O-AIRS: Optimized Artificial Immune Recognition System

Badr Merad[1,2*], Khaled Belkadi[3]
[1] Laboratory of Electrical Engineering and Materials (LEEM), Oran, Algeria
[2] Higher School in Electrical and Energetic Engineering of Oran (ESGEEO), Oran, Algeria
[3] Faculty of mathematics and computer sciences, University of Sciences and Technology of Oran Mohamed Boudiaf, Oran, Algeria
Email: merad_badr@esgee-oran.dz, belkadi1999@yahoo.com
[*]Corresponding author

Keywords: AIRS, Artificial Immune Systems, Artificial Intelligence, Classification, Clonal Selection, Supervised Learning

Received: February 15, 2023

Artificial Immune Recognition System (AIRS) offers a promising meta-heuristic approach inspired by the human immune system for classification tasks. However, limitations such as reliance on single-antigen activation and retention of untested memory cells can lead to inaccuracies. This paper proposes the Optimized Artificial Immune Recognition System (O-AIRS) to mitigate these issues. O-AIRS leverages Homogeneous Antigen Groups (HAGs) for refined memory cell activation, ensuring a precise threat response. Furthermore, O-AIRS incorporates a robust maturity mechanism to retain only validated memory cells, enhancing classification accuracy. The effectiveness of O-AIRS was assessed using established medical datasets: Liver Disorders (LD) and Haberman Surgery Survival (HSS). Experimental evaluation on both LD and HSS datasets establishes O-AIRS's superiority over AIRS and AIRS2 across various performance metrics. Notably, O-AIRS achieves this enhanced performance while utilizing approximately 50% fewer memory cells during classification due to its optimized activation mechanism. Importantly, O-AIRS guarantees the maturity of all memory cells, ensuring effective threat recognition.

Povzetek: Predstavljen je sistem za umetno prepoznavanje imunosti (O-AIRS) izboljšuje natančnost klasifikacije z uporabo homogenih antigenskih skupin (HAG) in mehanizma zrelosti za shranjevanje preverjenih spominskih celic, kar je dokazano učinkovitejše od AIRS in AIRS2 na medicinskih podatkovnih nizih, hkrati pa uporablja približno 50 % manj spominskih celic.

Povzetek: Predstavljen je sistem za umetno prepoznavanje imunosti (O-AIRS), ki izboljšuje klasifikacije z uporabo homogenih antigenskih skupin (HAG) in mehanizma zrelosti za shranjevanje preverjenih spominskih celic, kar je dokazano učinkovitejše od AIRS in AIRS2 na medicinskih podatkovnih nizih.

1 Introduction

The ability to learn and adapt is fundamental to human intelligence. It allows individuals to improve through experience, remember past decisions, and make better choices in similar situations. This concept underpins much of modern Artificial Intelligence (AI) research, which focuses on developing systems capable of learning and decision-making akin to human cognition.

Inspired by biological systems such as the human brain, neurons, and genetic processes, researchers develop bio-inspired approaches to AI. These methods seek to replicate the efficiency and adaptability observed in natural systems, paving the way for advancements in machine learning algorithms.

Among these bio-inspired approaches, Artificial Immune Systems (AIS) stand out for their emulation of the human immune system's ability to recognize and respond to threats. Originally conceived in the 1950s, AIS models have evolved to tackle complex computational challenges, including pattern recognition, anomaly detection, and optimization tasks.

Within the realm of AIS, the Artificial Immune Recognition System (AIRS) [1] and its successor, AIRS2 [2], have garnered significant attention for their effectiveness in supervised learning tasks. These algorithms are known for their good classification capabilities and strong capacities to support decision makers and resolve real-world issues [3].

While AIRS and AIRS2 have demonstrated success, their effectiveness can be hampered by inherent limitations. These limitations include dependence on single antigens for memory cell activation and the potential use of untested memory cells during classification.

This paper addresses these limitations by introducing a novel AIS algorithm: the Optimized Artificial Immune Recognition System (O-AIRS). O-AIRS leverages innovative methodologies like Homogeneous Antigen Groups (HAGs) and a refined memory cell activation mechanism to overcome these shortcomings. HAGs

enhance pattern recognition, and the refined activation mechanism ensures robust and reliable classifications. These advancements aim to improve classification accuracy and computational efficiency for diverse datasets.

To comprehensively evaluate O-AIRS's effectiveness, the authors conducted a series of rigorous experiments utilizing established benchmark datasets. These datasets, Liver Disorders (LD) and Haberman Surgery Survival (HSS), represent real-world scenarios commonly encountered in the medical domain. By employing a diverse range of metrics to assess O-AIRS's performance, the experiments provide a robust and generalizable understanding of its capabilities.

This paper is structured as follows: Section 2 discusses existing research relevant to Artificial Immune Systems. Section 3 provides a detailed explanation of the Artificial Immune Recognition System (AIRS). Section 4 explores the advancements made in AIRS2. Section 5 highlights the key innovations introduced in O-AIRS compared to AIRS/AIRS2. Section 6 presents the experimental study, including the use of the Haberman Surgery Survival (HSS) and Liver Disorder (LD) datasets, and summarizes the overall findings. Finally, Section 7 concludes the paper by summarizing the research findings and outlining future research directions.

2 Related works

Artificial Immune Systems (AIS) have been extensively researched, resulting in a variety of algorithms. These algorithms mainly fall into four categories: negative selection, immune network, danger theory, and clonal selection [4].

Negative selection algorithms, originating from the seminal work of Forrest et al. [5], emulate how the immune system identifies and eliminates antibodies that mistakenly recognize 'self' components as antigens. This process is crucial in data security contexts, where 'self' corresponds to the data to be protected and 'non-self' denotes potentially harmful data. Subsequent studies, such as [6], [7], and [8], have further refined and extended these mechanisms, leading to enhanced efficacy of self-nonself discrimination.

The concept of an immune network, a self-regulating system that distinguishes "self" from "non-self" without direct contact with antigens, was introduced by Jerne in 1974 [9]. This foundational idea laid the groundwork for Artificial Immune Networks (AINs). Building on this work, Varela and Coutinho [10] developed a second-generation model focused on enhancing adaptability and response dynamics within the network and De Castro and Von Zuben [11] proposed aiNet. These networks offer advantages similar to the human immune system, such as the ability to adapt to new threats. Recent research like [12], [13] and [14] continues to explore and optimize AIN frameworks for diverse applications, including pattern recognition and optimization problems.

The danger theory, proposed by Matzinger [15], suggests that immune responses are initiated not only through direct interactions between antibodies and antigens, but also by signals released during cellular damage or stress. This concept has played a crucial role in the evolution of Intrusion Detection Systems (IDSs), enabling the detection of potential threats while allowing harmless antibodies and antigens to coexist in the absence of danger signals. This approach has been successfully implemented and further developed in practical Intrusion Detection systems, exemplified by notable works such as those referenced in [16], [17] and [18].

The clonal selection principle, initially formulated by Burnet and refined through computational models, describes how antibodies proliferate upon encountering antigens, followed by mutations to enhance affinity [19]. Seminal contributions in this area include CLONALG (CLONal selection ALGorithm) [20] and AIRS (Artificial Immune Recognition System) [1]. CLONALG introduced fundamental concepts of clonal selection in computational models [21], laying the groundwork for subsequent advancements. Ongoing research [22], [2], [23] and [24] continues to explore and refine clonal selection mechanisms, with the aim of enhancing adaptability and efficiency across various problem domains. AIRS [1] and its successor AIRS2 [2] have gained recognition for their robust classification capabilities and practical application in decision support systems [3]. Table 1 provides a comprehensive overview of the key references discussed in this section.

Table 1: Summary of relevant literature.

	Algorithm	Dataset	Application area	Performance metric	Strengths	Limitations	Remarks
Negative Selection	Self-Nonself Discrimination [5]	Experimental data	Virus detection	· Probability of detection · Computational cost	Protects anti bodies from modification to conform to altered self	High computational cost for generating the initial "self" repertoire	First negative selection algorithm

	Improved Negative Selection Algorithm (Improved-NSA) [6]	Sensor data: three-tank system	Fault detection	· Coverage Rate · Overlap Rate · Detection Rate · False Alarm Rate	· Automatic adjustment of detector radius for optimal coverage · Reduces overlap between detectors	· Requires predefined "self" data for training · May not be suitable for complex anomaly patterns	Custom datasets for three-tank system (not public)
	Antigen Density Clustering - Negative Selection Algorithm (ADC-NSA) [7]	· Breast Cancer Wisconsin (BCW) · Knowledge Discovery and Data Mining Cup 99 (KDD-Cup99)	Anomaly detection: · Medical diagnosis · Network intrusion detection · Spam detection	· Detection Rate · False Positive Rate	· Addresses uneven antigen distribution · Improves detection efficiency · Reduces randomness in detector generation	· Requires tuning cutoff distance for clustering · Needs further research on identifying "loopholes" (undetected data)	-
	Improved Negative Selection Algorithm (INSA) [8]	Normal state sample library of distribution network dataset	High resistance fault identification in distribution network	· Detection Rate · Classification · Accuracy	Requires small number of samples for training	Vulnerable to black holes (cannot classify when there are many types of abnormal states)	Custom dataset (not public)
Artificial Immune Networks	Artificial Immune Networks (AINet) [11]	Unlabeled numerical datasets	Data clustering and filtering	· Compression Rate · Classification Accuracy · Reduction of Redundancy	· Reduces data redundancy · Identifies groups and subgroups in the data · Determines the number of clusters and their structure · Offers good compression rates	· High number of user-defined parameters · Computationally expensive ($O(p^3)$) · Sensitive to suppression threshold	50-sample classification (5 classes) and the two-donut problem
	Adaptive Artificial Immune Networks [12]	· KDD'99 · CAIDA'07 · CAIDA'08	Network security, specifically Denial-of-Service (DoS) flooding attack detection and mitigation	· True Positive Rate · False Positive Rate · Hit Rate · Entropy	· Adapts thresholds for anomaly detection · Implements quarantine zones to isolate threats	· Requires careful tuning of parameters for optimal performance · May have computational overhead	Other data generated for this study by the DDoSIM tool were used in simulations
	COVID-opt-aiNet [14]	· COVID-19 CT dataset · COVID-19 Radiography dataset · Chest X-ray dataset	COVID-19 detection	· Precision · Recall · F1-Score · Accuracy	· Improved accuracy compared to standalone DL/ML methods · Reduced training time	· Requires large datasets for training · Performance might be dependent on specific datasets	Hybrid approach using AINet with DL/ML techniques

Danger Theory	Deep Dendritic Cell Algorithm (DeepDCA) [16]	IoT-Bot dataset	IoT intrusion detection	· Accuracy · Precision · Recall · F1-Score · False Alarm Rate	· Detects various IoT attacks (DoS, DDoS, information gathering, theft) · High accuracy (over 98.73%) · Low False Positive Rate	Requires large datasets for training	-
	Multi-Level Intrusion Detection System Based on Immune Theory [17]	Custom data generated by cooja simulator	Intrusion detection in Wireless Sensor Networks (WSNs)	· Detection Rate · Packet Overhead · Energy Overhead	Distributed and lightweight approach	Detection probability varies for different attacks.	-
	Danger Theory - Dendritic Cell Algorithm (DT-DCA) [18]	Custom data generated by simulation and from real Wireless Sensor Network (WSN)	Wireless Sensor Networks (WSNs)	· True Positive (TP) · False Negative (FN) · True Negative (TN) · False Positive (FP)	· Low FP Rate · Low Energy Consumption	Lower TP rates compared to some other methods	Simulated and real-world WSN platform implementation
Clonal Selection	CLONALG [20]	Custom character set 30-city instance of the travelling salesman problem	· Pattern recognition · Optimization	· Accuracy (Pattern Recognition) · Fitness Function Value (Optimization)	· Efficient for multimodal problems · Good convergence speed	· Sensitive to parameters · May require significant memory for large datasets	-
	Improved Clonal Selection Algorithm with K-Nearest Neighbor (ICSAT-KNN) [22]	· Brain Tumor dataset · Leukemia dataset · Prostate Tumor dataset	Cancer classification	Classification Accuracy	Competitive accuracy (96.36% average)	Relies on a small subset of genes (16% average) Requires parameter tuning (Gsize, K) for optimal performance	-
	CLONALG-M [23]	Custom generated data	Wireless Sensor Networks (WSNs)	· First Node Dies (FNA) · Half of the Nodes Alive (HNA) · Total Remaining Energy (TRE)	Improved performance of fuzzy clustering algorithms	· Computationally expensive · Only approximates optimal solution	-

	Artificial Immune Recognition System (AIRS) [1]	· Two datasets of points in 10x 10 space · The Fisher Iris dataset · Pima diabetes dataset · The Sonar dataset · Ionosphere dataset	Classification	· Accuracy · Final memory cells number	· Good classification capabilities · Strong capacities to support decision makers and resolve real-world issues	· Memory cells activation mechanism solicited by a single antigen · Retention of memory cells even if they have never been into contact with antigens	· The first two datasets are randomly generated · Data space in the first dataset is linearly separable · In the second dataset data space is not linearly separable
	Artificial Immune Recognition System 2 (AIRS2) [2]	· The Fisher Iris dataset · Pima diabetes dataset · Ionosphere dataset · The Sonar dataset	· Signal processing · Medical diagnosis · Biology	· Accuracy · Final memory cells number	· Good classification capabilities · Refinement of AIRS	· Memory cells activation mechanism solicited by a single antigen · Retention of memory cells even if they have never been into contact with antigens	-
	Constant Length Multi-objective Clonal Selection Optimization Algorithm (CL-MCSOA) [24]	· Yeast Sporulation · Yeast Cell Cycle · Arabidopsis Thaliana · Human Fibroblasts Serum · Rat CNS · Colon Tumor	Gene expression clustering	· Silhouette width index · Deviation · Connectivity · Dunn-index · Execution time	· Multi-objective · Robust · Faster convergence	· Complex design · User needs to define final clustering	All 6 datasets used to evaluate the algorithm are publicly available

In summary, previous studies, such as those on AIRS [1] and AIRS2 [2], have extensively explored immune-inspired algorithms. However, optimizing their classification capabilities and practical applications remains challenging. This study introduces O-AIRS, an Optimized AIRS algorithm, which refines the clonal selection metaphor to enhance classification performance.

3 Artificial immune recognition system (AIRS)

"The artificial immune systems are computational models inspired by the biological immune system" [25]. This section presents AIRS, a widely used AIS algorithm that describes how memory cells recognize antigens [26].

AIRS is divided into two phases: a learning phase for generating memory cells and a classification phase that utilizes these cells.

3.1 The learning phase

3.1.1 Initialization step

The first step of initialization involves normalizing the data such that the Euclidean distance between any two vectors (antigens) falls within the interval [0, 1]. Following normalization, the initial set of memory cells

(MC) and the set (P) of Artificial Recognition Balls (ARBs), which represent lymphocytes with attributes of vector, resource, and class, are randomly created from the training dataset.

Subsequently, the affinity_threshold parameter is computed to represent the average affinity between all pairs of antigens using formula 1.

$$Affinity\_Threshold = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n} affinity(ag_i, ag_j)}{\frac{n(n-1)}{2}} \quad (1)$$

Where:

- n is the number of antigens.
- ag_i, ag_j are the i^{th} and j^{th} antigens.
- affinity(X,Y) represents the Euclidean distance between two vectors X and Y.

3.1.2 Learning by antigens

After initialization is complete, each element of the training set (each antigen) is presented to the AIRS algorithm to learn from its characteristics. For each antigen ag, a sequence of steps is repeated:

1. Selection of the mc_{match} cell: this is the closest cell to the antigen ag in the shape space. mc_{match} and ag belong to the same class (see formula 2).

$$mc_{match} = argmax_{mc \in MC_{ag.c}} stimulation\ (ag, mc) \quad (2)$$

Where:

- $stimulation\ (x, y) = \begin{cases} 1 - affinity(x, y)\ if\ x.c \equiv y.c \\ Affinity\ (x, y)\ else \end{cases}$
- MC_c represent MC set cells of the class c.
- ag.c is the class of ag.

The number of clones generated from mc_{match} depends on its affinity with the antigen (ag) (see formula 3)

$$nb\_clones = hyper\_clonal\_rate * clonal\_rate * stim \quad (3)$$

Where:

- nb_clones is the number of clones produced.
- hyper_clonal_rate is the maximum cloning rate.
- clonal_rate is the average cloning rate.
- stim = stimulation(mc_{match},ag).

2. Mutation of ARBs: each ARB generated by mc_{match} goes through a mutation function detailed in [27]. If the result is positive (mutation performed) the ARB is added to the set (P).
3. Calculation of resources for ARBs: resources are calculated for each element of (*P*), the cells closest to ag will experience higher stress and consequently receive more resources.
4. Clonage and mutation of ARBs: ARBs in this step are selected based on their affinity with the antigen ag.
5. Stimulation_threshold verification: the algorithm verifies that while the average stimulation of each group of ARBs of the same class is lower than a given value of "stimulation_threshold," it resumes from step (3).
6. Choice of the candidate cell: the candidate cell is selected from the set (P) based on its closest similarity to the antigen ag. This candidate cell is added to the set of memory cells (MC) only if its stimulation value with ag is higher than the stimulation value of mc_{match}.

The above steps are repeated for each antigen until the entire training set has been exhausted.

3.2 The classification phase

After the learning phase, the memory cells are ready to be used in the classification phase, which employs the K-Nearest Neighbors (KNN) algorithm. Each memory cell is presented to the data vector for stimulation. The classification system determines the vector's class based on the classes of its k-nearest memory cells.

4 AIRS2

AIRS2 is a refinements of AIRS developed in [2]. The changes made are:

- The initialization of the set of ARBs (P) is no longer necessary.
- Mutations now only concern data vectors and not classes.
- For the clonal selection and the criterion for stopping learning, only ARBs of the same class as ag are considered.

5 O-AIRS: an improvement of AIRS/AIRS2

In this section, we present O-AIRS, an innovative approach designed to refine Artificial Immune Systems (AIS) by addressing inherent limitations observed in traditional algorithms such as AIRS and AIRS2. Central to its improvements are two key concepts: Homogeneous Antigen Groups (HAGs) and a refined memory cell maturation mechanism.

AIRS algorithms faced two major hurdles: firstly, learning from individual antigens could lead to overfitting, where the system becomes overly influenced by random fluctuations in single data points. This skewed the development of memory cells, ultimately impacting classification accuracy. Secondly, not all memory cells were thoroughly evaluated during iterations. This results in a population of immature cells that couldn't handle diverse antigens, hindering classification performance.

5.1 Homogeneous antigen group

O-AIRS tackles the first issue by introducing HAGs. These represent groups of similar antigens belonging to the same class and residing close together within the data feature space. Unlike AIRS that dealt with single antigen one-at-a-time which may present noisy data [28], O-AIRS utilizes the collective information within HAGs during learning cycles.

This strategic shift reduces the impact of noise from isolated data points. By focusing on the average affinity between an antibody and an entire HAG, O-AIRS aims to make more robust and reliable classification decisions. Formula 4 presents the average affinity between an antibody and a HAG.

$$affinity(ac, GAH) = \frac{\sum_{i=1}^{n} affinity(ac, ag_i)}{n} \quad (4)$$

Where:

- ag_i is the i^{th} antigen in the group.
- n is the number of antigens in the group.
- ac is the antibody.

A parameter called GS (Group Scalar) controls the size and number of HAGs formed in O-AIRS. By adjusting GS, the system optimizes space allocation within the data space, ensuring cohesive HAGs with high internal similarity. This optimization aims to enhance classification accuracy by guaranteeing each HAG effectively represents a distinct antigenic subset.

Below, we propose a pseudo-code implementation (Algorithm1) for handling HAGs in O-AIRS.

```
1:  initialize_param (affinity_threshold, GS);
2:  all_antigen_group (0);
3:  AGSet ← all_antigen ();
4:  group ← 1;
5:  foreach (antigen_ag ∈ AGSet) do
6:      if (antigen_ag.group = 0) then
7:          nearest_ag ← best_affinity (ag, AGSet);
8:          if (affinity (ag, nearest_ag) < affinity_threshold * GS) then
9:              ag.group ← group;
10:             group ← group + 1;
11:             go_to_next_ag ();
12:         end if
13:         if (nearest_ag.group=0 ) then
14:             ag.group ← group;
15:             nearest_ag.group ← group;
16:             group ← group + 1;
17:         else
18:             ag.group ← nearest_ag.group;
19:         endif
20:     endif
21: done
```

Algorithm 1: Homogeneous Antigen Group Handling.

5.1.1 Initialization

Before constructing HAGs, O-AIRS lays the foundation by initializing two key parameters:

- Affinity Threshold: this parameter defines the minimum affinity required between antigens for them to be grouped together into a HAG.
- Group Scalar (GS): GS acts as a pivotal factor influencing the formation of HAGs. A higher GS value typically results in fewer, larger HAGs, whereas a lower GS value tends to produce more, smaller HAGs.

Additionally, the function all_antigen_group (0) is employed to initially assign all antigens in the dataset to an ungrouped state, represented by the value 0.

5.1.2 Identifying neighbors

Next, O-AIRS examines each antigen and identifies its closest "neighbor" within the entire dataset using a function called best_affinity. This neighbor is the antigen with the highest affinity, essentially the most similar one.

5.1.3 Building communities

The algorithm then evaluates the affinity between the antigen and its nearest neighbor using Euclidean distance. This measure constitutes a critical parameter for artificial immune system algorithms [29].

Here is where the affinity_threshold and GS come into play:

High affinity (lower distance): if the affinity (Euclidean Distance) between the antigen and its neighbor is lower than the threshold defined by affinity_threshold * GS, they are considered "close enough" and grouped together into the same HAG. This indicates a high degree of similarity.

Low affinity (higher distance): if the affinity value exceeds the threshold, it suggests a lack of sufficient similarity. In this case, the current antigen becomes the founding member of a new HAG. This indicates a lower degree of similarity.

This process continues for all antigens in the dataset.

5.2 Ensuring memory cell maturity

O-AIRS goes beyond HAGs for classification. It implements a memory cell maturity mechanism to address the issue of incomplete testing in AIRS/AIRS2. The process involves two stages.

5.2.1 Initial selection with K-Nearest Neighbors (KNN)

O-AIRS utilizes the KNN algorithm to refine the data into a memory cell set (MC). This identifies memory cells with the closest affinity to incoming antigens, paving the way for accurate classification.

5.2.2 Interaction and maturation testing with HAGs

O-AIRS tracks memory cell interaction with HAGs. After antigen interaction, memory cells that haven't engaged with any HAG are categorized:

- Mature cells: these cells demonstrate sufficient HAG interaction and remain in the MC for future classification tasks.
- Learning set: memory cells lacking adequate HAG interaction are placed in a dedicated set for further training with updated antigenic data. This iterative learning process refines the classification capabilities of these immature cells.

The process of maturity test is presented in the flowchart in Figure 1.

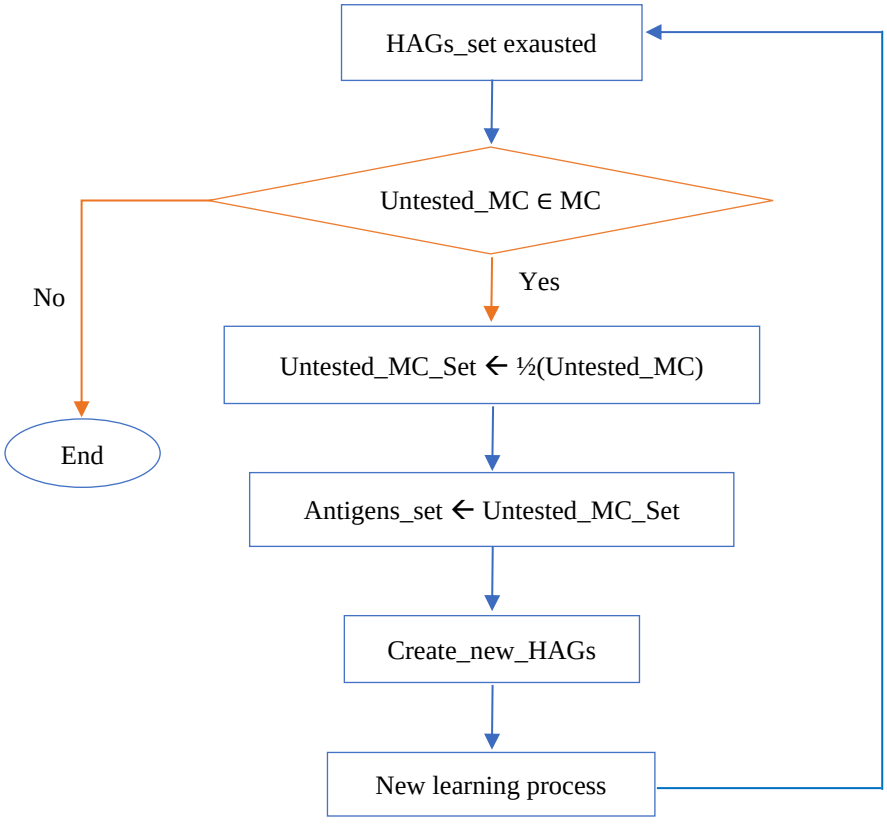

Figure 1: O-AIRS memory cell maturity Flowchart.

By leveraging HAGs and a refined memory cell maturity mechanism, O-AIRS aims to:

- Reduce memory footprint: by selectively retaining mature memory cells, O-AIRS optimizes memory usage, enhancing overall system efficiency.
- Enhance classification accuracy: through iterative interaction with HAGs and comprehensive memory cell testing, O-AIRS aims to achieve superior classification accuracy, even in noisy or complex datasets.
- Ensure comprehensive cell maturity: O-AIRS ensures all memory cells undergo rigorous testing and maturation, enhancing the system's resilience and adaptability in complex data environments.

6 Experimental study

This section presents an experimental comparison to evaluate O-AIRS, the improved version of AIRS and AIRS2. We implemented all three algorithms and tested their performance on two commonly used classification datasets: Haberman Surgery Survival (HSS) and Liver Disorders (LD).

6.1 Haberman surgery survival (HSS)

HSS is one of the hardest datasets for classification [30]. It contains case studies conducted between 1958 and 1970 at the University of Chicago Billings Hospital investigating the survival of patients who had undergone surgery for breast cancer. Characteristics of HSS dataset are depicted in Table 2.

Table 2: Characteristics of HSS dataset.

Number of attributes	Number of records	Number of positive records	Number of negative records
3	306	225	81

HSS dataset attributes are:

- Age of the patient during surgery
- Year of the surgery
- Number of axillary lymph nodes detected

HSS dataset class labels are:

- 1: Patient survived for 5 years or more
- 2: Patient died before 5 years

6.2 Liver disorder's (LD)

The Liver Disorders (LD) dataset contains 345 records of patients diagnosed with either confirmed liver disorders or no disorders. The characteristics of the LD dataset are presented in Table 3.

Table 3: Characteristics of LD dataset.

Number of attributes	Number of records	Number of positive records	Number of negative records
6	345	145	200

LD dataset attributes are:

- Mean Corpuscular Volume (MCV)
- Alkaline Phosphatase (ALP)
- Alanine Aminotransferase (ALT)
- Aspartate Aminotransferase (AST)
- Gamma-Glutamyl Transferase (GGT)
- Number of Alcoholic Drinks per Day

LD dataset class labels are:

- 1: Liver disorders
- 2: No liver disorders

6.3 Learning parameters

The initial learning phase involves creating memory cells (MC) and antigen recognition bodies (ARBs). The number of cells in these initial sets is determined randomly.

Subsequent learning steps utilize specific parameters defined in Table 4. These parameters influence the algorithm's behavior during the learning process.

Table 4: Learning parameters.

Parameter	Meaning	Domain of values	Chosen value
Hyper_clonal_rate	Maximum cloning rate	N	30
Clonal_rate	Average cloning rate	N	20
Mutate_rate	Mutation probability	[0,1]	0.1
Group_Scalar (GS)	Scale of distance between antigens in a HAG (O-AIRS parameter)	[0,1]	0.9

6.4 Results and discussion

Next, we evaluate O-AIRS against established algorithms on benchmark datasets (Liver Disorders and Haberman Surgery Survival) to assess its effectiveness in classification.

6.4.1 Experimental results with LD dataset

To compare the performance of AIRS, AIRS2, and O-AIRS, we used the LD dataset as input with unified common parameters. We assess the performance of the three algorithms using various metrics such as Accuracy, Precision, Recall, and F1-Score. These evaluation metrics are derived from confusion matrices, as shown in Tables 5, 6, and 7. The performance comparison of algorithms on the LD dataset is summarized in Table 8.

Table 5: Confusion matrix – AIRS/LD.

		Predicted	
		Liver disorders	**No liver disorders**
True	**Liver disorders**	47 (TP)	6 (FN)
	No liver disorders	27 (FP)	7 (TN)

Table 6: Confusion matrix – AIRS2/LD.

		Predicted	
		Liver disorders	**No liver disorders**
True	**Liver disorders**	50 (TP)	3 (FN)
	No liver disorders	25 (FP)	9 (TN)

Table 7: Confusion matrix – O-AIRS/LD.

		Predicted	
		Liver disorders	**No liver disorders**
True	**Liver disorders**	50 (TP)	3 (FN)
	No liver disorders	20 (FP)	14 (TN)

Table 8: Performance comparison of AIRS, AIRS2, and O-AIRS on the LD Dataset.

Algorithm	Accuracy	Precision	Recall	F1-Score	Number of memory cells used for classification	Percentage of tested memory cells
AIRS	62,07%	63.51%	88.67%	74.01%	100	68%
AIRS2	67,82%	66.66%	94.33%	78.12%	100	70%
O-AIRS	73,56%	71.42%	94.33%	81.30%	50	100%

The experimental evaluation on the Liver Disorders (LD) dataset revealed compelling insights into the performance of AIRS, AIRS2, and O-AIRS in classification tasks. 8 summarizes the key metrics obtained from the experiments, focusing on Accuracy, Precision, Recall, F1-Score, the number of memory cells used for classification, and the percentage of tested memory cells.

Accuracy and classification performance: O-AIRS demonstrated a significant improvement in Accuracy compared to AIRS and AIRS2, achieving 73.56% Accuracy. In contrast, AIRS and AIRS2 achieved 62.07% and 67.82% Accuracy, respectively. This enhancement underscores O-AIRS's capability to more effectively classify instances within the LD dataset, highlighting its refined learning and activation mechanisms.

Precision, Recall, and F1-Score: Precision measures the proportion of correctly identified positive instances among all instances predicted as positive. O-AIRS exhibited a Precision of 71.42%, while AIRS and AIRS2 showed 63.51% and 66.66%, respectively. This indicates O-AIRS's ability to minimize False Positives more effectively. Similarly, O-AIRS achieved a Recall of 94.33%, outperforming AIRS (88.67%) and comparable to AIRS2 (94.33%). The F1-Score, which balances Precision and Recall, was highest for O-AIRS at 81.30%, compared to 74.01% for AIRS and 78.12% for AIRS2. These results demonstrate that O-AIRS not only enhances Precision but also maintains a high Recall rate, contributing to its superior F1-Score.

Memory cells utilization: a critical observation lies in the number of memory cells utilized for classification. O-AIRS effectively reduced the number of memory cells to 50, half that used by AIRS and AIRS2 (100 memory cells each). This reduction optimizes computational resources and enhances the efficiency of the classification process in O-AIRS, reflecting its streamlined and selective memory cell activation mechanism.

Tested memory cells: importantly, O-AIRS ensures that all memory cells retained for classification have been rigorously tested by antigens, achieving 100% activation fidelity. In contrast, AIRS and AIRS2 activated a lower percentage of memory cells (68% and 70%, respectively), potentially leading to less reliable classifications due to the presence of untested cells.

6.4.2 Experimental results with HSS dataset

To validate the findings from the LD dataset, the same evaluation process was conducted on the Haberman Surgery Survival (HSS) dataset. The algorithms (AIRS, AIRS2, and O-AIRS) were again evaluated using performance metrics like Accuracy, Precision, Recall, and F1-Score. Similar to the LD dataset analysis, these metrics are derived from confusion matrices (presented in Tables 9, 10, and 11). Table 12 presents a comprehensive comparison of the algorithms' performance metrics on the HSS dataset

Table 9: Confusion matrix – AIRS/HSS.

		Predicted	
		Patient died before 5 years.	**Patient survived for 5 years or more**
True	**Patient died before 5 years.**	60 (TP)	3 (FN)
	Patient survived for 5 years or more	10 (FP)	4 (TN)

Table 10: Confusion matrix – AIRS2/HSS.

		Predicted	
		Patient died before 5 years.	**Patient survived for 5 years or more**
True	**Patient died before 5 years.**	62 (TP)	1 (FN)
	Patient survived for 5 years or more	10 (FP)	4 (TN)

Table 11: Confusion matrix – O-AIRS/HSS.

		Predicted	
		Patient died before 5 years.	**Patient survived for 5 years or more**
True	**Patient died before 5 years.**	63 (TP)	1 (FN)
	Patient survived for 5 years or more	9 (FP)	4 (TN)

Table 12: Performance comparison of AIRS, AIRS2, and O-AIRS on the HSS Dataset.

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Number of memory cells used for classification**	**Percentage of tested memory cells**
AIRS	77,92%	85.71%	95.23%	90.22%	100	58%
AIRS2	85,71%	86.11%	98.41%	91.85%	115	54,78%
O-AIRS	87,01%	87.50%	98.43%	92.64%	41	100%

Turning to the HSS dataset, Table 12 outlines the comparative performance of AIRS, AIRS2, and O-AIRS in terms of accuracy, precision, recall, F1-Score, the number of memory cells used for classification, and the percentage of tested memory cells.

Accuracy and classification performance: O-AIRS demonstrated robust classification Accuracy at 87.01%, surpassing both AIRS (77.92%) and AIRS2 (85.71%). This improvement underscores O-AIRS's efficacy in accurately predicting patient survival based on surgical outcomes in the HSS dataset.

Precision, Recall, and F1-Score: analyzing Precision, O-AIRS achieved 87.50%, outperforming AIRS (85.71%) and AIRS2 (86.11%). This higher Precision indicates O-AIRS's superior ability to identify true positive cases among all predicted positives. Moreover, O-AIRS maintained a commendable Recall of 98.43%, higher than AIRS (95.23%) and comparable to AIRS2 (98.41%). The F1-Score for O-AIRS stood at 92.64%, highlighting its balanced performance in Precision and Recall, essential for accurate classification in medical datasets.

Memory cells utilization: similar to its performance on the LD dataset, O-AIRS significantly reduced the number of memory cells utilized for classification in the HSS dataset to 41, compared to 100 for AIRS and 115 for AIRS2. This reduction not only optimizes computational efficiency but also enhances the interpretability and reliability of classification results in O-AIRS.

Tested memory cells: all memory cells retained for classification in O-AIRS were rigorously tested by antigens, ensuring 100% activation fidelity. In contrast, a substantial percentage of memory cells remained untested in AIRS (58%) and AIRS2 (54.78%), potentially leading to less reliable classifications and inconsistent performance.

6.4.3 Discussion and comparison

The experimental evaluation on both LD and HSS datasets establishes O-AIRS's superiority over AIRS and AIRS2. This is achieved through two key innovations: Homogeneous Antigen Groups (HAGs) and optimized memory cell utilization.

HAGs enhance pattern recognition by organizing antigens into groups based on similarity. This strategic

grouping minimizes redundancy and improves classification accuracy by concentrating memory cell activations on pertinent antigenic features. Consequently, O-AIRS achieves superior performance while simultaneously reducing computational complexity.

Moreover, O-AIRS introduces a rigorous testing protocol for memory cells, ensuring that each retained cell undergoes thorough validation by antigens. This meticulous testing process contrasts sharply with traditional methods where many memory cells remain untested, potentially compromising the reliability of classifications. By attaining 100% activation fidelity through tested memory cells, O-AIRS enhances the robustness of its classifications and elevates overall detection rates in diverse datasets.

O-AIRS transcends its role as a mere enhancement over AIRS and AIRS2, positioning itself as a formidable and versatile competitor in the expansive field of classification. By leveraging its innovative mechanisms, O-AIRS effectively competes with other AINs. Traditional Artificial Immune Networks (AINs), such as AINet [11] and Adaptive Artificial Immune Networks (AAIN) [12], excel in tasks such as data clustering and classification but face challenges like computational intensity and sensitivity to parameter settings. Similarly, conventional Negative Selection Algorithms (NSAs) like Self-Nonself Discrimination [5], INSA [6], and ADC-NSA [7] demonstrate proficiency in specific tasks but encounter issues such as high computational costs and the need for precise parameter tuning. Algorithms grounded in Danger Theory, such as DeepDCA [16] and Multi-Level Intrusion Detection System [17], prioritize specific domains and achieve high accuracy rates but are often limited by dataset size and varying detection probabilities. Clonal Selection Algorithms (CSAs), represented by CLONALG [20] and CLONALG-M [23], excel in pattern recognition and optimization tasks but can be sensitive to parameter settings and computational overhead.

O-AIRS leverages Homogeneous Antigen Groups (HAGs) alongside a refined maturity mechanism to significantly boost data classification efficiency. By organizing antigens into coherent groups based on similarity, HAGs streamline the identification of complex data structures without the need for extensive parameter adjustments. This strategic approach not only enhances computational efficiency but also improves result accuracy by mitigating challenges such as uneven antigen distributions and the randomness inherent in detector generation. Consequently, O-AIRS emerges as a robust solution for applications in immune-inspired algorithms. Moreover, O-AIRS dynamically adapts to evolving data environments, thereby enhancing adaptability and robustness. The refined maturity mechanism, coupled with the effective implementation of HAGs, ensures optimal solutions are achieved efficiently. This capability is particularly advantageous in scenarios requiring rapid decision-making and resource efficiency, underscoring O-AIRS's versatility and practical utility across various domains.

7 Conclusion

In conclusion, while current Artificial Immune Systems (AIS) algorithms such as AIRS and AIRS2 offer a distinctive approach to machine learning, their limitations compromise their efficacy. These include reliance on single antigen activation and the retention of untested memory cells, resulting in inaccurate classifications.

This paper introduces the Optimized Artificial Immune Recognition System (O-AIRS), a novel solution aimed at addressing these shortcomings. O-AIRS integrates advanced functionalities such as Homogeneous Antigen Groups (HAGs) and a refined memory cell activation mechanism to enhance its classification capabilities. By leveraging HAGs, which delineate specific subsets within a class, O-AIRS optimizes memory cell activation tailored to precise threat recognition. Moreover, O-AIRS selectively engages validated memory cells during classification, mitigating concerns about untested recruits and facilitating more precise classifications.

O-AIRS's effectiveness was rigorously evaluated using established benchmark datasets. Comparative analyses against its predecessors, AIRS and AIRS2, across multiple metrics including accuracy, precision, and recall consistently demonstrated O-AIRS's superior classification performance. Additionally, O-AIRS exhibited robust adaptability to diverse datasets while maintaining computational efficiency, affirming its utility in complex real-world applications.

This efficiency is further underscored by O-AIRS's optimized activation mechanism, ensuring activation of only pertinent memory cells during classification and reducing the required number of cells by approximately 50% compared to original algorithms. Critically, all memory cells retained in O-AIRS are validated, enhancing their contribution to accurate classification outcomes.

Looking ahead, O-AIRS presents promising avenues for future research. One compelling direction involves exploring O-AIRS's dynamic adaptation of HAGs during classification, enhancing its ability to detect evolving threats in real-time. Such adaptive strategies hold particular promise in dynamic domains like cybersecurity and financial anomaly detection.

Furthermore, expanding O-AIRS's applicability beyond current datasets would provide insights into its generalizability and suitability for broader deployment. Exploring diverse distance metrics for antigen comparisons within HAGs could also yield additional performance enhancements. Additionally, investigating alternative memory cell types or hybrid approaches within O-AIRS offers potential for further augmenting its efficacy in handling intricate classification tasks.

References

[1] A.B. Watkins, AIRS: A resource-limited artificial immune classifier, Mississippi State University, 2001.

[2] A. Watkins, J. Timmis, Artificial immune recognition system (AIRS): Revisions and refinements, AISB 2004 Convention, 2004, pp. 18.

[3] R. Abdelkhalek, Z. Elouedi, Optimized Evidential AIRS with Feature Selection and Genetic Algorithm, International Conference on Intelligent Systems Design and Applications, Springer, 2022, pp. 1088-1097. https://doi.org/10.1007/978-3-031-27440-4

[4] R.V. Melo, D.D. de Macedo, A.M. Dantas, C.L. de Bona, A novel immune detection approach enhanced by attack graph based correlation, 2019 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2019, pp. 1-6. https://doi.org/10.1109/ISCC47284.2019.8969772

[5] S. Forrest, A.S. Perelson, L. Allen, R. Cherukuri, Self-nonself discrimination in a computer, Proceedings of 1994 IEEE computer society symposium on research in security and privacy, Ieee, 1994, pp. 202-212. https://doi.ieeecomputersociety.org/10.1109/RISP.1994.296580

[6] Y. Ren, X. Wang, C. Zhang, A novel fault diagnosis method based on improved negative selection algorithm, IEEE Transactions on Instrumentation and Measurement 70 (2020) 1-8. https://doi.org/10.1109/TIM.2020.3031166

[7] C. Yang, L. Jia, B.-Q. Chen, H.-Y. Wen, Negative selection algorithm based on antigen density clustering, IEEE Access 8 (2020) 44967-44975. https://doi.org/10.1109/ACCESS.2020.2976875

[8] X. Song, F. Gao, Z. Chen, W. Liu, A negative selection algorithm-based identification framework for distribution network faults with high resistance, IEEE Access 7 (2019) 109363-109374. https://doi.org/10.1109/ACCESS.2019.2933566

[9] N.K. Jerne, Towards a network theory of the immune system, Ann. Immunol. 125 (1974) 373-389.

[10] F.J. Varela, A. Coutinho, Second generation immune networks, Immunology today 12 (1991) 159-166. https://doi.org/10.1016/S0167-5699(05)80046-5

[11] L.N. De Casto, F.J. Von Zuben, An evolutionary immune network for data clustering, Proceedings. Vol. 1. Sixth Brazilian Symposium on Neural Networks, IEEE, 2000, pp. 84-89. https://doi.ieeecomputersociety.org/10.1109/SBRN.2000.889718

[12] J.M. Vidal, A.L.S. Orozco, L.J.G. Villalba, Adaptive artificial immune networks for mitigating DoS flooding attacks, Swarm and Evolutionary Computation 38 (2018) 94-108. https://doi.org/10.1016/j.swevo.2017.07.002

[13] L. Zhang, Y. Wang, X. Zhou, W. Liu, Improved artificial immune algorithm based on dynamic shrinkage strategy, 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), IEEE, 2022, pp. 867-872. https://doi.org/10.1109/ICCECE54139.2022.9712692

[14] S. Kanwal, F. Khan, S. Alamri, K. Dashtipur, M. Gogate, COVID-opt-aiNet: A clinical decision support system for COVID-19 detection, International Journal of Imaging Systems and Technology 32 (2022) 444-461. https://doi.org/10.1002/ima.22695

[15] P. Matzinger, Essay 1: the Danger model in its historical context, Scandinavian journal of immunology 54 (2001) 4-9. https://doi.org/10.1046/j.1365-3083.2001.00974.x

[16] S. Aldhaheri, D. Alghazzawi, L. Cheng, B. Alzahrani, A. Al-Barakati, Deepdca: novel network-based detection of iot attacks using artificial immune system, Applied Sciences 10 (2020) 1909. https://doi.org/10.3390/app10061909

[17] V.T. Alaparthy, S.D. Morgera, A multi-level intrusion detection system for wireless sensor networks based on immune theory, IEEE Access 6 (2018) 47364-47373. https://doi.org/10.1109/ACCESS.2018.2866962

[18] H.M. Salmon, C.M. De Farias, P. Loureiro, L. Pirmez, S. Rossetto, P.H. de A Rodrigues, R. Pirmez, F.C. Delicato, L.F.R. da Costa Carmo, Intrusion detection system for wireless sensor networks using danger theory immune-inspired techniques, International journal of wireless information networks 20 (2013) 39-66. https://doi.org/10.1007/s10776-012-0179-z

[19] X. Zhou, K. Liu, Z. Jin, S. Tian, Y. Fu, L. Qin, Artificial Immune Based Cryptography Optimization Algorithm, Informatica 38 (2014).

[20] L.N. De Castro, F.J. Von Zuben, Learning and optimization using the clonal selection principle, IEEE transactions on evolutionary computation 6 (2002) 239-251. https://doi.org/10.1109/TEVC.2002.1011539

[21] R. Huang, H. Tawfik, A. Nagar, Licence plate character recognition using artificial immune technique, International Conference on Computational Science, Springer, 2008, pp. 823-832. https://doi.org/10.1007/978-3-540-69389-5

[22] E.D. Ülker, Gene Selection in Microarray Data Using an Improved Approach of CLONALG, The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Springer, 2019, pp. 466-472. https://doi.org/10.1007/978-3-030-36178-5

[23] S.A. Sert, A. Yazici, Increasing energy efficiency of rule-based fuzzy clustering algorithms using CLONALG-M for wireless sensor networks, Applied Soft Computing 109 (2021) 107510. https://doi.org/10.1016/j.asoc.2021.107510

[24] Z. Zareizadeh, M.S. Helfroush, A. Rahideh, K. Kazemi, A robust gene clustering algorithm based on clonal selection in multiobjective optimization framework, Expert Systems with Applications 113 (2018) 301-314. https://doi.org/10.1016/j.eswa.2018.06.047

[25] S. Shekhar, D. Sharma, D. Agarwal, Y. Pathak, Artificial immune systems-based classification model for code-mixed social media data, IRBM 43 (2020) 10. https://doi.org/10.1016/j.irbm.2020.07.004

[26] W. Nebili, B. Farou, Z. Kouahla, H. Seridi, Revised Artificial Immune Recognition System, IEEE Access

9 (2021) 167477-167488. https://doi.org/10.1109/ACCESS.2021.3133731

[27] A.B. Watkins, Exploiting immunological metaphors in the development of serial, parallel and distributed learning algorithms, University of Kent (United Kingdom), 2005.

[28] A.A. Freitas, J. Timmis, Revisiting the foundations of artificial immune systems for data mining, IEEE Transactions on Evolutionary Computation 11 (2007) 521-540. https://doi.org/10.1109/TEVC.2006.884042

[29] H. Merabti, B. Farou, H. Seridi, A segmentation-recognition approach with a fuzzy-artificial immune system for unconstrained handwritten connected digits, Informatica 42 (2018) 95-106.

[30] H.N.A. Pham, E. Triantaphyllou, An application of a new meta-heuristic for optimizing the classification accuracy when analyzing some medical datasets, Expert Systems with Applications 36 (2009) 9240-9249. https://doi.org/10.1016/j.eswa.2008.12.007