UNIVERSITY OF PIRAEUS - DEPARTMENT OF INFORMATICS

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

MSc «Informatics»

ΠΜΣ «Πληροφορική»

MSc Thesis

Μεταπτυχιακή Διατριβή

Thesis Title: Τίτλος Διατριβής:	**Artificial Immune Systems for Sentiment Analysis** Τεχνητά Ανοσοποιητικά Συστήματα για την Ανάλυση Συναισθήματος
Student's name-surname: Ονοματεπώνυμο φοιτητή:	**DAFNI VIDALI-SOULA** ΔΑΦΝΗ ΒΙΔΑΛΗ-ΣΟΥΛΑ
Father's name: Πατρώνυμο:	**MICHAIL** ΜΙΧΑΗΛ
Student's ID No: Αριθμός Μητρώου:	**ΜΠΠΛ/14006**
Supervisor: Επιβλέπων:	**Dionisios Sotiropoulos, Assistant Professor** Διονύσιος Σωτηρόπουλος, Επίκουρος Καθηγητής

Μάϊος 2021/ May 2021

3-Member Examination Committee

Τριμελής Εξεταστική Επιτροπή

Dionisios Sotiropoulos **Assistant Professor**	**Georgios Tsihrintzis** **Professor**	**Efthimios Alepis** **Associate Professor**
Διονύσιος Σωτηρόπουλος Επίκουρος Καθηγητής	Γεώργιος Τσιχριντζής Καθηγητής	Ευθύμιος Αλέπης Αναπληρωτής Καθηγητής

Acknowledgements

First of all, I would like to thank Professor Tsihrintzis and Assistant Professor Sotiropoulos for introducing me to the field of Artificial Immune Systems. Without their insight, I may have overlooked this exciting new field of computational intelligence.

Assistant Professor Sotiropoulos deserves further thanks for freely providing his time and extensive knowledge to explain difficult concepts and offer ideas, as well as his support and encouragement, which proved invaluable to the completion of this thesis.

Table of Contents

Abstract

This thesis aims to present AIRS-x, an alternative formulation to the Artificial Immune System (AIRS) algorithm. AIRS is a supervised learning classification algorithm, which has emerged from the field of Artificial Immune Systems (AIS) – or algorithms inspired by immunological concepts. We present a broad overview of machine learning and immunology; we describe how AIS emerged as a computational paradigm based on biological metaphors; and we explore the scope and the wide range of applications of this relatively novel field. Furthermore, we introduce the AIRS algorithm, describing its workflow, its advantages, and its limitations. In addition, the thesis showcases a novel formulation submitted by Giatzitzoglou (2018) and Giatzitzoglou et al. (2019), the AIRS-x algorithm, indicates how it enhances the original, explores its newly introduced parameters and its promise as a sentiment analysis tool. Finally, it suggests paths for future exploration.

Στόχος της παρούσας εργασίας είναι να παρουσιάσει τον αλγόριθμο AIRS-x, ο οποίος έχει προταθεί ως εναλλακτική διαμόρφωση του αλγορίθμου Artificial Immune System (AIRS). Ο AIRS αποτελεί αλγόριθμο που πραγματοποιεί ταξινόμηση έπειτα από μάθηση με επίβλεψη και ο οποίος προέρχεται από τον τομέα των Τεχνητών Ανοσοποιητικών Συστημάτων (AIS), αλγορίθμων που εμπνέονται από το ανοσοποιητικό σύστημα. Παρουσιάζεται μια ευρεία επισκόπηση της μηχανικής μάθησης και της ανοσολογίας, περιγράφεται η ανάδυση των Τεχνητών Ανοσοποιητικών Συστημάτων ως υπολογιστικού παραδείγματος βασισμένου σε μεταφορές από τη βιολογία και εξερευνείται το αντικείμενο και το εύρος των εφαρμογών του νέου αυτού πεδίου. Στη συνέχεια, εισάγεται ο αλγόριθμος AIRS και περιγράφεται η ροή εργασιών του, τα πλεονεκτήματα και οι αδυναμίες του. Επίσης, η εργασία προβάλλει τη νέα διαμόρφωση που προτάθηκε από τους Giatzitzoglou (2018) και Giatzitzoglou et al. (2019), δηλαδή τον αλγόριθρμο AIRS-x, αναδεικνύει τις βελτιστοποιήσεις που παρέχει, εξερευνεί τις μεταβλητές του και την πολλά-υποσχόμενη χρήση του ως εργαλείο ανάλυσης συναισθημάτων. Τέλος, προτείνονται θέματα για περαιτέρω έρευνα.

1. Objective

This thesis aims to provide a broad overview of the emergent field of Artificial Immune Systems (AIS), a machine learning technology inspired by theoretical immunology. Secondly, this work aspires to showcase AIRS-x, a supervised learning classification algorithm. Proposed in 2018 by Giatzitzoglou and expanded on by Giatzitzoglou et al. (2019), AIRS-x expands Watkins' Artificial Immune Recognition System, a well-recognized and often-used immune classification framework. Our goal is to highlight the newly proposed features, the flaws they aim to correct and the advantages the novel algorithm provides over the original. Finally, we present its potential use as a sentiment analysis classifier and suggest alterations and avenues for future research.

2. Structure

The thesis is structured as follows: **Sections 1** and **2** provide the thesis' Objective and Structure. **Section 3** introduces the theoretical basis on which AIS are built, namely theoretical immunology, as well as machine learning in general. **Section 4** describes the general AIS framework and various AIS models. **Section 5** outlines the scope and range of AIS applications and highlights select literature references. **Sections 6** and **7** present the AIRS and AIRS-X algorithms, respectively. **Sections 8** summarizes and concludes the thesis, while **Section 9** proposes future research. Finally, **Section 10** is dedicated to literature references.

3. Theoretical Background

According to de Castro and Timmis (2000), "Artificial Immune Systems are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving." This section aims to introduce the theoretical foundations of AIS, specifically machine learning and immunology.

3.1 Machine Learning

Learning is a distinctive ability of human and generally intelligent behavior (Giatzitzoglou, 2018). De Castro and Timmis (2000) define it as the process of acquiring knowledge through experience and adapting, in order to successfully interact with the environment.

Machine Learning, on the other hand, involves the study of how to create systems capable of learning and modeling the learning processes (Sotiropoulos and Tsihrintzis, 2018). Mitchell (1997) provided a formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks T, as measured by P, improves with experience E".

Machine Learning algorithms are applied to a broad range of fields, from data mining to computer vision and sentiment classification. However, most have developed around specific research lines (Sotiropoulos and Tsihrintzis, 2018):

- **Task-oriented studies:** focused on improving performance in a specific task.
- **Cognitive simulation:** simulation of human learning processes
- **Theoretical analysis:** algorithm development independent of application domain
- **Algorithms based on biological metaphors**; AIS is a characteristic example.

Furthermore, machine learning algorithms can be categorized according to 1) the type of inference or 2) the amount of inference. We will briefly discuss each category in order to gain a clear perspective on the position of AIS in the field of machine learning.

3.1.1 Type of Inference

Machine learning algorithms can be divided in two groups:

1) Algorithms based on **model identification**: these focus on solving problems by identifying the model behind the phenomenon under examination.
2) Algorithms based on **model prediction**: these forgo identifying the model, under the assumption that this is an ill-posed problem, and instead aim at discovering a rule for adequate prediction for the problem at hand. This approach proves quite useful in situations where information is limited (Sotiropoulos and Tsihrintzis, 2018).

3.1.2 Amount of inference

While the above categorization is not commonly used, machine learning algorithms are habitually classified according to the amount of inference performed by the machine during the learning progress. Specifically, we encounter the following types of learning:

1) **Supervised Learning:** In this type of learning, the machine is provided with training data, namely vectors of input-output pairs. Each input is associated with its correct output; the machine needs to infer the function underlying the relationship between the input and output data. Supervised learning is used in classification (predicting a label) and regression (predicting a continuous value) tasks.

2) **Unsupervised Learning:** the machine is only provided with data without input-output associations; it must discover patterns in the collection of sample objects and extract the underlying structure.
3) **Reinforcement Learning:** the machine does not have access to train data and is not instructed in which actions to take. Usually referred to as an *agent* in this context, the machine explores its environment and takes actions, with the goal of maximizing a numerical, long-term reward through trial and error (Mo, 2009; Sotiropoulos and Tsihrintzis, 2018).

The algorithm under consideration, AIRS-X, is based on a supervised learning approach and addresses classification tasks (Giatzitzoglou, 2018).

3.2 Immunology

Artificial immune systems are machine learning paradigms based on immunological metaphors. Therefore, an overview of theoretical immunology is critical to understanding their nature.

Immunology is the science of the immune system, the mechanism that protects its host against infections by pathogens (de Castro and Timmis, 2000). It is a relatively new science, developed after the discovery of vaccination, the existence of pathogens and the recognition that they are vectors of disease in the 19th and 20th centuries (de Castro and Timmis, 2000).

Below we will introduce an overview of the immune system as it is understood by contemporary immunology and present competing theoretical perspectives that have been adopted as AIS metaphors.

3.2.1 Overview

As stated above, the immune system is the defense mechanism of an organism, such as the human body, against infection by pathogens, such as parasites, viruses, bacteria, or fungi. It consists of two layers, the so-called *innate* and *adaptive* immune systems, which coordinate to recognize and eliminate pathogens threatening the host (Mo, 2009).

3.2.2 The innate immune system

The innate immune system constitutes the first line of defense against invading microorganisms. It is a general, i.e. not specialized, system available for immediate action against all pathogens. Responsible for controlling the infection until the adaptive immune system takes over, it consists of a complex array of physical barriers, biochemical barriers and most importantly, white blood cells such as macrophages and neutrophils (de Castro and Timmis, 2000; Dasgupta and Niño, 2009).

Macrophages and neutrophils are able to recognize non-self molecules, named *pathogen associated molecular patterns* (PAMPs); to engulf and eliminate the pathogens; to secrete molecules that alert other defense molecules, such as lymphocytes and adaptive immune system cells; and, finally, to present antigen molecules on their surface, which in turn are recognized by the adaptive immune system and which lead to the production of specialized antibodies (Dasgupta and Niño, 2009).

While the innate immune system displays a capability of discerning self from non-self molecules, it is not capable of learning and memory; that is the purview of the adaptive immune response (Dasgupta and Niño, 2009).

3.2.3 The adaptive immune system

The adaptive immune response mainly depends on the type of cells known as *lymphocytes*, specifically B-cells, produced in the bone marrow, and T-cells, produced in the thymus.

All lymphocytes carry receptors matching a single molecular antigenic pattern, i.e. they are *monospecific*. Specifically, the molecules of the receptor bind to a specific antigenic sequence, called an *epitope*, based on chemical complementarity. The structure of each receptor and thus the antigenic pattern it binds to depends on the gene sequence of each cell; the random rearrangement, mutation and splicing of genes during cell production, combined with the number of cells in circulation gives rise to virtually infinite pattern recognition capabilities (Mo, 2009).

B-cells and T-cells thus are able to recognize, bind to, neutralize, and eliminate antigens, while also activating the complement system, or plasma proteins promoting the opsonization of pathogens (Dasgupta and Niño, 2009).

3.2.4 The Clonal Selection Principle

According to the *clonal selection principle*, the common model used to describe the adaptive immune response, the lymphocytes with the highest affinity to the antigens will be able to proliferate. B-cells that recognize antigenic patterns will start secreting antibodies and will undergo mitosis (cell division). During mitosis, B-cells will also mutate at high rates (*somatic hypermutation*), thus increasing diversity and perhaps producing new cells with even higher affinities. Eventually, the dividing B-cells will mature into plasma cells, no longer dividing but secreting great numbers of antibodies. Some of them will also mature into *memory cells*, which remain in circulation, in order to rapidly eliminate pathogens in case of a second infection.

T-cells, too, undergo clonal selection, where the highest affinity cells will proliferate. However, unlike B-cells, they do not suffer mutation. Mature T-cells, also known as effector T-cells, do not produce antibodies, but T-killer or cytotoxic cells, responsible for antigen destruction.

In summary, only the lymphocytes that present the correct specificity will undergo clonal expansion, i.e. mature and proliferate (de Castro and Timmis, 2000).

In contrast, cells that react to self-molecules or *self-antigens* are culled early in the process; this *negative selection* protects the organism from being attacked by its own immune system. Failure to remove self-reactive lymphocytes leads to autoimmune diseases.

Similarly, cells with low affinities to the presented antibodies may die through apoptosis (de Castro and Timmis, 2000).

3.2.5 Learning and Memory

Negative selection and clonal selection constitute mechanisms through which the immune system adapts to its antigenic environment, learning to quickly exterminate any foreign antigens while preserving self-molecules.

Nonetheless, such strategies would be short-lived without memory. The adaptive immune system stores information in order to react to later infections by the same pathogen in a fast and accurate way. Specifically, some B-cells that were selected during the clonal expansion phase remain in circulation; if the same pathogen is encountered, these memory cells quickly recognize and eliminate the threat. Thus, a secondary infection, against which the immune system has the advantage of not only numbers but already specialized cells, will be resolved sooner and more efficiently.

In addition, memory cells will not only react to antigens they are specific against; similarly structured molecules will also cause a partial response, minimizing the time required to eliminate the infection. This immunological cross-reaction carries properties resembling the associative memory and generalization powers of neural networks and showcases how valuable a metaphor the immune system may be for artificial intelligence (de Castro and Timmis, 2000).

3.2.6 Immune Network Theory

A different approach to immunology, *immune network theory*, was proposed by N. K. Jerne in the 1970's. Though not widely accepted among immunologists, the theory provides useful metaphors for computer science and often forms the theoretical basis of artificial immune system algorithms.

Immune network theory, as the name suggests, does not view the immune system as a collection of independent molecules. Instead, it poses that immune components form a network of molecules capable of recognizing not only antigens, but each other (Mo, 2009).

According to Jerne (1974), the antibody region recognizing an antibody's epitope is called a *paratope*. Additionally, an *idiotype* is the set of epitopes carried by antibodies. Each individual antibody epitope is called an *idiotope*, and its formation is governed by the genes that code for the paratopes.

Thus, the immune system consists of a dynamic network of molecules with paratopes and epitopes, each recognizing and being recognized by each other, the recognition followed by cell activation or suppression. Even in the absence of antigens, the molecules of the system react to each other, causing a dynamic reallocation of system resources, as certain cells are suppressed and decay, while others are activated and proliferate (de Castro and Timmis, 2000; Dasgupta and Niño, 2009).

As mentioned earlier, immune network theory has proven fruitful for computer science and specifically for artificial immune systems, giving rise to paradigms such as the optimization algorithm of Mori et al. (1993), Hunt and Cooke's (1996) unsupervised learning algorithm, and AIRS (Watkins, 2001).

3.2.7 Properties of the Immune System for Artificial Intelligence

Why is the immune system chosen as a paradigm for artificial intelligence algorithms? The answer is that it exhibits powerful features that the algorithms endeavor to take advantage of. Below, we will explore some of these properties, as exposed by de Castro and Timmis, (2000).

- ***Pattern recognition*:** Immune cells are able to recognize molecular patterns, such as antigens or self- surface molecules and chemical signals.
- ***Self-identity*:** Immune cells can differentiate between self and non-self.
- ***Noise tolerance*:** immune cells activate even when pathogens are partially recognized; noise is ignored.
- ***Learning and memory*:** Negative selection, clonal selection and memory cells ensure that the immune system learns to react fast and accurately to specific antigens in its environment and ignore self-molecules.
- ***Uniqueness*:** Each immune system is unique to its organism, representing its adaptation to self-molecules and environmental antigens.
- ***Diversity*:** The immune system orchestrates various defense mechanisms, cells, and molecules to ward off infection by pathogens.
- ***Disposability*:** Each immune component is individually disposable; cells and molecules are constantly replaced.

- ***Autonomy*:** No control system administers the immune response. The immune components act autonomously and dynamically to coordinate the recognition and elimination of antigens, as well as sustain the population of immune cells.
- ***Multiple layers*:** The immune response does not depend on one particular mechanism or layer of defense. Multiple specialized layers coordinate a secure and robust defense against pathogens.
- ***Dynamic system*:** Immune repertoires constantly adapt to the presence of pathogens and self-molecules. Self-reactive cells are culled while antigen-specific antibodies proliferate. The population changes with each successive infection to reflect the pattern of known antigens.
- ***Anomaly detection*:** Immune cells are able to recognize non-self-molecules, even if not been encountered previously.
- ***No secure layer*:** Immune cells may attack any cell or molecule, including self-molecules or even other immune cells, if infected or otherwise harmful to the organism.
- ***Distributivity*:** Complementary to autonomy, distributivity ensures the immune cells are present in all regions and not subject to centralized concentration or control.
- ***Resilience*:** The immune system is able to perform its role, if less effectively, even in case of disturbances such as malnutrition or stress.
- ***Fault tolerance*:** This property emerges as a result of disposability. Even if one immune component fails, other immune cells or molecules will fill in the role. For example, antibodies will be assisted by macrophages and natural killer cells or other high affinity antibodies.

- ***Robustness*:** Similarly, the distribution of multiple types of immune cells and molecules that act complementarily ensures a robust response to pathogens.
- ***Predator-prey pattern of response*:** The number of immune cells increases in proportion to the number of pathogens, thus preventing the pathogens from overwhelming the organism's defenses. Conversely, the population of antibodies will be reduced after a successful immune response.
- ***Self-organization*:** No information directs the immune response. The immune reaction is guided only by the intrinsic properties of clonal and negative selection.
- ***Integration*:** The immune system does not act in vitro, but interacts, regulates, and is regulated by other bodily systems.

In summary, it becomes clear that the immune system presents an exceedingly attractive metaphor for artificial intelligence and machine learning. In the following chapters, we will explore how researchers have exploited these desired properties to create high-functioning artificial immune system algorithms and to solve problems in a wide range of fields.

4. Framework and models

4.1 A general framework

Before we delve into the various models inspired by the immune system, it would be useful to describe the general framework to which artificial immune systems adhere. When we have defined the main skeleton of our paradigms, we will be better equipped to understand the variations in modelling according to the specific domain areas and problems they are applied to.

According to de Castro and Timmis (2000), an artificial immune system should consist of:

1. A representation for the immune components.
2. A mechanism or set of mechanisms to evaluate their affinity to their environment and each other.
3. Rules of adaptation that describe how the system will evolve with time.

In addition to this high-level framework, we should introduce some terms that define common components in AIS.

4.2 Shape-Spaces

The term shape-space was coined by Perelson and Oster (1979)[1], to describe the generalized features of an antibody's binding region. In more detail, the researchers posited that a region capable of pattern recognition is defined by a set of parameters, such as shape, charge, complementary chemical groups, length, etc. These *L* parameters form an L-dimensional space, a

[1] As quoted in Sotiropoulos and Tsihrintzis (2017).

point in which is called *space-shape S*, and which defines the pattern recognition capability of a specific cell or molecule (Sotiropoulos and Tsihrintzis, 2017).

Consequently, an organism with N antibodies will have a shape-space of N points, which lie within a finite volume V. Assuming a *cross-reactivity threshold* ϵ, which defines the complementary region, i.e. the region each antibody interacts with, then the resulting volume Vϵ defines the *recognition region*. Any molecule occurring in the recognition region is recognized by the complementary antibodies.

A molecule m may be represented as an attribute string of length L. For example:

$$m = (m_1, m_2, \dots, m_L)$$

Depending on the domain the AIS is applied to and according to the problem at hand, the algorithms may use real-valued, integer, symbolic, or Hamming space-shapes (Sotiropoulos and Tsihrintzis, 2017).

Furthermore, the degree of activation between two attribute strings or their *affinity measure* may be defined as following[2]:

$$D : S^L \times S^L \rightarrow R^+$$

The above expression maps the two interacting molecules and returns a nonnegative real value representing their distance. An *affinity landscape* is the (graphic) representation of collective antibody affinities with a given antigen.

[2] In Hamming and symbolic space-shapes, the affinity measure is defined differently, as it depends on the degree of complementarity between the attribute strings.

4.3 Models

Having introduced the general framework of AIS, we can now move to presenting various high-level models, based on different components of the immune system. We will follow de Castro and Timmis (2000) and categorize them into 1) bone marrow models, 2) thymus models, 3) clonal selection algorithms and 4) immune network models.

4.3.1 Bone marrow models

The bone marrow is the site of hematopoiesis, i.e. blood cell generation. This is where lymphocytes are born and where B-cells differentiate independent of antigens. Algorithms based on a bone marrow model generate populations of immune components -cells or receptors-, usually by creating attribute strings. In less sophisticated algorithms, the strings are generated randomly, whereas more advanced ones use gene libraries to mimic genetic recombination.

4.3.2 Thymus models

The thymus is a gland where antigen-independent T-cells mature and differentiate. T-cells that recognize self-MHC molecules are considered immunocompetent and allowed to circulate (positive selection), whereas T-cells that react to self-antigens are eliminated (negative selection). Accordingly, thymus models are divided into positive and negative selection algorithms.

In positive selection algorithms, a population of T-cells is generated and their affinity to certain molecules is measured. If the affinity measure supersedes a threshold, the T-cell is considered immunocompetent and introduced into the population (cf. Seiden and Celada, 1992).

On the other hand, in negative selection algorithms, the population of T-cells is evaluated for their affinity to self-molecules and are accordingly eliminated. The first negative algorithm application was performed by Forrest et al. (1994) in the field of computational security.

4.3.3 Clonal selection algorithms

As the name suggests, these algorithms are based on the principle of clonal selection, that is the specific maturation and differentiation of B-cells according to their antigenic affinity. The algorithm starts by initializing a population and presenting an antigen pattern. It then selects the individuals with the highest affinities for clonal expansion and mutates them in inverse proportion to their affinity. Finally, the least reactive B-cells are replaced by randomly generated ones and the cycle begins anew. CLONALG by de Castro and Von Zuben (2000a) employs these steps for pattern recognition and function optimization.

Other examples include the B-Cell Algorithm by Kelsey and Timmis (2003), the Multi-objective Immune System Algorithm by Coello Coello and Cortés (2002) or the Simple Immunological Algorithm by Cutello and Nicosia (2007).

4.3.4 Immune network models

Finally, immune network models receive their inspiration in immune network theory by Jerne (1974). As mentioned earlier, the theory is based on the premise that even in the absence of antigenic stimuli, the components of the immune system are capable of recognizing and reacting to each other.

Farmer et al. (1986; 1987) pioneered the field. Their model consists of binary strings in a Hamming space-shape, with concatenated idiotopes and paratopes. If a paratope recognizes an idiotope, it proliferates; in contrast, the recognized molecule bearing the idiotope may be eliminated, along with any antibodies with low antigenic affinities.

Varela and Coutinho (1991) proposed a second-generation immune network, emphasizing its structure or connections between the components, its dynamics, or variation with time and its

metadynamics, or the continuous elimination and replacement of immune components according to their selective affinities. Both of the above systems do not take foreign antigens into consideration, instead focusing on the interactions between self-elements.

A decade later, Timmis (2000) presented the Resource Limited Artificial Immune Network or RAIN. Its novelty lies on the fact that B-cells with low affinities are eliminated based on a limited amount of resources. The higher the affinity, the higher amount of resources each cell collects; at the end of the cycle, the B-cells with the least resources are removed from the system.

Finally, de Castro and Von Zuben (2000b) proposed the Artificial Immune NETwork or AINET the same year, with similar properties: clonal expansion, clonal memory, clonal suppression and metadynamics.

5. Scope

Considering the amount of desirable computational features the immune system displays, it is not surprising that AIS have been applied to a wide range of sectors, from modelling the immune system to solving real problems (Sotiropoulos and Tsihrintzis, 2017). This section is devoted to presenting a broad overview of AIS applications. While we cannot discuss all works in detail within the confines of a dissertation, we will aim to showcase the most important categories.

5.1 Computer and Network Security

According to Brownlee (2011), the earliest formative works of AIS were computer security algorithms, based on the notion that anti-virus and other defensive programs could be developed around metaphors from the immune system. The idea is direct and intuitive and is still inspiring further research (de Castro and Timmis, 2000).

Kephart (1994) and Kephart et al. (1997; 1999) proposed a computer immune system, the Immune Anti-Virus, able to recognize and eliminate previous unknown threats, such as viruses. The implemented recognized viruses based on full or partial matches to virus signatures and contained control mechanisms to limit potential "autoimmune" responses. Furthermore, the authors equipped the system with immune memory and repair mechanisms: the system would not only remember the viral signature in future infections, but would try and patch the vulnerabilities to preclude them altogether (de Castro and Timmis, 2000; Brownlee, 2011).

A similar approach has been taken by Forrest et al. (1994; 1996; 1997) and Hofmeyr and Forrest (1999; 2000), based on her and her collaborators' work on negative selection algorithms.

Somayaji et al. (1997), on the other hand, argued that a complete mapping of the immune system and computer security is not possible. Specifically, the immune system has evolved to ensure

integrity and availability, but does not possess mechanisms to address confidentiality, accountability, and correctness.

Kim and Bentley (1999a; 1999b; 2001), Dasgupta (1999a; 2000), Skormin et al. (2001) and Gu et al. (2000) proposed network intrusion detection and elimination systems based on metaphors from and properties of the immune system, such as distributivity, self-organization, evolvable antibodies, co-stimulation etc. (de Castro and Timmis, 2000).

More recent approaches include works by Anchor et al. (2002), Aickelin et al. (2004), Haag et al. (2007), Shafiq et al. (2007) and Tedesco et al. (2010).

5.2 Anomaly Detection

Partially overlapping with computer and network security and pattern recognition, the field of anomaly detection has attracted several AIS researchers (de Castro and Timmis, 2000; Sotiropoulos and Tsihrintzis, 2017). Early in the history of AIS, Aisu and Mizutani (1996) created an immune network for image inspection, while McCoy and Devarajan (1997) applied a negative selection algorithm to aerial image segmentation. Dasgupta and Forrest (1996), meanwhile, adopted negative selection to examine time series data and detect new elements. Anomaly detection techniques based on immune systems has been further applied in hardware fault detection (Canham et al. 2003), malicious network node detection (Trapnell, 2005), in foraging swarm robotic systems (Lau et al., 2009) or even spam detection (Oda et al., 2005; Bezerra et al., 2006).

5.3 Optimization

Optimization constitutes another sector where AIS research has flourished. Their purpose may generally be stated to be to design a system as effective or functional as possible (de Castro and Timmis, 2002). Most authors implement the clonal selection principle to solve uni- or multi-modal,

time dependent or combinatorial problems (Sotiropoulos and Tsihrintzis, 2017). The seminal paper by de Castro and Zuben (2000a) and the following version (2002) introduced the CLONALG algorithm, which we have mentioned earlier, in Section 4.3.3. Bersini and Varela (1990) tackled the problem of function optimization using an immune-evolutionary hybrid algorithm, while Hajela and Yoo (1999) addressed constrained optimization. Joshi (1995) mapped the immune network model to a production system and applied it to inventory optimization. Combinatorial optimization research through AIS was explored by Toma et al. (1999) and time dependent optimization by Gaspard and Collard (2000). Other noteworthy approaches include the B-cell algorithm (Timmis, 2004) and aiNet by de Castro (2002).

5.4 Robotics

Owing to the exciting properties of the immune system, such as robustness, pattern recognition and detection, learning and coordination, AIS have been widely implemented in the field of robotics (Sotiropoulos and Tsihrintzis, 2017). Examples include robot navigation (Ishiguro et al., 1996;1998; Villalobos-Arias et al., 2004; Luh et al., 2004; Ko et al. 2005; Lee et al., 2007), control and coordination of agents (Bersini, 1991; Ootsuki and Sekiguchi, 1999; Takahashi and Yamada, 1997; Ko et al., 2004; Lau et. Al., 2005, Lu et al., 2009), as well as collective behavior (Lee and Sim, 1997; Jun et al., 1999; Lee et al., 1999).

5.5 Machine Learning

Machine learning is generally defined by Sotiropoulos and Tsihrintzis (2017) as “the process of acquiring knowledge from experience and being able to generalize that knowledge to previously unseen problem instances”. Therefore, the immune system, with its capability of detecting, learning, memory and generalization, seems an ideal metaphor for machine learning problems.

Indeed, researchers have devoted considerable effort in applying AIS techniques to pattern recognition and learning problems (de Castro and Timmis, 2002; Sotiropoulos and Tsihrintzis, 2017).

Most proposed systems fall under the sub-categories of classification and clustering (Sotiropoulos and Tsihrintzis, 2017). AIS-based unsupervised learning, clustering examples include the Artificial Immune NETwork (aiNet) by de Castro and Von Zuben (2002), the adaptive radius immune algorithm by Bezerra et al. (2005) and various approaches to the B-cell algorithm (Clark et al., 2005; Bull et al., 2006). Furthermore, we may include the biclustering text-mining approach by de Castro et al. (2007), immune network data mining by Ciesielski et al. (2006), Hart and Ross' (2002) sparse distributed memory techniques for clustering non-stationary data, as well as Reche and Reinherz's (2004) research on MHC molecules.

Supervised learning and specifically classification tasks have also been tackled with success by AIS algorithms (Sotiropoulos and Tsihrintzis, 2017). AIRS, the Artificial Immune Recognition System, is a noteworthy example of AIS classification (Watkins et al., 2004) and will be explored further in Section 6.

Extensive literature explores approaches to AIS classification by formalizing and benchmarking immune classification algorithms (Ceong et al., 2003; de Castro et al., 2005; Chen and Zang, 2006; Garain et al., 2006; Polat et al., 2006; Figueredo et al., 2007; Oates et al., 2007; McEwan and Hart, 2009).

5.6 Further applications

While, as mentioned before, we cannot possibly include all applications of AIS within the confines of a dissertation, we would be remiss to mention the vibrant fields of scheduling (Mori et al., 1998; Lambert et al., 1999; Cui et al., 2000), protein structure prediction (Michaud et al., 2001), smart home applications (Dilger, 2006; 2007; Lehman and Dilger, 2006), recommender systems (Cayzer and Aickelin, 2008; Morrison and Aickelin, 2008) and bioinformatics (Tarakanov et al., 2002; Bezerra and de Castro, 2003; Goncharova et al., 2003).

6. Artificial Immune Recognition System (AIRS)

In the present section, we will go over the Artificial Immune Recognition System or AIRS algorithm, the predecessor of the main subject of this thesis, AIRS-x. We will describe its main features and the biological metaphors it is based upon, as well as its aim and its efficiency.

6.1 Introduction

According to Brownlee (2005), AIRS is a supervised learning classification algorithm, one of the first to be applied to such problems. Designed by Watkins (2001), it consists of memory cells competing for system-wide resources (Sotiropoulos and Tsihrintzis, 2017). The memory cells best suited for classifying data are allowed to survive and proliferate, through an evolutionary mechanism that eliminates those antibodies that were not fit enough (i.e., did not exhibit high affinity to the antigen) to collect system resources (Golzari et al., 2008).

6.2 Characteristics and efficiency

AIRS has been one of the most common used classification AIS algorithms (Golzari et al., 2008) as it exhibits important characteristics (Brownlee, 2005; Sotiropoulos and Tsihrintzis, 2017):

- **Self-regulation:** the appropriate architecture is automatically decided upon by the evolutionary algorithm during training. The user does not have to select it themselves, thus eliminating the possibility of an unsuited pick. Furthermore, this feature enables the algorithm to adjust and run consistently against different datasets.
- **Performance:** AIRS is a highly competent algorithm, as shown by empirical research against other common classifiers, such as neural networks. For specific datasets, the algorithm outperforms other classifiers (cf. Watkins and Boggess, 2002; Goodman et al., 2003; Golzari et al., 2008.; Giatzitzoglou, 2018).

- **Generalization:** AIRS performs data reduction during training, thus representing the data with a reduced number of exemplars.
- **Parameter stability:** AIRS is capable of maintaining accuracy even if its parameter values vary to adapt to different datasets.

6.3 Immune metaphors

AIRS borrows ample metaphors from the immune system; however, it is important to remember that they are mere abstractions chosen for desirable features and not intended accurate representations of the immune system or other biological components (Brownlee, 2005).

First, the main component of the algorithm, the classifier, is an abstraction of B-cells (Giatzitzoglou, 2018). Just as the B-cells are responsible for recognizing and binding to antigenic patterns, the AIRS classifier recognizes presented patterns. Moreover, the classifier commits the pattern to memory and can recognize it after training, just like memory cells during immune learning and memory. In addition, classifier-antibodies adopt T-cell behavior, such as presentation of the antigen to B-cells for recognition (Sotiropoulos and Tsihrintzis, 2017).

Furthermore, the algorithm displays properties akin to immunological concepts such as affinity maturation, clonal expansion and somatic hypermutation (Brownlee, 2005). The classifiers with the highest affinity or recognition ability proliferate (clonal expansion) and mutate (somatic hypermutation) according to their stimulation levels. Thus, the classifiers achieve affinity maturation; in other words, they have adapted to efficiently recognize the presented pattern. (Sotiropoulos and Tsihrintzis, 2017)

A key difference between the immune system and the AIRS metaphor is the concept of the artificial recognition ball or ARB. Essentially, the ARB represents a collection of B-cells with the same

pattern recognition “antibodies”. The ARB is characterized by a single stimulation value and a certain amount of resources (Giatzitzoglou, 2018).

6.4 An overview of the algorithm

The main aim of AIRS is to create a pool of ARBs that can recognize and classify target data (Brownlee, 2005). The diagram below presents its lifecycle:

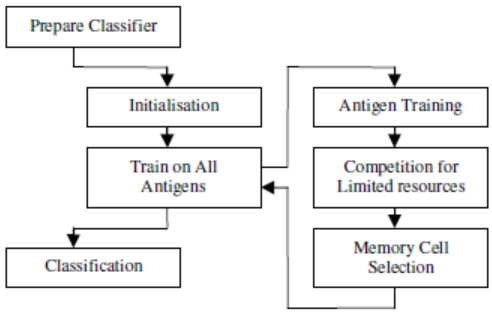

Figure 1 - AIRS Lifecycle (Source: Brownlee, 2005)

6.4.1 Initialization

Data preparation: All training data is prepared and normalized. Any training data attribute, as well as the Euclidean distance between any two data vectors (antibody or antigen), lies in the range [0,1]. Any nominal attributes must be converted to binary data.

B-cell generation: During this optional phase, a random number of data vectors are chosen to act as memory cells.

Affinity threshold computation: This threshold is set by calculating the average affinity of all or part of the training set (Brownlee, 2005; Sotiropoulos and Tsihrintzis, 2017).

6.4.2 Antigen Presentation

AIRS presents antigens to the memory cells one at a time and assigns a stimulation value to each cell, according to their recognition ability; the cell with the highest value undergoes cloning and element-wise mutation. The number of clones and the rate of mutation is proportional to the cell's stimulation value. The clones are then placed in the ARB pool (Brownlee, 2005; Sotiropoulos and Tsihrintzis, 2017).

6.4.3 Competition

Once the clones have been added to the ARB pool, they are presented with the antigen and their stimulation level is calculated. Then, a series of sub-steps are repeated until a predetermined average stimulation threshold is reached:

1. Stimulation levels are normalized.
2. Resources are allocated to each antibody, according to its stimulation level.
3. If the number of allocated resources is greater than a predefined limit, cells with the lowest stimulation level lose resources, until the total allocation falls within the allowed range. Afterwards, antibodies that have zero resources are removed from the pool.
4. Stimulation levels are calculated once more, and the best candidate is chosen as the memory cell (Sotiropoulos and Tsihrintzis, 2017).

The following diagram presents a high-level overview of the competition process:

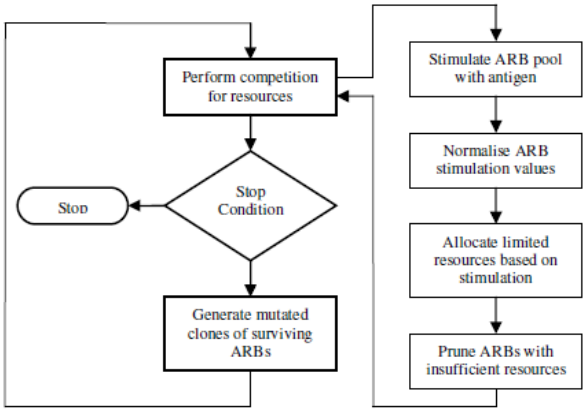

Figure 2 - ARB Competition (Source: Brownlee, 2005)

6.4.4 Memory cell selection

The highest scoring ARB candidate is then placed in the memory cell pool if its stimulation value is higher than that of the original memory cell. The original may even be eliminated if its affinity differs greatly from the newly introduced candidate's affinity. The now trained ARB classifier performs its task using a k-neighbors approach, i.e. by a majority vote on the outputs of the k best matches (Brownlee, 2005).

6.4.5 Conclusion

As mentioned earlier, AIRS has been a fairly successful algorithm, often outperforming alternative classification techniques in a wide range of datasets. Furthermore, it features a simple, easy to understand immune metaphor; generalization, associative memory, and continuous learning; the ability to handle numeric, nominal, and discrete values and offers predictable training times (Brownlee, 2005). However, since its conception, a series of extensions, modifications and alteration proposals have been made (Giatzitzoglou, 2018). In the following section, we will examine one of them, the AIRS-x algorithm, in detail, and analyze its advantages over the original AIRS algorithm.

7. AIRS-x

AIRS-x has been proposed as an alternative formulation to the AIRS algorithm, partly inspired by negative selection algorithms. Its main aim is to increase the overall accuracy of AIRS while improving its efficiency. The algorithm under investigation was first presented as iAIRS by Giatzitzoglou (2018) in his master's dissertation, and further developed by Giatzitzoglou et al. in the 2019 paper. Before diving into the algorithm's specifics, let us first take a look at the deficiencies of AIRS the authors aimed to correct, or improve upon, in their proposal.

7.1 AIRS Deficiencies

While AIRS has proven effective in the task of classification, several researchers have pinpointed deficiencies since its inception.

7.1.1 Stimulation Threshold

McEwan and Hart (2009) underlined that AIRS is elitist, in the sense that it selects the candidate cell closest to the antigen to seed the initial ARB pool and consequently only chooses the closest mutant as the candidate memory cell. This process does not certify that the selected cell is the fittest. Furthermore, there is no feedback between different classes of antibodies, further limiting the algorithm's ability to accurately generate a representation of the target data (Giatzitzoglou, 2018). The authors suggested that the stimulation threshold should not be constant across all regions, but instead should vary according to the region's homogeneity or heterogeneity.

As a reminder, the stimulation threshold is a user-defined parameter between [0, 1] in the original AIRS algorithm; it controls when the training process stops and remains fixed throughout (see Section 6.4.3). It controls how refined (i.e. close to the antigen) the resulting ARBs will be (Brownlee, 2005). A fixed stimulation threshold, especially in the absence of feedback between

different classes of antibodies, may distort the final representation, even with a high selection criterion (Giatzitzoglou et al., 2019).

7.1.2 Mutation Process

Another area where scholars have uncovered flaws is the mutation process, which relies in a stochastic search. According to Watkins (2002) and McEwan and Hart (2009), it renders the algorithm computationally complex. In more detail, the computational complexity is calculated by Sotiropoulos and Tsihrintzis (2017) to equal to O(NL), where is the total number of antibodies and L is the number of shape-space features. Furthermore, it does not seem that this complexity results in better classification accuracy, since mutations are purely random, often resulting in inferior offspring, which will then be discarded by the algorithm's inherent elitism. The problem becomes even more pronounced in high-dimensional vectors (Giatzitzoglou, 2018).

7.1.3 Other flaws

Researchers have proposed additional changes to the algorithm. Brownlee (2005), for example, suggests that there are not sufficient grounds to mandate data normalization. He further proposes a restructuring of the sample cell generation process, which has been proven to be computationally costly and unnecessary for the algorithm's accuracy (Goodman et al., 2003). Finally, a tree structure is proposed to replace the current classifier data structure to reduce complexity (Brownlee, 2005).

While these proposals merit further examination and research, they will not be the focus of this thesis, as they are not directly tackled by the AIRS-x algorithm.

7.2 AIRS-x

7.2.1 Notation

In this section, we will present the notation used in the AIRS-x proposal, as it is necessary to understand the introduced functions and diagrams (Giatzitzoglou, 2018):

- AT: affinity threshold
- ATS: affinity threshold scalar
- CR: clonal rate
- MR: mutation rate
- HCR: hyper clonal rate
- TR: total resources
- RA: resources allocated
- NRR: number of resources to be removed
- C: number of classes
- DT: distance threshold
- DTS: distance threshold scalar
- Ag ϵ $M_{M\times I}$: matrix of training antigenic patterns, constructed by concatenating the matrices:

$$Ag = [Ag^{(1)}; \dots ; Ag^{(C)}],$$

where $Ag^{(k)} \in M_{M_a\times L}$ the sub-matrix of the kth class, such that:

$$M = \sum_{k=1}^{C} M_k$$

while $Ag^{(k\prime)}$: the sub-matrix storing training instances for all classes minus the kth class.

- $Ab \in M_{N_k \times L}$: the matrix storing the total antibody pool. This matrix is constructed by concatenating the following matrices:

$$Ab = [Ab^{(1)}; \dots; Ab^{(C)}],$$

where $Ab^{(k)} \in M_{N_k \times L}$ the sub-matrix of antibodies for the k[th] class, such that:

$$N = \sum_{k=1}^{C} N_k$$

- $S^{(k)} \in M_{1 \times N}$: the vector storing the stimulation levels for the currently presented antigen, for all the antibodies. It is a concatenation of vectors:

$$S = [S^{(1)}, \dots, S^{(C)}]$$

where $S^{(k)} \in M_{1 \times N_k}$: the sub-vector containing the stimulation levels of the k[th] class antibodies.

- $s \in M_{1\times C}$: the vector storing the average stimulation level of all antibodies for each class, such that:

$$s = [s^{(1)}; \dots; s^{(C)}]$$

- $R \in M_{1\times N}$: the vector storing the resources allocated to all antibodies after the antigenic presentation. It is a concatenation of vectors:

$$R = [R^{(1)}, \dots, R^{(C)}]$$

where $R^{(k)} \in M_{1\times N_k}$: the sub-vector containing the resources allocated to the k[th] class antibodies.

- $M \in M_{m\times L}$: the matrix storing the memory antibodies for all classes, resulting by the concatenation of the matrices:

$$M = [M^{(1)}, \dots, M^{(C)}]$$

where $M^{(k)} \in M_{m\times L}$: the sub-matrix containing the memory antibodies for the k[th] class, such that:

$$m = \sum_{k=1}^{C} mk$$

7.2.2 Alterations

In this section, we will explore the alterations to the AIRS algorithm proposed by Giatzitzoglou (2018) and Giatzitzoglou et al. (2019). Specifically, we will present novel implementations of 1) the stimulation threshold and 2) the mutation rate.

7.2.2.1 Variable Stimulation Threshold

Giatzitzoglou (2018) and Giatzitzoglou et al. (2019), inspired in part by the Forrest et al. (1994) paper and the concept of negative selection, proposed that an antibody should not be selected only based on its stimulation level to the presented antigen, but also to antigens of different classes.

In more detail, the authors introduced two novel parameters: the *Distance Threshold* (DT) and the *Distance Threshold Scalar* (DTS). The stimulation threshold equals the normalized Euclidean distance d between the antigen and its closest antigen of a different class, multiplied by the Distance Threshold; furthermore, the closest the antigen to an antigen of a different class, the higher the Distance Threshold. This results in stricter selection criteria for antibodies that recognize antigenic patterns of different classes, minimizing the chances of overlapping during classification. On the contrary, an antibody that is not stimulated by patterns of a different class will not face an overly strict criterion, which makes generalization difficult in the original AIRS algorithm (Giatzitzoglou, 2018).

Furthermore, the researchers propose the addition of an artificial antigenic pattern sAg_j for every antigen $Ag_j^{(k)}$, so that their Euclidean distance equals $DTS \times DT \times d$. Meanwhile, the artificial antigen's distance from the closest antigen from a different class equals $DTS \times DT \times d + DT \times d$. The training concludes only if the antibody stimulation levels are greater than the stimulation threshold ST_j for both the artificial and the original antigen (Giatzitzoglou et al., 2019).

Mathematically, the Euclidean distance d between the antigen and its closest antigen of a different class is computed as follows:

$$Ag'^{(k')}_j = \arg \min_{ag \in Ag^{(k')}} D\left(Ag_j^{(k)}, ag\right), \text{ where } k \neq k'$$

$$d = \min D(Ag_j^{(k)}, Ag'^{(k')}_j)$$

The artificial antigen is computed by the equation below:

$$sAg_j = Ag_j^{(k)} + DT \times DTS \times Diff(Ag_j^{(k)}, Ag'^{(k')}_j)$$

where $Diff\left(Ag_j^{(k)}, Ag'^{(k')}_j\right) \in M_{1\times L}$ the element-wise subtraction array of $Ag_j^{(k)}$ and $Ag'^{(k')}_j$, calculated as follows:

$$Diff(A, B) =< A^{(1)} - B^{(1)}, \dots, A^{(L)} - B^{(L)} >$$

The antibody's stimulation level is given by the equation below,

$$s_j^k = 0.5 \times stimulation(Ag_j^{(k)}, \widehat{m}_j^k) + 0.5 \times stimulation(sAg_j, \widehat{m}_j^k)$$

and the sub-matrix $S_j^{(k)}$ is populated with the antibody stimulations for the kth class to the jth antigen of that class and the sAg_j, as follows:

$$S_j^{(k)}(i) = 0.5 \times stimulation\left(Ag_j^{(k)}, Ab_i^{(k)}\right) + 0.5 \times stimulation\ (sAg_j, Ab_i^{(k)})$$

The average stimulation level $s_j(k)$ to antigen $Ag_j^{(k)}$ and the average stimulation level $s'_j(k)$ to the artificial antigen sAg_j are given by the equations below (Giatzitzoglou, 2018):

$$s_j(k) = \sum_{i=1}^{N_k} stimulation\ (Ag_j^{(k)}, Ab_i^{(k)})$$

$$s'_j(k) = \sum_{i=1}^{N_k} stimulation\ (sAg_j, Ab_i^{(k)})$$

The figures below showcase the target stimulation areas for AIRS and AIRS-x:

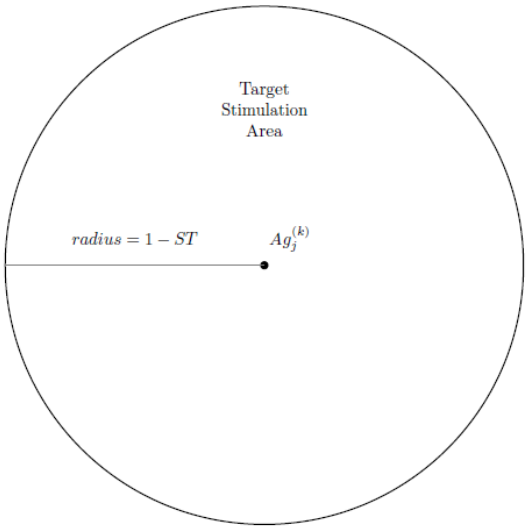

Figure 3 - Target Stimulation Area for AIRS (Source: Giatzitzoglou, 2018)

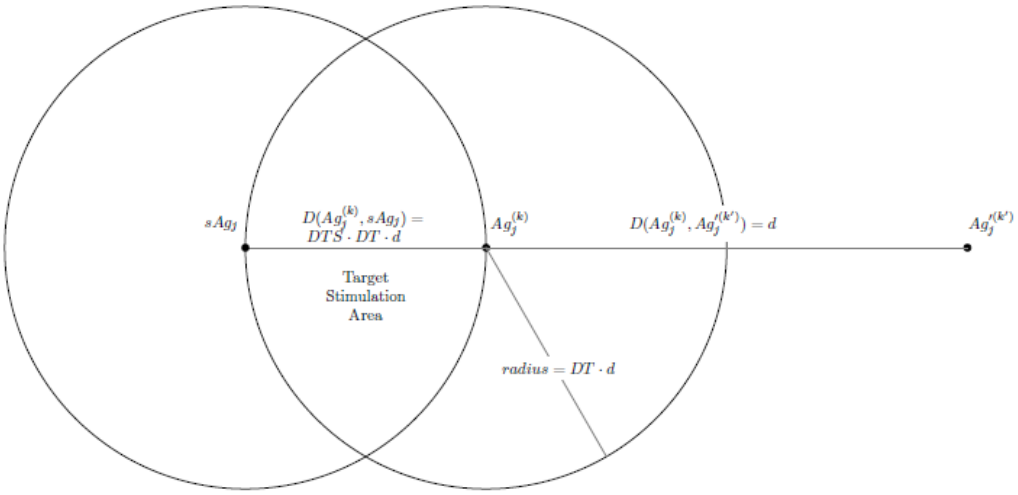

Figure 4 - Target Stimulation Area for DT = 0.5 and DTS = 1 (Source: Giatzitzoglou, 2018)

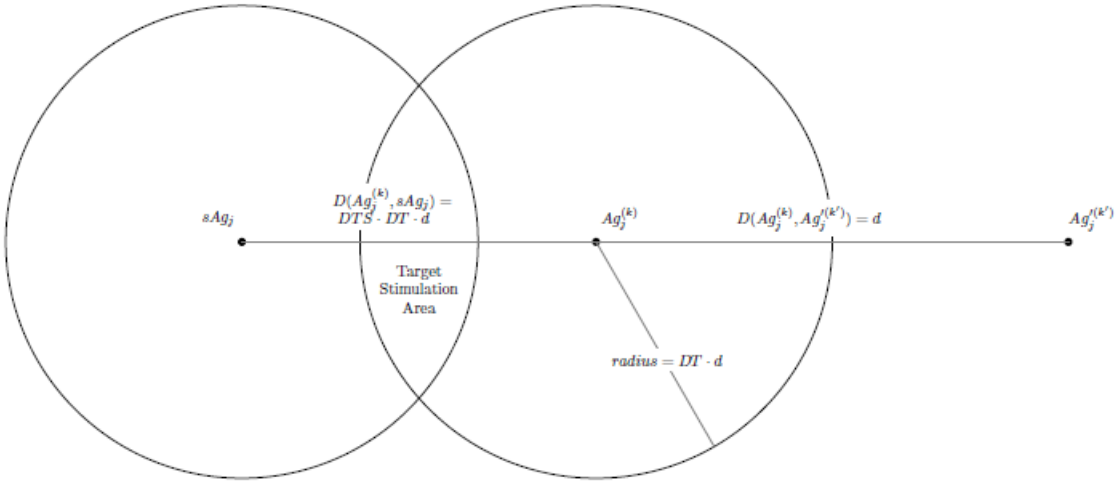

Figure 5 - Target Stimulation Area for DT = 0.5 and DTS = 1.5 (Source: Giatzitzoglou, 2018)

In the graphs below, we notice that the stimulation threshold rises with lower DT values, as expected:

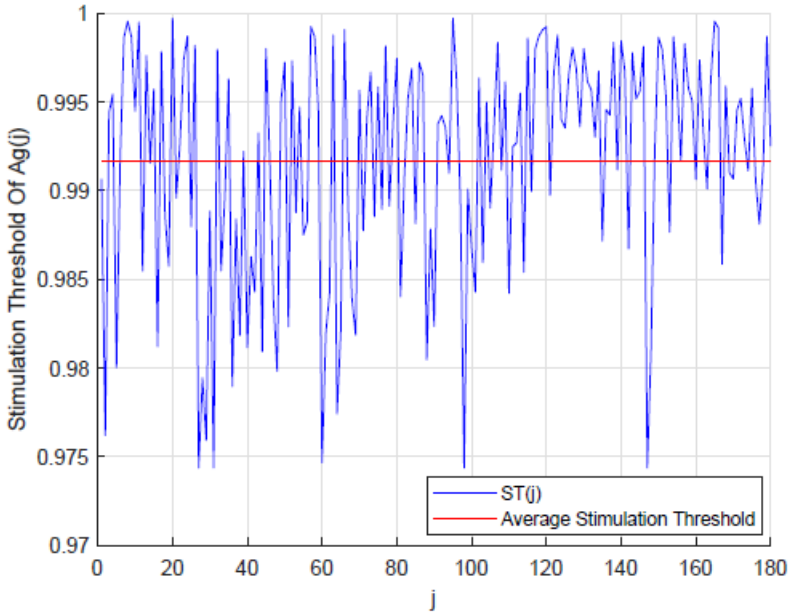

Figure 6 - Stimulation Threshold, DT = 0.1 (Source: Giatzitzoglou, 2018)

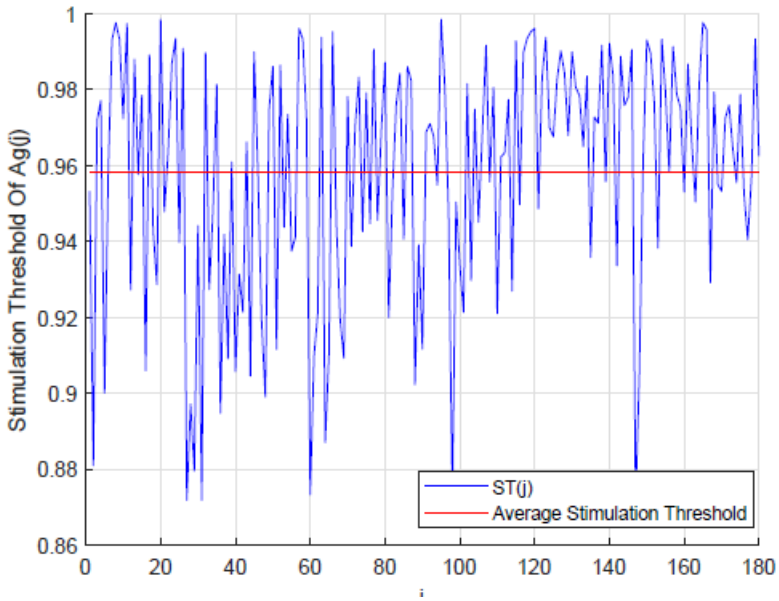

Figure 7 - Stimulation Threshold, DT = 0.5 (Source: Giatzitzoglou, 2018)

7.2.2.2 Directed Mutation Process

As mentioned earlier, the mutation routine of AIRS is purely stochastic, resulting in high computational complexity and low effectiveness (McEwan and Hart, 2009). To remedy the process, the authors propose that the matrix of mutated clones is calculated as follows:

$$C^{(k)}(l,p) = \begin{cases} \widehat{m}_j^k(p) + 0.5 \times P_j(l,p) \times \delta_j^k(p), Q_j(l,p) < MR; \\ \widehat{m}_j^k(p), otherwise \end{cases}$$

where $\delta_j^k(p) = 0.5 \times Diff\left(sAg_j(p), \widehat{m}_j^k(p)\right) + 0.5 \times Diff(Ag_j^{(k)}(p), \widehat{m}_j^k(p))$

Essentially, the mutation range equals the average of two distances:

1. The distance between the antigen $Ag_j^{(k)}$ and the memory cell and
2. The distance between the artificial antigen sAg_j and the memory cell

This proposal does not modify the magnitude of the mutation range; as in AIRS, it consists of a random value between [0, 1] (Giatzitzoglou, 2018).

The novel formulation, while still employing stochastic processes, results in a better aimed mutation process and antibody calibration, without abandoning potentially beneficial diversification (Giatzitzoglou et al., 2019).

The following diagrams present the mutation range areas of AIRS and AIRS-x:

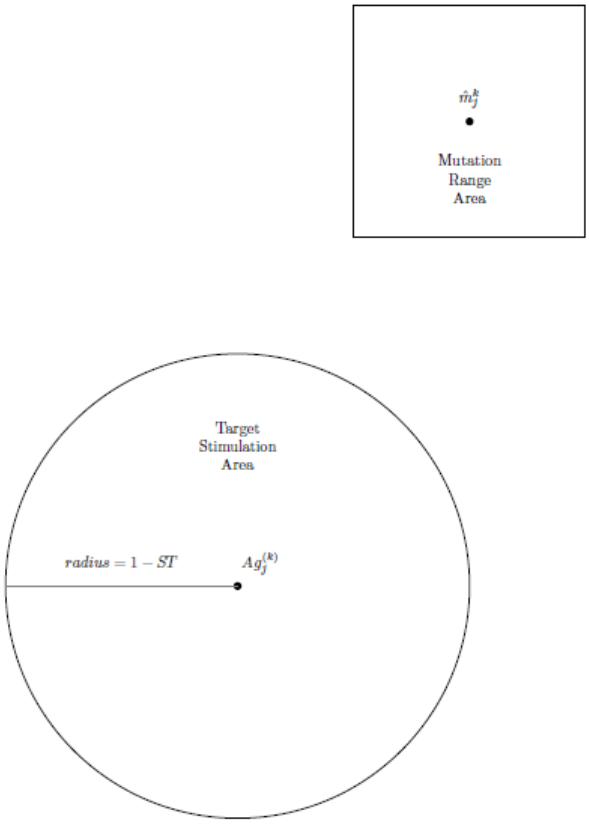

Figure 8 - Mutation range area for AIRS (Source: Giatzitzoglou, 2018)

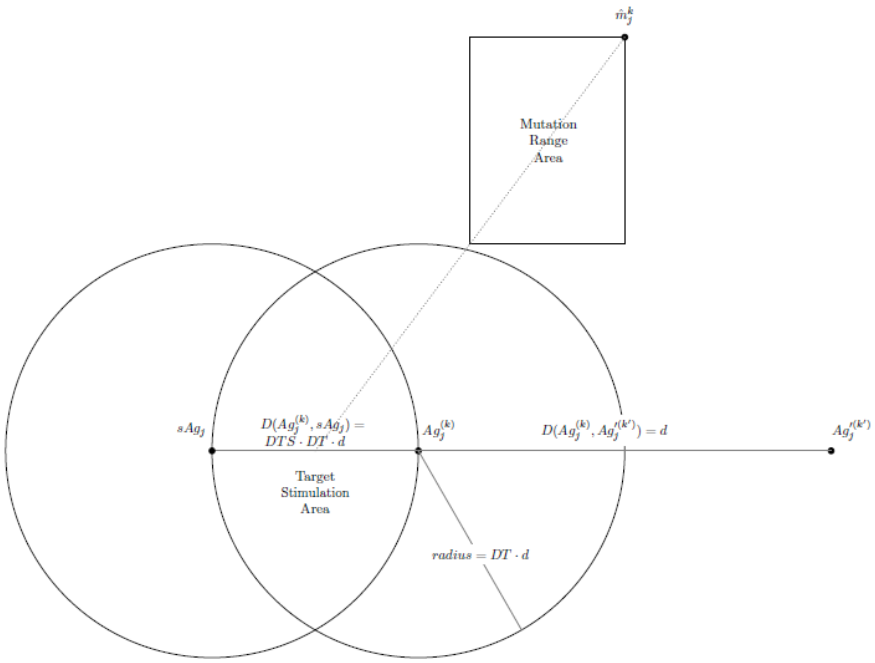

Figure 9 - Mutation Range Area for AIRS-x (Source: Giatzitzoglou, 2018)

7.2.3 Evaluation

Both Giatzitzoglou (2018) and Giatzitzoglou et al. (2019) conducted experimental testing, to validate the proposed formulation and compare its accuracy, data reduction and efficiency with the original.

The data sets used consisted of synthetic – MATLAB generated multivariate normal distributed random numbers – and real-world data, including 1000 pieces of multi-class modern Western music, the Wisconsin Diagnostic Breast Cancer (WDBC) data set, as well as the Ionosphere data set. The authors reported greater classification accuracy, increased efficiency, and higher data reduction.

To summarize, the proposed alterations to AIRS appear promising, as the novel formulation exhibits significant enhancements over the original algorithm. In the next section, we will delve deeper into the algorithm, run several training sessions, and observe the results.

7.3 Further Exploration: Sentiment Analysis

7.3.1 Sentiment Analysis

Sotiropoulos et al. (under publication) have proposed the use of AIS as classifiers in the relatively new field of sentiment analysis. Sentiment analysis aims to classify text-based sentiment, in order to computationally understand expressed opinion (Soleymani, 2017). Text data is usually extracted through data mining sources such as social media and the resulting datasets often suffer from class imbalance (Sotiropoulos et al., under publication). Consequently, an AIS classifier would be an ideal candidate for the task, since – as state previously – they excel at anomaly detection.

7.3.2 Dataset

The dataset used in this thesis contains 9,000 tweets randomly collected by Sotiropoulos et al. (under publication) between 2013-2015. The authors processed the raw data according to the Vector Space Model (Salton, 1975), to produce a corpus of predominantly negative sentiment tweets (6460:3132), individually labelled by students.

7.3.3 Results

We proceeded to run multiple training sessions to evaluate how accurate the algorithm is in the task of sentiment analysis, especially compared to the performance of the original AIRS. Furthermore, we tweaked parameters – the distance threshold, affinity threshold scalar, clonal rate, stimulation threshold – to observe any resulting differences in accuracy.

Table 1 presents the overall accuracy of AIRS and AIRS-X after seven sessions. AIRS noticeably underperforms in comparison, with a difference in accuracy ranging between 11 to 37%.

AIRS	**AIRS-X**	+/-
67%	81%	**+14%**
48%	79%	**+31%**
46%	77%	**+31%**
68%	79%	**+11%**
42%	79%	**+37%**
54%	80%	**+26%**
55%	82%	**+27%**
68%	79%	**+11%**
67%	82%	**+15%**
67%	79%	**+12%**

Table 1 – Comparison of AIRS and AIRS-X

Further tweaking of specific parameters results in interesting observations. We compare sessions with variable parameters (in bold) below:

Sessions 1 and 2 differ by a small variation in the stimulation threshold (+0.7 in Session 2), yielding a slight accuracy increase

	Session 1	**Session 2**
Affinity threshold scalar	0.4	0.4
Clonal Rate	10	10
Hyperclonal rate	10	10
Stimulation Threshold	**0.92**	**0.99**
Distance Threshold	0.4	0.4

Table 2 – Parameters: Sessions 1 & 2

Session 1 (ST = 0.92)	**Session 2 (ST = 0.99)**	+/-
83%	85%	+2%
83%	84%	+1%
84%	85%	+1%
83%	83%	
83%	83%	
83%	84%	+1%
84%	83%	-1%
84%	82%	-2%
88%	89%	+1%
84%	86%	+2%

Table 3 – Higher ST results in slight accuracy increase

Sessions 2 and 3, meanwhile, differ in their affinity threshold scalars (+0.3):

	Session 2	**Session 3**
Affinity threshold scalar	**0.4**	**0.7**
Clonal Rate	10	10
Hyper clonal rate	10	10
Stimulation Threshold	0.99	0.99
Distance Threshold	0.4	0.4

Table 4 – Parameters: Sessions 2 & 3

The higher affinity threshold scalar is accompanied by a barely noticeable accuracy decrease:

Session 2 (ATS = 0.4)	**Session 3 (ATS = 0.7)**	+/-
85%	85%	
84%	81%	-3%
85%	84%	-1%
83%	84%	+1%
83%	83%	
84%	83%	-1%
83%	84%	+1%
82%	82%	
89%	88%	-1%
86%	85%	-1%

Table 5 - Higher ATS results in slight accuracy decrease.

Next, comparing sessions 3 and 4, we notice that an increased distance threshold (+0.3) decreases the accuracy by an average of -5%.

	Session 3	**Session 4**
Affinity threshold scalar (ATS)	0.7	0.7
Clonal Rate	10	10
Hyperclonal rate	10	10
Stimulation Threshold	0.99	0.99
Distance Threshold	**0.4**	**0.7**

Table 6 – Parameters: Sessions 3 & 4

Session 3 (DT = 0.4)	**Session 4 (DT = 0.7)**	+/-
85%	78%	-7%
81%	79%	-2%
84%	78%	-6%
84%	79%	-5%
83%	79%	-4%
83%	80%	-3%
84%	78%	-6%

82%	80%	-2%
88%	79%	-9%
85%	79%	-6%

Table 7 - Higher DT results in considerable accuracy decrease.

Finally, we double the clonal rate in session 7. In comparison to session 2, the algorithm's accuracy displays negligible fluctuations:

	Session 2	**Session 7**
Affinity threshold scalar (ATS)	0.4	0.4
Clonal Rate	**10**	**20**
Hyperclonal rate	**10**	**20**
Stimulation Threshold	0.99	0.99
Distance Threshold	0.4	0.4

Table 8 – Parameters: Sessions 2 and 7

Session 2 (CR = 10)	**Session 7 (CR = 20)**	**+/-**
85%	84%	-1%
84%	84%	
85%	83%	-2%
83%	82%	-1%
83%	84%	+1%
84%	83%	-1%
83%	84%	+1%
82%	82%	
89%	88%	-1%
86%	84%	-2%

Table 9 – Doubling the clonal rate does not seem to affect the overall accuracy.

8. Conclusion

This thesis aimed to present AIRS-x, a novel formulation to the well-known immune classification algorithm, AIRS. Our work presented an analytical description of the algorithm, explored how different parameter values affect its results and described its position within the universe of machine learning algorithms and artificial immune learning algorithms in particular.

In more detail, we have introduced the theoretical fields of machine learning on the one hand and artificial immune systems on the other, and described how artificial immune learning algorithms, such as AIRS, emerged in the intersection of the two fields.

Moreover, this thesis provided a detailed description of the immune system, where AIS algorithms draw their biological metaphors from. We discussed the broad scope of AIS applications, from computer security to robotics and specifically focused on the sub-field of artificial immune learning, where AIRS and AIRS-x sprang from.

Furthermore, we presented AIRS and AIRS-x in detail, describing their goals, their workflows, their parameters and the immunological concepts they employ, as well as how they translate into their classification framework. Additionally, we highlighted the proposed modifications to AIRS and novel the AIRS-x framework by Giatzitzoglou (2018) and Giatzitzoglou et al. (2019), discussed the experimental results provided by the authors and compared the two algorithms on accuracy, data reduction and efficiency grounds. Finally, this thesis concluded with further experimentation on sentiment analysis accuracy and parameter tweaking.

9. Limitations and Future Work

The goal of this thesis was not to suggest alterations or modifications to the AIRS algorithm, rather merely to describe a proposed formulation. As such, this work is de facto limited to presentation and does not enhance the above frameworks, other than by exploring parameters and thus adding to experimental results.

It is clear that, while AIRS-x performs better than its original counterpart, there is still room for improvement. First of all, even though time complexity decreased for AIRS-x, the algorithm has not been evaluated for overall computational complexity. Further work could mathematically explore the issue, compare the results to AIRS complexity evaluations and, if required, propose modifications. Furthermore, while data reduction was overall greater for AIRS-x, the results are not consistent for all datasets; the subject merits additional exploration.

We may also turn for guidance to previous researchers' suggestions to improve other problematic areas of AIRS. For example, Brownlee (2005) has suggested that normalization is not necessary and may be adding to needless computational complexity. Additionally, he has indicated that the Euclidean distance metric may not be idea for every type of dataset and that the algorithm appears to use superfluous parameters. McEwan and Hart (2009) have also shown that AIRS lacks in compression abilities, perhaps due to the constant affinity threshold. Modifying the algorithm to work with a variable threshold may exhibit interesting results.

Moreover, we have only performed a shallow exploration of how tweaking parameters affects the algorithm's accuracy. Dedicated research could potentially produce interesting observations through rigorous testing of various parameter settings. Performance areas such as data reduction and efficiency need to be explored.

Furthermore, AIRS-X should undergo comparison with algorithms already in use for sentiment analysis; its ability to overcome class imbalance appears promising, but has not been satisfactorily explored in the present thesis.

Finally, since Goodman et al. (2003) have revealed that the sample cell generation process, a great burden on the algorithm's computational complexity – is not the "source of power" or the reason for AIRS's classification accuracy, it would be worthwhile to explore less computationally demanding alternatives to this mechanism.

10. Literature

- Aickelin, U., Greensmith, J. and Twycross, J. (2004) 'Immune system approaches to intrusion detection – a review'. *Proceedings of the 3rd International Conference on Artificial Immune Systems* 3239. Springer, pp. 316-329.
- Aisu, H. & Mizutani, H. (1996) 'Immunity-Based Learning – Integration of Distributed Search and Constraint Relaxation'. *Proceedings of the ICMAS Workshop on Immune-Based Systems*, pp. 124-135.
- Anchor, K., Zydallis, J., Gunsch, G. and Lamont, G. (2002) 'Extending the computer defense immune system: network intrusion detection with multiobjective evolutionary programming approach'. *ICARIS 2002: 1st International Conference on Artificial Immune Systems Conference Proceedings*, pp. 12-21.
- Bersini, H. (1991) 'Immune Network and Adaptive Control'. *Proceedings of the First European Conference on Artificial Life*. MIT Press, pp. 217-226.
- Bersini, H. and Varela, F.J. (1994) 'The Immune Learning Mechanisms: Reinforcement, Recruitment and Their Applications'. Paton, R. (ed.) *Computing with Biological Metaphors*. Chapman & Hall, pp. 166-192.
- Bezerra, G.B., Barra, T.V., de Castro, L.N. and Zuben, F.J.V. (2005) 'Adaptive radius immune algorithm for data clustering'. *ICARIS*, pp. 290-303.
- Bezerra, G.B., Barra, T.V., Ferreira, H.M., Knidel, H., de Castro, L.N. and Zuben, F.J.V. (2006) 'An immunological filter for spam'. *ICARIS*, pp. 14-27.
- Bezerra, G.B. and de Castro, L.N. (2003) 'Bioinformatics data analysis using an artificial immune system network'. *ICARIS*, pp. 22-33.

- Brownlee, K. (2005) *Artificial Immune Recognition System (AIRS): A Review and Analysis*. Technical Report No. 1-02, Centre for Intelligent Systems and Complex Processes (CISCP), Faculty of Information & Communication Technologies (ICT), Swinburne University of Technology (SUT).
- Brownlee, J. (2011) *Clever Algorithms: Nature-Inspired Programming Recipes*. First Edition. Australia: LuLu, pp. 265-299.
- Bull, P., Knowles, A., Tedesco, G. and Hone, A. (2006) 'Diophantine benchmarks for the b-cell algorithm'. *ICARIS*, pp. 267-279.
- Canham, R.O. and Tyrrell, A.M. (2003) 'A hardware artificial immune system and embryonic array for fault tolerant systems'. *Genetic Programming and Evolvable Machines* 4(4), pp. 359-382.
- Cayzer, S. and Aickelin, U. (2008) 'On the effects of idiotypic interactions for recommendation communities in artificial immune systems'. *CoRR*, abs/0801.3539.
- Ceong, H.T., Kim, Y.-I., Lee, D. and Lee, K.H. (2003) 'Complementary dual detectors for effective classification'. *ICARIS*, pp. 242-248.
- Chen, B. and Zang, C. (2009) 'Unsupervised structure damage classification based on the data clustering and artificial immune pattern recognition'. *ICARIS'09: Proceedings of the 8th International Conference on Artificial Immune Systems*. Berlin, Heidelberg: Springer, pp. 206-219.
- Ciesielski, K., Wierzchon, S.T. and Klopotek, M.A. (2006) 'An immune network for contextual text data clustering'. *ICARIS*, pp. 432-445.

- Clark, E., Hone, A. and Timmis, J. (2005) 'A markov chain model of the b-cell algorithm'. *Proceedings of the 4th International Conference on Artificial Immune Systems* LNCS 3627. Springer, pp. 318-330.
- Coello Coello, C.A. and Cortés, N.C. (2002) 'An approach to solve multiobjective optimization problems based on an artificial immune system'. Bentley, P.J. (ed.) *First International Conference on Artificial Immune Systems*, pp. 212-221.
- Cui, X., Li, M. and Fang, T. (2001) 'Study of Population Diversity of Multiobjective Evolutionary Algorithm Based on Immune and Entropy Principles'. *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 1316-1321.
- Cutello, V. and Nicosia, G. (2005) Chapter VI: 'The clonal selection principle for in silico and in vivo computing'. Von Zuben, F. (ed.) *Recent Developments in Biologically Inspired Computing*. Hershey, London, Melbourne, Singapore: Idea Group Publishing, pp. 104-146.
- Dasgupta, D. (1999a) 'Immunity-Based Intrusion Detection System: A General Framework'. *Proceedings of the 22nd NISSC* 1, pp. 147-160.
- Dasgupta, D. (2000) 'An Immune Agent Architecture for Intrusion Detection'. *Proceedings of the Genetic and Evolutionary Computation Conference*, Workshop on Artificial Immune Systems, pp.42-44.
- Dasgupta, D. and Forrest, S. (1996) 'Novelty Detection in Time Series Data Using Ideas from Immunology'. *Proceedings of the 5th International Conference on Intelligent Systems*.
- Dasgupta, D. and Niño, L.F. (2009) *Immunological Computation: Theory and Applications*. USA: Taylor and Francis Group LLC.
- De Castro, L.N. (2003) 'The immune response of an artificial immune network (ainet)'. *IEEE Congress on Evolutionary Computing* 1, pp. 146-153.

- De Castro, L.N. and Timmis, J. (2002) *Artificial Immune Systems: A New Computational Intelligence Approach*. Great Britain: Springer.
- De Castro, L.N. and Von Zuben F.J. (2000a) 'The Clonal Selection Algorithm with Engineering Applications'. *Proceedings of the Genetic and Evolutionary Computation Conference*, Workshop on Artificial Immune Systems and their Applications, pp. 36-37.
- De Castro, L.N. and Von Zuben F.J. (2000b) 'An Evolutionary Immune Network for Data Clustering'. *Proceedings of the IEEE Brazilian Symposium on Artificial Neural Networks*, pp. 84-89.
- De Castro, L.N. and Zuben, F.J.V. (2002) 'Learning and optimization using the clonal selection principle'. *IEEE Transactions on Evolutionary Computation* 6(3), pp. 239-251.
- De Castro, P.A.D., Coelho, G.P., Caetano, M.F. and Zuben, F.J.V. (2005) 'Designing ensembles of fuzzy classification systems: an immune-inspired approach'. *ICARIS*, pp. 469-482.
- De Castro, P.A.D., de França, F.O., Ferreira, H.M. and Von Zuben, F.J. (2007) 'Applying biclustering to text mining: an immune-inspired approach'. *ICARIS'07: Proceedings of the 6th International Conference on Artificial Immune Systems*. Berlin, Heidelberg: Springer, pp.83-94.
- Dilger, W. (1996) 'The Immune System of the Smart Home'. Ishida, Y. (ed.), *Proceedings of the ICMAS International Workshop on Immunity-Based Systems*, pp. 72-81.
- Dilger, W. (1997) 'Decentralized Autonomous Organization of the Intelligent Home According to the Principles of the Immune System'. *Proceedings of the IEEE System, Man, and Cybernetics Conference*, pp. 351-356.

- Farmer, J.D., Kauffman S.A., Packard, N.H. and Perelson, A.S. (1987) 'Adaptive Dynamic Networks as Models for the Immune System and Autocatalytic Sets'. *Annuals of the New York Academy of Science* 504, pp. 118-131.
- Farmer, J.D., Packard, N.H. and Perelson, A.S. (1986) 'The Immune System, Adaptation and Machine Learning'. *Physica* 22D, pp. 187-204.
- Figueredo, G.P., Ebecken, N.F.F., Barbosa, H.J.C. (2007) 'The supraic algorithm: a suppression immune based mechanism to find a representative training set in data classification tasks'. *ICARIS*, pp. 59-70.
- Forrest, S., Hofmeyr, S.A. and Somayaji, A. (1997) 'Computer Immunology'. *Communications of the ACM* 40(10), pp. 88-96.
- Forrest, S., Hofmeyr S.A., Somayaji, A. and Longstaff, A. (1996) 'A Sense of Self for Unix Processes'. *Proceedings of the IEEE Symposium on Computer Security and Privacy*, pp. 120-128.
- Forrest, S., Perelson, A., Allen, L. and Cherukuri, R. (1994) 'Self-Nonself Discrimination in a Computer'. *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pp. 202-212.
- Garain, U., Chakraborty, M.P. and Dasgupta, D. (2006) 'Recognition of handwritten indic script using clonal selection algorithm'. *ICARIS*, pp. 256-266.
- Gaspard, A. and Collard, P. (2000) 'Immune Approaches to Experience Acquisition in Time Dependent Optimization'. *Proceedings of the Genetic and Evolutionary Computation Conference*, Workshop on Artificial Immune Systems, pp. 49-50.
- Giatzitzoglou, D. (2018) *An Alternative Formulation of the Artificial Immune Recognition System Learning Algorithm*, M.S. Dissertation, University of Piraeus, Greece.

- Giatzitzoglou, D., Sotiropoulos, D. and Tsihrintzis, G. (2019) 'AIRS-x: An eXtension to the Original Artificial Immune Recognition Learning Algorithm'. *2019 International Conference on Computer, Information and Telecommunication Systems (CITS).* Beijing, China, pp. 1-5.
- Golzari, S., Doraisamy, S., Sulaiman, M.N. and Udzir, N.I. (2009) 'Improving the accuracy of AIRS by incorporating real world tournament selection in resource competition phase'. *2009 IEEE Congress on Evolutionary Computation*. Trondheim, pp. 3040-3044.
- Golzari, S., Doraisamy, S., Sulaiman M., Udzir, N. and Norowi, N. (2008) 'Artificial Immune Recognition System with Nonlinear Resource Allocation Method and Application to Traditional Malay Music Genre Classification'. Bentley P.J., Lee D. and Jung S. (eds.) *Artificial Immune Systems. ICARIS 2008. Lecture Notes in Computer Science* 5132. Berlin, Heidelberg: Springer, pp. 132-141.
- Goncharova, L.B., Melnikov, Y. and Tarakanov, A.O. (2003) 'Biomolecular immunocomputing'. *ICARIS*, pp. 102-110.
- Goodman, D., Boggess, L. and Watkins, A. (2003) 'An investigation into the source of power for airs, an artificial immune classification system'. *Proceedings of the International Joint Conference on Neural Networks* (IJCNN 03), pp. 1678-1683.
- Gu, J., Lee, D., Park, S. and Sim, K. (2000) 'An Immunity-Based Security Layer Model'. *Proceedings of the Genetic and Evolutionary Computation Conference*, Workshop on Artificial Immune Systems and Their Applications, pp.47-48.
- Haag, C.R., Lamont, G.B., Williams, P.D. and Peterson, G.L. (2007) 'An artificial immune system-inspired multiobjective evolutionary algorithm with application to the detection of distributed computer network intrusions'. *GECCO'07: Proceedings of the 2007 GECCO*

Conference Companion on Genetic and Evolutionary Computation. New York, USA: ACM, pp. 2717-2724.

- Hajela, P. and Yoo, J.S. (1999) 'Immune Network Modeling in Design Optimization'. Corne, D., Dorigo, M. and Glover, F. (eds.) *New Ideas in Optimization*. London: McGraw Hill, pp. 203-215.
- Hart, E. and Ross, P. (2002) 'Exploiting the analogy between immunology and sparse distributed memories: A system for clustering non-stationary data'. *ICARIS'01: Proceedings of the 1st International Conference on Artificial Immune Systems,* pp. 49-58.
- Hofmeyr, S. and Forrest, S. (1999) 'Immunity by Design: An artificial immune system'. Daida, J. et al. (eds.) *Proceedings of the Genetic and Evolutionary Computation Conference* 2, pp. 1289-1296.
- Hofmeyr, S.A. and Forrest, S. (2000) 'Architecture for an artificial immune system'. *Evolutionary Computation* 8(4), pp.443-473.
- Hunt, J.E. and Cooke, D.E. (1996) 'Learning Using an Artificial Immune System'. *Journal of Network and Computer Applications* 19, pp. 189-212.
- Ishiguro, A., Ichikawa, S., Shibata, T. and Uchikawa, Y. (1998) 'Moderationism in the Immune System: Gait Acquisition of a Legged Robot Using the Metadynamics Function'. *Proceedings of the IEEE System, Man, and Cybernetics Conference*, pp. 3827-3832.
- Joshi, R.R. (1995) 'Immune Network Memory: An Inventory Approach'. *Computer Operations Research* 22(6), pp. 575-591.
- Jun, J.-H. Lee, D.-W. and Sim, K.-B. (1991) 'Realization of Cooperative and Swarm Behavior in Distributed Autonomous Robotic Systems Using Artificial Immune System'. *Proceedings of the IEEE System, Man and Cybernetics Conference* 4, pp. 614-619.

- Kelsey, J. and Timmis, J. (2003) 'Immune inspired somatic contiguous hypermutation for function optimization'. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, pp. 207-218.
- Kephart, J.O. (1994) 'A Biologically Inspired Immune System for Computers'. Brooks, R.A. and Maes, P. (eds.) *Artificial Life IV Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. MIT Press, pp. 130-139.
- Kephart, J.O., Sorkin, G.B. and Swimmer, M. (1997) 'An Immune System for Cyberspace'. *Proceedings of the IEEE System, Man and Cybernetics Conference*, pp. 879-884.
- Kephart, J.O., Sorkin, G.B., Swimmer, M. and White, S.R. (1999) 'Blueprint for a Computer Immune System'. Dasgupta, D. (ed.) *Artificial Immune Systems and Their Applications*. Springer-Verlag, pp. 242-259.
- Kim, J. and Bentley, P. (1999a) 'The Human Immune System and Network Intrusion Detection'. *Proceedings of the EUFIT'99*, CD ROM.
- Kim, J. and Bentley, P. (1999b) 'Negative Selection and Niching by an Artificial Immune System for Network Intrusion Detection'. *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 149-158.
- Kim, J. and Bentley, P.J. (2001) 'Evaluating Negative Selection in an Artificial Immune System for Network Intrusion Detection'. *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1330-1337.
- Ko, A., Lau, H.Y.K. and Lau, T.L. (2004) 'An immuno control framework for decentralized mechatronic control'. *ICARIS*, pp. 91-105.

- Ko, A., Lau, H.Y.K. and Lau, T.L. (2005) 'General suppression control framework: application in self-balancing robots'. *ICARIS*, pp. 375-388.
- Lambert, A., King, R.L., Russ, S.H., Rajan, R. and Reese, D.S. (1999) 'Adaptive Analysis for the Design of Hardware Agents Using the Artificial Immune System Model for Resource Management of Heterogeneous Systems'. *Miss. State Technical Report n. MSSU-COE-ERC-98-10*, August.
- Lau, H., Bate, I. and Timmis, J. (2009) 'An immune-engineering approach for anomaly detection in swarm robotics'. *ICARIS*, pp. 136-150.
- Lau, H.Y.K. and Wong, V.W.K. (2004) 'Immunologic responses manipulation of ais agents'. *ICARIS*, pp. 65-79.
- Lee, D.-W., Jun, H.-B. and Sim, K.-B. (1999) 'Artificial Immune System for Realization of Cooperative Strategies and Group Behavior in Collective Autonomous Mobile Robots'. *Proceedings of the AROB'99*, pp. 232-235.
- Lee, D.-W. and Sim, K.-B. (1997) 'Artificial Immune Network-Based Cooperative Control in Collective Autonomous Mobile Robots'. *Proceedings of the IEEE International Workshop on Robotics and Human Communication*, pp. 58-63.
- Lee, J., Roh, M., Lee, J. and Lee, D. (2007) 'Clonal selection algorithms for 6-dof pid control of autonomous underwater vehicles'. *ICARIS*, pp. 182-190.
- Lehman, M. and Dilger, W. (2006) 'Controlling the heating system of an intelligent home with an artificial immune system'. *ICARIS*, pp. 182-190.
- Luh, G.-C. and Liu, W.-W. (2004) 'Reactive immune network based mobile robot navigation'. *ICARIS*, pp. 28-41.

- Lu, S.Y.P. and Lau, H.Y.K. (2009) 'An immunity inspired real-time cooperative control framework for network multi-agent systems'. *ICARIS*, pp. 234-247.
- Marwah, G. and Boggess, L. (2002) 'Artificial immune systems for classification: Some issues'. *1st International Conference in Artificial Immune Systems 1*. Springer, pp. 149-153.
- McCoy, D.F. and Devarajan, V. (1998) 'Artificial Immune Systems for Multispectral feature Extraction'. *Proceedings of SPIE Conference on Algorithms for Multispectral and Hyperspectral Imagery* IV 3372, pp. 241-249.
- McEwan, C. and Hart, E. (2009) 'On airs and clonal selection for machine learning'. *ICARIS*, pp. 67-79.
- Michaud, S.R., Zydallis, J.B., Lamont, G.B., Harmer, P.K. and Patcher, R. (2001) 'Protein Structure Prediction with EA Immunological Computation'. *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1367-1374.
- Mitchell, T.M. (1997) *Machine Learning*. New York: McGraw Hill.
- Mo, H. (2009) *Handbook of Research on Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies*. New York: Hershey.
- Mori, M., Tsukiyama, M. and Fukuda, T. (1998) 'Adaptive Scheduling System Inspired by Immune System'. *Proceedings of the IEEE Systems, Man and Cybernetics Conference*, pp. 3833-3837.
- Morisson, T. and Aickelin, U. (2008) 'An artificial immune system as a recommender system for web sites'. *CoRR*, abs/0804.0573.
- Oates, R., Greensmith, J., Aickelin, U., Garibaldi, J. and Kendall, G. (2007) 'The application of a dendritic cell algorithm to a robotic classifier'. *ICARIS'07: Proceedings of*

the 6th International Conference on Artificial Immune Systems. Berlin, Heidelberg: Springer, pp. 204-215.

- Oda, T. and White, T. (2005) 'Immunity from spam: An analysis of an artificial immune system for junk email detection'. *Artificial Immune Systems, Lecture Notes in Computer Science*. Springer, pp. 276-289.
- Ootsuki, J.S., Fujii, Y., Mizutani, H. and Sekiguchi, T. (1998) 'Immune System Derived Approach for Finding Firing Sequences of a Sub-Class of Petri Nets'. *Transactions of IEE-J* 188-C, pp. 419-427.
- Polat, K., Kara, S., Latifoglu, F. and Günes, S. (2006) 'A novel approach to resource allocation mechanism in artificial immune recognition system: Fuzzy resource allocation mechanism and application to diagnosis of atherosclerosis disease'. *ICARIS*, pp. 244-255.
- 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, pp. 645-670.
- Reche, P.A. and Reinherz, E.L. (2004) 'Definition of mhc supertypes through clustering of mhc peptide binding repertoires'. *ICARIS*, pp. 189-196.
- Salton, G. et al. (1975) 'A Vector space model for automatic indexing. *Communications Journal* 18(11), pp. 613-620.
- Seiden, P.E. and Celada, F. (1992) 'A Model for Simulating Cognate Recognition and Response in the Immune System'. *Journal of Theoretical Biology* 158, pp. 329-357.
- Shafiq, M.Z. and Farooq, M. (2007) 'Defense against 802.11 dos attacks using artificial immune system'. *ICARIS*, pp.95-106.

- Skormin, V.A., Delgado-Frias, J.G., McGee, D.L., Giordano, J.V., Popyack, L.J., Gorodetski, V.I. and Tarakanov, A.O. (2001) 'BASIS: A Biological Approach to System Information Security'. *Proceedings of the International Workshop MMM-ACNS 2001*, pp. 127-142.
- Soleymani, M. et al. (2017) 'A survey of multimodal sentiment analysis'. *Image and Vision Computing* 65, pp. 3-14.
- Sotiropoulos, D.N. and Tsihrintzis, G.A. (2017) *Machine Learning Paradigms: Artificial Immune Systems and their Applications in Software Personalization*. Springer International Publishing.
- Sotiropoulos et al. (under publication) 'SVM-Based Sentiment Classification: A Comparative Study against State-of-the-Art Classifiers'. *International Journal A Hoc and Ubiquitous Computing*.
- Takahashi, K. and Yamada, T. (1997) 'A Self-Tuning Immune Feedback Controller for Controlling Mechanical Systems'. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, CD-ROM#101.
- Tarakanov, A., Goncharova, L., Gupalova, T., Kvachev, S. and Sukhorukov, E. (2002) 'Immunocomputing for Bioarrays'. *ICARIS*.
- Tedesco, G. Twycross, J. and Aickelin, U. (2010), 'Integrating innate and adaptive immunity for intrusion detection'. *CoRR*. abs/1003.1256.
- Timmis, J. (2004) 'Assessing the performance of two immune inspired algorithms and a hybrid genetic algorithm for optimization'. *Proceedings of Genetic and Evolutionary Computation Conference, GECCO 2004*. Springer: pp. 308-317.

- Timmis, J. and Andrews, P. (2007) 'A Beginners Guide to Artificial Immune Systems'. Flower, D. and Timmis, J. (eds.) *In Silico Immunology*. Boston: Springer, pp. 47-62.
- Timmis, J. and Neal, M. (2000) 'Investigating the Evolution and Stability of a Resource Limited Artificial Immune System'. *Proceedings of the Genetic and Evolutionary Computation Conference*, Workshop on Artificial Immune Systems and Their Applications, pp. 40-41.
- Timmis, J. and Neal, M. (2001) 'A resource limited artificial immune system for data analysis'. *Knowledge Based Systems* 14 (1), pp. 121-130.
- Toma, N., Endo, S. and Yamada, K. (1999) 'Immune Algorithm with Immune Network and MHC for Adaptive Problem Solving'. *Proceedings of IEEE System, Man and Cybernetics* 4, pp. 271-276.
- Trapnell, B.C. Jr. (2005) 'A peer-to-peer blacklisting strategy inspired by leukocyte-endothelium interaction'. *ICARIS*, pp. 339-352.
- Varela, F.J. and Coutinho, A. (1991) 'Second Generation Immune Networks'. *Immunology Today* 12(5), pp. 159-166.
- Vargas, P.A., de Castro, L.N., Michelan, R. and Zuben, F.J.V. (2003) 'An immune learning classifier network for autonomous navigation'. *ICARIS*, pp. 69-80.
- Watkins, A. (2001) *AIRS: A Resource Limited Artificial Immune Classifier*. M.S. Dissertation, Mississippi State University, USA.
- Watkins, A. (2002) 'Artificial immune recognition system (airs): Revisions and refinements'. *Genetic Programming and Evolvable Machines*, pp. 173-181.

- Watkins, A. and Boggess, L. (2002) ‘A resource limited artificial immune classifier’. Ebherhart, R. (ed.), *Congress on Evolutionary Computation*. Part of the World Congress on Computational Intelligence, pp. 926-931.
- 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), pp. 291-317.