# FdDCA : A Novel Fuzzy Deterministic Dendritic Cell Algorithm

Nura Mukhtar
r01nm13@abdn.ac.uk

George M.Coghill
g.coghill@abdn.ac.uk

Wei Pang
pang.wei@abdn.ac.uk

School of Natural and Computing Sciences
University of Aberdeen
United Kingdom

## ABSTRACT

The Dendritic Cell Algorithm (DCA) and its improved version: Deterministic Dendritic Cell Algorithm (dDCA) are essentially binary classification algorithms based on the behavior of Dendritic Cells (DCs) in the immune system. Both DCA and dDCA collect and process the data in form of signals, and produce output signal. The signals are divided in two types: danger and safe signals, and the output signal is determined by the values of the danger and safe signals. However, both DCA and dDCA suffer from data misclassification due to their sensitivity to data order. In this research we proposed a Fuzzy Deterministic Dendritic Cell Algorithm (FdDCA), which combines dDCA, fuzzy sets, and K-means clustering. The main objective of this research is to smooth the sharp boundaries between signals since we cannot always identify a clear boundary between the values of the signals. Our approach fuzzifies the signal values using linguistic variables, and a rule base is built to support fuzzy inference. The experimental results based on real data sets show that our approach shows a promising results compared to DCA and dDCA.

## Keywords

artificial immune systems; deterministic Dendritic Cell Algorithm; fuzzy logic

## 1. INTRODUCTION

The immune system (IS) consists of two layers of defence, namely the innate immune system and the adaptive immune system. The innate immune system is the first tier of defence, and it provides a general (non-specific) and immediate defence against pathogens. However, it cannot recognise the same pathogen it recognised should the body be exposed to the same pathogen for the second time. The adaptive immune system is the second line of defence. It has high level of specificity towards intruding pathogens (specific) and also can recognise the same pathogen it recognised earlier should the body be exposed to the same pathogen for the second time. "The powerful information processing capabilities of the immune system such as feature extraction, pattern recognition, learning, memory, and its distributive nature provide rich metaphors for its artificial counterpart" [1]. The immune system is a biological systems that inspired the development of artificial immune systems (AIS).

AIS mimic the behaviour of the IS to develop algorithms, and it uses the ideas from theoretical immunology to develop computational systems that have the ability of performing several tasks in different feilds [2]. The fact that the IS is able to recognise, identify, and eliminate intruders inspired AIS researchers to mimic these processes and develop immune-inspired algorithms to solve real world problems. This led to the development of the first generation algorithms such as negative selection [3] , clonal selection [2], and aiNet algorithms [4], and they showed promising results when applied in Network Intrusion Detection [1], Pattern Recognition [5], and Optimization [6]. But these algorithms, especially the negative selection algorithm have scaling and high rate of false negatives generation issues [7]. This motivated the development of the second generation AIS algorithm called the Dendritic Cell Algorithm (DCA) [8], that it incorporates more sophisticated properties such as robustness and self organization. Despite the success of DCA in computer security, the original algorithm was highly stochastic, which made it difficult to understand and analyse due to random-based elements. In view of that a deterministic dendritic cell algorithm (dDCA) was proposed [9], and the algorithm's performance in terms of computational analysis was improved. DCA and dDCA show evidence of their ability to process data for classification. However, they are sensitive to data order [7, 9]. This is due to crisp separation between mature (abnormal) and semi-mature (normal) context. Regarding this, Chelly and Elouedi [10] developed a fuzzy Dendritic cell method (FDCM). The aim of FDCM is to improve the classification accuracy of DCA by smoothing the abrupt separation between normal and abnormal cell contexts. Nonetheless, FDCM suffers from some limitations [11] . For instance, the midpoints of the membership functions are user defined, and this affects the classification accuracy of the system. An improved version of FDCM was developed to handle the problem by incorporating fuzzy c-means clustering using euclidean distance (FDCM-EUC) [11] was developed to handle the problem by incorporating fuzzy c-means clustering using euclidean distance, and



*GECCO'16 Companion, July 20-24, 2016, Denver, CO, USA*



DOI: http://dx.doi.org/10.1145/2908961.2931662

it is an improved version of FDCM. FDCM-EUC produces good results against DCA and FDCM, but there is a high rate of false negative in some cases, and this may be due to the crisp nature of input data. The artificial DCs in DCA and FDCM-EUC represent the signal concentration in crisp form, and the concentration is marked as either high or low; this sharp division does not work well in describing the real world domains. Fuzzy set theory is well known for handling uncertain and imprecise knowledge [12].

In this research, we develop a Fuzzy Deterministic Dendritic Cell Algorithm (FdDCA). Our algorithm adopts dDCA and incorporates fuzzy set theory and k-means clustering. The aim of this research is to smooth the boundaries between danger and safe signals and also between the output signals, this may improve the classification performance of our algorithm. In our proposed FdDCA, the above mentioned boundary smoothing is achieved by converting the crisp values of the two signals (danger and safe) and the output value into fuzzy intervals. To achieve this Mamdani Fuzzy Inference Engine (MFI) [13] incorporated to infer from a set of fuzzy rules to determine the class of a cell's context, as it is commonly used fuzzy methodology

## 1.1 Dendritic Cell Algorithm (DCA)

The dendritic cell algorithm is an immune inspired algorithm proposed by Greensmith [7]. DCA is an abstraction of the biological DC model based on the danger theory. DCA is a population based classification algorithm which receives two types of inputs, namely antigen and signal. The antigens are the candidates to be classified while signals are associated with the antigens [14]. These signals are categorised either as PAMP, danger, or safe.

The DCs exist in one of three states, immature, semi-mature, or mature. Each artificial DC calculates the input signal values at each iteration to produce three temporary output signal values; costimulation molecules (*csm*), semi-matured DC (smDC), and matured DC (mDC). Each cell is assigned a migration value, and once the *csm* value exceeds the migration value the immature DC matures to be either a semi-mature or a mature DC. A DC becomes mature if the antigen from DC presents more danger signals than the safe signals; conversely, if the antigen from a DC is presented by more safe signals than the danger signals, it is labeled as semi mature. When a particular antigen is measured by several DCs, the mature context antigen value (MCAV) is calculated by dividing the number of antigens presented by mature DCs with total number of antigens presented by all cells, and compared with the anomaly threshold to determine whether the antigen is anomalous. DCA has been successfully applied in computer security [10]. In the dDCA, a minimum of two signals categories and antigens are required for the system to work properly. A uniform distribution of lifespan values is used across the population, the lifespan is the amount of signals a DC process during it lifetime. Each DC in the repertoire is exposed to similar signals and these processes these signals in similar way. Another modification is the incorporation of antigen profile to replace random sampling and storage in the previous implementation of the DCA.

# 2. THE FUZZY DETERMINISTIC DENDRITIC CELL ALGORITHM (FDDCA)

The dDCA have two input signals (danger and safe) as

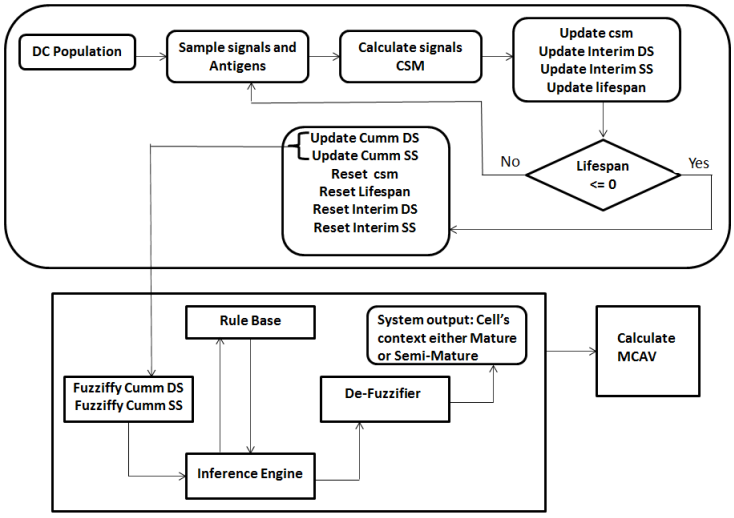


**Figure 1: Fuzzy dDCA Flow chart**

mentioned in Section 2, and these signals combine to generate two intermediate output values *csm* and K. The *csm* measures the overall concentration of signals a cell is exposed to during its life time, while the K value measures the normality or abnormality of the cell. When the cell exhausts its lifespan it will migrate and be ready to classify the antigens it collected during its lifetime as normal or anomalous. The summation of danger and safe signals forms the *csm* value while K value is derived by the difference between danger signal and twice of safe signal as shown in Equations (1) and (2).

$$csm_i = DS_i + SS_i \tag{1}$$

$$K_i = DS_i - 2SS_i \tag{2}$$

The proposed Fuzzy Deterministic Dendritic Cell (FdDCA) converts the crisp values of the cumulative danger signals (DS) and cumulative safe signals (SS) into fuzzy numbers, and the cumulative signals are the total amount of safe and danger signals that the cells have been exposed to during their lifespan. Figure 1 gives an illustration of FdDCA. The signals are used to determine the concentration of both the signals and K as shown in Equations (1) and (2), where $DS$ and $SS$ are danger and safe signals respectively: i = (1...N).

## 2.1 Initialization and signal processing

### 2.1.1 *Initialization*

The DC population and its parameters are initialized. The size of DC population is set up to a maximum of 100 cells as suggested in [7] . A parameter named antigen array size is set to store antigens, this allows the sampling of a number of antigens per iteration. Moreover, the lifespan is set uniformly, and the lifespan of each DC is uniformly distributed. The output parameters K and csm are initialised to be zero.

### 2.1.2 Signal Processing and Update

- **Costimulation (CSM)**

The costimulation is the accumulation of signal concentration over a period of DC's lifetime within its environment. When a DC's life span expires, it migrates to the lymph node and presents antigens under a context. Equation (1) shows the calculation of *csm*.

- **Lifespan**

The lifespan of a DC is the amount of time this DC spend collecting signal concentrations within its environment before it migrates to the lymph node. The lifespan of the DC is subtracted from the accumulated concentration of signals over time until the value of lifespan is less than the sum of the concentration. In this case, lifespan is a fixed value, however, its value is decreasing overtime as shown in Equation (3).

$$lifespan = lifespan - (SS_i + DS_i) \quad (3)$$

*where i = (1...N).*

## 2.2 Fuzzification

Two signals (danger and safe) and one output signal K are defined, and each of the input and output crisp values fuzzified into linguistic variables. Each of the input and output crisp values needs to be fuzzified into linguistic variables, and a membership a function is used to determine the range which each linguistic variable belongs.

- **Linguistic Variables**

Setting linguistic variables is one of the basic tools in fuzzy logic, SS and DS are classified into three categories: low, medium, and high. The output signal is classified into two classes: mature and semi-mature. $K_{Maturity}$ is a variable that interprets the state of the K value as either mature or semi-mature DC. To construct a membership function, there is a need to specify the range of each linguistic variable. dDCA was first run to generate the values of K, DS, and SS. k-means clustering was then used to determine the ranges (as clusters) and core values in the membership function as mid points. The membership functions of input variables were designed to be trapezoidal . This applies to both signals. The membership functions of output variable K were also designed using trapezoidal function.

## 2.3 The Rule Base

The set of rules built to support the fuzzy inference are adopted from [10]. The input signals DS and SS are combined by "AND" to produce output as the context of the cell, as shown in table 1.

## 2.4 Context Assessment

The output context value for each DC is in a fuzzy form, and it has to be converted to a crisp value for context assessment. Center of Gravity (COG) defuzification is adopted as it is widely used and easy to work with, and the middle value of the output range $K_{maturity}$ is taken. If the centroid generated by COG is greater than the middle value then it indicates that the DC is mature, and also the antigen collected may be anomalous.

**Table 1: Set of Rules, where Low, Medium and High represents the concentration of the signals, mature and semi-mature = The context of the cells**

| | DS is Low | DS is Medium | DS is High |
|---|---|---|---|
| SS is Low | mature | mature | mature |
| SS is Medium | semi-mature | semi-mature | mature |
| SS is High | semi-mature | semi-mature | mature |

# 3. EXPERIMENTAL SETUP

Two Experiments are conducted using two different data orders of Wisconsin Breast Cancer (WBC), Blood Transfusion Center (BTSC) and Haberman's Survival (HS) data sets [15], Experiment 1 (one-step order) uses all class 1 data items followed by all class 2 data items, while Experiment 2 (two-step order) uses part of class 1 items then all class 2 items followed by remaining class 1 item. Each experiment is performed 10 times, and each run samples the antigen once. Resulting in 700,748 and 306 antigen presentations per run of the used data sets. The final clsaa of each antigen is determined by anomaly treshold, and its is defined by the total number of malign divide by sum of both malign and bening classes. The threshold for classification is set to 0.66, 0.7 and 0.7 for WBC, BTSC and HS respectively. Items whose MCAV value is above the threshold are classified as anomalous and below are labeled as normal.
The classification accuracy of our FdDCA is assessed using Accuracy and F- Score.

# 4. RESULTS

The results of Experiment 1 shows good classification 99.43% accuracy and F-score 0.99, five errors out of a total of 700 data items. These errors are items that are in class 1 but classified as class 2, and this error is termed false positive. Experiment 2 recorded two errors (false positive) out of 700 items, and this yields a lower error rate than Experiment 1 with 99.71% accuracy and f score of 1. Figures 2 and 3 show graphs that represent MCAV values per antigen on one-step and two-step data order. This shows that FdDCA is able to switch between semi-mature and mature contexts with respect to changes related to input data order.
The results presented in [7, 9] shows that both DCA and dDCA misclassifications occur mainly at the transition boundaries, and this proves that these algorithms makes error in classification when the context changes more than one time in quick succession. However, our FdDCA improves the classification result compared with DCA and dDCA in terms of disordered contexts in Experiment 2. Tables 2 and 3 presents a comparison of the percentage accuracy and F-Score of the three algorithms. This table shows that our FdDCA produced better results in terms of classification accuracy, and this shows that randomisation effect on our algorithm. For instance, the accuracy of FdDCA in Experiment 2 in WBC dataset is 99.71% while DCA, dDCA have 91%, 98%, 97.57% and 98.14% respectively.

# 5. CONCLUSION AND FUTURE WORK

In this research a fuzzy deterministic version of the dDCA is presented, and the algorithm effectively classifies antigens

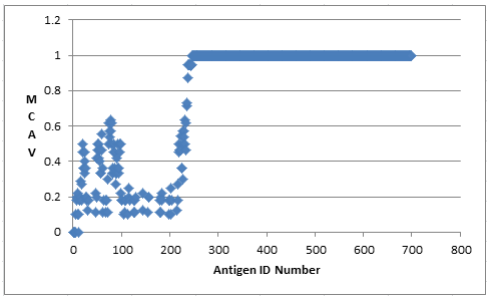


**Figure 2: Classification of the 700 items for Experiment 1 (WBC Dataset)**

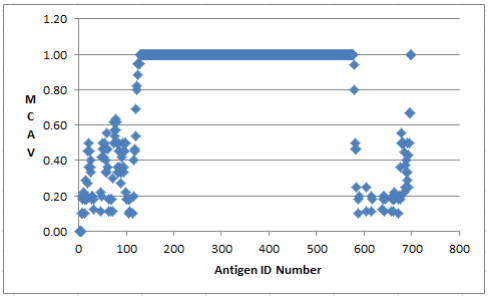


**Figure 3: Classification of the 700 items for Experiment 2 (WBC Dataset)**

into two classes (either class 1 or 2). The introduction of fuzzy system to DCA improves the classification of the algorithm, and our algorithm demonstrates that it has the potential to be applied in a real world domains for classification.

We intend to carryout further experiments to validate our system, and we will make further comparisons between our approach and other approaches such as DCA, dDCA, FDCM, and FDCM-EUC. Our future work will involve the analysis of input data from multiple source instead of one as we used in our FdDCA.

## 6. REFERENCES

[1] J.Kim and P.J.Bentley. Towards an artificial immune system for network intrusion detection: An investigation of dynamic clonal selection. In *In Proceedings of Congress Evolutionary Computation*, pages 1244–1252. IEEE, 2001.

[2] L.N. de Castro and F.J.Von Zuben. The clonal selection algorithm with engineering applications. In *Proceedings of GECCO Workshop on Artificial Immune Systems and Their Applications*, pages 36–37, July 2000.

[3] S.Forrest, A.S.Perelson, L.Allen, and R.Cherukuri. Self-nonselfdiscrimination in a computer. In *In Proceedings of IEEE Symposium on Research in Security and Privacy*, pages 36–37. IEEE, 1994.

[4] L.N. de Castro and F.J.Von Zuben. ainet: An artificial immune network for data analysis. In *In In Data Mining: A Heuristic Approach, H.A.Abbass and R.A.Sarker, and C.S.Newton*. Idea Group Publishing, USA, March 2001.

**Table 2: Comparing the F-Score of DCA, dDCA and FdDCA. Where E1 = Experiment 1 and E2 = Experiment 2**

| Datasets | DCA | | dDCA | | FdDCA | |
|---|---|---|---|---|---|---|
| | E1 | E2 | E1 | E2 | E1 | E2 |
| WBS | 0.93 | 0.83 | 0.98 | 0.97 | 0.99 | 1.00 |
| BTSC | 0.83 | 0.80 | 0.86 | 0.65 | 0.81 | 0.80 |
| HS | 0.75 | 0.31 | 0.21 | 0.20 | 0.92 | 0.91 |

**Table 3: Comparing the Percentage Accuracy of DCA, dDCA and FdDCA. Where E1 = Experiment 1 and E2 = Experiment 2**

| Datasets | DCA | | dDCA | | FdDCA | |
|---|---|---|---|---|---|---|
| | E1 | E2 | E1 | E2 | E1 | E2 |
| WBS | 99.4 | 91.0 | 98.0 | 98.0 | 99.4 | 99.7 |
| BTSC | 91.8 | 93.3 | 73.9 | 49.7 | 74.5 | 51.1 |
| HS | 83.0 | 17.3 | 30.3 | 35.0 | 88.2 | 87.3 |

[5] J.E. Hunt and D.E.Cooke. Learning using an artificial immune system. *Journal of Network and Computer Applications*, 19(2):189–212, April 1996.

[6] L.N. de Castro and F.J.Von Zuben. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251, June 2002.

[7] Julie Greensmith. *The Dendritic Cell Algorithm*. University of Nottingham, 2007.

[8] F.Gua, J.Greensmith, and U.Aickelin. Theoretical formulation and analysis of the deterministic dendritic cell algorithm. *Biosystems*, 111(2):127–135, February 2013.

[9] J.Greensmith and U.Aickelin. The deterministic dendritic cell algorithm. *Artificial Immune system*, 5132:291–302, 2008.

[10] Z.Chelly and Z.Elouedi. Fdcm: A fuzzy dendritic cell method. *Artificial Immune system*, 6209:102–115, 2010.

[11] Z.Chelly and Z.Elouedi. Further exploration of the fuzzy dendritic cell method. *Artificial Immune system*, 6825:419–432, 2011.

[12] T.Stibor, R.Oates, G.Kendall, and J.M.Garibaldi. Geometrical insights into the dendritic cell algorithm. In *In Proceedings of GECCO*, pages 8–12. ACM, July 2009.

[13] E. H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. In *In Proceedings of the Institution of Electrical Engineers*, pages 15851–1588. IEEE, December 1967.

[14] R.Bi, J.Timmis, and A.Tyrrell. The diagnostic dendritic cell algorithm for robotic systems. In *IEEE Congress on Evolutionary Computation*, pages 18–23. IEEE, July 1967.

[15] A .Asuncion and D. D. Newman. Uci machine learning repository. *http://www.ics.uci.edu/ mlearn/.*, 2007.