SVM) in classification task for Persian handwritten digit recognition. They used the sequential minimal optimization method (SMO), a binary classifier. Smith et al. [47], proposed a method of Nearest Neighbor method for classify handwritten character for large database. Three distance metrics are investigated, which are simple Hamming distance metrics, pixel distance metrics and metrics based on the extraction of pen-strokes features.

There are also hybrid approaches of classier in pattern recognition. The idea of hybrid approach is based on the fact that a given classifier can achieve very good performances in terms of correct recognition rate [48]. Some of the related works are Teredesai et al. in [45], present a two-step classification strategy to improve the recognition accuracy of a digit recognizer. They used K-nearest-neighbor template-based recognizers and genetic programming (GP)-based recognizers. First classifier using the full feature set, the high confidence recognition result will lead to an end of the recognition process. Then, secondary classifier designed using a subset of the original feature set. Zhang et al. in [32], proposed a hybrid neural network namely ANN coarse classifier and three decision tree classification for handwritten numeral. Meanwhile, Prevost et al. in [48], present an original two stages classifiers which are model based classifier (to generate a probability vector while finding the most relevant pair of classes) and discriminative classifier (to improve the accuracy of first classifier by separating the most ambiguous pairs of classes) to handwritten character recognition.

5. Conclusion

This overview, in general attempts to bring out the capabilities of AIS in pattern recognition. Basic immune system and NSA of AIS have been discussed in general. Related work with AIS in pattern recognition also briefly described. Handwritten pattern recognition and method uses are mentioned as well. In this study, an approach of biological concept, i.e., an artificial immune classifier will be explored to generate better accuracy for handwritten classification.
Reference


