

---

# Intelligent Arabic Writer Identification Using Artificial Immune System Algorithms: A Bio-Inspired Approach for Smart Pattern Recognition

Fahad Ghabban<sup>1,\*</sup>

<sup>1</sup>College of Computer Science and Engineering, Taibah University, Medina, 41477, Saudi Arabia

Email: [fghaban@taibahu.edu.sa](mailto:fghaban@taibahu.edu.sa)

## Abstract

Artificial immune systems (AIS) represent an emerging facet of artificial intelligence, offering innovative solutions to a spectrum of problems. It draws inspiration from the biological immune system's fascinating properties, mechanisms, and principles, resulting in mathematical and computer-based implementations. In this paper, we aim to assess the accuracy of artificial immune systems as classification tools in the realm of Arabic handwriting recognition. Among the repertoire of immune-computing models, we focus on the Artificial Immune Recognition System (AIRS), Immunos, Clonal Selection Algorithm (CLONALG), and Clonal Selection Classification Algorithm (CSCA), which have garnered significant attention for their prowess in pattern recognition applications. To conduct this investigation, we leverage the comprehensive IFN-INIT Arabic handwriting database, which comprises contributions from 411 distinct writers. Feature selection plays a pivotal role in enhancing classification performance, and for this purpose, we harness the grey level co-occurrence matrix. In pursuit of a thorough comparative analysis, we also employ well-established classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Naive Bayes. The obtained results exhibit the promising potential of AIS-based classifiers in the context of Arabic handwriting recognition, offering insights into the evolving landscape of AI solutions in this domain.

Received: January 31, 2025 Revised: April 21, 2025 Accepted: June 30, 2025

**Keywords:** Handwritten analysis; Arabic Writer identification; Artificial Immune Systems; Image processing; Pattern Recognition

## 1. Introduction

Handwriting behaviour, which forms part of the writer identification analytics, has huge potential in pattern recognition. This is done by comparing a set of unknown handwriting text with database of known people, which is similar to solving a jigsaw puzzle. Including writer identification can be used in classification of historical documents, identification in hand writing recognition systems, in biometric identification, and in forensic document examination. It is possible to distinguish two main categories of recognising approaches text-independent and text-dependent methods. Text-independent methods identify writers irrespective of content, useful in forensic investigations. Text-dependent methods require identical content, e.g., signature verification. Writer identification can also be online or offline, with the former analysing dynamic elements and the latter relying solely on written text. Over the decades, research in writer identification has focused on diverse features. While traditional studies introduced novel features with traditional classifiers, exploring innovative classification techniques remains enticing. Bio-inspired classification, particularly Artificial Immune Systems (AIS), holds potential for substantial contributions to writer identification. AIS's memorisation, learning, and recognition align well with writer identification needs.

Despite AIS's success in various domains, its application in writer recognition is limited. This paper aims to objectively validate widely used immunological models in pattern recognition, focusing on writer identification. CLONALG, CSCA, AIRS and Immunos. We propose a number of immune classifier models and perform Arabic Writer Identification tasks using these models and comparing them with some traditional classification methods. The methods of feature selection are examined to achieve the set of computational efficiency and improved performance.

The elaborate discussion of the different aspects within the scope of identification of Arabic author's offline is seamlessly presented in the rest of the paper. The literature review in Section 2 is highly detailed with the author going into the specific contribution of each work.

In the course of the story, being Section 3, Theoretical Foundations, an in-depth discussion is able to be had concerning the foundation of AIS and the Clonal Selection Algorithm. The article makes a masterstroke in the revelation of the complexities involved in the approach suggested to the Arabic Writer Identification. This section is the core of the study in which the theoretical backgrounds meet the practice providing a clear guide on how further analyses can be conducted. The section 4, known as, "Feature Extraction," retains its own self and gives a detailed look into how feature extraction uses the techniques they use.

As the story advances to Section 5, Experiments and Results, there is a critical description of the results and findings of the experiments made.

Finally, the concluding Section 6, Conclusion and future trends continues to retain the leading position by having the essential summary of findings and suggestions as to prospective directions of research. The reflective conclusion on the implications of the research provides the grounds upon which further developments in the direction of offline Arabic writer identification can be realized.

## **2. Literature Review**

### **2.1 Arabic writer identification: Related Works**

Arabic Writer Identification is one unique area of study in handwriting identification and started its adventure in 2005. One of the first attempts was made by Al Zoubeidy et al. [1] who identified 92.8 percent of texts on a database that contained 20 writers who each provided 25 samples of Arabic text. The approach to their methodology involved the global approach involving the use of grey-level co-occurrence matrices and multi-channel Gabor filters. Further, Gazzah et al. [2] combined local and global characteristic sets, whereas the local features were derived by the morphological variation of writing consisting of ascender inclination and line height. The lifting scheme was used to acquire global characteristics. They used multilayer perceptron classifiers on database of 180 samples of text by the members of 60 different writers and achieved the average accurate identification rate of 95.68%. In [3], genetic algorithms together with support vector machine methods have been used to derive discriminant characteristics. This research conducted on four classifiers, that is, a linear discriminant classifier (LDC), SVM, KNN and weighted Euclidean distance (WED) had an identification rate of 90% in a dataset with 20 writers. In other words, the utilisation of co-occurrence matrices and Gabor filtering [4] extended to Farsi handwriting, achieving an identification rate of 88% across 25 writers [5]. Feddaoui et al. introduced a method in [6] capable of identifying writers and distinguishing between printed and handwritten texts. Their approach employed a set of 16 Gabor filters to analyse writing texture. It was tested using weighted Euclidean distance on two databases: one containing handwritten text samples from 22 individuals and the other housing printed texts featuring 20 different character fonts. An innovative approach in [7] combined allographic with textural features and was evaluated on the IFN-ENIT database [8], resulting in an identification rate of 84%. Authors in their study [9] used Gabor energy-based and moment-based features to come up with a text dependent scheme of calculating feature, where they achieved a Top-1 recognition rate of 75 percent and Top-5 rate of 97.5 percent on a dataset consisting of 40 writers. A text-dependent identification scheme was thus presented in [10], which involved the use of moment invariants as well as edge-based directional probability distributions during the feature extraction phase. These experiments carried out on a database of 100 writers reported Top-10 identification rates at 90%. Fractal and multi-fractal analysis in [11], Chaabouni et al. investigated a new use of fractal and multi-fractal processing to text-dependent writer identification. They used Diffusion Limited Aggregates (DLA) in computing multi-fractal dimensions and fractal dimension using the Box-counting method. The classification was performed with a k-nearest neighbour approach, which has good prospects of identification. Abdi et al. [12] used six feature vectors built using the minimum perimeter polygon (MPP) contours of Arabic words where probability distribution functions direction, length, curvature measurements and angle analysis were used in computing feature vectors. They achieved an identification rate of 90% on 82 writers from the IFN-ENIT database through the Borda ranking algorithm. Some authors, as seen in [13], extended their writer identification processes to samples of Arabic handwritten digits. They employed a fusion of concavity features, horizontal and vertical run lengths [14], [15], curvature, gradient, and density for writer characterisation, achieving an identification rate of 88% using a k-nearest neighbour classifier. Djeddi et al. [16] proposed a comprehensive method for text-independent writer

verification and identification. Their approach relied on a set of features extracted from Grey Level Run Length (GLRL) Matrices. Identification was performed using nearest-neighbour classification, with evaluation conducted on 650 handwritten documents from 130 different individuals, yielding promising results. In [17], Jin et al. introduced a method for detecting and removing ruling lines from handwritten documents. Their study highlighted the substantial improvements such an approach could bring to Arabic writer identification performance.

## **2.2 AIS for pattern recognition & classification**

The natural immune system, a remarkable and intricate mechanism, serves as the body's defence against foreign pathogens. It possesses the either extraordinary ability to distinguish molecules within the body as self-cells or non-self-cells. Researchers in the field of artificial intelligence recognise the biological immune system as a source of inspiration, owing to its adaptability, memory mechanisms, learning capabilities, and recognition prowess, which can be applied to a wide range of computational tasks and the solution of complex problems [18]. While replicating the entirety of the natural immune system in an artificial process is challenging, researchers have successfully simulated its relevant functions for artificial intelligence tasks. This endeavour gave rise to the field of "artificial immune systems" (AIS).

AIS are computational and evolutionary algorithms stimulated by the mechanisms and progressions of the natural immune system, particularly the differentiation of self (components belonging to the host organism) from potential non-self (pathogens). Dasgupta defined AIS as a collection of intelligent methodologies inspired by the natural immune system, designed to solve real-world problems [19]. Timmis and De Castro, on the other hand, view AIS as adaptive systems inspired by immunological theories, observations of immune functions, and immunological models, which can be applied to problem solving. A seminal reference in this field is the book edited by de Castro and Timmis [20].

In recent years, AIS has made significant inroads in various domains of artificial intelligence, piquing the interest of researchers. The appeal of natural mechanisms for recognition, identification, and post-processing, which underpin the human body's immune functions, has given rise to new ideas in learning, pattern recognition, and communication [21]. Some applications of AIS include handwritten character recognition based on the clonal selection principle [22], a negative selection-based system for speaker recognition [23], audio content-based retrieval [24], English grammar checking [25], and text categorisation [26]. Similarly, the clonal selection algorithm (CLONALG) has been applied to shape recognition [27]. More recently, CLONCLASS, an improved version of CLONALG, has been successfully used in handwritten word recognition [28]. Artificial Immune Recognition Systems (AIRS) have found success in various pattern recognition problems, such as music genre classification [29], genre and author detection in Turkish manuscripts [30], DNA microarray data analysis [31], and semantic document classification [32]. Furthermore, in [33], [34], a negative selection theory-based system was proposed for writer identification using a small dataset of 200 handwritten words contributed by 10 writers.

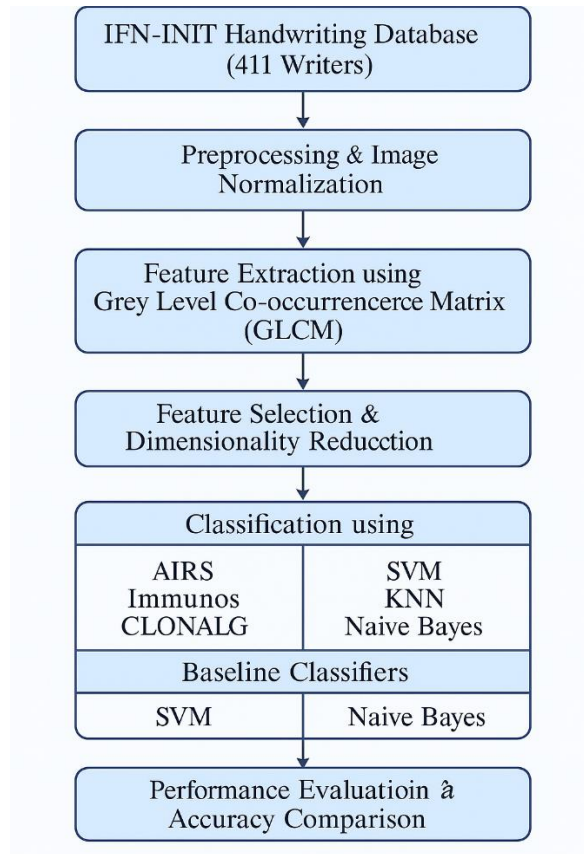
Our motivation for this study stems from the noticeable gap in research that explores the use of the immune system metaphor in handwriting-based writer recognition. While there have been numerous investigations into the application of immune mechanisms in other areas, such as classification and detection, there remains a limited body of work that explores the potential of immune-inspired techniques for Arabic manuscript-based writer recognition.

In the following sections, we will introduce the key learning algorithms that constitute the core components of an artificial immune system, including antigens, antibodies, and memory B cells.

## **3. Theoretical Foundations**

AIS signify a fascinating paradigm in the realm of computational intelligence, drawing inspiration from the intricate and adaptive mechanisms inherent in the natural immune system. This part is a deeper look at the theoretical background of the AIS, its development as a computational algorithm and evolutionary algorithm. Then we discuss at length a young model in AIS the CLONALG, based on the clonal selection theory.

AIS as a discipline is an attempt to take advantage of the highly successful biological immune system adaptive capabilities, memory and recognition skills to solve complex computer-related problems and to cope with the real world. This type of exploration about AIS has resulted in the birth of some great learning algorithms that can be applied in various forms such as pattern recognition, classification among others. In our discussion below, we shall undo the mysteries of the fundamental elements included in AIS, namely antigens, antibodies and memory B cells, throwing their light into the functions regarding pattern recognition and grouping. Moreover, we shall have a close look at the CLONALG algorithm, breaking it down to its component parts of initialisation and classification and how it uses the clonal selection theory to achieve so. Figure 1 demonstrates an overview of the whole methodology, which the present study has followed.



**Figure 1.** Overview of the proposed method

### 3.1 Clonal Selection Algorithm (CLONALG)

The CLONALG is an early method in AIS, based on the biological theory of clonal selection, which is a model of the adaptive immune response of B-cells. CLONALG models the selection, cloning and maturation in response to foreign antigens, and it is highly useful in the classification and optimization problems [20]. The algorithm has four main steps:

**Initialization:** The initialization is initiated by the creation of an initial antibody population of size  $N$ , which is separated into two different subsets: the RAP and the MAP. The RAP is a working population, the MAP is the best-performing antibodies, and it acts as the core classifier.

**Training and Affinity Calculation of Antibodies:** Every antibody is subjected to antigens through  $G$  generations. The similarity between an antibody and an antigen is computed by a distance measure-usually Euclidean distance, which is the distance between two points in a Euclidean space:

$$Affinity(A_i, A_g) = \sqrt{\sum_{j=1}^d (A_{ij} - A_{gj})^2} \quad (1)$$

where  $A_i$  is the  $i^{th}$  antibody,  $A_g$  is the antigen, and  $d$  is the dimensionality of the feature space. Based on affinity scores, a selection of high-affinity antibodies undergoes cloning. The clonal expansion is governed by a user-defined clonal factor  $F$ , producing  $n_i$  clones for each selected antibody  $A_i$  as:

$$n_i = \left\lfloor F \cdot \frac{1}{Affinity(A_i, A_g)} \right\rfloor \quad (2)$$

This ensures that better-matching antibodies produce more clones, reflecting biological principles of immune response.

**Clones Evaluation and Replacement:** The generated clones undergo mutation (or hypermutation), producing variations evaluated against the antigen. Their affinities are recalculated, and only the clones with the highest affinity are maintained. These better clones are then thought of to replace into the Memory Antibody Pool (MAP) pushing out members with lower affinity. This substitution enhances the memory of the system and improves the classification ability as generations pass by.

Classification: Following training, the MAP works as the CLONALG classifier. For an unknown antigen, the classification is done by comparison with memory antibodies. Majority voting mechanism is used: each memory antibody vote based on its class label, and the majority determines the final classification.

Parameter Tuning: The performance of CLONALG is very sensitive to a number of user-defined parameters, such as the number of generations  $G$ , the clonal factor  $F$  and the sizes of the antibody pools (RAP and MAP). These parameters require careful calibration in order to perform optimally in the process of classification.

### 3.2 Clonal Selection Classification Algorithm (CSCA)

The CSCA is an extension of the CLONALG and improves its effectiveness in the classification tasks by adding mechanisms to reduce misclassifications and to optimize the learning performance [35]. It brings a more structured training pipeline that enhances antibody diversity and model generalization.

The algorithm starts by initialisation stage where a set number of antibodies are randomly created. These antibodies are possible classifiers and they have class labels. The training phase entails a number of important steps namely: selection of high-affinity antibodies, culling of ineffective ones, cloning of selected antibodies, mutation of clones to add diversity, and a final pruning step to leave only the most pertinent classifiers. This systematic training process guarantees that the memory pool develops in a more precise and representative antibodies.

Classification: Once training is complete, classification of a new antigen  $A_{g_{new}}$  is performed using the trained memory antibodies. Affinities are computed as before, and a majority vote mechanism assigns the class label:

$$\hat{y} = mode(\{y_i | A_i \in MAP, Affinity(A_i, A_{g_{new}}) < \delta\}) \quad (3)$$

where  $y_i$  is the class label of antibody  $A_i$ , and  $\delta$  is an affinity threshold.

#### Immunos Algorithm

Immunos algorithm is a biologically inspired classifier, which models the cooperation of T-cell and B-cell where B-cells learn to respond to class specific antigens and are influenced by the T-cell signals. The improved version, Immunos-99, adds cloning and mutation, so that the classifier becomes more adaptive [36].

Initialization: class labels divide the training dataset. For each class  $C_k$ , a seed population of B-cells  $B^k = \{B_1^k, B_1^k, \dots, B_{n_k}^k\}$  is created. The Seed Population Percentage (SPP) determines the number of initial B-cells:

$$n_k = \lfloor SPP \cdot N_{train}^{(k)} \rfloor \quad (4)$$

where  $N_{train}^{(k)}$  is the number of training samples in class  $C_k$ .

B-cell Training: Each B-cell in group  $B^k$  is exposed to antigens from class  $C_k$ . Affinity is computed, and based on it, clones are generated using a rule similar to CSCA. Hypermutation is again applied with mutation rate inversely related to affinity. Fitness of a B-cell  $B_i^k$  is defined as:

$$Fitness(B_i^k) = \frac{1}{1 + Affinity(B_i^k, A_g)} \quad (5)$$

B-cells with low fitness are pruned, and the best-performing ones are retained for the next generation. This process is repeated for Total Generations (TG).

Classification: When an unknown antigen  $A_{g_{test}}$  is presented, each B-cell group  $B^k$  computes an Avidity Index to determine its match with the antigen. The avidity index for class  $x=C_k$  is computed as the average fitness of the top  $p$  % B-cells:

$$Avidity = 1(C_k) = \frac{1}{p} \sum_i^p Fitness(B_i^k) \quad (6)$$

The final classification is given by the class with the highest avidity:

$$\hat{y} = arg \max_k (Avidity(C_k)) \quad (7)$$

### 3.3 The Artificial Immune Recognition System (AIRS)

The Artificial Immune Recognition System (AIRS), introduced by Watkins [37], is an immune-inspired supervised learning algorithm that mimics the adaptive behaviour of the biological immune system. It has since evolved into more advanced variants, including AIRS2 [38] and AIRS2-Parallel [39], to enhance performance and scalability. The algorithm operates through several key stages: initialisation, antigen training, memory cell identification competition, memory cell construction, candidate memory cell development, and ultimately data classification.

Antigen-Antibody Matching and Affinity Calculation AIRS uses an antigen-antibody metaphor where: Antigens represent input data instances. Antibodies or memory cells represent learned solution prototypes. The affinity between an antigen  $Ag$  and an antibody  $Ab$  is a measure of similarity, typically computed using Euclidean distance:

$$Affinity(Ag, Ab) = \sqrt{\sum_{i=1}^n (A_{gi} - Ab_i)^2} \quad (8)$$

Alternatively, in some implementations, cosine similarity may be used. A lower affinity score indicates a better match.

Stimulation Calculation The stimulation of a memory cell is calculated as an inverse function of the affinity, normalised by a maximum affinity value:

$$Stimulation(Ag, Ab) = 1 - \frac{Affinity(Ag, Ab)}{MaxAffinity} \quad (9)$$

This ensures stimulation values lie in the range [0,1], with higher values representing stronger activation.

Clonal Expansion and Mutation. Each candidate memory cell is cloned based on its stimulation level:

$$N_{clones}(C_i) = \alpha \cdot Stimulation(C_i) \quad (10)$$

Where:  $N_{clones}(C_i)$  is the number of clones for the candidate cell  $C_i$   $\alpha$  is a cloning factor

Mutation introduces variability and is inversely related to stimulation:

$$MutationRate(C_i) = \beta \cdot (1 - Stimulation(C_i)) \quad (11)$$

Where  $\beta$  is the mutation rate constant. Highly stimulated cells undergo fewer mutations, preserving their structure.

Artificial Recognition Ball (ARB) AIRS utilises the concept of the Artificial Recognition Ball (ARB) to model the recognition range of memory cells. Each ARB is a hypersphere defined by a centre (memory cell) and a radius  $r$ . An antigen  $x$  is said to be recognised by an ARB if:

$$Affinity(x, MC) \leq r \quad (12)$$

Where  $MC$  is a memory cell. This mechanism simulates the antigen-antibody binding process, essential for pattern matching.

Classification Using k-Nearest Neighbours. In the classification phase, AIRS leverages the k-Nearest Neighbour (k-NN) approach. Given an input, the k most activated memory cells are chosen and the class label of the output of the K most activated cells is the end class label with a majority vote:

$$\hat{y} = majority\_vote(\{y_i\}_{i=1}^k) \quad (13)$$

Where  $y_i$  is the class label of the  $i^{th}$  most stimulated memory cell.

Use in writing for Arabic Writer Identification the presence of the biological motif in AIRS, specifically the learning ability, flexibility, and memory creation, offers the applicability of the method in complex pattern recognition activities (e.g. Arabic writer identification). Through good modelling of feature similarities and competitive learning, AIRS-based models offer good performance in the ability of recognizing slight variation of handwriting.

#### 4. Feature Extraction

This is one of the key aspects in writer identification since feature extraction algorithm reduces the complexity of handwritten scripts into patterns that can be used to distinguish the writer. In Arabic hand writing recognition, feature extraction is very critical since there are many intricacies in the script to be extracted and such features are very vital to define features and segment difference between one writer and another. A properly designed feature extraction scheme minimizes the dimension of data, improves the performance of the classifier and helps in the correct identification of the script authors.

The modern literature places several features to examine handwritings, but a high percentage of them are structured to study western and Chinese writing, and they are far much different than the complex Arabic script that is being examined in this case. Letter infusion, connection, diacritics and variants of vowels and other similarities in Arabic language make it sophisticated to analyse. This research pursues a global analysis strategy; here Arabic handwriting is regarded as a texture image and the extraction of characteristics with the help of the Grey Level Co-occurrence Matrix (GLCM) is performed [40].

The choice of GLCM is grounded in several justifications, collectively enhancing Arabic writer identification:

- **Texture-Based Approach for Arabic Script:** GLCM aligns with a holistic texture-based approach, capturing the underlying textural characteristics of Arabic handwriting, bypassing the need to dissect individual characters.
- **Spatial Distribution of Textural Patterns:** GLCM excels in quantifying the spatial distribution of textural patterns, crucial for Arabic script, where character arrangement conveys valuable information about the writer's style.
- **Quantitative Discrimination of Textures:** GLCM estimates second-order statistics, facilitating discrimination between different textures, providing a robust foundation for identifying distinct script authors.
- **Adaptability to Binary Images:** GLCM is efficient for binary handwriting images, ensuring compatibility with the nature of images common in document analysis.
- **Multi-Scale and Multi-Directional Analysis:** The ability to construct GLCM at different distances and directions allows performing a thorough examination of textural characteristics, which can hold the shapes of local letters and general features of page layout.

In summary, the choice of GLCM is deliberate, addressing the intricacies of Arabic script. Its adaptability, capacity for discriminating features, and multi-scale analysis make it a judicious choice for writer identification in this study [41].

## 5. Experimental results and discussion

In this section, we shall first present the database used in Section 10.1 then the results shall be presented and analysed in Section 10.2.

### 5.1 Database

To perform our experiments, we used Arabic handwriting that is the only publicly known database of handwriting in Arabic language IFN/ENIT [8]. This database consists of forms with written in the hand as Arabic town and village names collected by 411 participants. The given information is in form of binary images with a resolution of 300 dots per inch (dpi) and the vast majority of writers completed 5 forms altogether.

The IFN/ENIT dataset [8] is renowned and extensively employed in the field of handwriting recognition. It has been used in previous research [7], [12], and [16] for tasks such as writer identification, owing to the inclusion of recorded writer information. Our experimentation involved the extraction of handwriting from the scanned forms. The content of the text varies, and the samples encompass a imperfect amount of handwriting, featuring only 12 names and 12 postal codes for Tunisian towns and villages. In the context of writer identification experiments, we worked with data from 130 writers, each having 5 samples. Figure 2 shows a sample of the dataset used for the evaluation.

3124	قمر التمش
8069	حليبة الشنة
4160	بن قردان
4237	كلمين
7072	منزل بوزوية السحاب
9030	نهره مدين
5154	المحارزة 18
1039	حوض السابن
6076	سيدي التوايتي
1215	تند بنت
9176	الرضاء
1005	الدمران

Figure 2. A sample of the used dataset.

## 5.2 Results and Analysis

To evaluate the performance of the proposed classifiers for the writer identification task, the dataset was partitioned into two subsets. The first subset, comprising 390 handwritten documents (the initial three samples from each writer), was used as the training set. The second subset, consisting of 260 handwritten documents (the remaining two samples per writer), was designated as the testing set.

In our initial approach, we employed three AIRS algorithms, namely AIRS1, AIRS2, and AIRS2-Parallel, using all 640 features. The results, as shown in Table 1, provide a comprehensive overview of the performance accuracy of various AIS-based classifiers utilising all GLCM extracted features. These results were assessed using two distinct distance metrics, Euclidean and Cityblock. The outcomes yield valuable insights into the efficiency of each classifier concerning Arabic writer identification.

Among the classifiers that come out such an analysis is the CSCA, which scored very well in relation to accuracy in both non-parametric distance measures. In particular, it achieved the accuracy of 95.38% and the Euclidean distance measure and a marginally less great reputation of 95.00% and the Cityblock distance measure. These findings highlight the strength of CSCA in the characteristic of capturing unique writer-specific patterns in the GLCM Data set.

**Table 1:** AIS-based classifiers' performance accuracy using all the features extracted from GLCM)

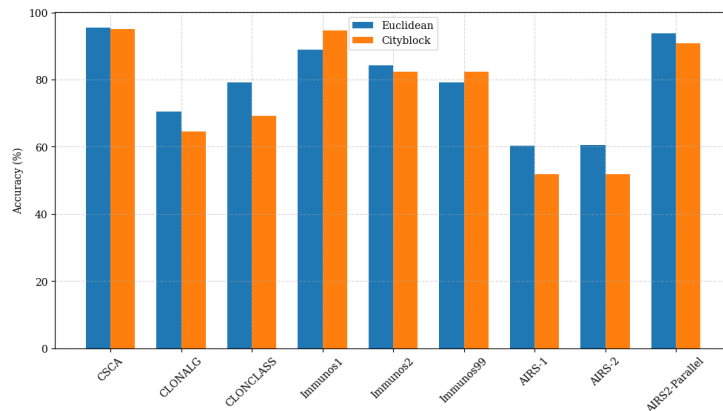
Classifiers/ Distance	Euclidean	Cityblock
CSCA	95.38%	95.00%
CLONALG	70.38%	64.61%
CLONCLASS	79.23%	69.23%
Immunos1	88.85%	94.61%
Immunos2	84.23%	82.31%
Immunos99	79.23%	82.31%
AIRS-1	60.38%	51.92%
AIRS-2	60.42%	51.92%
AIRS2-Parallel	93.85%	90.77%

Comparatively, Clonal Selection Algorithm (CLONALG) produced results of lower exceedingly accuracy. It had an accuracy of 70.38 percent with Euclidean metric and that of 64.61 percent with Cityblock metric. This result indicates the possible differences in the adaptation and evolution of the population of antibodies through the application of CLONALG and these differences may result in different performance with CSCA. The Clonal Selection Classifier (CLONCLASS) was not the best in terms of performance but it had a fair level of accuracy. When using the Euclidean measure it achieved the accuracy of 79.23 but when using Cityblock measure it got the accuracy of 69.23, which was a little bit lower. The differences in the accuracy might be explained by the sensitivity of the features to types of metrics. The Immunos algorithm demonstrated good accuracy outcomes, and especially shows reliable results when the Cityblock measure is used since an accuracy rate of 94.61 can be achieved. Even on Euclidean metric, Immunos performed quite well and achieved an accuracy of 88.85%. These results support the effectiveness of Immunos at detecting characteristics of distinctive patterns in the GLCM properties, and thus, it is a good selection as far as this is concerned. In contrast, the variants of the AIRS, both, AIRS-1 and AIRS-2, showed relatively reduced level of accuracy with each measure of distance. AIRS-1 and AIRS-2 have the same level of accuracy, but with the Euclidean metric, this accuracy is a bit higher and is around 60 percent. In the Cityblock metric, accuracy did not very much, which was about 51.92%. These results indicate that their peculiar adaptations and antibodies selection approaches might not suit this its specific writer identification task as well.

It is quite interesting that the best, both based on Euclidean and Cityblock, score was the Artificial Immune Recognition System - Parallel (AIRS2-Parallel) system. It also showed high levels of accuracy with 93.85 % accuracy with Euclidean metric and 90.77 % accuracy with city block metric. These findings, which focus on the

immune algorithm because of parallelism, point out that running its operations in parallel is a major contribute to determining and categorising writers using GLCM characteristics.

As a conclusion of this, these findings imply that the AIS-based classifiers differ in performance in recognizing the writers. CSCA, Immunos and AIRS2-Parallel are the best contenders that tend to prove themselves at the given task. These disparities in the levels of accuracy are indicative of the individual strengths and weakness of these algorithms and provide insight on further research in optimising AIS-based classifier to achieve greater accuracy of identifying writers.



**Figure 3.** AIS-Based Classifiers Accuracy

Figure 3 indicates the result of variety of AIS-based classifiers involving the use of all GLCM features in identification of Arabic writers. The level of accuracy of each classifier is presented with the help of Euclidean and Cityblock distances. These results emphasize that CSCA has the best accuracy on both the metrics, 95.38(%) and 95.00(%) (Euclidean and Cityblock, resp.) and this indicates that CSCA possesses high classification ability even when the dimension of the data space is high. AIRS2-Parallel also performs notably well, especially with Euclidean distance (93.85%). However, several other AIS models like AIRS-1 and AIRS-2 show significantly lower accuracy, suggesting that the use of all features may lead to overfitting or noise interference in certain models. The impact of distance metrics is noticeable but not dominant, as both Euclidean and Cityblock yield relatively similar trends in performance.

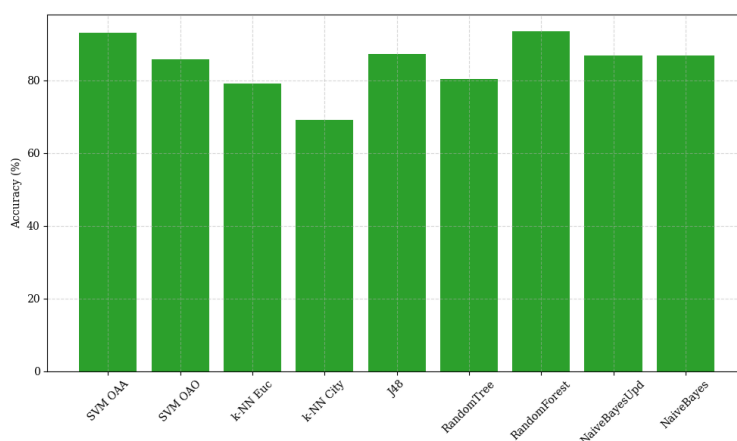
**Table 2:** Performance of Conventional Classifiers Using the Complete Set of Features Extracted from the Grey-Level Co-occurrence Matrix

Classifiers	Accuracy
k-NN with Cityblock distance	69.23%
SVM Multiclass One Against All	93.08%
SVM Multiclass One Against One	85.77%
J48	87.31%
RandomTree	80.38%
Random Forest	93.46%
k-NN with Euclidean distance	79.23%
NaiveBayesUpdateable	86.92%
NaiveBayes	86.92%

In Table 2, we delve into the performances of traditional classifiers using the same 640 features extracted from GLCM. Here, Support Vector Machines (SVM) exhibit robust performance, with SVM Multiclass One against All achieving an impressive accuracy of 93.08%, and SVM Multiclass One against One reaching 85.77%. These results emphasise that SVM-based classifiers present stiff competition for AIS-based classifiers in Arabic handwriting recognition.

K-Nearest Neighbours (k-NN) classifiers maintain respectable performance, with accuracy values of 79.23% and 69.23% for Euclidean and Cityblock distances, respectively. Decision tree-based classifiers, including J48, RandomTree, and Random Forest, perform credibly, achieving accuracy scores ranging from 80.38% to 87.31%. These points out those such conventional classifiers could prove to be successful in identifying Arabic scriptwriters. Other rather attractive type is called Naive Bayes classifiers, and two of its representatives, NaiveBayesUpdateable and NaiveBayes, performed with accuracy of 86.92%. These findings emphasize the flexibility of more conventional classifiers herein.

Finally, the paper can finer conclude that we evaluated the AIS-based and conventional classifiers with the entire set of features after their extraction based on GLCM to identify the Arabic scriptwriters. The findings indicate that CSCA, an AIS-based classifier, is remarkably accurate, and thus the AIS-based methods should have potential in this area. Nevertheless, conventional classifiers like SVM, k-NN, Naive Bayes also perform competitively, which means that the classifier ought to be tailored to the needs of a task and a dataset at hand. A deeper analysis of parameter optimisation and feature selection should lead to better performance that would be helpful in future researches in this field.



**Figure 4.** Traditional Classifiers Accuracy

Figure 4 shows the accuracy of the traditional machine learning classifiers overall feature set. Random Forest and SVM One-vs-All are out-performers, where the former has 93.46 and the latter 93.08. Other algorithms such as J48, Naive Bayes, and RandomTree also have competitive levels of accuracy of more than 85 %. In contrast, the lowest accuracy (69.23%) is reported by k-NN with Cityblock distance, and it means that using a high-dimensional feature of GLCM, not all the classical methods are effective. In general, ensemble models and use of kernel would be superior in such a setting compared to distance-based classifiers, especially in the absence of feature selection.

**Table 3:** AIS-based classifiers using selected features extracted from GLCM

Classifiers	Euclidean	Cityblock
CLONALG	61.54%	61.92%
CLONCLASS	76.54%	78.08%
Immunos1	74.61%	79.23%
CSCA	76.54%	78.08%
Immunos2	77.31%	79.61%

Immunos99	74.23%	76.54%
AIRS-1	83.46%	76.15%
AIRS-2	85.77%	76.54%
AIRS2-Parallel	85.77%	85.00%

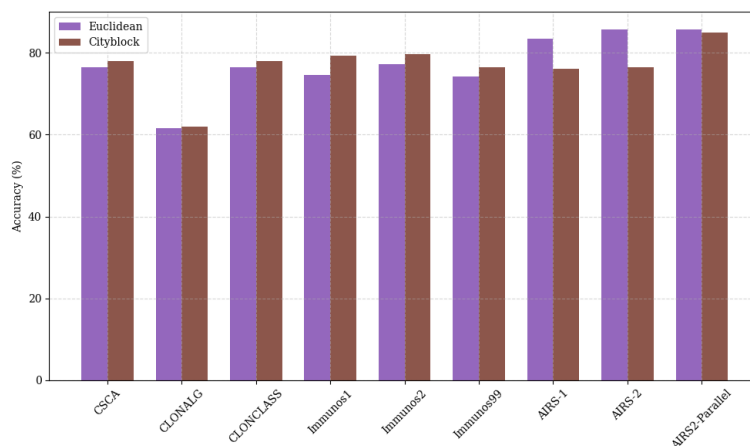
The given results in Table 3 provide us with valuable overview of AIS-based classifier performance applying the chosen features obtained working with the GLCM data set. These classifiers were tried with two different distance measures, Euclidean and Cityblock. The results provide insight on how best each classifier works when the problem is to identify Arabic writers.

Clonal Selection Classification Algorithm (CSCA) demonstrated a significant degree of accuracy using both the distance measures with a particular accuracy of 76.54 and a better accuracy of 78.08 of the euclid and the cityblock respectively. These findings lend credence to the fact that feature selection resulted in the improvement of the performance of the classifier and that the Cityblock measure may be effective in revealing applicable writer-specific patterns in selected features. Comparatively, the Clonal Selection Algorithm (CLONALG) failed to reach the same level of accuracy with 61.54 percent using the Euclidean as the metric and a slight better 61.92 percent accuracy using the Cityblock metric. This implies that CLONALG might not be advantageous to feature selection and that it holds strong performance concerning both metrics.

The accuracy of the Immune-Based Classifier, CLONCLASS was similar to CSCA, and it performed with 76.54 percent in Euclidean and a better 78.08 percent with the Cityblock measure. The obtained findings underline a stable characteristic of CLONCLASS with chosen peculiarities and suggest the effectiveness of the Cityblock measure in the writer identification task. Immunos is the other Immune-Based Classifier that reported accuracy of 74.61 percent with Euclidean and significantly higher 79.23 percent with Cityblock with selected features. Such great rise in accuracy points to the importance of feature selection. The effectiveness of the Cityblock metric that captures useful patterns in the chosen features is also demonstrated in the results. With either measure of distance, the versions of AIRS, especially AIRS-2 exhibited excellent levels of accuracy. AIRS-1 had the highest percentage of accuracy of 83.46% on Euclidean but the AIRS-2 portrayed the best accuracy of 85.77% on Cityblock. Such findings show that the AIRS classifiers are strongly positively impacted by feature selection. These good results were greatly influenced by a distance metric used, as the Cityblock was a bit better than the Euclidean.

AIRS2-Parallel, a parallelisation of AIRS was shown to be highly accurate when tested with both measures of distance. It showed an accuracy of 85.77, 85.00% using Euclidean, and Cityblock respectively. These observations confirm again the impressive uniformity of AIRS2-Parallel in writer identification tasks.

To conclude, the table 3 diagram has demonstrated the great influence of feature selection on the functioning of AIS-based classifiers. Furthermore, the choice of distance metric significantly influences the accuracy achieved. The study emphasises the need to optimise feature selection methods and select the most suitable distance metric to enhance the effectiveness of AIS-based classifiers in Arabic writer identification tasks.



**Figure 5.** AIS-Based Classifiers Accuracy

Figure 5 shifts focus to AIS-based classifiers, but now using only a selected subset of features—those identified as most relevant through feature selection techniques. Interestingly, performance becomes more balanced across classifiers. While the top scores slightly drop compared to Figure 3, models like AIRS2-Parallel still maintain strong accuracy (85.77% with Euclidean, 85.00% with Cityblock). This suggests that dimensionality reduction may help mitigate overfitting and improve model generalizability. Additionally, classifiers like CLONCLASS and Immunos2 show marked improvements compared to their all-feature counterparts, confirming the benefit of using a leaner, more informative feature set. The difference between Euclidean and Cityblock distances becomes subtler, reinforcing the robustness of the selected features across distance metrics.

**Table 4:** Performance of Traditional Classifiers Using Selected Features Extracted from the Grey-Level Co-occurrence Matrix (GLCM)

Classifiers	Accuracy
SVM Multiclass One Against All	89.61%
k-NN with Euclidean distance	76.54%
J48	39.23%
SVM Multiclass One Against One	84.23%
Random Tree	33.46%
Random Forest	79.23%
k-NN with Cityblock distance	78.08%
Naïve Bayes Updateable	51.54%
NaiveBayes	51.54%

On the other hand, as shown in Table 4, traditional classifiers offer a diverse performance spectrum. "SVM Multiclass One against All" emerges as a robust performer with an accuracy rate of 89.61%, affirming the effectiveness of the Support Vector Machine (SVM) approach. "Random Forest" also proves highly competent, achieving a commendable accuracy rate of 79.23%, making it a promising choice for Arabic writer identification.

The k-Nearest Neighbours (k-NN) method demonstrates consistency in performance, with "k-NN with Euclidean distance" and "k-NN with Cityblock distance" both exhibiting accuracies of 76.54% and 78.08%, respectively. This suggests their reliability in handling the complexity of Arabic script.

However, some traditional classifiers, including "J48," "Random Tree," "Naïve Bayes Updateable," and "NaiveBayes," yield relatively lower accuracy percentages. These results point to the necessity for further optimisation or the potential need to explore other techniques when using GLCM features for Arabic writer identification.

Concluding, this comparative study makes apparent the importance of feature extraction and the selection of the classifier in the task of identifying Arabic writers. Although a number of the AIS-based classifiers perform at praiseworthy levels of accuracy, especially the AIRS-2 and AIRS2-Parallel, the performance of traditional classifiers, such as the SVM Multiclass One Against All and the Random Forest, proves to be better in regards to accuracy, thus, indicating their efficiency in addressing the peculiarities of the Arabic script.

Whether one should use AIS based or traditional classifiers could also depend on the targeted objectives of the task and the available computational resources. Traditional SVM approach is the best when it comes to finding highly accurate results especially when accuracy is the most important factor. AIS-based classifiers (e.g. the AIRS variants), on the other hand are promising, and they might have an application in cases where a trade-off between precision and originality is desirable.

Finally, the outcomes given in this paper give enlightenments on the weaknesses and strengths of both Arabic writer identification of AIS-based and traditional classifier features. The knowledge can be used by those who are doing research and applied work because it provides them with a basis to make informed choices when deciding which classification method is the most appropriate to work with in their particular case on the rising sea of AI solutions in this area.

## 6. Conclusion and future trends

Here we have set out to take a systematic review of artificial immune systems (AIS) as possible useful tools of classification within the area of Arabic handwriting recognition. Our two algorithms that are based on AIS, namely, Clonal Selection Algorithm (CLONALG), Clonal Selection Classification Algorithm (CSCA), Artificial Immune Recognition System (AIRS), and Immunos showed possible potential in this area. The findings demonstrated the data dependent nature of AIS classifiers and supported in the fact that they are most appropriate in some situations.

Prospectively, our results provide a good starting point of further researches in the field. We can observe that there is a number of promising research perspectives, such as the detailed examination of the methods of feature selection. It could be suggested that even more efficient AIS classifiers can be developed with the help of further investigation, and they could be adjusted to particular needs and issues. In addition, the idea of hybrid models where AIS will be implemented along with another machine learning algorithms presents itself as the interesting path to follow. These hybrids can increase the level of precision and strength in the Arabic writer identification area.

The next direction we can also choose in the future concerns the application of our analysis to the case of online Arabic writer detection, in which dynamic features (the speed of writing, the force of pressures applied to the pen) is taken into account. Passing into online means may unlock opportunities of new uses and higher knowledge. Use of bigger size data and additional development of AIS algorithms especially to the characteristics of Arabic script also holds great promise. Such measures will enable to have a more detailed and detailed idea of what AIS is truly capable of when it comes to recognizing Arabic handwriting chaos.

Finally, our study preconditions the further use of immunological models and helps to provide shape to the emerged environment of the identification of Arabic writers. This is a very dynamic area and there are a good deal of virgin grounds to be covered.

## References

- [1] L. M. Al-Zoubeidy and H. F. Al-Najar, "Arabic writer identification for handwriting images," in *International Arab Conference on Information Technology*, Amman, Jordan, 2005, pp. 111-117.
- [2] S. Gazzah and N. E. Ben Amara, "Arabic handwriting texture analysis for writer identification using the DWT-lifting scheme," in *9th ICDAR*, vol. 2, 2007, pp. 1133–1137.
- [3] A. Al-Dmour and R. A. Zitar, "Arabic writer identification based on hybrid spectral-statistical measures," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 19, pp. 307–332, 2007.
- [4] H. E. Said, T. Tan, and K. Baker, "Personal identification based on handwriting," *Pattern Recognition*, vol. 33, no. 1, pp. 149-160, 2000.
- [5] F. Shahabi Nejad and M. Rahmati, "Comparison of Gabor-based features for writer identification of Farsi/Arabic handwriting," in *IWFHR '2006*, La Baule, France, 2006, pp. 545-550.
- [6] N. Feddaoui and K. Hamrouni, "Personal identification based on texture analysis of Arabic handwriting text," in *IEEE ICTTA*, vol. 1, 2006, pp. 1302–1307.
- [7] M. Bulacu, L. Schomaker, and A. Brink, "Text-independent writer identification and verification on off-line Arabic handwriting," in *9th ICDAR*, vol. 2, 2007, pp. 769–773.
- [8] M. Pechwitz, S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, "IFN/ENIT-database of handwritten Arabic words," in *Proc. of CIFED 2002*, pp. 129-136, 2002.
- [9] F. Shahabi Nejad and M. Rahmati, "A new method for writer identification of handwritten Farsi documents," in *ICDAR 2009*, Spain, 2009, pp. 426 - 430.
- [10] S. Al-Ma'adeed, E. Mohammed, D. Al Kassis, and F. Al-Muslih, "Writer identification using edge-based directional probability distribution features for Arabic words," in *IEEE/ACS International Conference on Computer Systems and Applications*, 2008, pp. 582–590.
- [11] A. Chaabouni, H. Boubaker, M. Kherallah, A. M. Alimi, and H. El Abed, "Fractal and multi-fractal for Arabic offline writer identification," in *International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 1051-4651.
- [12] M. N. Abdi, M. Khemakhem, and H. Ben-Abdallah, "Off-line text-independent Arabic writer identification using contour-based features," *International Journal of Signal and Image Processing*, vol. 1, pp. 4–11, 2010.

- [13] S. M. Awaida and S. A. Mahmoud, "Writer identification of Arabic handwritten digits," in *First International Workshop on Frontiers in Arabic Handwriting Recognition*, Istanbul, Turkey, 2010.
- [14] M. M. Galloway, "Texture analysis using grey level run lengths," *Computer Graphics and Image Processing*, vol. 4, pp. 172–179, 1975.
- [15] X. Tang, "Texture information in run-length matrices," *IEEE Transactions on Image Processing*, vol. 7, no. 11, pp. 1602-1609, 1998.
- [16] C. Djeddi and L. Souici-Meslati, "A texture-based approach for Arabic writer identification and verification," in *IEEE International Conference on Machine and Web Intelligence, ICMWI'2010*, Algiers, Algeria, 2010, pp. 115–120.
- [17] J. Chen, D. Lopresti, and E. Kavallieratou, "The impact of ruling lines on writer identification," in *Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition*, Kolkata, India, November 2010, pp. 439-444.
- [18] L. N. de Castro and J. I. Timmis, "Artificial immune systems as a novel soft computing paradigm," *Soft Computing*, vol. 7, pp. 526-544, 2003.
- [19] D. Dasgupta and L. N. Fernando, "Immunological computation, theory and application," Auerbach, 2009.
- [20] L. N. de Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239-251, 2002.
- [21] U. Garain, M. P. Chakraborty, and D. Dasgupta, "Recognition of handwritten Indic script digits using clonal selection algorithm," in *Lecture Notes in Computer Science (LNCS)*, vol. 4163, eds. Bersini and Carneiro, Springer Berlin/Heidelberg, 2006, pp. 256-266.
- [22] C. Yuefeng et al., "A handwritten character recognition algorithm based on artificial immune," in *International Conference on Computer Application and System Modeling*, vol. 12, 2010, pp. V12-273.
- [23] K. M. Faraoun and A. Boukelif, "Artificial immune systems for text-dependent speaker recognition," *Journal of Computer Science*, vol. 5, no. 4, pp. 19-26, Dec. 2006.
- [24] N. A. Draman Muda, C. C. Wilson, and S. Ling, "Bio-inspired audio content-based retrieval framework," in *Proceedings of the World Academy of Science, Engineering and Technology*, Turkey, 2009, vol. 53, pp. 791-796.
- [25] A. Kumar and S. Nair, "An artificial immune system based approach for English grammar checking," in *LNCS, Artificial Immune Systems, ICARIS 2007*, Springer Berlin/Heidelberg, vol. 4628, 2007, pp. 348-357.
- [26] Q. Zhang et al., "Multi-class text categorization based on immune algorithm," *ettandgrs*, vol. 1, pp. 749-752, 2008.
- [27] N. Isa et al., "Application of the clonal selection algorithm in artificial immune systems for shape recognition," in *International Conference on Information Retrieval & Knowledge Management 2010*, Malaysia, pp. 223–228.
- [28] C. Andrzej et al., "An immune approach to recognition of handwritten words," in *2009 International Conference on Biometrics and Kansei Engineering*, pp. 49-54, 2009.
- [29] Golzari et al., "A hybrid approach to traditional Malay music genre classification: Combining feature selection and artificial immune recognition system," in *International Symposium on Information Technology*, vol. 2, pp. 1–6, 2008.
- [30] K. Zafer and D. Banu, "Gender and author detection in Turkish texts using artificial immune recognition systems," in *IEEE 16th Signal Processing, Communication and Applications Conference*, Turkey, 2008, pp. 1–4.
- [31] C. Chuanliang et al., "Artificial immune recognition system for DNA microarray data analysis," in *Fourth International Conference on Natural Computation, ICNC '08*, China, vol. 6, pp. 633–637.
- [32] G. Julie and C. Steve, "An artificial immune system approach to semantic document classification," in *International Conference on Artificial Immune Systems*, vol. 2787, pp. 136-146, 2003.
- [33] A. K. Muda, S. M. Shamsuddin, and M. Darus, "Bio-inspired generalized global shape approach for writer identification," *Transactions on Engineering, Computing and Technology*, vol. 16, pp. 55-59, 2006.

- [34] A. K. Muda and S. M. Shamsuddin, "A framework of artificial immune system in writer identification," in *BIC'05, International Symposium on Bio-Inspired Computing*, Johor, Malaysia, 2005.
- [35] J. Brownlee, "Clonal selection theory & CLONAG. The clonal selection classification algorithm (CSCA)," Technical Report No. 2-02, Centre for Intelligent Systems and Complex Processes (CISCP), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology, Victoria, Australia, 2005.
- [36] J. Brownlee, "Immunos-81 - The misunderstood artificial immune system," Centre for Intelligent Systems and Complex Processes (CISCP), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology, Victoria, Australia, Technical Report ID: 3-01, Feb. 2005.
- [37] A. Watkins, J. Timmis, and L. Boggess, "Artificial immune recognition system (AIRS): An immune inspired supervised machine learning algorithm," *Genetic Programming and Evolvable Machines*, 2004, pp. 291-317.
- [38] A. Watkins and J. Timmis, "Exploiting parallelism inherent in AIRS," in *3rd International Conference on Artificial Immune Systems (ICARIS 2004)*, Catania, Sicily, 2004, pp. 427-438.
- [39] A. Watkins, "Exploiting immunological metaphors in the development of serial, parallel, and distributed learning algorithms," PhD thesis, Computer Science, University of Kent, Canterbury, England, 2005.
- [40] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.
- [41] M. Bulacu, "Statistical pattern recognition for automatic writer identification and verification," PhD thesis, University of Groningen, 2007.