International Journal of Digital Information and Wireless Communications (IJDIWC) 1(1): 284-292
The Society of Digital Information and Wireless Communications, 2011(ISSN 2225-658X)

Recognition of Plant Leaves Using the Dendritic Cell Algorithm

Esma Bendiab and Mohamed Kheirreddine Kholladi
MISC Laboratory, Department of Computer Science,
Mentouti University of Constantine, 25017, Algeria
Bendiab_e@yahoo.fr, Kholladi@yahoo.com

ABSTRACT

In this paper, we propose a new approach of plant classification based on leaves recognition. The approach uses the dendritic cell algorithm, from the danger theory, as a classifier and the wavelet transform as a features space. The dendritic cell algorithm is a new immune inspired paradigm. It has demonstrates a potential proprieties as a classifier. As for the wavelet transform, it provides a powerful tool to capture localized and attractive features for texture classification.
The experimental results about the data set leaf images show that the method is feasible and it gives an interesting recognition rate.

KEYWORDS

Artificial Immune System (AIS), Danger Theory, Dendritic Cell Algorithm (DCA), wavelet transform, leaves recognition.

1 INTRODUCTION

Artificial immune system (AIS) is a terminology that refers to adaptive systems inspired by theoretical and experimental immunology with the goal of solving problems. They encompass any system or computational tool that extracts ideas and metaphors from the biological immune system in order to solve real world problems[1-4].

Artificial immune systems incorporate the ability to learn new information, recall previously learned information and perform pattern matching in highly diversified manner. Researchers are making efforts to apply successfully AIS paradigm to various domains [1-3].

The first algorithms have been presented as implementations of artificial immune systems are [1,2]: the immune network, the clone selection algorithm and the negative selection algorithm.

Recently, a new paradigm, the Dendritic Cell Algorithm (DCA) was developed based on the danger theory [3]. It has become a popular immune inspired approach to solve a wide range of applications, particularly data classification.

The danger theory was born to address some issues raised concerning the validity of self non-self paradigm because of its inability to explain a number of documented phenomena, which will be listed in Section 2 [4]. The self non self paradigm discriminates between self entities and foreign ones.

The danger theory proposes that the immune system responds when damage to the host is detected, rather than discriminating between self and non-self proteins[4,6]. In this context, it is a matter of distinguishing non self but harmless from self but harmful invaders, termed: antigen. If the labels self and non self were to be replaced by interesting and non interesting data, a distinction would prove beneficial. In this case, the AIS is being applied as a classifier [6].

Besides, plant recognition is an important and challenging task [7-10] due to the lack of proper models or representation schemes. Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for plant classification. Sampling leaves and photogening them are low-cost and convenient. Moreover, leaves can be very easily found and collected everywhere.
By computing some efficient features of leaves and using a suitable pattern classifier, it is possible to recognize different plants successfully.

The literature reveals that the leaf features are very useful for plant recognition. We can especially mention [7-10]. In [7], authors proposed a classification method of plant classification based on wavelet transforms and support vector machines. The approach is not the first in this way, as authors in [8] have earlier used the support vector machines as an approach of plants recognition but using the colour and the texture features space. In [9], a method of recognizing leaf images based on shape features using and comparing three classifiers approaches was introduced. In [10], the author proposes a method of plants classification based on leaves recognition. Two methods called the gray-level co-occurrence matrix and principal component analysis algorithms have been applied to extract the leaves texture features.
This paper presents a new approach for classifying plant leaves based on the dendritic cell algorithm, from the danger theory, as a classifier and the wavelet transform as a space features. The wavelet transform [11] provides a powerful tool to capture localized features and gives developments for more flexible and useful representations.
Also, it presents constant analysis of a given signal by projection onto a set of basic functions that are scaled by means of frequency variation. Each wavelet is a shifted scaled version of an original or mother wavelet. These families are usually orthogonal to one another, important since this yields computational efficiency and ease of numerical implementation [7].

The material in this paper is arranged as the following order. In Section 2, the danger theory is briefly reviewed. In section 3, the dendritic cell algorithm is described. In section 4, the wavelet transform is presented. This is followed by Sections 5, where the proposed approach is described. The results obtained are discussed in section 6. Finally section 7 concludes this paper.

2 THE DANGER THEORY

The main goal of the immune system is to protect our bodies from invading entities, called: antigens, which cause damage and diseases.
Initially, immunologists believe that the protection was done by distinguishing self and non self inside the body and by eliminating the non self.
Matzinger's Danger Theory debates the self-nonself point of view. She points out that there must be discrimination

happening that goes beyond the self-non-self distinction, for instance [3]:

- There is no immune reaction to foreign bacteria in the guts or to the food which we eat although both of them are foreign entities.
  - The system does not govern to body changes, even self changes as well.
  - On the other hand, there are certain auto immune processes which are useful like some diseases and certain types of tumours that are fought by the immune system (both attacks against self) and successful transplants.

So, a new field in AIS emerges, baptized the danger theory, which offers an alternative to self non self discrimination approach. The danger theory stipulates that the immune response is done by reaction to a danger not to a foreign entity. In the sense, that the immune system is activated upon the receipt of molecular signals, which indicate damage or stress to the body, rather than pattern matching in the self non self paradigm. Furthermore, the immune response is done in reaction to signals during the intrusion and not by the intrusion itself.
These signals can be mainly of two nature [3,4]: safe and danger signal. The first indicates that the data to be processed, which represent antigen in the nature, were collected under normal circumstances; while the second signifies potentially anomalous data. The danger theory can be apprehended by: the Dendritic Cells Algorithm (DCA), which will be presented in the following section.

3 THE DENDRITIC CELL ALGORITHM

The Dendritic Cell Algorithm (DCA) is an immune inspired algorithm developed by Greensmith in [6,12]. The DCA which is an abstract model of dendritic cells (DCs) behavior is applied to a wide range of applications [12-17]. It is also used as a classifier for a static machine learning data set [12,13], where it was shown that the algorithm can process data classification.
Initial implementations of the DCA have provided promising classification accuracy results on a number of benchmark datasets. However, the basic DCA uses several stochastic variables which make its systematic analysis and functionality understanding very difficult. In order to overcome those problems, a DCA improvement was proposed [17]: the dDCA (deterministic DCA). In the this paper, we focus on the new version. Its Pseudo code can be found in [17].

The dDCA is based population algorithm in which each agent of the system is represented by a virtual cell.
The first stage of the DCA is an anomaly detection phase, where the population's classification decisions are monitored in order to identify anomalies within a given dataset. The second phase attempts to correlate the antigen sampled by the cells with the occurrence of detected anomalies [17].

The dDCA receives two types of input, namely signal and antigen. Signals are represented as vectors of real-valued numbers and are periodic measurements of features within the problem environment. All assumption made by the algorithm is that the presence or absence of an anomaly can be detected by observing these features. Antigen are symbols (typically represented as an enumerated type), which represent items of interest within the environment. It is assumed that some of the antigens have a causal relationship with observed anomalies. Signals can be on two kinds: 'safe' and 'danger' signal.

At each iteration t, the dDCA inputs consist of the values of the safe signal St, the danger signal Dt and antigens At. The dDCA proceeds on three steps as follows:

3.1 Initialization

The DC population and algorithm parameters are initialized and initial data are collected.

3.2 Signal Processing and Update phase

All antigens are presented to the DC population so that each DC agent samples only one antigen and proceeds to signal processing. At each step, each single cell *i* calculates two separate cumulative sums, called *CSMi* and *Ki*, and it places them in its own storage data structure. The values *CSM* and *K* can be given by Eq.(1) and (2) respectively :

$$\mathrm{CSM} = \mathrm{St} + \mathrm{Dt} \tag{1}$$

$$\mathrm{K} = \mathrm{Dt} - 2\mathrm{St} \tag{2}$$

This process is repeated until all presented antigens have been assigned to the population.
At each iteration, incoming antigens undergo the same process. All DCs will process the signals and update their values *CSMi* and *Ki*. If the antigens number is greater than the DC number only a fraction of the DCs will sample additional antigens.
The DC*i* updates and cumulates the values *CSMi* and *Ki* until a migration threshold *Mi* is reached. Once the *CSMi* is greater than the migration threshold *Mi*, the cell presents its temporary output *Ki* as an output entity *Kout*. For all antigens sampled by DC*i* during its lifetime, they are labeled as normal if *K*out < 0 and anomalous if *K*out > 0.
After recording results, the values of *CSMi* and *Ki* are reset to zero. All sampled antigens are also cleared. DC*i* then continues to sample signals and collect antigens as it did before until stopping criterion is met.

3.3 Aggregation phase

At the end, at the aggregation step, the nature of the response is determined by measuring the number of cells that are fully mature. In the original DCA, antigens analysis and data context evaluation are done by calculating the mature context antigen value (MCAV) average. A representation of completely mature cells can be done. An abnormal *MCAV* is closer to the value 1. This value of the *MCAV* is then thresholded to achieve the final binary classification of normal or anomalous. The $K\alpha$ metric, an alternative metric to the *MCAV* , was proposed with the dDCA in [21]. The $K\alpha$ uses the average of all output values *Kout* as the metric for each antigen type, instead of thresholding them to zero into binary tags.

4 THE WAVELET TRANSFORM

Over the last decades, the wavelet transform has emerged as a powerful tool for the analysis and decomposition of signals and images at multi-

resolutions. It is used for noise reduction, feature extraction or signal compression. The wavelets transform proceeds by decomposing a given signal into its scale and space components.
Information can be obtained about both the amplitude of any periodic signal as well as when/where it occurred in time/space. Wavelet analysis thus localizes both in time/space and frequency [11].
The wavelet transform can be defined as the decomposition of a signal g (t) using a series of elemental functions called: wavelets and scaling factors.
The approximation is then split itself into a second level of approximation and detail. The image is usually segmented into a so-called approximation image and into so-called detail images. The transformed coefficients in approximation and detail sub images are the essential features, which are as useful for image classification. A tree wavelet package transform can be constructed [11]. Where *S* denotes the signal, *D* denotes the detail and *A* the approximation, as shown in figure 1.

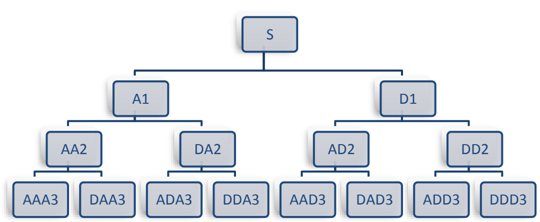

Figure 1. The tree-structured wavelets transform

For a discrete signal, the decomposition coefficients of wavelet packets can be computed iteratively by Eq. (3) and (4):

$$x_{2n,j+1}^{k} = \sum_{l} h_{l-2k}\, x_{n,j}^{l} \qquad (3)$$

$$x_{2n+1,j+1}^{k} = \sum_{l} g_{l-2k}\, x_{n,j}^{l} \qquad (4)$$

Where: $x_{n,j} = \{x_{n,j}^{k}\}$ is the decomposition coefficient sequence of the nth node at level j of the wavelet packet tree.

5 A METHOD OF LEAF CLASSIFICATION

An approach based on artificial immune system ought to describe two aspects:

1. The projection and models advocating of immune elements in the real world problem.
2. The use of the appropriate immune algorithm or approach to solve the problem

.
These two aspects are presented in following sections.

5.1 Immune Representation Using the dDCA

For sake of clarity, before describing the immune representation, we must depict the feature space. In this paper, we consider the decomposition using the wavelet package transform in order to get the average energy [11]. This is as follows:
The texture images are decomposed using the wavelet package transform. Then, the average energy of approximation and detail sub-image of two level decomposed images are calculated as features using the formulas given in (5) as follows:

$$E = \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{j=1}^{N} |f(x,y)| \qquad (5)$$

Where: N denotes the size of sub-image, f (x, y) denotes the value of an image pixel.
Now, we describe the different elements used by the dDCA for image classification:

Antigens: In AIS, antigens symbolize the problem to be resolved. In our approach, antigens are leaves images set to be classified. We consider the average energy of wavelet transform coefficients as features.

For texture classification, the unknown texture image is decomposed using wavelet package transform and a similar set of average energy features are extracted and compared with the corresponding feature values which are assumed to be known in a priori using a distance vector formula, given in Eq.6:

$$D(j) = \sum_{i=1}^{N} abs\,[f_i(x) - f_j(j)] \quad (6)$$

Where; fi (x) represents the features of unknown texture, while fi(j) represents the features of known *jth* texture. So:

Signals: Signals input correspond to information set about a considered class. In this context, we suggest that:

1. Danger signal: denote the distance between an unknown leaf texture features and known j texture features.
2. Safe signal: denote the distance between an unknown leaf texture features and known j' texture features.

The two signals can be given by Ddanger and Dsafe as described in Eq. (7) and (8) at the manner of Eq. (6)

Danger signal =

$$D_{danger} = \sum_{i=1}^{N} abs\,[f_i(x) - f_j(j)] \quad (7)$$

Safe signal=

$$D_{safe} = \sum_{i=1}^{N} abs\,[f_i(x) - f_j(j')] \quad (8)$$

5.2 Outline of the Proposed Approach

In this section, we describe the proposed approach in the context of leaves image classification. The approach operates as follows:

5.2.1 Initialisation

At first, the system is initialized by setting various parameters, such: Antigens collection and signals input construction. At the same time of collecting leaves image, signals are constructed progressively.
The known leaves images, selected from labelled set, are decomposed using the wavelet package transform. Then, the average energy of approximation and detail sub-image of two level decomposed images are calculated as features using the formulas given Eq. (5).

Each leaf image (antigen), collected from the leaves image collection, is decomposed using wavelet package transform and a similar set of average energy features are extracted, (two labelled images selected randomly) and compared with the corresponding feature values which are assumed to be known in a priori using a distance vector formula, given in Eq. 6, in order to construct danger Ddanger and the safe Dsafe signals as in Eq. 7 and 8. Both streams are presented to the dDCA.

5.2.2 Signal Processing and Update Phase

Data Update:
we collect leaves image and we choose randomly two images from labelled images set. Then, we assess the danger Ddanger and the safe Dsafe signals, as given in Eq.7 and 8. Both streams are presented to the dDCA. (This process is repeated until the number of images present at each time *i*, is assigned to all the DC population).

Cells Cycle: The DC population is presented by a matrix, in which rows correspond to cells. Each row-cell *i* has a maturation mark *CSMi* and a temporary output *Ki*.

For each cell *i*, a maturation mark *CSMi* is evaluated and a cumulatively output signal *Ki* is calculated as follows:

CSMi = Ddanger *t* + Dsafe *t* and
Ki = Ddanger *t* − 2 Dsafe *t*

When data are present, cells cycle is continually repeated. Until the maturation mark becomes greater than a migration threshold *Mi* (*CSMi* > *Mi*). Then, the cell prints a context: *Kout*, it is removed from the sampling population and its contents are reset after being logged for the aggregation stage. Finally, the cell is returned to the sampling population.
This process is repeated (cells cycling and data update) until a stopping criteria is met. In our case, until the iteration number is met.

5.2.3 Aggregation Phase

At the end, at the aggregation phase, we analyse data and we evaluate their contexts.

In this work, we consider only the *MCAV* metric (the Mature Context Antigen Value), as it generates a more intuitive output score. We calculate the mean mature context value (MCAV: The total fraction of mature DCs presenting said leaf image is divided by the total amount of times by which the leaf image was presented. So, semi mature context indicates that collected leaf is part of the considered class. While, mature context signifies that the collected leaf image is part of another class.
More precisely, the MCAV can be evaluated as follows: for all leaves images in the total list, leaf type count is incremented. If leaf image context equals one, the leaf type mature count is incremented. Then, for all leaves types, the MCAV of leaf type is equal to mature count / leaf count.

6 RESULTS AND DISCUSSION

In our approach, the classifier needs more information about classes in order to give a signification indication about the image context. For this, we have used a set of leaves images. The samples typically include different green plants, with simple backgrounds, which imply different colour and texture leaves, with varying lighting conditions.
Thus, in order to form signals inputs. The collection is presented during the run time with the image to be classified.
During the experiment, we select 10 kinds of plants with 100 leaf images for each plant. Leaves images database is a set of web collection, some samples are shown in figure 2. The size of the plant leaf images is 240∗240. The following experiments are designed for testing the accuracy and efficiency of the proposed method. The experiments are programmed using Matlab 9.

Algorithm parameters are important part in the classification accuracy. Hence, we have considered 100 cells agent in the DC population and 100 iterations as stopping criteria which coincides to the leaves images number. The maturation mark is evaluated by *CSMi*. For an unknown texture of a leaf image, if *CSMi=Ddanger+Dsafe=Ddanger*. the unknown texture have a high chance to be classified in the *jth* texture, if the distance *D* (*j*) is minimum among all textures.

As far as, if *CSMi=Ddanger+Dsafe =Dsafe* the unknown texture have a high chance to be classified in the *j' th* texture, if the distance *D* (*j'*) is the minimum.

To achieve a single step classification, a migration threshold *Mi* is introduced that can take care of data in overlapping the different leaves texture. The migration threshold *Mi* is fixed to one of the input signals. In the sense that if *CSMi* tends towards one of the two signals, this is implies that one of the two signals tends to zero.

So, we can conclude that the pixel have more chance to belong to one of the signals approaching zero.

In order to evaluate the pixel membership to a class, we assess the metric *MCAV*.

Each leaf image is given a *MCAV* coefficient value which can be compared with a threshold. In our case, the threshold is fixed at 0,90. Once a threshold is applied, it is then possible to classify the leaf. Therefore, the relevant rates of true and false positives can be shown.

The use of the wavelet transform to evaluate texture features and the proposed classifier, enhance the performance of our system. The recognition accuracy is about 94%.

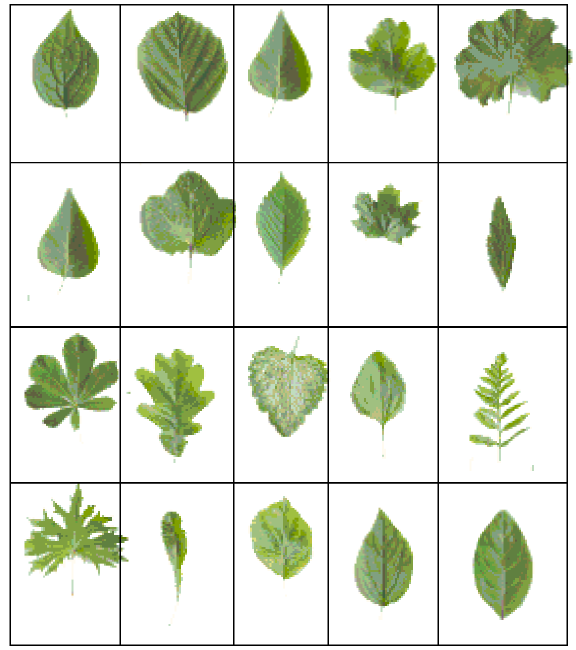

Figure 2. Samples of images used in tests

7 CONCLUSION AND FUTURE WORKS

In this paper, recognition of plant leaves was proposed. The approach is based on the deterministic dendritic cell algorithm, a new immune inspired algorithm which shows interesting proprieties as classifier. The classification process is performed at the basis of texture features which are given using the average energy of approximation and detail sub-image of two level decomposed leaf images, obtained by the wavelet transform.

We have presented our preliminary results obtained in this way. Different species of plants was classified and the rate of recognizing using this method is satisfying. The experimental results indicate that our algorithm is workable with a recognition rate greater than 94% on 10 kinds of plant leaf images.

Our future work includes selecting most suitable parameters and texture features, as well as the leaves shapes beside leaves textures.

8 REFERENCES

1. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006).

1. De Castro, L., Timmis, J. (eds.): Artificial Immune Systems: A New Computational Approach. Springer, London (2002)

2. Hart, E., Timmis, J.I.: Application Areas of AIS: The Past, The Present and The Future. In: Jacob, C., Pilat, M.L., Bentley, P.J., Timmis, J.I. (eds.) ICARIS 2005. LNCS, vol. 3627, pp. 483–497. Springer, Heidelberg (2005)
3. Aickelin, U., Bentley, P.J., Cayzer, S., Kim, J., McLeod, J.: Danger theory: The link between AIS and IDS? In: Timmis, J., Bentley, P.J., Hart, E. (eds.) ICARIS 2003. LNCS, vol. 2787, pp. 147–155. Springer, Heidelberg (2003)
4. Aickelin, U., Cayzer, S.: The danger theory and its application to artificial immune systems. In: The 1th International Conference on Artificial Immune Systems (ICARIS 2002), Canterbury, UK, pp. 141–148 (2002)
5. Dasgupta, D.: Artificial Immune Systems and their applications. Springer, Heidelberg (1999)
6. Greensmith, J.: The Dendritic Cell Algorithm. University of Nottingham (2007)
7. Liu, J., Zhang, S., Deng, S.: A Method of Plant Classification Based on Wavelet Transforms and Support Vector Machines. In: Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J., Bevilacqua, V. (eds.) ICIC 2009. LNCS, vol. 5754, pp. 253–260. Springer, Heidelberg (2009)
8. Man, Q.-K., Zheng, C.-H., Wang, X.-F., Lin, F.-Y.: Recognition of Plant Leaves Using Support Vector Machine. In: Huang, D.-S., et al. (eds.) ICIC 2008. CCIS, vol. 15, pp. 192–199. Springer, Heidelberg (2008)
9. Singh, K., Gupta, I., Gupta, S.: SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape. International Journal of Signal Processing, Image Processing and Pattern Recognition 3(4) (December 2010)
10. Ehsanirad, A.: Plant Classification Based on Leaf Recognition. International Journal of Computer Science and Information Security 8(4) (July 2010)
11. Zhang, Y., He, X.-J., Huang, J.-H.H.D.S., Zhang, X.-P., Huang, G.-B.: Texture Feature-Based Image Classification Using Wavelet Package Transform. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 165–173. Springer, Heidelberg (2005)
12. Greensmith, J., Aickelin, U., Cayzer, S.: Introducing Dendritic Cells as a Novel Immune- Inspired Algorithm for Anomaly Detection. In: Jacob, C., Pilat, M.L., Bentley, P.J., Timmis, J.I. (eds.) ICARIS 2005. LNCS, vol. 3627, pp. 153–167. Springer, Heidelberg (2005)
13. Oates, R., Greensmith, J., Aickelin, U., Garibaldi, J., Kendall, G.: The Application of a Dendritic Cell Algorithm to a Robotic Classifier. In: The 6th International Conference on Artificial Immune (ICARIS 2006), pp. 204–215 (2007)
14. Greensmith, J., Twycross, J., Aickelin, U.: Dendritic Cells for Anomaly Detection. In: IEEE World Congress on Computational Intelligence, Vancouver, Canada, pp. 664–671 (2006b)
15. Greensmith, J., Twycross, J., Aickelin, U.: Dendritic cells for anomaly detection. In: IEEE Congress on Evolutionary Computation (2006)
16. Greensmith, J., Aickelin, U., Tedesco, G.: Information Fusion for Anomaly Detection with the Dendritic Cell Algorithm. Journal Information Fusion 11(1) (January 2010)
17. Greensmith, J., Aickelin, U.: The deterministic dendritic cell algorithm. In: Bentley, P.J., Lee, D., Jung, S. (eds.) ICARIS 2008. LNCS, vol. 5132, pp. 291–302. Springer, Heidelberg (2008)