Article

A Coordinated Air Defense Learning System Based on Immunized Classifier Systems

Sulemana Nantogma, Yang Xu * and Weizhi Ran

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 201814080012@std.uestc.edu.cn (S.N.); 202012081713@std.uestc.edu.cn (W.R.)
* Correspondence: xuyang@uestc.edu.cn; Tel.: +86-189-809-478-76

Abstract: Autonomous (unmanned) combat systems will become an integral part of modern defense systems. However, limited operational capabilities, the need for coordination, and dynamic battlefield environments with the requirement of timeless in decision-making are peculiar difficulties to be solved in order to realize intelligent systems control. In this paper, we explore the application of Learning Classifier System and Artificial Immune models for coordinated self-learning air defense systems. In particular, this paper presents a scheme that implements an autonomous cooperative threat evaluation and weapon assignment learning approach. Taking into account uncertainties in a successful interception, target characteristics, weapon type and characteristics, closed-loop coordinated behaviors, we adopt a hierarchical multi-agent approach to coordinate multiple combat platforms to achieve optimal performance. Based on the combined strengths of learning classifier system and artificial immune-based algorithms, the proposed scheme consists of two categories of agents; a strategy generation agent inspired by learning classifier system, and strategy coordination inspired by Artificial Immune System mechanisms. An experiment in a realistic environment shows that the adopted hybrid approach can be used to learn weapon-target assignment for multiple unmanned combat systems to successfully defend against coordinated attacks. The presented results show the potential for hybrid approaches for an intelligent system enabling adaptable and collaborative systems.

Keywords: air defense system; artificial immune system; command and control; learning classifier system; multi-agent systems; weapon-target assignment

1. Introduction

In recent years, the challenges arising from the deployment of robotic systems in defense and security applications have been addressed from several different perspectives. As a typical multi-agent system, the complexity does not only emanate from the dynamics of the application domain, but also the limited and uncertainty in perceived information. In an attempt to handle these uncertainties, the research community has proposed several theories [1–3].

In the domains of defense application, unmanned combat platforms may be deployed to provide defense against targets and coordinated attacks by enemy adversaries. In these regards, an important functionality required by the multi-agent system is the threat evaluation and weapon assignment to targets that pose threats to its survival and/or the assets being protected. The threat evaluation and weapon assignment functionality is one of the fundamental problems in Command and Control Systems [4]. Threat evaluation and weapon assignment includes the evaluation and assignment of defensive weapons to intercept detected hostile threats with the primary objective of minimizing the total expected survivability of the threat [5,6]. Threat evaluation and weapon assignment functionalities are a fundamental part of achieving successful outcomes in naval combat scenarios.

To assist with the ever more complex and rapidly changing combat environment, combat systems require access to real-time adaptable control systems that can handle uncertain

information to provide intelligent decision-making when executing combat tasks. Research work in this area seeks to realize the effective utilization of combat resources by minimizing defensive assets loss, amount of resource utilized during interception, and maximizing the effective kill probability while operating within the multiple constraints of the operating platforms [5,7–10]. Effective combat systems in a multi-agent setting must be able to solve conflicts of interest and optimize resource allocation.

Conflicts of interest and optimized resource allocation in general have received extensive attention in recent decades through the application of various approaches [11–13]. However, the limited operational capabilities and sophisticated battlefield environments with the requirement of timeless in decision-making are peculiar difficulties to be solved by Unmanned Combat Systems in modern battles. The dynamic and sophistication in missions and situations encountered in the battlefield makes it challenging to design intelligent and self-adaptive decision-making agents that are robust and fit for the battlefield since the urgency associated with combat decision-making has many consequences. Hence, the need to design methods and models to enable self-learning and coordinated decision-making that is capable of adapting online by employing learning algorithms and simulation-based training.

Multiple unmanned combat systems performing air defense operation is a typical multi-agent system. In these type of missions, multiple entities are connected in a network, with each system capable of computing and communicating autonomously and concurrently. Decision-making is based on the agents' knowledge of the environment and various entities in their locality of the network. The actions of individual agents in a multi-agent system are suppressed or stimulated during the lifetime of the agents. This suppression and stimulation are key characteristics of the artificial immune system.

The Artificial Immune System is an example of multi-agent and decentralized information processing system that is capable of learning and remembering [14]. The artificial immune system is based on working mechanisms exhibited by the Biological Immune System. The complexity exhibited by the immune system has inspired various theories and models which represent the different aspects proposed under the artificial Immune system such as the immune network [15], Clonal Selection [16], Negative Selection, and Danger Theory [17]. Several applications have been demonstrated based on these theories [18]. Despite the distributed nature of the immune system, their many engineering implementations are centralized.

On the other hand, the learning classifier system is one of the machine learning techniques introduced by Holland in 1976. It combines evolution and learning in an integrated way. A Learning classifier system evolves a group of $IF - THEN$ rules called classifiers [19] to find a suitable response of a learning system (agent) to incoming sensory inputs. As opposed to other learning approaches such as neural networks, rules evolved by a learning classifier system and its variants are easily understandable and general enough for a wide range of learning tasks such as traffic control problems [20,21], multi-agent control problems [22–25], and robotic control [26–28] etc. In the learning classifier system approach, agents acquire appropriate strategies while interacting with an environment by updating their classifier sets. The adaptation of classifiers is mostly done through evolutionary computing based on genetic algorithms.

In this paper, a hybrid of artificial immune and learning classifier system is utilized to design a self-learning system that implements autonomous cooperative threat evaluation and weapon assignment learning. Based on the combined strengths of the two approaches, we seek to realize an intelligent and adaptive system for a team of combat surface platforms to successfully perform air defense weapon-target allocation against a coordinated attack of heterogeneous targets in a realistic maritime environment. By considering the air defense operation from a force coordination perspective, heterogeneous distributed combat systems are required to make effective weapon-target assignment decisions to achieve interception of fast and multi-batch targets. To meet the requirement of real-time decision-making, combat systems are controlled by intelligent agents to coordinate and engage threats in

stages based on the observed outcome of previous engagements. Threat evaluation and weapon assignment is done in a distributed manner to reduce communication complexity. Each agent only communicates and receives initial action strategies it generates with its neighbors. This approach can increase the adaptability and success rate of the overall mission. An experiment in a realistic environment shows that the method can be utilized to learn weapon-target assignment for groups of combat platforms to successfully defend against coordinated attacks. This research presents the following general contributions:

1. We investigate and demonstrate the applicability of a hybrid artificial immune and Learning Classifiers System for realizing air defense intelligence.
2. A hierarchical self-learning scheme for multiple unmanned combat systems air defense weapon-target allocation that integrates artificial Immune based algorithms with learning classifier system is presented.
3. We propose an approach to facilitate learning by applying a negative selection concept to filter out and condense situations from individual decision units.

The rest of this paper is presented as follows. In Section 2, the related work of this study and a brief background of the learning classifier system and the artificial immune system algorithms are discussed briefly. In Section 3, a description, analysis, and requirement of the air defense system are described and the air defense problem formulated. Based on this analysis and requirement, the framework allowing the multiple unmanned combat systems to learn an air defense strategy using artificial Immune system and learning classifier system is presented in Section 4. Having presented the framework for the autonomous air defense strategy learning, experimental results in a realistic combat environment are presented in Section 5. Finally, we present our concluding remarks by summarizing our contribution and experimental results.

2. Background and Related Work

In this section, we first present some related works in unmanned combat vehicles autonomous decision-making and learning approaches. Next, the biological principles of the artificial immune system that are relevant to this work are presented. Finally, the background knowledge of learning classifier system and it's working mechanism is presented.

2.1. Related Work

It is expected that unmanned combat systems such as unmanned combat aerial vehicles, unmanned combat surface vehicles, unmanned combat ground vehicles, etc., will be deployed in diverse missions in the near future. However, the insufficient and noisy information available for decision-making and control, simultaneous control variables, several conflicting objectives toppled with a dynamic and complex operating environment presents a challenge.

Several algorithms and approaches have been applied to realize control systems that enable autonomous systems to execute complex missions [29–35]. For instance, an autonomous decision-making for unmanned combat aerial vehicles (UCAVs) using genetic fuzzy trees is presented by authors in [29]. The proposed system shows the capability of obtaining strategies that are robust, aggressive and responsive against opponents. In [30], an intelligent air combat learning system based on the learning dynamics of the human brain is designed and its strategy acquisition without prior rules is demonstrated. The authors in [34] used deep q-neural networks to obtain combat strategies in an attack-defense pursuit-warfare of multiple UCAVs in a simplified environment. Alpha C2, an intelligent Air Defense Commander operations using a deep reinforcement learning framework was presented in [10]. The proposed system makes combat decision independent of human decision-making. A multi-agent-based training system for USVs was presented in [35] and a combat task for combat USVs was used to evaluate the system performance. Other methods such as optimizations and game theoretic approaches have been reported [31–33].

This work contributes to these body of literature by hybridising two known control and decision-making approaches; Learning Classifier Systems and Artificial Immune Systems.

2.2. Learning Classifier Systems

In addition to the different variations proposed since its first introduction by Holland, learning classifier systems have witnessed some improvement and real-world applications in recent years [36–41]. One of the most main-stream and widely applied learning classifier systems is the extended classifier system [42]. In the extended classifier system, classifiers that always receive the same reward are sort after. As in a standard learning classifier system, the extended classifier system consists of three main components; Performance, Reinforcement, and Discovery components.

$$P(a_i) = \frac{\Sigma_{cl_k \in [M]|a_i} cl_k.p \times cl_k.F}{\Sigma_{cl_l \in [M]|a_i} cl_l.F} \tag{1}$$

In the Performance Component, actions are selected and executed. To do this, a matched set $[M]$ of classifiers that match the current input messages in a population $[P]$ are formed and the acquisition reward for each action a_i is calculated to obtain a prediction array $P(a_i)$ using Equation (1). An action inferred from the prediction array is selected and performed in the environment. Meanwhile, classifiers in $P(a_i)$ that have the selected action are stored in an action set $[A]$. After the execution of the selected action, execution is passed on to the Reinforcement component.

In the Reinforcement Component, the parameters of the classifiers in $[M]$ are updated. These parameters include the prediction p, the error ϵ which is the difference between the reward received R and classifier prediction $cl.p$ and the fitness of classifiers. To update the p, the reward received from the environment is used as shown in Equation (2).

$$cl.p = cl.p + \beta(R - cl.p); \tag{2}$$

where β is the learning rate. Consequently, the prediction error, ϵ is updated using Equation (3)

$$cl.\epsilon = cl.\epsilon + \beta(|R - cl.p| - cl.\epsilon); \tag{3}$$

$$cl.F = cl.F + \beta(\hat{\lambda}(cl) - cl.F); \tag{4}$$

$$\lambda(cl) = \begin{cases} 1 & if\ \epsilon < \epsilon_0 \\ \alpha(\frac{\epsilon}{\epsilon_0})^{-v} & otherwise \end{cases}; \tag{5}$$

$$\hat{\lambda} = \frac{cl.n \times \lambda cl}{\sigma_{cl_b \in [A]} \lambda(b) \times b.n} \tag{6}$$

On the other hand, the fitness of a classifier is updated based on its accuracy λ of the classifier cl as represented in Equation (4), where $\hat{\lambda}$ is the relative accuracy of the classifier. And $\hat{\lambda}$ is computed using Equation (6) by first computing the accuracy, λ of the classifier with Equation (5). In Equation (5), ϵ_0 is an accuracy criterion constant. A classifier cl is deemed accurate when $cl.\epsilon$ is less than ϵ_0. The variables α and v are used to control the rate of reduction of accuracy.

Finally, the Discovery component applies Genetic Algorithms (GA) to generate new rules. This is usually done by selecting two parents from $[A]$ using a selection probability to produce new offspring through a crossover process. The new off-springs further undergo some mutation with a probability before they are added back to $[P]$. To keep the population fixed, classifiers with lower fitness are deleted if the number of classifiers in $[P]$ exceeds a certain value. A subsumption process can be applied to replace classifiers whose condition parts are included in more accurate and experienced classifiers.

A learning classifier system largely employs genetic algorithm for rules discovery. Genetic Algorithm is a gradient-free and parallel optimization algorithms. Genetic Algorithm

searches for global optimum using performance criteria for evaluation and population of possible solutions. Due to its ability to handle complex and irregular solution spaces, various difficult optimization problems have been studied with genetic algorithm. Generally, in a genetic algorithm, parameters of a possible solution are encoded as chromosomes which are manipulated using genetic operators. In its simplest form, selection, crossover, and mutation are the main operators applied in a genetic algorithm.

During Selection operation, higher fitted chromosomes in the population are selected for reproduction. Thus, the fitter a chromosome, the more chances for it to be selected to reproduce.

Crossover on the other hand, enables the creation of new offspring. By biologically mimicking recombination between two single chromosome organisms, this process leads to the exchange of sub-sequences between two chromosomes to create two offspring.

Lastly, Mutation operator changes the values of genes in the chromosome. In the case of binary encoded chromosomes, some of the bits are randomly flipped. Mutation can occur on any of the genes of the chromosome. In this paper, genetic algorithm is applied as in standard learning classifier system for rules discovery at the strategy generation level while the clonal selection and immune network theory are applied at the strategy coordination level.

2.3. Artificial Immune System

The domain of artificial immune system involves the study and application of various computational techniques based on the biological immune system [43]. The biological immune system provides defense mechanisms that protect a host organism against harmful foreign agents such as viruses and bacteria. The two main biological defense mechanisms influencing each other include the Innate Immunity and the Adaptive Immunity that utilizes a set of cells referred to as T-cell to recognize harmful foreign agents (antigens) and B-cells to eliminate these harmful foreign agents. The acquired immune system located in the vertebrates of the organism retains a memory of encountered exposures which is recalled when reinfection occurs to cure the infection.

The dynamism exhibited by the biological immune system has inspired various theories and models which represent the different aspects proposed under the artificial Immune system such as the Immune Network [15], Clonal Selection [16], and the Danger Theory [17]. These under-listed theories which are relevant to our proposed approach are explained below.

2.3.1. Clonal Selection

The Clonal Selection Based Algorithms are inspired by the working mechanisms of antigen-antibody recognition, binding, cell propagation, and separation into memory cell [16]. The clonal selection algorithm, CLONALG [44] based on clonal selection and affinity maturation principles with engineering applications are one such algorithm. According to the clonal selection theory, the presence of affinity between stimulated antigen's epitope and B-cell receptors causes the B-cells to divide. These divided cells then undergo maturity to become plasma cells that secrete antibodies. Subsequently, the Antibodies with higher affinities undergo reproduction through hyper-mutation of B-cells. Some matching B-cells are retained by the Immune System. The adaptation feature emanates from the mechanism of building up of concentrations of the B-cells and the diversity in mutation of the cells in the bone marrow. The interaction in the clonal selection theory is between the antigens and the antibodies only.

2.3.2. Danger Theory

Among the mechanisms that inspire the immune system theories is its ability to identify events that signify danger and those that do not. The concept of responding to these harmful events that cause damage form the hypothesis of the danger theory [17]. The idea of danger theory is that the immune systems does not just respond to foreign

antigens but harmful ones. Which introduces the concept of discrimination on which non-self/antigens to respond to. Hence, the immune system chooses to respond to danger instead. Danger in this case, is measured by cells' damage, indicated by distress signals emanating from unnatural deaths of cell. These danger signals essentially establish a danger zone. Consequently, only B-cells that match antigens within this danger zone are stimulated and undergo the clonal expansion process [45].

Similarly, in an immunized air defense system, determining which threat to prioritize is key to a successful defense. Since defense systems employ other mechanisms such as jamming, weapon assignment decision should be made taking into concentration the threatening level of the target. For instance, successfully jammed weapons should not be intercepted by the defense system.

2.3.3. Immune Network Concepts

The network theory of the immune system proposed by Jerne in [15] presents a generic view of the working mechanism of the biological immune system. It is premised on the concept that the immune system is constructed as a large-scale closed system of lymphocytes through mutual interaction between different species of lymphocytes. The interaction of cells happens regardless of the presents of harmful foreign agents or not. Jerne's theory stipulates that, the antibody of an immune cell possess paratope and idiotope. The idiotope of one antibody can be recognized by the paratope of another antibody with or without the presence of an antigen that posses an epitope (analogous to an idiotope). This recognition and interaction results in a network that is dynamic and leads to stimulation and suppression. The recognised antibody is suppressed while the recognizer antibody is simulated. This process of antibodies being able to recognize each other drives the adoption of this theory to heterogeneous robotic applications.

3. Air Defense System Model

In this paper, the problem of threat evaluation and weapon assignment decision-making for air defense operations is considered. In particular, this work considers the problem where a team of unmanned combat systems is deployed to protect a high-value target or contested region. The group consists of a heterogeneous set of unmanned combat systems with different capabilities and multiple types of air-to-surface weapon systems. In this scenario, high valued asset is to be protected from enemy air units trying to penetrate the defense area and destroy the defended asset. This problem is challenging as the characteristics of the scenario can scale to an unlimited number of stages based on the following.

- The number of detected targets at the decision time.
- The types of targets (attackers),
- The weapon capability vector of attackers,
- The number of air defense platforms,
- The weapon capability vector of defenders
- The state of high valued assets being protected, etc.

In all the different stages that the mission can be modified, the objective is to destroy enemy targets and protect the asset from being destroyed by enemy units. The objective is in conflict with the enemy targets who seek to penetrate the defense area. Hence, an efficient threat evaluation and weapon assignment system is required by both sides to achieve their mission objectives. For every detected target, the control system should make its decision based on the following.

- The defense system evaluation of the threat.
- The high valued asset state if any.
- Weapon capability and state of the unit.
- Available weapons and teammates.

Taking these into consideration, the control system decides on whether to deploy a weapon against the threat or not. If yes,

- the type of weapon to use against the target and
- quantity to be fired.

Formally, at a certain time T, the multi-agent system ag detects N targets H which can be a weapon (e.g., missile) or hostile flying unit. Meanwhile, each agent has C categories of weapons. Each category C_p has j quantity of ammunition, $C_p = \{c_{p1}, c_{p2}, ...c_{pj}\}$ to intercept targets at different ranges and efficiency. Here, the objective is to survive the attack with minimal loss and cost of resource utilization by maximizing targets being destroyed while minimizing the cost of operations in other to be able to defend against subsequent engagements. Hence, the objective can be formulated as:

$$
O_b = \begin{cases}
min \sum_{j=1}^{H'(t)} \sum_{i=1}^{C} \beta_i q_{ij} \delta_{ij} \\
\\
max \sum_{j=1}^{H'(t)} v_j (1 - \prod_{i=1}^{C} (1 - p_{ij})^{\delta_{ij}(s)})
\end{cases}
\tag{7}
$$

1. W is the total remaining number of employable weapons of the combat platform. i.e., weapon that can be used against the contact by any member of the group.
2. H' is the remaining number of unassigned detected targets
3. δ_{ij} is the minimum delay before weapon i can be deployed against target j based on ready time and current allocation of the weapon of a unit.
4. q_{ij} is the quantity of ammunition of weapon of type i allocated to target j,
5. β_i represents the unit cost of the ammunition of weapon i
6. v_j means the threat value of target j
7. p_{ij} is the weapon kill probability

In this mission, the two main problems to be solved are: action strategy selection and cooperation. Action strategy selection in this case, involves deciding the type of weapon to deploy, the salvo or number of weapons. Cooperation is the coordination of the multi-agent system so that the assignment is distributed as much as possible among several agents to minimize the time between target detection and interception.

4. Approach

In this section, we present the approach and scheme adopted in this paper to realize a robust control system. The focus is to design a self-learning system to dynamically acquire threat evaluation and weapon assignment and coordination strategies. In this section, the details of system model are presented.

4.1. System Architecture

The framework which is an integration of artificial Immune system and learning classifier system proposed in this work is shown in Figure 1. The general framework is a typical multi-agent system consisting of two main agents; the Strategy Generation Agent and Strategy Coordination Agent. The Strategy Generation Agent is a low level agent that learns the appropriate weapons and quantity to deploys against a given target using learning classifier mechanism. The Strategy Coordination Agent receives proposed assignments from other agents and takes as input the generate classifiers as antigens from the Strategy Generation Agent perform immune network dynamics to determine the actions of the agent.
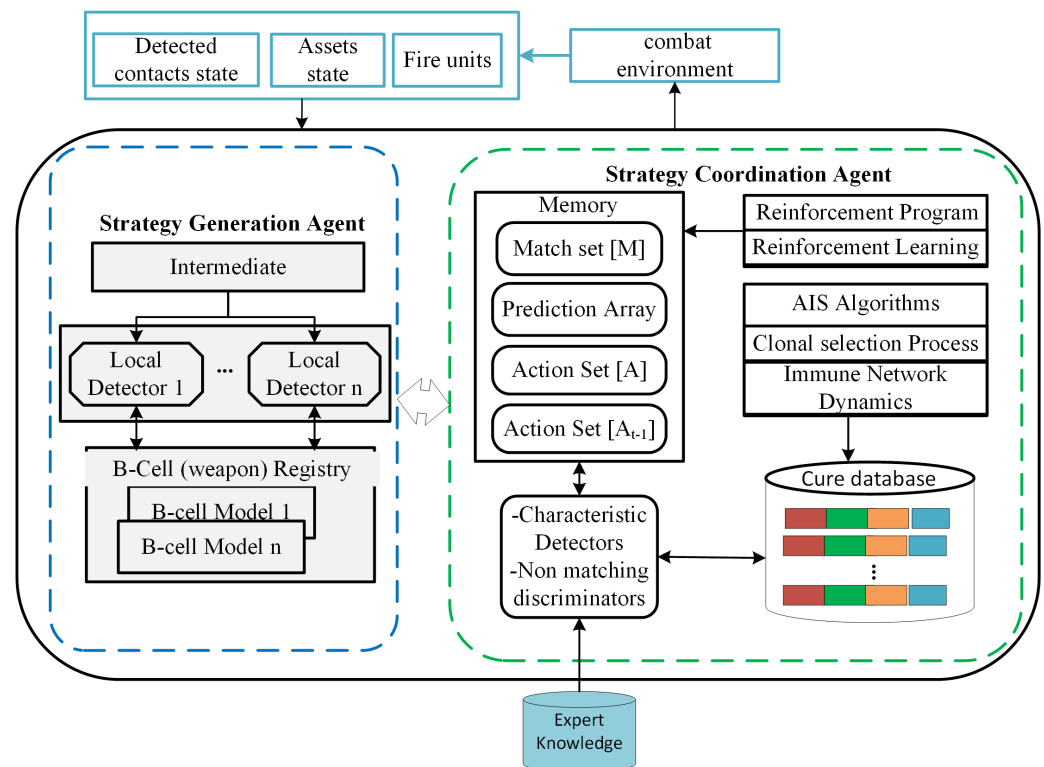
Figure 1. Immunized air Learning Classifier System Model.

In the Strategy Generation Agent, each weapon type is an abstract B-cell (decision unit) which is a vessel for carrying antibodies. Antibodies are classifiers that contain the control decisions as its consequent. While detected contacts are modeled as antigens, whose epitopes are expressed in terms of their intent, capability, and opportunity. The characteristic mapping based on the local detectors of antigen and individual B-cells and their activation leads to the generation of possible classifier sets that forms a sub-population of classifiers according to the specific targets. The competition between these sub-populations of classifiers produces a single classifier whose action part is applied against a target. The Strategy Generation Agent interfaces with the environment through the intermediate module which provides local detectors for B-cells.

Strategy Coordination Agent performs the final actions selection based on the concentration of classifiers in the Immune network by applying the AI algorithms. The Immune network models a collection of classifiers of the various B-cells. Each classifier has a connections part that is used to determine its appropriateness in a given situation with respect to the other classifiers. Strategy Coordination Agent receives classifiers from the B-cells, established their connections to form a network, and apply the immune network dynamics. The individual classifiers that undergo genetic operations and updates are based on the output classifiers of the immune network dynamics.

4.2. Classifier Representation and Encoding

In order to be able to apply immune network dynamics and coordinate classifiers generated by different B-cells, the form of classifiers is changed. In this case, classifiers generated by a b-cell are partitioned into 4 parts that corresponds to the paratope and idiotope of a cell. These connections are synonymous with idiotope and indicate which other classifiers it is connected to. The Identifier is used to indicate the particular b-cell that the classifier belongs. The antecedent indicates when the B-cell and weapon type is applicable while the consequent specifies quantity to fire. The parameters of standard extended classifier still applies. The antecedent part is encoded as singletons of fuzzy sets that represent the states of the variables of decision elements. The consequence part of each classifier is the number of ammunition to fire. The weapon to employ can directly

be inferred from the posted classifiers since classifiers are identified by B-cells which represents the agent and weapon type. Figure 2 shows a schema of a classifier in our system. Both unit and target states are vectors of the individual variable representations of the classifier.
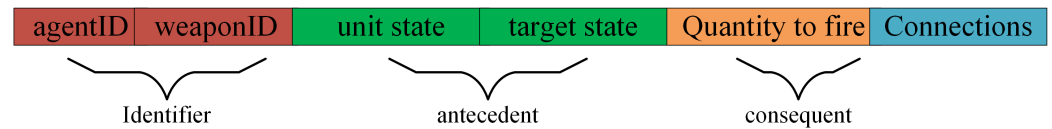
agentID	weaponID	unit state	target state	Quantity to fire	Connections

Identifier   antecedent   consequent

Figure 2. A schematic representation of a classifiers.

We design an intermediate module to further quantize all variable and apply binary coding scheme for the antecedent with size equal to the sum of the states of quantization in each input variable. The consequent part employs an integer coding scheme where each gene contains the index of the state used for the corresponding output variable. The levels of quantization of decision variables and continuous variables' value ranges are mapped to seven states from 1 to 7 of the corresponding value bits. For instance, consider the continues variable of 'velocity, distance, heading' of a detected target. If we are to quantize them to 3 states (fuzzy sets) each as shown in Equations (8)–(10). The indexes are 1, 2, and 3 from *low* to *high*. By this quantization, the velocity variable state can be represented by 3-bits only, where allele *'1'* means that the corresponding state/index is used in the velocity variable. That is, the encoding *'001'* means the velocity of a contact as read from the environment is *high*.

$$
velocity = \begin{cases} high & when\ velocity\ > 500\ m/s \\ medium & when\ 250 < velocity \leq 500\ m/s \\ low & when\ 0 \leq velocity \leq 250\ m/s \end{cases} \tag{8}
$$

$$
distance = \begin{cases} high & when\ distance\ > 1000\ km \\ medium & when\ 500 < velocity \leq 1000\ km \\ low & when\ 0 \leq distance \leq 500\ km \end{cases} \tag{9}
$$

$$
heading = \begin{cases} high & when\ heading\ > 90\ deg \\ medium & when\ 45 < heading \leq 90\ deg \\ low & when\ 0 \leq heading \leq 45\ deg \end{cases} \tag{10}
$$

Hence, a B-cell with sub characteristic detectors of distance, velocity and heading respectively for a target as an input, can have a classifier as *'1;2|100;010;001|2|1;3;5'* conforming schema as shown in Figure 2. The individual parts are separated by *'|'* and can be interpreted as; "IF *distance* IS *low* AND *velocity* IS *medium* AND *heading* IS *high* THEN fire 2 ammunition USING weapon 2". Also, in this particular example, the classifier is connected to classifiers 1, 3, 5.

It is worth noting that each hypothetical B-cell (decision unit) implement an extended classifier system for action selection. However, contrary to the many approach, the entire prediction array from individual classifiers are synthesized with respect to other B-cells predictions using the artificial immune network dynamics to obtain an optimal weapon-target assignment for the current environment state.

4.3. Coping with Multiple Targets

Learning to prioritizing contacts to counter first by an autonomous defense system when multiple threats are detected is key to a successful defense. Since each weapon is a B-cell with specific characteristics with respect to the detected threats, the concept of danger signals and affinity can be modeled. Based on the hypothesis of the danger theory, when a target is detected, a danger zone is constructed based on the trajectory of the detected target.

The value of the combat units within the danger zone signals how much priority should be given in intercepting the target. This is important since combat systems with jamming capability will succeed in jamming some incoming targets. Therefore it is unnecessary to intercepts those targets after it has been jammed.

4.4. Action Strategy Selection

At time t during an engagement, an agent receives observation from the environment. The trajectories of the targets are use to construct danger zone and the danger level of all detected targets determined. The intermediate module process the targets information and encode as antigens to the individual B-cells in range of the targets. The input information is of two types: (1) The external observation from the battlefield which include the observed state of detected targets, high value asset and teammate agents information. This information includes basic information such as distance, heading, speed, and contact type, etc. for detected targets as observed by a unit; (2) The characteristic internal state of decision units, including weapon types and states, range, quantity, etc.

Based on the quantized variables and characteristics of each B-cell, a matched set is generated by each B-cell for each target. To form the prediction array $[P_a]$, the expected action payoff for every action strategy of B-cells is calculated using Equation (11). ϖ is the affinity between B-cell of classifiers and antigens.

$$P(a_i) = \frac{\Sigma_{cl_k \in [M] | a_i} cl_k.p \times cl_k.F * \varpi}{\Sigma_{cl_l \in [M] | a_i} cl_l.F * \varpi} \tag{11}$$

The affinity between B-cell's generated classifiers and antigens is estimated base on Equation (12). In Equation (12), q is the quantity of ammo suggested by the classifier, d_g is normalized distance between target c and combat unit (B-cell), p_g is the speed advantage of fire unit weapon against target. w_e is the effectiveness (kill probability) of the weapon suggested, r_t is the ready time of weapon if it were to be deployed and v_c is the quantified target type that estimates the value of the target.

$$\varpi = \left[1 - \prod_{i=1}^{q} (1 - d_g * p_g * w_e * r_t) \right] * v_c \tag{12}$$

The classifiers are condensed by applying the negative selection described in the next section. Next, the condensed classifier set is transmitted to its neighbors and execution is passed to the Strategy Coordination Agent.

In the Strategy Coordination Agent, the connections of the classifiers are established based on the classifiers from the Strategy Generation Agent and neighbor units. Classifier i is said to be connected to classifier j if they produce antibodies for the same target. Based on the classifiers' connections the immune network dynamics are executed to obtain the classifiers whose action will be posted to the environment by a combat unit. The classifiers with the higher concentrations for each target are allowed to post their actions.

Since classifiers must cooperate and compete with other classifiers, each classifier has a concentration level based on it connectivity and affinity to the observed environment state. This concentration is determined by a system of ordinary differential equations, which corresponds to the immune network dynamics introduced in [46]. Equation (13) is used to control the network model's dynamics in this work.

$$\frac{da_i(t+1)}{dt} = \left(\alpha \sum_{j=1}^{N} m_{ji} a_j(t) - \beta \sum_{j=1}^{N} m_{ik} a_k(t) + \gamma m_i - k \right) a_i(t) \tag{13}$$

where:

- N is the number of classifiers (antibodies) that composes the sub-population of classifiers dealing with a target.

- m_i is the affinity between classifier i and current stimuli (antigen).
- m_{ji} is the mutual stimulus coefficient of antibody j into classifier i.
- m_{ki} represents the inhibitory effect of classifier k into classifier i.
- k is the rate of natural death rate of classifier i;
- $a(t)$ is the bounded concentrations imposed on classifiers;
- the coefficients α, β and γ is weight factor that determines the significance of the individual terms.

4.5. Methods for Coordinated Learning and Knowledge Sharing

The training of classifiers is done through the conventional extended classifier system mechanisms and facilitated by the negative selection and immune network mechanism of the artificial immune system. Figure 3 shows the top-level view of the negative selection mechanism. The training and learning procedure are as follows.
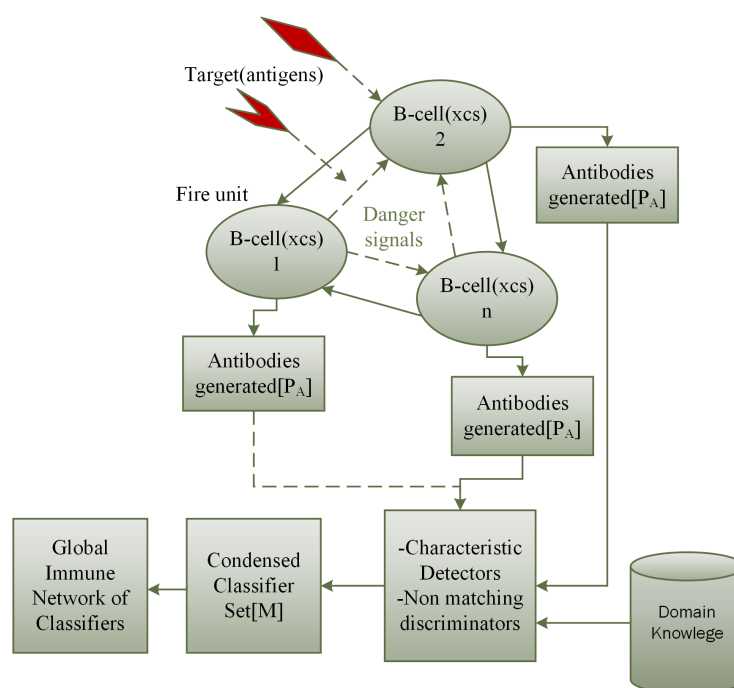

Figure 3. Negative selection Mechanism during action strategy selection.

1. Before the training begins each B-cells' population of individuals are initialized randomly. In this work, partial Pittsburgh-Style is adopted. That is, classifiers are treated both at the classifier level and as individuals. An individual is a collection of classifiers. Each classifier connection part is initialized as empty since the connections of a classifier are dynamically established during action selection. For each generation, an individual from each agent B-cells' population is selected and used to control the agent's actions. By initializing each population independently, a diverse population of classifier are generated collectively. Based on the characteristics detectors and quantized variables using the classifier encoding and representations above, the size of classifiers in a population can be determined a prior. However, the initial number of classifiers of an individual is chosen at the beginning of the training.

2. When individual B-cells generate their classifier sets based on the current environment, Non-matching and characteristics detectors in a form of rules obtained from an expert are used to filter out redundant classifiers in the matched set of the individual B-cells after merging to produce a condensed set of classifiers that undergo further processing for actions selection. In this case, manual rules are encoded to discriminates certain classifiers in the match set from being processed further. For example, if the status of a weapon is damage or a weapon has no ammunition remaining, all classifiers of that

particular weapon are filtered out. This forms a first phase of matching and message processing the system.

3. Next, the action selection mechanism in Section 4.4 is applied to obtain the weapon-target assignment of the agent. After executing the actions in the environment, the agent receives a reward based on the targets that were successfully intercepted as against the resources utilized. The reward obtained and concentration value of the classifier is used to update the fitness of classifiers within the individual under evaluation whose actions resulted in the reward. Also, in order to properly evaluate the classifiers of an individuals and their connections, each individual is simulated a predetermined number of times in each generation. While the immune network is applied each episode, genetic algorithm is applied on each extended classifier system at the individual level.

4. Finally, classifiers of individual with high accuracy are cloned and merged with other individuals of the same type. Two types of cloning are adopted: whole classifier cloning and merging and classifier action cloning, and replacement with higher accuracy classifiers' actions after all the individuals in the populations are evaluated.

5. Experimental Setup

In order to realize autonomous combat platforms operations, intelligent agents are required to interact with the battlefield environment during training. This is unrealistic in the physical environment as the process can be dangerous and costly. Hence, for facilitating the development of military intelligence, the physical environments are normally recreated in a virtual environment through modeling and simulation. To provide a realistic environment for training and evaluation, a scenario of digital battlefield is established in a real-time wargame, "Command: Modern Operations". The system generates battlefield data in real-time and log the damages of combat units after every engagement both in stages and end of engagement. In this scenario, there are two parties involved the ground combat units herein referred to as the ally (blue) faction who are the defending side and the air combat units herein referred to us the enemy (red) faction which are the invaders. The experiments setup for this work are described as follows.

5.1. Configuration of Ally Faction

The Ally faction consists of 7 combat ships. These include two (2) class A combat ships with 4 different weapons of ranges and kill probabilities, four (4) class B combat ships with two weapons types on board and one (1) carrier ship with 2 short ranges weapons on board. The number of weapon types of all the units, $M = 5$, the main difference between these weapons is the different interception range and probability of kill which is shown in Table 1. The per salvo is the number that can be fired in succession without delay. The subscripts for weapon types is the *ID* of the weapon used in classifiers.

Table 1. Information on weapons deployed on combat units.

Weapon Type	Min. Range (nm)	Max. Range (nm)	Per Salvo	Type
$Weapon_1$	2	80	1	missile
$Weapon_2$	2	21	1	missile
$Weapon_3$	0.2	4	1	missile
$Weapon_4$	0	1	5	Gun
$Weapon_5$	1	1.8	40	Gun

The air defense weapons are deployed on the seven combat ships in different defensive formation. The types and quantities of weapons equipped on each combat platform of the Ally faction are indicated in Table 2. Zero (0) means that the unit is not equipped with this type of weapon. Figure 4 depicts a typical confrontational setup of the digital battle

environment. The survival of middle and large ships is the most critical requirement of a successful defense in this work. In this defense formation, $Unit_7$ is the center ship, and the task of the entire formation is to intercept as many missiles as possible before the center warship is destroyed.

Table 2. The number of weapons equipped on each warship.

Combat Platform ID	$Weapon_1$	$Weapon_2$	$Weapon_3$	$Weapon_4$	$Weapon_5$
$Unit_1$	48	0	24	20	220
$Unit_2$	48	0	24	20	220
$Unit_3$	0	32	0	20	0
$Unit_4$	0	32	0	20	0
$Unit_5$	0	32	0	20	0
$Unit_6$	0	32	0	20	0
$Unit_7$	0	0	72	20	0
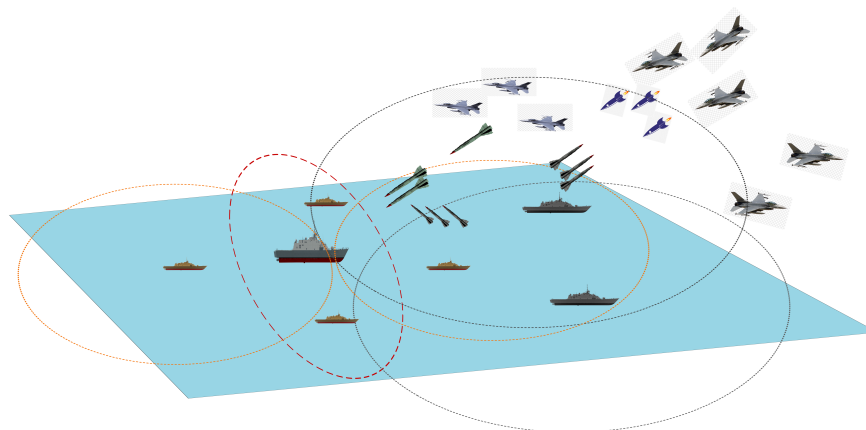
Figure 4. A typical air maritime combat digital battlefield scenario.

5.2. Reinforcement Program

In order for a learning system to improve its performance, reward function design is of great importance. Hence, much attention is given to designing the reward and punishment mechanisms employed in this work. In our case the destruction of $Unit_7$ is deemed as mission failure. However, the system must be encouraged to search for the best strategy that can enable it to succeed. So, we define two types of wins; Confident Win and Weak Win. The Ally Faction is said to win Confidently if more than two of its units survived in addition to $Unit_7$ after a confrontation in an episode. On the other hand, a weak win is encountered if at most 2 other units survived in addition to $Unit_7$. In addition to these specifics, each weapon of the ally and the enemy factions are assigned a cost and reward values respectively. The cost of losing a unit of the ally faction is function of the total number of ammunition on the unit at the time of destruction. However, the cost of losing $Unit_7$ is always higher in any case. Also, each of the 3 types of enemy units are assigned values which are rewarded to the ally faction when the enemy unit is destroyed. Table 3 shows the costs of weapons consumption. In addition, the interception of a bomber aircraft, fighter jet $T2$ and UAVs $T1$ rewards the Ally Faction with $0.2, 0.15$ and 0.1 respectively. For the weapons deployed by these units of the enemy faction the rewards for intercepting them are $0.25, 0.2$ and 0.15 respectively. Also, the Ally Faction is reward 0 when $Unit_7$ is destroyed and $+10$ when it survived. A squashed function is used to normalize the final rewards to a range between 0 and 1.

Table 3. Costs of Ally Faction's weapons consumption.

Weapon Type	Cost of Consumption
$Weapon_1$	0.004
$Weapon_2$	0.002
$Weapon_3$	0.001
$Weapon_4$	0.0001
$Weapon_5$	0.0001

5.3. Configuration of Enemy Faction

The Enemy faction consist of multiple hostile air units approaching the Ally faction in formations. A formation consists of a maximum of 8 fighter units and can comprise of different types of units. Each unit has a maximum of 8 ammunition of its weapon type that it can deploy on a defensive unit. However, a unit may run out of fuel and may need to return to based. Hence, not all weapons maybe deployed during its initial engagement. It is assumed that the enemy faction already knows the location of the ally faction, this means they only need to send out attackers. The enemy faction consists of different units comprising bomber aircraft, fighter jet $T2$ and unmanned combat aerial vehicles $T1$. The bombers have the shortest firing range and proximity hence must get closed enough to drop the bombs. In doing so the enemy unit is exposed to the firing range of the ally units. The other two types of units consist can also deploy short to medium range missiles.

5.4. Scenarios Setup

The engagements and battle rounds are design based on three scenarios as follows:

Scenario 1: In this scenario, the defense units know the approach direction of the enemy units. So, the defense units form a formation that ensures that the defended unit is put at the rear of the formation. Also, the attack in this scenario is executed in a single formation deployment. Each of the enemy units can fire only 3 of its ammunition before it must return to base for refueling and there is no second wave of attack.

Scenario 2: In this scenario, a formation is formed around the defended unit which is placed at the center. The enemy units can appear from multiple direction. The attack, in this case, is done on 3 confrontations. In each confrontation, the formations can consist of different units and each unit can fire as many weapons as possible within its maximum allowed capacity. Also, they attack the ally units on different rounds. Each round presents more challenges by increasing the number of enemy units and number of weapons fired.

Scenario 3: In the final scenario, the entire enemy units are divided into 4 groups of 8 each with each group approaching from the north, south, east and west of the ally faction. However, the formation of the ally faction is similar to scenario 2. Also, jamming is enabled in all scenarios except scenario one.

Each of these strategies are employed to train a different model independently from scratch. During training, parallel simulations were run on three different machines simultaneously for almost 2400 training instances (generations). The population of individuals was 20 for each classifier system and maximum number of classifiers in an individual was set to 1000 with Mutation probability = 0.15. The results of the 3 scenarios are shown in Figures 5–7. In each case, the score, win rate and rate of resource utilization is plotted as the training progresses. As can be seen, the win rate for the ally faction increases whiles the total quantity of weapons deployed reduces as training progresses. Except for scenario 3 which obtained a final win rate of around 50%, the other two models of scenario 2 and 3 attained a win rate of about 75%. Similarly, the resource utilization in all scenarios saw a reduction of at least 30% at end of training Figure 8. The comparison of the 3 models with regards to battle costs is presented in Figure 8d.
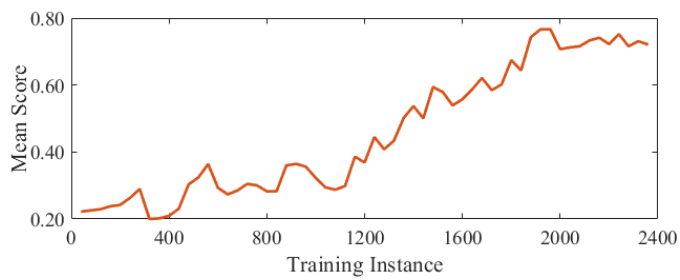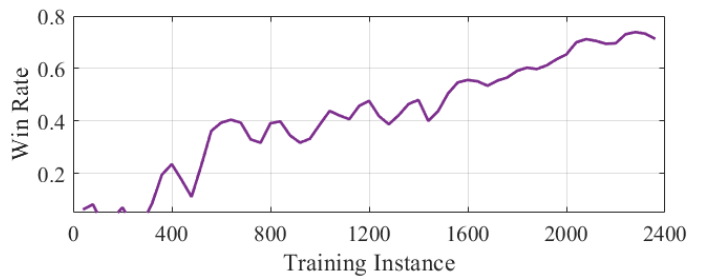
(**a**) Model training performance score
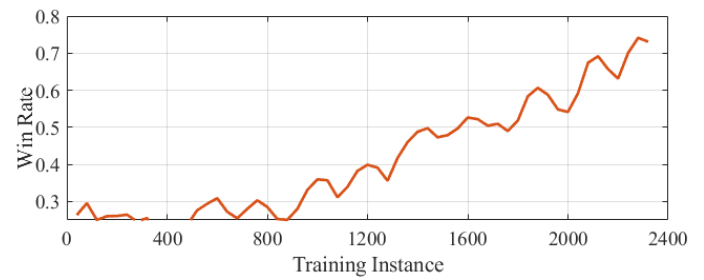
(**b**) Winning rate of model during training

Figure 5. Training performance on scenario 2.
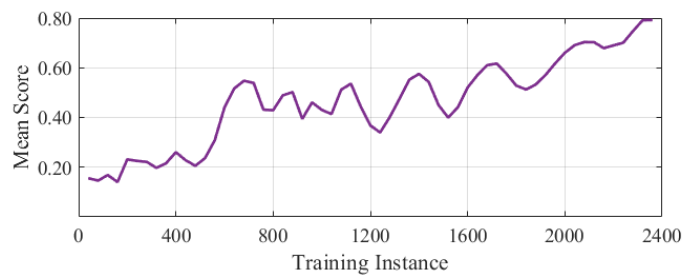
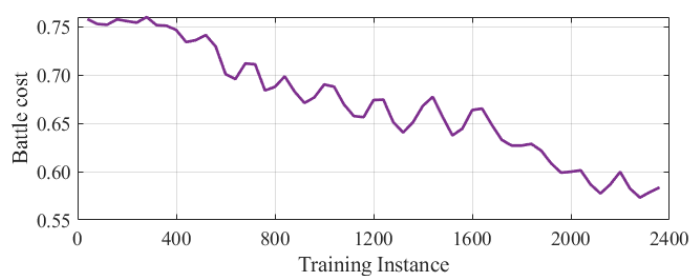(**a**) Model training performance score

(**b**) Winning rate of model during training

Figure 6. Training performance on scenario 1.
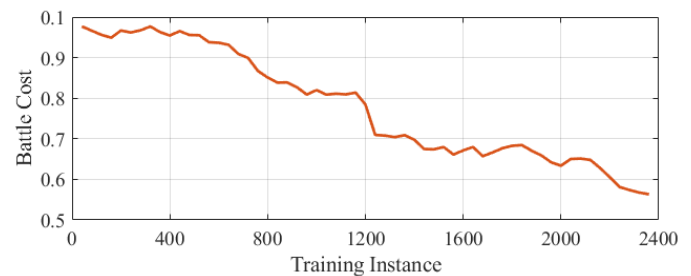

(**a**) Model training performance score
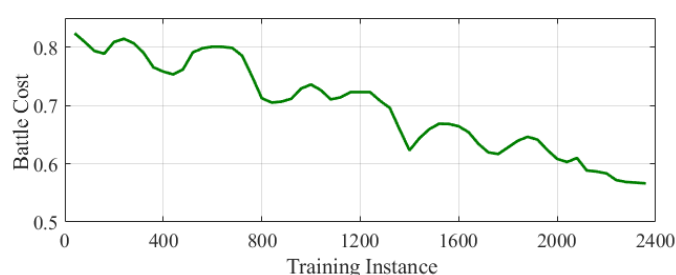
(**b**) Winning rate of model during training

Figure 7. Training performance on scenario 3.

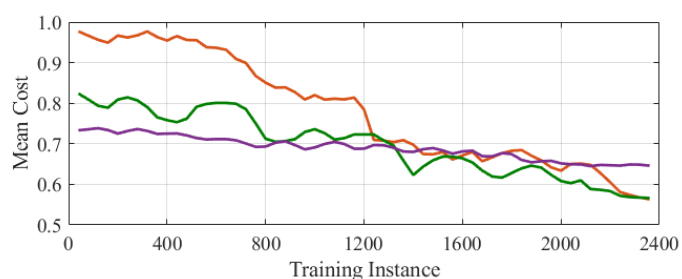

(**a**) Rate of resource consumption in scenario 1

(**b**) Rate of resource consumption in scenario 2

Figure 8. *Cont*.

(**c**) Rate of resource consumption in scenario 3   (**d**) Resource utilization of the 3 models during training

Figure 8. Individual and Compared resource utilization of the 3 scenarios models during training.

5.5. Baseline Heuristics

In order to evaluate the performance of the obtained strategies or models of the system, a heuristic assignment strategy was developed. This heuristic model employs multiple rules to govern the decision-making of the system to serve as a credible baseline to benchmark this work.

Hence, in the heuristic baseline,

1. Weapon-target assignment is performed based on priority.
2. To achieve priority-based assignment, the threat value of targets is evaluated based on the assigned targets values, heading and the distance to the $Unit_7$.
3. Based on the computed values, a sorting algorithm is used to sort the targets in ascending order. Targets with low computed values are considered to pose higher threat to the ally faction.
4. After the threat levels are determined and sorted, for each target we select the closest ally unit to attack it with any weapon within range. The number of ammunition to fire is set to a maximum of 4.
5. When intercepting targets, targets identified as weapons are intercept first. This is different from the threat level computation of targets. In other words, targets identified as weapons are assigned first based on their threat level before non-weapons are also assigned based on their threat levels.
6. Also, when a target is within multiple weapon range of a unit, the shortest ranged weapon is utilized.

To evaluate the final models with the baseline, we run the simulations for 30 rounds each. In this case, a round consists of 10 episodes of simulations. The results of the win rate and average costs are presented in Figure 9. From the results, it can be seen that even though model 2 took longer time to make a significant increase in win rate it turns our to perform well in all the scenarios. On the other, model 3 turns to perform better in the other scenarios than it performs in the trained scenario. The three models generalized well to scenarios they were not trained in. Table 4 shows the summary of the percentage wins in the final experiment conducted. Figure 10 shows a successful defense scenario run that employs model 1 of in the first scenario. In this case, the yellow colored units means that those units are not detected by any of the ally units.
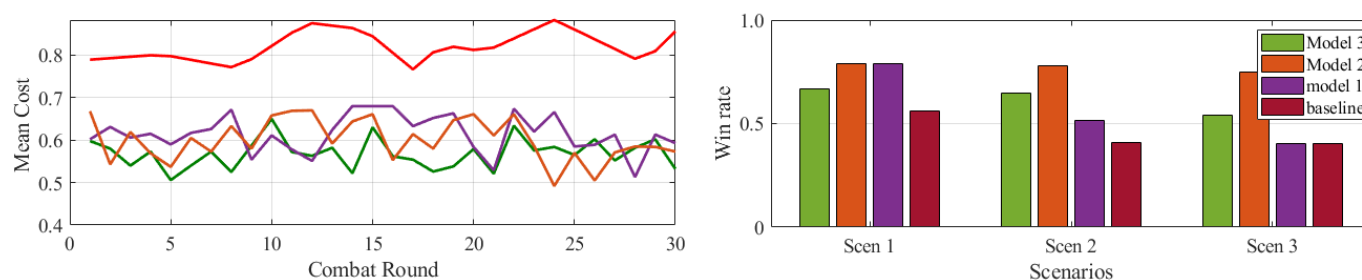
(**a**) Resource consumption for the 3 models and baseline model

(**b**) Win rate for the 3 models and baseline model

Figure 9. Comparison of performance of the 3 models and baseline model.

Table 4. Detail Performance of trained models.

Scenario/Model	Scenario 1 (% Win)			Scenario 2 (% Win)			Scenario 3 (% Win)		
	Confident	**Weak**	**Total**	**Confident**	**Weak**	**Total**	**Confident**	**Weak**	**Total**
Model 1	44	35	79	23.8	28.2	52	20.3	19.7	40
Model 2	40.3	38.8	79.1	44.7	33.4	78.1	39.3	35.8	75.1
Model 3	28.7	37.9	66.7	33.9	30.8	64.7	29.6	24.6	54.2
Baseline	26.6	29.7	56.3	21.5	19.5	41	19.3	21.3	40.6

(**a**) Battle state At the beginning of confrontation (t = 10 s)

(**b**) Battle state Mid-way during enemy confrontation (t = 20 s)

(**c**) Battle state towards the end of confrontation t = 30 s

(**d**) At the end of confrontation (t = 40 s)

Figure 10. A typical confrontation details of model 1 on scenario 1.

5.6. Discussion

From the results presented, the trained models show a more efficient use of resources. This can be realized from the models' winning rate and rate of resource consumption graphs of the respective models. In all cases, in addition to the ally faction winning mores as training progresses, it does so with less resources. The trained models show better

weapon distribution and target prioritization strategy compared with the baseline and target prioritization approach. At the initial stages of training, the system will assign weapons to targets that does not pose immediate danger to its survival. This behavior was also peculiar to the baseline heuristic model. However, at the end of the training process targets such as non-weapon units are ignored and only assigned in some instances when no weapon targets were detected. Figure 11 depicts the two situations as indicated. Also, as can be seen in Figure 11b, the model exhibits better cooperative behaviors as the threat level of targets increases. In this case, detected targets are distributed among suitable units for interceptions.
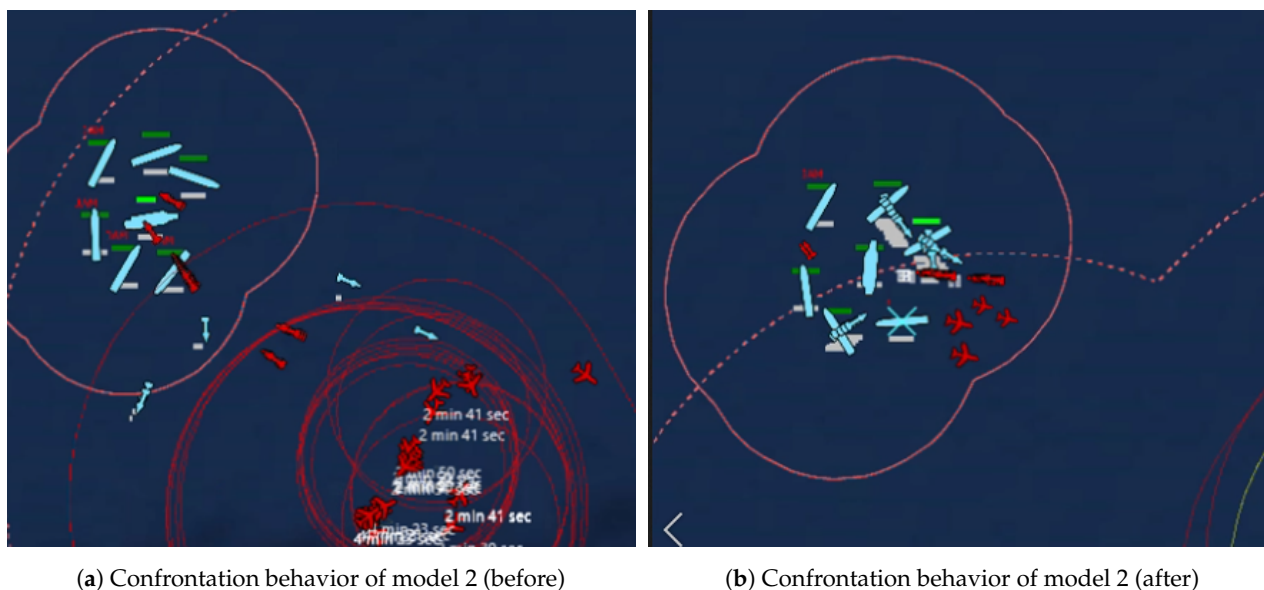

(**a**) Confrontation behavior of model 2 (before)   (**b**) Confrontation behavior of model 2 (after)

Figure 11. A typical confrontation details of model 2 before and after training.

In comparison with other approaches such as neural network based models, the system realized by this approach is interpretable and can be refined using prior knowledge. The learning process is facilitate by prior specification and quantization of fuzzy sets for classifier variables and the classifiers filtering introduced by the negative selection mechanism. From the data presented, the performances and winning rates appear to increase more sharply after 400 generations except for model 2. This might be attributed to the complexity level of the scenario model 2 was trained on. Even though model 2 turns out to be the most optimal of the 3. Therefore, even though more complex scenarios might take time for performance to start manifesting, the resulting controllers can achieve more optimal performances.

However, this approach also has its challenges, one of which is the memory consumption to keep track of classifiers and their connections. Also, the computations of classifiers concentrations with respect to other classifiers add additional computational burden on the approach as the size and number of classifiers increase.

6. Conclusions

In this paper, the air defense problem is studied. We explore the application of classifier system and artificial immune models for coordinated self-learning air defense. In particular, this paper presents an approach for multiple unmanned combat systems for coordinated air defense that implements autonomous cooperative threat evaluation and weapon assignment learning approach. We investigate and demonstrate the applicability of a hybrid artificial immune and learning classifiers system for realizing air defense intelligence and presents a hierarchical self-learning approach for multiple unmanned combat systems air defense operations that integrates artificial Immune based algorithms with classifier systems. We further design a mechanism to speed up learning by applying

a negative selection mechanism to filter out and condense situations from individual decision units that proposed solutions for the situations in the battlefield using domain knowledge. We evaluate the proposed approach by designing several simulation scenarios. The experimental results of training multiple combat systems demonstrate the learning ability of the approach. Finally the proposed scheme is compared with baseline heuristics by comparing the resource utilization and win rate of the different models obtained and the baseline. Based on the data presented, the most optimal model (model 2) was able to increase its win rate to about 75% and reduce its resource utilization by 30%. The approach is easily scalable since the addition of new units will only require the transfer of knowledge from similar decision units. The ability to transfer and encode prior knowledge facilitates transfer learning and prevent the system from learning from scratch when new units are added to the team. This will speed up the learning process of the system. The model is interpretable by humans compared to 'blackbox' approaches, which is a desirable feature for military applications. The generalization capability of the approach was validated by testing them in the scenarios they were not trained in.

However, more work is still required to improve the performance of the system and approach adopted in this work. For instance, there is a significant level of disparity between the physical world and the simulation environment. In the physical environment, there might not be an immediate confirmation of a successful interception. Hence, the performance in the real battlefield might be degraded to some extend. Also, it will be interesting to consider tuning or learning of the input variables and not just the appropriate consequent as adopted in this approach. In this regard, the learning of the input variables for each classifier system can either be done along with the rule base learning or one after the other. Similarly, further investigation as to how to identify useful knowledge, and condense and merge with other learning agents is needed. We plan to extend this framework to develop a general training system for autonomous decision-making of unmanned combat vehicles for cooperative missions.

Author Contributions: Conceptualization, Supervision and funding acquisition, Y.X.; methodology, software, validation, writing—original draft preparation, S.N.; software and validation, writing—review and editing W.R. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhang, B.; Zhang, M.; Song, Y.; Zhang, L. Combining Evidence Sources in Time Domain With Decision Maker's Preference on Time Sequence. *IEEE Access* **2019**, *7*, 174210–174218. [CrossRef]
2. Zhou, Y.; Tang, Y.; Zhao, X. A Novel Uncertainty Management Approach for Air Combat Situation Assessment Based on Improved Belief Entropy. *Entropy* **2019**, *21*, 495. [CrossRef]
3. Posen, B.R. Foreword: Military doctrine and the management of uncertainty. *J. Strateg. Stud.* **2016**, *39*, 159–173. [CrossRef]
4. Manne, A.S. A Target-Assignment Problem. *Oper. Res.* **1958**, *6*, 346–351. [CrossRef]
5. Karasakal, O. Air defense missile-target allocation models for a naval task group. *Comput. Oper. Res.* **2008**, *35*, 1759–1770. [CrossRef]
6. Bogdanowicz, Z.R. A new efficient algorithm for optimal assignment of smart weapons to targets. *Comput. Math. Appl.* **2009**, *58*, 1965–1969. [CrossRef]
7. Li, X.; Zhou, D.; Pan, Q.; Tang, Y.; Huang, J. Weapon-Target Assignment Problem by Multi-objective Evolutionary Algorithm Based on Decomposition. *Complexity* **2018**, *2018*, 8623051. [CrossRef]
8. Chang, Y.Z.; Li, Z.W.; Kou, Y.X.; Sun, Q.P.; Yang, H.Y.; Zhao, Z.Y. A New Approach to Weapon-Target Assignment in Cooperative Air Combat. *Math. Probl. Eng.* **2017**, *2017*, 2936279. [CrossRef]
9. Naseem, A.; Shah, S.T.H.; Khan, S.A.; Malik, A.W. Decision support system for optimum decision making process in threat evaluation and weapon assignment: Current status, challenges and future directions. *Annu. Rev. Control* **2017**, *43*, 169–187. [CrossRef]

10. Fu, Q.; Fan, C.L.; Song, Y.F.; Guo, X.K. Alpha C2—An Intelligent Air Defense Commander Independent of Human Decision-Making. *IEEE Access* **2020**, 8, 87504–87516. [CrossRef]

11. Li, Y.; Du, W.; Yang, P.; Wu, T.; Zhang, J.; Wu, D.; Perc, M. A Satisficing Conflict Resolution Approach for Multiple UAVs. *IEEE Internet Things J.* **2019**, *6*, 1866–1878. [CrossRef]

12. Pappas, G.J.; Tomlin, C.; Sastry, S.S. Conflict resolution for multi-agent hybrid systems. In Proceedings of the 35th IEEE Conference on Decision and Control, Kobe, Japan, 13 December 1996; Volume 2, pp. 1184–1189. [CrossRef]

13. Sislak, D.; Volf, P.; Pechoucek, M.; Suri, N. Automated Conflict Resolution Utilizing Probability Collectives Optimizer. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2011**, *41*, 365–375. [CrossRef]

14. Hunt, J.E.; Cooke, D.E. Learning using an artificial immune system. *J. Netw. Comput. Appl.* **1996**, *19*, 189–212. [CrossRef]

15. Jerne, N. Towards a network theory of the immune system. *Collect. Ann. Inst. Pasteur* **1974**, *125 C*, 373–389.

16. Burnet, F.M. *The Clonal Selection Theory of Acquired Immunity*; Vanderbilt University Press: Nashville, TN, USA, 1959; p. 232. Available online: https://www.biodiversitylibrary.org/bibliography/8281 (accessed on 10 December 2020).

17. Matzinger, P. The danger model: A renewed sense of self. *Science* **2002**, *296*, 301–305. [CrossRef]

18. Kong, X.; Liu, D.; Xiao, J.; Wang, C. A multi-agent optimal bidding strategy in microgrids based on artificial immune system. *Energy* **2019**, *189*, 116154. [CrossRef]

19. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*; MIT Press: Cambridge, MA, USA, 1992.

20. Cao, Y.; Ireson, N.; Bull, L.; Miles, R. *Design of a Traffic Junction Controller Using Classifier Systems and Fuzzy Logic*; Springer: Cham, Switzerland, 1999; pp. 342–353.

21. Rochner, F.; Prothmann, H.; Branke, J.; Müller-Schloer, C.; Schmeck, H. An Organic Architecture for Traffic Light Controllers. In Proceedings of the Informatik 2006—Informatik für Menschen, Dresden, Germany, 2–6 October 2006; pp. 120–127.

22. Gershoff, M.; Schulenburg, S. Collective behavior based hierarchical XCS. In Proceedings of the Tenth International Workshop on Learning Classifier Systems (IWLCS 2007), London, UK, 8 July 2007; pp. 2695–2700. [CrossRef]

23. Inoue, H.; Takadama, K.; Shimohara, K. Exploring XCS in multiagent environments. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2005, Washington, DC, USA, 25–26 June 2005; pp. 109–111. [CrossRef]

24. Wang, C.; Chen, H.; Yan, C.; Xiang, X. Reinforcement Learning with an Extended Classifier System in Zero-sum Markov Games. In Proceedings of the 2019 IEEE International Conference on Agents (ICA), Jinan, China, 18–21 October 2019. [CrossRef]

25. Bonarini, A.; Trianni, V. Learning fuzzy classifier systems for multi-agent coordination. Recent Advances in Genetic Fuzzy Systems. *Inf. Sci.* **2001**, *136*, 215–239. [CrossRef]

26. Bacardit, J.; Bernadó-Mansilla, E.; Butz, M. *Learning Classifier Systems: Looking Back and Glimpsing Ahead*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4998, pp. 1–21.

27. Stalph, P.; Butz, M.; Pedersen, G. *Controlling a Four Degree of Freedom Arm in 3D Using the XCSF Learning Classifier System*; Springer: Cham, Switzerland, 2009; pp. 193–200.

28. Studley, M.; Bull, L. X-TCS: Accuracy-based learning classifier system robotics. In Proceedings of the 2005 IEEE Congress on Evolutionary Computation, Edinburgh, UK, 2–5 September 2005; Volume 3, pp. 2099–2106.

29. Ernest, N. Genetic Fuzzy Trees for Intelligent Control of Unmanned Combat Aerial Vehicles. Ph.D. Thesis, University of Cincinnati, Cincinnati, OH, USA, 2015. [CrossRef]

30. Zhou, K.; Wei, R.; Xu, Z.; Zhang, Q.; Lu, H.; Zhang, G. An Air Combat Decision Learning System Based on a Brain-Like Cognitive Mechanism. *Cogn. Comput.* **2020**, *12*. [CrossRef]

31. Başpınar, B.; Koyuncu, E. Assessment of Aerial Combat Game via Optimization-Based Receding Horizon Control. *IEEE Access* **2020**, *8*, 35853–35863. [CrossRef]

32. Changqiang, H.; Kangsheng, D.; Hanqiao, H.; Shangqin, T.; Zhuoran, Z. Autonomous air combat maneuver decision using Bayesian inference and moving horizon optimization. *J. Syst. Eng. Electron.* **2018**, *29*, 86–97. [CrossRef]

33. Yang, Z.; Zhou, D.; Kong, W.; Piao, H.; Zhang, K.; Zhao, Y. Nondominated Maneuver Strategy Set With Tactical Requirements for a Fighter Against Missiles in a Dogfight. *IEEE Access* **2020**, *8*, 117298–117312. [CrossRef]

34. Ma, X.; Xia, L.; Zhao, Q. Air-Combat Strategy Using Deep Q-Learning. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 3952–3957. [CrossRef]

35. Han, W.; Zhang, B.; Wang, Q.; Luo, J.; Ran, W.; Xu, Y. A Multi-Agent Based Intelligent Training System for Unmanned Surface Vehicles. *Appl. Sci.* **2019**, *9*, 1089. [CrossRef]

36. Holmes, J.; Lanzi, P.L.; Stolzmann, W.; Wilson, S. Learning classifier systems: New models, successful applications. *Inf. Process. Lett.* **2002**, *82*, 23–30. [CrossRef]

37. Brownlee, J. Learning classifier systems. In *Encyclopedia of Machine Learning and Data Mining*; Springer: Cham, Switzerland, 2017.

38. Stein, A. Learning Classifier Systems: From Principles to Modern Systems. In Proceedings of the GECCO '19, Prague, Czech Republic, 13–17 July 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 747–769. [CrossRef]

39. Abbasi, M.; Yaghoobikia, M.; Rafiee, M.; Jolfaei, A.; Khosravi, M.R. Efficient resource management and workload allocation in fog–cloud computing paradigm in IoT using learning classifier systems. *Comput. Commun.* **2020**, *153*, 217–228. [CrossRef]

40. Liang, M.; Palado, G.; Browne, W.N. Identifying Simple Shapes to Classify the Big Picture. In Proceedings of the 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), Dunedin, New Zealand, 2–4 December 2019; pp. 1–6. [CrossRef]

41. Nordsieck, R.; Heider, M.; Angerer, A.; Hähner, J. Towards Automated Parameter Optimisation of Machinery by Persisting Expert Knowledge. In Proceedings of the ICINCO 2019, Prague, Czech Republic, 29–31 July 2019.
42. Butz, M.; Wilson, S. An Algorithmic Description of XCS. *Soft Comput. Fusion Found. Methodol. Appl.* **2001**, *6*. [CrossRef]
43. De Castro, L.; Timmis, J. *Artificial Immune Systems: A New Computational Intelligence Approach*; Springer: Cham, Switzerland, 2002.
44. De Castro, L.; Von Zuben, F. The Clonal Selection Algorithm with Engineering Applications. *Artif. Immune Syst.* **2001**, *8*. [CrossRef]
45. Aickelin, U.; Cayzer, S. The Danger Theory and Its Application to Artificial Immune Systems. *SSRN Electron. J.* **2008**. [CrossRef]
46. Farmer, J.; Packard, N.H.; Perelson, A.S. The immune system, adaptation, and machine learning. *Phys. D Nonlinear Phenom.* **1986**, *22*, 187–204. [CrossRef]