

# Artificial Bee Colony Optimization based Negative Selection Algorithms to Classify Iris Plant Dataset

Prashant Kamal Mishra  
M.tech Student  
Dept. of Computer Science & Engg,  
Ajay Kumar Garg Engineering College, UPTU,  
Ghaziabad, India

Mamta Bhusry  
Professor  
Dept. of Computer Science & Engg,  
Ajay Kumar Garg Engineering College, UPTU,  
Ghaziabad, India

## ABSTRACT

This paper presents a new technique for classification of data. Artificial Immune System is the best technique to classify the data. Three main algorithms came under Artificial Immune System are - (1) Clonal selection algorithms (CLONALG), (2) Negative selection algorithms (NSA), (3) Artificial immune networks (AINE). Negative selection algorithms is one of the best technique to classify the data. NSA works in two phases Training and Testing. Training is an optimization task so it is required to get the optimal value. In tradition training process NSA have some drawbacks like local minima and computational complexity. So to overcome this problem optimized data is to be used. Many optimization algorithms have been investigated, Artificial Bee Colony (ABC) optimization algorithm is one of the best algorithm.

The proposed hybrid ABC and NSA can be applied to improve the global convergence behavior of the algorithm. The experimental results focus on Iris dataset plant and show that the proposed algorithm is more effective in classification of iris dataset when compared with other approaches. This method is more effective for random search and an effective hybridized method for artificial immune system optimization problem.

## Keywords

Artificial Bee Colony optimization algorithm, Clonal Selection Algorithm, Negative selection algorithm, IRIS Plant Dataset.

## 1. INTRODUCTION

Artificial Immune system (AIS) is a promising and novel research area in the 21st century. It is inspired from the Natural immune system (NIS). The main goal of AIS is to understand and gain the knowledge and working of NIS. As in the NIS if any pathogen enters in the body, the body's immune system starts working to identify them and generate antigen to remove or oppose these pathogens. Main problem with biological data in extracting knowledge is highly advanced technologies, tools and algorithms is to be provided for it. Some problems like prediction of protein structure, phylogenetic inferences and multiple alignments etc. are non-deterministic problems that are hard to solve in polynomial-time. To solve these kinds of problems Artificial Immune System is one of them. Artificial Immune techniques are commonly in use because of its ability to use the functionalities of NIS in scientific issues. In classification process we maps the data into predefined groups. It is one of the major task of data mining process.

In this paper, we propose an Artificial Bee Colony Optimization Based Negative Selection Algorithm to Classify

Iris Plant Dataset. Using the optimization process we can find the affinity level and the best dataset will be given to the negative selection algorithm for the training process. After the training process is completed data is provided to the negative selection algorithm that will classify the data. The computational complexity of finding the accurate classification can decline. The result shown that the proposed method of finding the classification has achieved good accuracy as compared to other methods of classification. Our optimized data classification techniques are compare with the previously used technique. The experimental result proved that the optimized data classification gives the minimum false values. The dataset on which the experiment performs is taken from the UCI machine learning repository.

The rest of paper is organized as follows. Section 2 deals with IRIS plant dataset followed by Section 3 and 4, which presents the essential concepts of Negative Selection Algorithm and ABC Optimization Algorithm respectively Section 5 illustrates the performance of the proposed approach and finally results, conclusions and future scope are provided towards the end.

## 2. IRIS PLANT DATASET

IRIS plant dataset is taken from UCI Machine Learning Repository. It is the best database for the application of the Artificial Immune System. The dataset is originally created by Ronald Fisher and submitted to UCI Repository by Michael Marshall in 1988.

The IRIS dataset consist of three classes and each class contains the 50 instances and each class refer to a type of IRIS plant. So total 150 instances are there in the dataset. Each class of the dataset consist of four attributes:

1. Sepal length
2. Sepal width
3. Petal length
4. Petal width
5. Predictive attribute

It shows the name of the class which belong to any of the following three: Setosa, Versicolour, and Virginica. The main goal of examining the IRIS plant dataset is to find out the pattern by analyzing sepal and petal size of IRIS plant and make the predictions by examining the pattern to make the class of IRIS plant.

It is already mentioned that the relationship which is to be find out using the IRIS plant data set is to be used as a classification model and this model is used to classify the type IRIS plant dataset by evaluating the size of sepal and petal.

There is a positive relationship between the length and width of both sepal and petal. This relationship is easily identified with nude eyes or without using any tools and formulas. It is realized that the sepal length is always larger than sepal width and petal length is also larger than petal width.

### 3. NEGATIVE SELECTION ALGORITHM

Artificial Immune System is inspired from Natural immune System. Many applications like fault diagnosis, anomaly detection, optimization and computer security is to be done by Artificial Immune System. One of the type of Artificial Immune System is Negative Selection Algorithm (NSA), Initially NSA is proposed for distinguishing self from others by Forrest et al [8]. Many modified algorithms are proposed by researchers[8-11]. In this the T-cells are generated in the immune system. If a T-cell identifies any self-cell then it is discarded from maturing to antibody and the others are applied in to the immune system to identify and oppose the pathogens.

In the same way, negative selection algorithm creates detector. If any detector candidate match from the group of self samples then those detectors is to be discarded from the set. In this the central goal of this algorithm is to randomly generate candidates then eliminate those cells which identify the self-data and then these detector is used to find anomaly.

### 4. ABC OPTIMIZATION ALGORITHM

Karaboga proposed the Algorithm of Artificial bee colony (ABC). It is a swarm intelligence algorithm which is inspired by the nature of honey bees.

ABC optimization technique is used to simulate the intelligent foraging behavior of honey bees. Bulk of honey bees are known as swarm. Honey bees can complete the task by the social co-operation. Three types of bees are used in the ABC algorithm: Employed bees, Onlooker bees and scout bees. Initially the employed bees find the food source available in their memory and checks the value of the food source and their neighborhood. After finding the position and food source value they share this information to the on looker bees. Onlooker bees select the best food source which is find out by the employed bees. The food source have the higher food quality (fitness value) will have the great chance to be selected by onlooker bees as compare to the lower one. Some employed bees left their food source and start searching the new food source, these type of bees are known as scout bees or we can say that employed bees are converted into scout bees.

In the ABC optimization algorithm the employed bees start searching the good food source they memorize the location and fitness value, on which food source they found the higher fitness value they come back to hive and start dancing at the node. The onlooker bees watching the dance and select the food source depend on the dance performed by employed bees. Finally the onlooker bees find out the probability of the data set and best data is to be selected.

### 5. PROPOSED METHODOLOGY

The objective of this approach is to find the high accuracy of classification. The proposed method of finding the classification accept a training dataset and gives high accuracy classifier. The NSA algorithm proceeds in two phases: Training and Testing. In the training phase huge number of detector set is to be generated due to the increment in the size of self data, the redundant detectors is ought to be generated.

So to overcome this problem optimization is needed. ABC optimization algorithm is used to optimize the data and find the objective value (fitness). This objective value is use to find the minimum number of detectors.

We have used an Artificial Bee Colony optimization method based on Negative Selection Algorithm in order to classify the iris data set. In the ABC optimization process following steps are to be taken:

Randomly choose the initial food source by following eq.

$$X_{ij} = X^{\min} + \text{rand}(0, 1)(X^{\max} - X^{\min})$$

Where  $i$  and  $j$  = Number of food sources and initially  $X^{\max}, X^{\min} = 0,1$  respectively for normalization. Now local searching is started by employed bees and thus, finding the quality of the existing food source by:

$$V_{ij} = X_{ij} + \text{rand}(-1, 1)(X_{ij} - X_{kj})$$

Where  $V_{ij}$  is neighbor source and  $k$  is random number.  $k$  and  $i$  value should not be same.

Now, onlooker bees are selected on the basis of fitness value. More the fitness value, more the chances of the selection of the bee. Fitness value shows the nectar amount of food source. And then probability is calculated by:

$$P = 0.9 * (\text{fitness})_i / \max(\text{fitness}) + 0.1$$

And if the food source is not improved then it is discarded and employed bees becomes scout and these scout start finding new food source by following eq.

$$X^i = X^i \min + \text{rand}[0, 1](X^i \max - X^i \min)$$

Finally the optimized data is generated whose objective value is best. Now this optimized data is given to the Negative Selection Algorithm. In NSA the random population of detector is to be generated and each detector find the shortest distance to any self-point. It checks the distance of the self-point with their radius. If distance is less than it is discarded otherwise find the direction by using following eq.

$$\frac{\sum_{j=1}^n (d_i - c^{\text{nearest}})}{|\sum_{j=1}^n (d_i - c^{\text{nearest}})|}$$

Where  $d_i$  is the current position of detector and  $c$  is the nearest point. Then calculate the updation using following rule:

$$\eta_i = \eta_0 e^{\frac{-i}{\tau}}$$

Where  $\eta_0$  is initial step size,  $\tau$  control the decay and  $i$  is age detector. Now detector is moved by:

$$d(i+1) = d_i + \eta_i * \text{dir}$$

Where,  $d(i+1)$  is the next position of the detector and again find the distance of self to detector. This condition continues until the maturity condition is not met. As the adaptation rate reduces with each movement so the detector may never move far away from the self-subspace.

By definition of NSA, if any data is detected then it is consider as Non-self otherwise it is consider as self.

### 5.1 Proposed Method Flowchart

We can summarize figure1 as the proposed strategy for finding optimized solution for dataset by taking the input as artificial immune network. On that input, we will apply ABC and Negative Selection algorithm in a hybrid form where randomly generated data is achieved by ABC algorithm and data optimization is applied on that independent generated data. Now, thread bee algorithm is to be implemented where fitness and probability calculation is done using ABC algorithm and further data classification is done using NSA approach. Thus, we may come with the best optimized solution for the dataset.

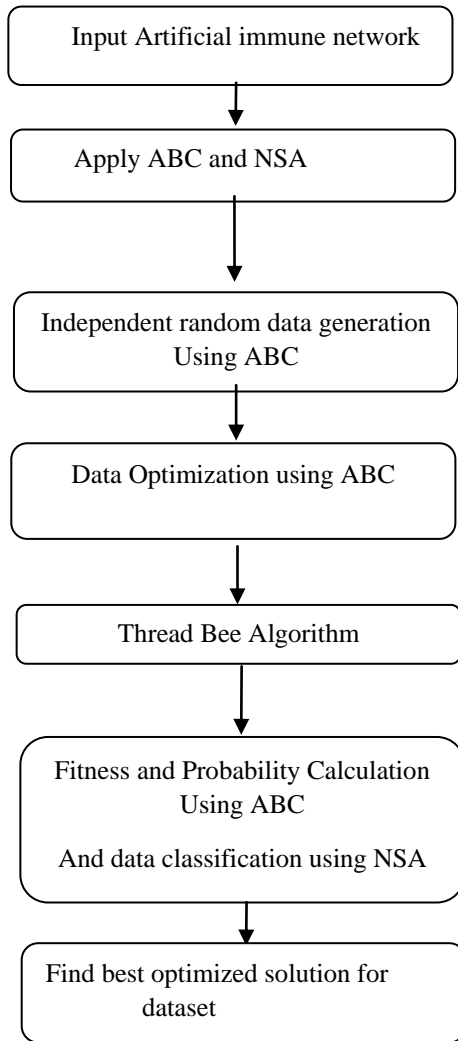


Figure 1: Proposed strategy

### 6. RESULTS

To perform this work we use MATLAB. MATLAB is a high-level programming language that can be used to visualize the data, numerical computation and optimization and development of various algorithms. For the analysis of the algorithm we use 50 instances from all three classes.

For the evaluation of the effectiveness of NSA two parameters is to be used, the *detection rate* and *false alarm rate*. Detection rate shows that how many non self points are correctly classified.

$$\text{Detection Rate} = \frac{\text{Correctly identified non self point}}{\text{Total non self point}} * 100$$

The false Alarm Rate is calculated on self point. It shows that how many self points are incorrectly classified.

$$\text{False Alarm Rate} = \frac{\text{Incorrectly identified self point}}{\text{Total self point}} * 100$$

For the overall performance of the algorithm figure of merit (FOM) is calculated the final score, which shows the difference between detection rate and false alarm rate.

$$\text{Figure of Merit (FOM)} = \text{Detection Rate} - \text{False Alarm Rate}$$

Table 1. Table contains the objective value of ABC for all three classes and then classification accuracy.

IRIS plant (150)	ABC objective value	Total non self	Detection Rate
Setosa	4.9759	98	98%
Versicolor	1.2076	96	96%
Virginica	1.2076	99	99%

So overall 97.67% classification accuracy is to be achieved with this method. Below figure shows the result of one of the IRIS plant data set i.e. virginica. In this the Total\_Self\_Incorrect shows that the number of elements incorrectly classified from the self-sub-space and the Total\_Nonself shows that the correctly classified the elements from the non-self sub-space. And finally detection rate is the difference between the Total\_Nonself and Total\_Self\_Incorrect. This detection rate shows the accuracy of the proposed method.

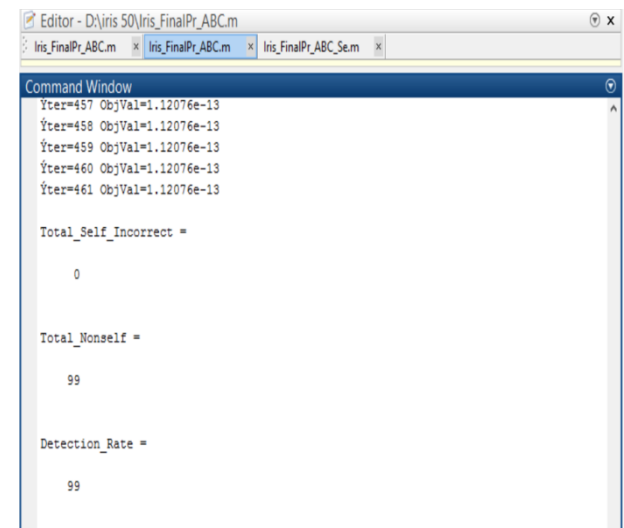


Figure 2: Expected snapshot of the objective value of ABC and detection rate of negative selection algorithm of one of the iris plant class (virginica) data set.

**Table: Comparison Of Accuracy Between Our Approach And Other Classifiers**

TECHNIQUES OF CLASSIFICATION	% ACCURACY
<b>Optimized negative selection algorithm</b>	<b>97.67%</b>
AIRS (Artificial Immune Recognition System)	96.7%
MLFFNN (Multi Layer Feed Forward Neural network)	96.66%
MINSA (Multi-Class Iteratively Refined Negative Selection Classifier)	96%
Nearest neighbor	96%
ANSC(Artificial Negative Selection Classifier)	95.8%
DAIS	95.8%
M-NSA(Modified Negative Selection Algorithm)	95.33%
Bayes net	94%
C 4.5	94%
RBF(Radial Basis Function Neural Network)	93%

Above table represents the comparison of our approach and other approaches of classifications on the basis of accuracy of result.

## 7. FUTURE SCOPE AND CONCLUSION

In the work, the performance of Artificial bee colony optimization based Negative Selection Algorithm was investigated. As result shows that the optimized Negative selection classification system gives the good accuracy of classification on IRIS plant dataset so this method of classification is used on different datasets.

For future scope in Negative Selection Algorithm the size of the detectors may also vary. Different optimization operator is also be used for finding the optimized data.

## 8. REFERENCES

[1] Yunfeng Xu, Ping Fan, and Ling Yuan. A Simple and Efficient Artificial Bee Colony Algorithm in Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2013.

[2] Ahmet Ozkis, Ahmet Babalik, 2014. Performance Comparison of ABC and A-ABC Algorithms on Clustering Problems. In Proceedings of the International Conference on Machine Vision and Machine Learning Prague, Czech Republic.

[3] Akay B., Karaboga D. 2012. A modified Artificial Bee Colony algorithm for real-parameter optimization, Information Sciences 192,120–142.

[4] Karaboga D., Basturk, B. 2007. A powerful and efficient algorithm for numerical function optimization: artificial

bee colony (ABC) algorithm in Journal of Global Optimization, 39(3), 459–471.

[5] Basturk B., Karaboga D. 2006. An Artificial Bee Colony (ABC) Algorithm for Numeric function Optimization in IEEE Swarm Intelligence Symposium, Indianapolis, Indiana, USA.

[6] Dervis Karaboga and Bahriye Basturk. 2007. Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems in Springer-Verlag Berlin Heidelberg.

[7] Pei-Wei Tsai, Jeng-Shyang Pan, Bin-Yih Liao<sup>1</sup>, and Shu-Chuan Chu. 2009. ENHANCED ARTIFICIAL BEE COLONY OPTIMIZATION in International Journal of Innovative Computing, Information and Control Volume 5, Number 12.

[8] Forrest S., Perelson A., Allen L.R. 1994. Self-nonsel discrimination in a computer, In Proceedings of the IEEE Symposium on Research in Security and Privacy, IEEE Computer Society Press, Los Alamitos, CA. pp 202–212.

[9] F. Gonzalez, D. Dasgupta. 2003. Anomaly detection using real-valued negative selection in Genetic Programming and Evolvable Machine 383–403.

[10] M. Bereta, T. Burczynski, 2009. Immune K-means and negative selection algorithms for data analysis. In Information Sciences 179 (10) 1407–1425.

[11] Victor. Onomza Waziri, Ismaila Idris, Mohammed Bashir Abdullahi, Hahimi Danladi, Audu Isah. 2013. A Negative Selection Algorithm Based on Email Classification Techniques in World of Computer Science and Information Technology Journal (WCSIT) Vol. 3, No. 3, 56-59.

[12] Ilhan Aydin, Mehmet Karakose, Erhan Akin. 2010. Chaotic-based hybrid negative selection algorithm and its applications in fault and anomaly detection in Expert Systems with Applications 37, 5285–5294.

[13] L. N. de Castro & F. J. V. Zuben. 2002. Learning and optimization using the clonal selection principles” in IEEE Transactions on Evolutionary Computation, 6(3), 239–251.

[14] K. Igawa & H. Ohashi. 2010. a negative selection algorithm for classification and reduction of noise effect” in Applied Soft Computing, 9(1), 431–438.

[15] Maoguo Gong, Jian Zhang, Jingjing Ma, Licheng Jiao. 2012. An efficient negative selection algorithm with further training for anomaly detection in Knowledge-Based Systems, 30, 185–191.

[16] D. Dasgupta, S. Yu, F. Nino. 2011. Recent advances in artificial immune systems in Applied Soft Computing 11 (2) 1574–1587.

[17] Z. Ji, D. Dasgupta. 2007. Revisiting negative selection algorithms in Evolutionary Computation 15 (2) 223–251.

[18] Z. Ji, D. Dasgupta. 2009. V-detector: an efficient negative selection algorithm with probably adequate detector coverage in Information Sciences 179 (10) 1390–1406.