A Fuzzy-Rough Data Pre-processing Approach for the Dendritic Cell Classifier

Zeineb Chelly and Zied Elouedi

LARODEC, University of Tunis, High Institute of Management of Tunis, Tunisia
zeinebchelly@yahoo.fr, zied.elouedi@gmx.fr

Abstract. The Dendritic Cell Algorithm (DCA) is an immune inspired classification algorithm based on the behavior of natural dendritic cells. The DCA performance relies on its data pre-processing phase based on the Principal Component analysis (PCA) statistical method. However, using PCA presents a limitation as it destroys the underlying semantics of the features after reduction. One possible solution to overcome this limitation was the application of Rough Set Theory (RST) in the DCA data pre-processing phase; but still the developed rough DCA approach presents an information loss as data should be discretized beforehand. Thus, the aim of this paper is to develop a new DCA data pre-processing method based on Fuzzy Rough Set Theory (FRST) which allows dealing with real-valued data with no data quantization beforehand. In this new fuzzy-rough model, the DCA data pre-processing phase is based on the FRST concepts; mainly the fuzzy lower and fuzzy upper approximations. Results show that applying FRST, instead of PCA and RST, to DCA is more convenient for data pre-processing yielding much better performance in terms of accuracy.

Keywords: Dendritic cell algorithm, Fuzzy rough set theory, Feature selection, Classification.

1 Introduction

The Dendritic Cell Algorithm (DCA) [1] is a bio-inspired classification binary algorithm derived from behavioral models of natural dendritic cells (DCs) [2]. DCA has the ability to combine a series of informative signals with a sequence of repeating abstract identifiers, termed "antigens", to perform anomaly detection. To achieve this and through the pre-processing phase, DCA selects a subset of features and categorizes each selected feature into one of three signal types which are defined as "Danger Signal" (DS), "Safe Signal" (SS) and as "Pathogen-Associated Molecular Pattern"(PAMP). The resulting combination signal values are then classified to form an anomaly detection style of two-class classification.

Initially, in [3], the principal component analysis (PCA) statistical method was introduced in the DCA data pre-processing phase which is composed of two main sub-steps; namely feature reduction and signal categorization. The use of PCA aims to automatically reduce data dimension by generating new

L.C. van der Gaag (Ed.): ECSQARU 2013, LNAI 7958, pp. 109–120, 2013.
© Springer-Verlag Berlin Heidelberg 2013

features to retain, which is achieved throughout the first sub-step, and to perform their categorization to their specific signal types (SS, DS, and PAMP), which is achieved throughout the second sub-step. However, applying PCA in the DCA data pre-processing step, destroys the underlying meaning behind the features present, initially, in the input database; seen as an undesirable property for the DCA [4].

To have a more reliable data pre-processing phase and to overcome the PCA limitation, in [4], a rough DCA version was introduced. The algorithm, named RC-DCA, is based on the application of Rough Set Theory (RST) [5] for the DCA data pre-processing task. To select features and based on the RST concepts, RC-DCA selects the most informative attributes, a subset termed *reduct*, that preserve nearly the same classification power of the original database. Furthermore, in RC-DCA, the signal categorization step is based on the RST *Reduct* and *Core* concepts. It was shown, in [4], that applying RST, instead of PCA, to DCA is more convenient for data pre-processing yielding much better performance in terms of accuracy.

However, based on rough set theory and to perform feature selection, the attribute values of the input database should be discretized beforehand. Thus, important information may be lost as a result of quantization [6]. Formally, in most databases, the attribute values may be real, and this is where RST encounters a problem. It is not possible within this theory to say whether two attribute values are similar and to what extent they are the same [6]. For instance, two close values may only differ as a result of noise, but in RST they are considered to be as different as two values of a different order of magnitude. One answer to this problem has been to discretize the dataset beforehand, producing a new database with crisp values. This is often still inadequate as it is a source of information loss; which is against the rough set objective of retaining information content [6]. This information loss may influence the RC-DCA feature selection process by generating an incorrect set of selected features; as a consequence, this will misguide the algorithm categorization phase by categorizing the features to erroneous signal categories. As a result, this will influence the algorithm classification process by generating unreliable classification results.

To overcome the RST applicability restriction, Fuzzy Rough Set Theory (FRST) [7] was introduced as a data reduction technique dealing with crisp and real-valued attributed datasets. FRST, which utilizes the extent to which values are similar, encapsulates the related but distinct concepts of vagueness (for fuzzy sets) and indiscernibility (for rough sets), both of which occur as a result of uncertainty in data; a method employing fuzzy-rough sets can handle this uncertainty. We, therefore, in this paper, propose to develop a novel fuzzy-rough DCA model based on a new feature selection and signal categorization technique. Our fuzzy-rough DCA classification model, named FBR-DCA, is based on the use of fuzzy rough set theory and more precisely on the use of the fuzzy boundary region (FBR); to guarantee a more rigorous data pre-processing phase.

The major contributions of this paper are to introduce the concept of FRST in the DCA data pre-processing phase and to show how FRST can be applied to search for the most convenient set of features to select. Additionally, we aim to show how the application of FRST can be appropriate for the categorization of each selected feature to its right type of signal. This will be achieved by avoiding the information loss already discussed, by keeping the semantics of the initial attributes and with no need for a quantization process beforehand.

2 The Dendritic Cell Algorithm

DCA is a population based system, with each agent in the system is represented as a cell. Each cell has the capacity to collect data items, termed *antigens*. Formally, the DCA initial step is the automatic data pre-processing phase where feature selection and signal categorization are achieved. More precisely, DCA selects the most important features, from the initial input database, and assigns each selected attribute to its specific signal category (SS, DS or PAMP). To do so, the PCA was used. Once data pre-processing is achieved and after calculating the values of the safe, PAMP and DS signals [8], DCA adheres these three signal categories and antigen to fix the context of each object (DC) which is the step of *Signal Processing*.

In fact, the algorithm processes its input signals (already pre-categorized) in order to get three output signals: costimulation signal (Csm), semi-mature signal (Semi) and mature signal (Mat) [8]. A migration threshold is incorporated into the DCA in order to determine the lifespan of a DC. As soon as the Csm exceeds the migration threshold; the DC ceases to sample signals and antigens. The migration state of a DC to the semi-mature state or to the mature state is determined by the comparison between cumulative $Semi$ and cumulative Mat. If the cumulative $Semi$ is greater than the cumulative Mat, then the DC goes to the semi-mature context, which implies that the antigen data was collected under normal conditions. Otherwise, the DC goes to the mature context, signifying a potentially anomalous data item. This step is known to be the *Context Assessment* phase.

The nature of the response is determined by measuring the number of DCs that are fully mature and is represented by the Mature Context Antigen Value (MCAV). $MCAV$ is applied in the DCA final step which is the *Classification* procedure and used to assess the degree of anomaly of a given antigen. The closer the $MCAV$ is to 1, the greater the probability that the antigen is anomalous. By applying thresholds at various levels, analysis can be performed to assess the anomaly detection capabilities of the algorithm. Those antigens whose $MCAV$ are greater than the anomalous threshold, which can be automatically generated from the input data, are classified as anomalous while the others are classified as normal. For a detailed description of the DCA and its implementation, please, refer to [8].

3 Rough Sets and Fuzzy-Rough Sets for Feature Selection

3.1 Fundamentals of Rough Set Theory

In Rough Set Theory (RST) [5], an *information table* is defined as a tuple $T = (U, A)$ where U and A are two finite, non-empty sets, U the *universe* of primitive objects and A the set of attributes. A may be partitioned into C and D, called *condition* and *decision* attributes, respectively. Let $P \subset A$ be a subset of attributes. The indiscernibility relation, $IND(P)$, is an equivalence relation defined as: $IND(P) = \{(x, y) \in U^2 : \forall a \in P, a(x) = a(y)\}$, where $a(x)$ denotes the value of feature a of object x. The family of all equivalence classes of $IND(P)$ is denoted by $U/IND(P)$. Equivalence classes $U/IND(C)$ and $U/IND(D)$ are respectively called *condition* and *decision* classes. For any concept $X \subseteq U$ and attribute subset $R \subseteq A$, X could be approximated using only the information contained within P by constructing the *P-lower* and the *P-upper approximations* of X defined as $\underline{P}(X) = \{x \in U | [x]_p \subseteq X\}$ and $\overline{P}(X) = \{x \in U | [x]_p \cap X \neq \emptyset\}$, respectively. The lower approximation of X is the set of objects of U that are surely in X and the upper approximation of X is the set of objects of U that are possibly in X. The tuple $< \underline{P}(X), \overline{P}(X) >$ is called a *rough set*. Let P and Q be sets of attributes inducing equivalence relations over U, then the *positive region* can be defined as: $POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}(X)$. The positive region contains all objects of U that can be classified into classes of U/Q using the information in attribute P.

For feature selection, RST defines the *core* and the *reduct* concepts. The core is equivalent to the set of features which are *indispensable* attributes that cannot be removed without loss of prediction accuracy of the original database. The reduct is a combination of all these features and some features that can sometimes contribute to prediction accuracy. In RST, a subset $R \subseteq C$ is said to be a D-*reduct* of C if $POS_R(D) = POS_C(D)$ and there is no $R' \subset R$ such that $POS_{R'}(D) = POS_C(D)$. There may exist a family (F) of reducts, $RED_D^F(C)$, in T. The core is the set of attributes that are contained in all reducts, defined as: $CORE_D(C) = \bigcap RED_D^F(C)$.

3.2 Fundamentals of Fuzzy Rough Set Theory

Fuzzy Rough Set Theory (FRST) [7] comes as an extension to RST as this latter theory can only operate effectively with datasets containing discrete values. As most datasets contain real-valued attributes, it is necessary to perform a discretization step beforehand. To avoid this information loss, fuzzy rough set theory is applied.

Basic Concepts. In the same way that crisp equivalence classes are central to rough sets, *fuzzy equivalence classes* are central to the fuzzy-rough set approach. For typical applications, this means that the decision values and the conditional values may all be fuzzy. The concept of crisp equivalence classes can be extended by the inclusion of a fuzzy similarity relation S on the universe, which

determines the extent to which two elements are similar in S. The fuzzy lower and fuzzy upper approximations become $\mu_{\underline{R_P}X}(x) = inf_{y \in U} I(\mu_{R_P}(x,y), \mu_X(y))$ and $\mu_{\overline{R_P}X}(x) = sup_{y \in U} T(\mu_{R_P}(x,y), \mu_X(y))$. In the presented formulae, I is a fuzzy implicator and T is a t-norm. R_P is the fuzzy similarity relation induced by the subset of features P: $\mu_{R_P}(x,y) = \bigcup_{a \in P}\{\mu_{R_a}(x,y)\}$ where $\mu_{R_a}(x,y)$ is the degree to which objects x and y are similar for feature a. A fuzzy similarity relation can be constructed for this purpose, defined as: $\mu_{R_a}(x,y) = max(min(\frac{(a(y)-(a(x)-\sigma_a))}{(a(x)-(a(x)-\sigma_a))}, \frac{((a(x)+\sigma_a)-a(y))}{((a(x)+\sigma_a)-a(x))}), 0)$ where σ_a is the standard deviation of feature a. The tuple $< \underline{P}(X), \overline{P}(X) >$ is called a *fuzzy-rough set*. The difference between the fuzzy lower approximation, containing information regarding the extent of certainty of object membership to a given concept, and the fuzzy upper approximation, containing information regarding the degree of uncertainty of objects, generates the *fuzzy boundary region*; defined as: $\mu_{BND_{R_P}(X)}(x) = \mu_{\overline{R_P}X}(x) - \mu_{\underline{R_P}X}(x)$. This subset contains objects within the boundary region with less uncertainty.

Reduction Process. To search for the optimal subset of features, the fuzzy-rough reduct, the uncertainty for every concept has to be calculated. The uncertainty for a concept X using features in P can be calculated as follows: $U_P(X) = \frac{\sum_{x \in U} \mu_{BND_{R_P}(X)}(x)}{|U|}$. This is the average extent to which objects belong to the fuzzy boundary region for the concept X. The total uncertainty degree for all concepts, given a feature subset P, is defined as: $\gamma'_P(Q) = \frac{\sum_{X \in U/Q} U_P(X)}{|U/Q|}$.

A Fuzzy-Rough QuickReduct algorithm, defined in Fig.1, can be constructed for locating a fuzzy-rough reduct based on this measure. The task of the algorithm is to minimize the total uncertainty degree. When this reaches the minimum for the dataset, a fuzzy-rough reduct has been found. A worked example on how to compute a fuzzy-rough reduct using the Fuzzy-Rough QuickReduct algorithm, based on the fuzzy boundary region, can be found in [9].

FRQUICKREDUCT($\mathbb{C}$,$\mathbb{D}$).
$\mathbb{C}$, the set of all conditional attributes;
$\mathbb{D}$, the set of decision attributes.

(1) $R \leftarrow \{\}$; $\gamma'_{best} = 0$; $\gamma'_{prev} = 0$
(2) **do**
(3) $T \leftarrow R$
(4) $\gamma'_{prev} = \gamma'_{best}$
(5) **foreach** $x \in (\mathbb{C} - R)$
(6) **if** $\gamma'_{R \cup \{x\}}(\mathbb{D}) > \gamma'_T(\mathbb{D})$
(7) $T \leftarrow R \cup \{x\}$
(8) $\gamma'_{best} = \gamma'_T(\mathbb{D})$
(9) $R \leftarrow T$
(10) **until** $\gamma'_{best} == \gamma'_{prev}$
(11) **return** R

Fig. 1. Fuzzy-Rough QuickReduct algorithm

4 FBR-DCA: The Fuzzy-Rough Solution Approach

In this Section, we focus mainly on our FBR-DCA data pre-processing step as the rest of the fuzzy-rough FBR-DCA steps including Signal Processing, Context Assessment and the Classification procedure are performed the same as the standard DCA and as described, previously, in Section 2.

4.1 The FBR-DCA Signal Selection Process

For antigen classification, our learning problem has to select high discriminating features from the original input database which corresponds to the antigen information dataset. We may formalize this problem as an information table, where universe $U = \{x_1, x_2, \ldots, x_N\}$ is a set of antigen identifiers, the conditional attribute set $C = \{c_1, c_2, \ldots, c_A\}$ contains each feature of the information table to select and the decision attribute D of our learning problem corresponds to the class label of each sample. As FBR-DCA is based on the standard DCA concepts, except for the data pre-processing phase, and since DCA is applied to binary classification problems; then our developed FBR-DCA will be, also, applied to two-class datasets. Therefore, the decision attribute, D, of the input database of our FBR-DCA has binary values d_k: either the antigen is collected under safe circumstances reflecting a normal behavior (classified as normal) or the antigen is collected under dangerous circumstances reflecting an anomalous behavior (classified as anomalous). The condition attribute feature D is defined as follows: $D = \{normal, anomalous\}$.

For feature selection, FBR-DCA has to determine, first of all, the fuzzy boundary region for both concepts, the two-class labels, d_k. To do so, the fuzzy lower and the fuzzy upper approximations of each concept d_k for each feature c_i and for all objects x_j must be calculated. The fuzzy boundary region, the fuzzy lower and the fuzzy upper approximations are denoted by: $\mu_{BND_{R_{c_i}}(d_k)}(x_j)$, $\mu_{\underline{R_{c_i}}(\{d_k\})}(x_j)$ and $\mu_{\overline{R_{c_i}}(\{d_k\})}(x_j)$, respectively. Once the fuzzy boundary regions are measured, FBR-DCA calculates the uncertainty degrees for each attribute c_i for each concept d_k, denoted by $U_{c_i}(d_k)$, as presented in Section 3.

To find the fuzzy-rough reduct, FBR-DCA starts off with an empty set and moves to calculate the total uncertainty degrees for each feature c_i; defined as $\gamma'_{c_i}(D)$. The attribute c_m having the smallest total uncertainty degree among all the calculated total uncertainty degrees of the remaining features is added to the empty fuzzy-rough reduct set. Once the first attribute c_m is selected, FBR-DCA adds, in turn, one attribute to the selected first attribute and computes the total uncertainty degrees of each obtained attributes' couple $\gamma'_{\{c_m,c_i\}}(D)$. The algorithm chooses the couple having the smallest total uncertainty degree. The process of adding each time one attribute to the subset of the selected features continues until the total uncertainty degree of the obtained subset results in the minimal uncertainty for the dataset.

The generated subset of the selected features, constituting the fuzzy-rough reduct, shows the way of reducing the dimensionality of the original dataset by eliminating those conditional attributes that do not appear in the set. Those discarded attributes are removed in each FBR-DCA computation level since they do not add anything new to the target concept nor help the FBR-DCA to perform well its classification task. In fact, the obtained fuzzy-rough reduct includes the most informative features that preserve nearly the same classification power of the original dataset. Using the fuzzy-rough reduct concept, our method can guarantee that attributes of extracted feature patterns will be the most relevant for the FBR-DCA classification task.

4.2 The FBR-DCA Signal Categorization Process

The second step of our FBR-DCA data pre-processing phase is signal categorization. More precisely, our method has to assign for each selected attribute, produced by the previous step and which is included in the generated fuzzy-rough reduct, its definite and specific signal category. The general guidelines for signal categorization are based on the semantic of each signal type [1]:

- Safe signals: Certainly indicate that no anomalies are present.
- PAMPs: Usually mean that there is an anomalous situation.
- Danger signals: May or may not show an anomalous situation, however the probability of an anomaly is higher than under normal circumstances.

From the definitions stated above, both PAMP and SS are positive indicators of an anomalous and normal situation while the DS is measuring situations where the risk of anomalousness is high, but there is no signature of a specific cause. In other words, PAMP and SS have a certain final context (either an anomalous or a normal behavior) while the DS cannot specify exactly the final context to assign to the collected antigen. This is because the information returned by the DS is not certain as the collected antigen may or may not indicate an anomalous situation. This problem can be formulated as follows:

Based on the semantics of the mentioned signals, a ranking can be performed for these signals. More precisely, both SS and PAMP are more informative than DS which means that both of these signals can be seen as indispensable attributes; reflecting the first and the second ranking positions. To represent this level of importance, our method uses the first obtained couple of features through the fuzzy-rough reduct generation. On the other hand, DS is less informative than PAMP and SS; reflecting the last and third ranking position. Therefore, our method applies the rest of the fuzzy-rough reduct attributes, discarding the two first selected attributes that are chosen to represent the SS and PAMP signals, to represent the DS. More precisely, our method processes as follows:

As FBR-DCA has already calculated the total uncertainty degree of each attribute c_i a part, $\gamma'_{c_i}(D)$, FBR-DCA selects the first attribute c_m having the smallest total uncertainty degree to form the SS as it is considered the most informative first feature added to the fuzzy-rough reduct. With no additional

computations and since FBR-DCA has already computed the total uncertainty degree of each attributes' couple $\gamma'_{\{c_m,c_i\}}(D)$ when adding, in turn, one attribute c_i to the selected first attribute c_m that represents the SS, FBR-DCA chooses the couple having the smallest total uncertainty degree. More precisely, FBR-DCA selects that second attribute c_r having the smallest $\gamma'_{\{c_m,c_r\}}(D)$ among the calculated $\gamma'_{\{c_m,c_i\}}(D)$; to form the PAMP signal. Finally, the rest of the fuzzy-rough reduct attributes are combined and affected to the DS as it is less than certain to be anomalous.

Once the selected features are assigned to their suitable signal types, our method calculates the values of each signal category using the same process as the standard DCA [8]. The output is, thus, a new information table which reflects the signal database. In fact, the universe U of the induced signal dataset is $U = \{x'_1, x'_2, \ldots, x'_N\}$ a set of antigen identifiers and the conditional attribute set $C = \{SS, PAMP, DS\}$ contains the three signal types: SS, PAMP and DS. Once data pre-processing is achieved, FBR-DCA processes its next steps which are the Signal Processing, the Context Assessment and the Classification phase as the DCA does and as described in Section 2.

5 Experimental Setup

To test the validity of our FBR-DCA fuzzy-rough model, our experiments are performed on two-class, real-valued attributes, databases from [10]. The used databases are described in Table 1.

Table 1. Description of Databases

Database	Ref	♯ Instances	♯ Attributes
Sonar	SN	208	61
Molecular-Bio	Bio	106	59
Spambase	SP	4601	58
Cylinder Bands	CylB	540	40
Chess	Ch	3196	37
Ionosphere	IONO	351	35
Sick	Sck	3772	30
Mushroom	Mash	8124	23
Horse Colic	HC	368	23
German-Credit	GC	1000	21
Red-White-Win	RWW	6497	13

It is likely that not all of the attributes presented in the mentioned databases, are required to determine the class of each instance. Hence, feature selection, which is the first sub-step of the DCA data pre-processing phase, is needed. In [4], this is achieved by applying RST. However, as the datasets are entirely composed of real-valued attributes, discretization had to be performed. This is clearly a potential source of information loss. By applying the present work,

FBR-DCA, such loss can be reduced as attribute values are kept unchanged; no quantization is performed on the original databases. We try to show that our FBR-DCA can operate well in case of real-valued attributes avoiding the mentioned information loss while generating better classification results than when applying the crisp rough set theory. Thus, we will compare our FBR-DCA model to the crisp rough DCA approach, RC-DCA. Note that FBR-DCA and RC-DCA are based on the same concepts, except for the data pre-processing phase, as the standard DCA version, PCA-DCA. For data pre-processing, FBR-DCA applies FRST, RC-DCA applies RST and the standard DCA applies PCA.

For the DCA approaches, namely FBR-DCA, RC-DCA and PCA-DCA, each data item is mapped as an antigen, with the value of the antigen equal to the data ID of the item. For all DCA algorithms, a population of 100 cells is used. The migration threshold of an individual DC is set to 10. To perform anomaly detection, a threshold which is automatically generated from the data is applied to the MCAVs. The MCAV threshold is derived from the proportion of anomalous data instances of the whole dataset. Items below the threshold are classified as class one and above as class two. The resulting classified antigens are compared to the labels given in the original datasets. For each experiment, the results presented are based on mean MCAV values generated across a 10-fold cross validation.

We evaluate the performance of the DCA methods in terms of number of extracted features, running time, sensitivity, specificity and accuracy which are defined as: $Sensitivity = TP/(TP + FN)$; $Specificity = TN/(TN + FP)$; $Accuracy = (TP + TN)/(TP + TN + FN + FP)$; where TP, FP, TN, and FN refer respectively to: true positive, false positive, true negative and false negative. We will, also, compare the classification performance of our FBR-DCA method to well known classifiers which are the Support Vector Machine (SVM), Artificial Neural Network (ANN) and the Decision Tree (DT) and to the standard DCA version, PCA-DCA. The parameters of SVM, ANN and DT are set to the most adequate parameters to these algorithms using the Weka software. All experiments are run on a Sony Vaio G4 2.67 Ghz machine.

FRST has been experimentally evaluated with other leading feature selection techniques, such as Relif-F and entropy-based approaches in [11], and has been shown to outperform these in terms of resulting classification performance. Hence, only comparison to fuzzy rough set theory and rough set theory are given here. In addition, in [4], it was already shown that RC-DCA outperforms PCA-DCA. Thus, comparisons are made between FBR-DCA and RC-DCA.

6 Results and Analysis

Let us remind that the first step of the DCA classification algorithm is data pre-processing which is based on the use of PCA [3]. In [4], results showed that applying PCA for both feature selection and signal categorization is not convenient for the DCA as both phases are not consistent. It was, also, shown that applying rough set theory with DCA is a good alternative leading to a

better classification performance. However, the developed RC-DCA rough model suffers from a main limitation which is the performance of data discretization beforehand.

Table 2. Comparison Results of DCA Approaches

Database	Specificity(%)		Sensitivity(%)		Accuracy(%)		Time(s)		♯ Attributes	
	DCA		DCA		DCA		DCA		DCA	
	RC	FBR	RC	FBR	RC	FBR	RC	FBR	RC	FBR
SN	93.82	97.93	90.10	97.29	91.82	97.59	1705.79	14.87	20	9
Bio	79.24	92.45	77.35	86.79	78.30	89.62	1679.53	13.58	19	9
SP	98.49	99.89	98.40	99.77	98.45	99.84	3184.83	2119.95	8	8
CylB	97.75	98.39	97.00	97.00	97.46	97.85	1441.93	29.06	7	5
Ch	98.88	98.82	98.80	99.40	98.84	99.12	1779.83	714.95	11	4
IONO	97.33	99.11	96.82	98.41	97.15	98.86	668.32	41.12	19	9
Sck	97.68	99.09	96.96	96.53	97.64	98.93	1401.43	704.95	20	14
Mash	99.76	99.95	99.51	99.92	99.64	99.93	4567.34	4092.6	6	3
HC	94.73	97.36	93.05	96.29	93.75	96.73	260.08	39.84	14	5
GC	90.77	90.35	89.05	87.95	90.30	89.70	533.72	196.9	17	10
RWW	99.49	99.37	99.22	99.18	99.29	99.23	2201.98	1599.11	6	3

In this Section, we aim to show that applying FRST, instead of RST, can avoid the information loss caused by the mandatory step of data quantization. We, also, aim to show that by leaving the attribute values unchanged, our proposed FBR-DCA algorithm is able to select fewer features than the crisp rough RC-DCA approach, leading to better guide the FBR-DCA algorithm classification process. This is confirmed by the results presented in Table 2. For instance, from Table 2, we can notice that our new fuzzy-rough DCA model, FBR-DCA, has fewer features than the rough DCA model, RC-DCA. This is explained by the fact that FBR-DCA, by applying the Fuzzy-Rough QuickReduct algorithm, incorporates the information usually lost in crisp discretization by utilizing the fuzzy boundary region to provide a more informed technique. The results show that FBR-DCA selects features without much loss in information content. Our FBR-DCA new approach performs much better than traditional RST on the whole, in terms of both feature selection and classification quality. For instance, applying FBR-DCA to the Bio database, the number of selected attributes is 9. However, when applying RC-DCA to the same database, the number of selected features is set to 19. A second example can be the HC dataset where the number of selected features, by applying FBR-DCA, is reduced by more than 50% (5 features) in comparison to the number of features selected by the crisp rough DCA model, RC-DCA, which is set to 14.

Furthermore, from Table 2, we can notice that our FBR-DCA outperforms RC-DCA in terms of classification accuracy. For instance, when applying the algorithms to the SN dataset, the classification accuracy of FBR-DCA is set to 97.59%. However, when applying RC-DCA to the same database, the accuracy is set to 91.82%. Same remark is observed for the specificity and the sensitivity

criteria. When comparing the results in terms of running time, we can notice that the time taken by our FBR-DCA to process is less than the time needed by RC-DCA to function. This is explained by the fact that our FBR-DCA generates only one fuzzy-rough reduct as it is based on the Fuzzy-Rough QuickReduct algorithm. In contrast, RC-DCA generates all possible reducts that can be produced from data. Obviously, this is an expensive solution to the problem. Most of the time only one reduct is required as, typically, only one subset of features is used to reduce a dataset, so all the calculations involved in discovering the rest are pointless. Moreover, RC-DCA proposes different solutions for signal categorization; in case where the algorithm generates one reduct and when the algorithm generates a family of reducts; which is seen as a time consuming task. For example, when applying the algorithms to the Bio database, the amount of time taken by our FBR-DCA to process is 13.58(s) which is much less than the time taken by RC-DCA which is set to 1679.53(s).

We have, also, compared the performance of our FBR-DCA to other classifiers which are SVM, ANN and DT. The comparison made is in terms of the average of accuracies on the databases presented in Table 1. Fig.2 shows that the standard PCA-DCA has nearly the same classification performance as SVM and a better one than ANN and DT. It, also, shows that RC-DCA outperforms all the mentioned classifiers including the PCA-DCA in terms of overall accuracy. This is explained by the fact that RC-DCA applies rough set theory, instead of PCA, in the algorithm data pre-processing phase. Most importantly, the highest classification accuracy is noticed for our fuzzy-rough DCA new model, FBR-DCA. These promising FBR-DCA results are explained by the appropriate application of FRST to the DCA data pre-processing phase. This makes the algorithm a better classifier by generating more reliable and more pertinent results.

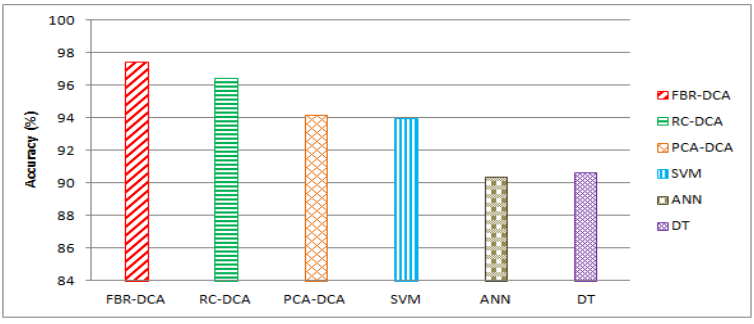

Fig. 2. Classifiers' Average Accuracies

To summarize, we have shown, in this Section, that our proposed FBR-DCA has the advantages of selecting fewer features than our proposed first work; RC-DCA. FBR-DCA is capable of avoiding the information loss caused by the use of the crisp rough set theory. The application of FBR-DCA to the unchanged attribute values led our new fuzzy-rough model to better guide its classification task yielding better performance in terms of classification accuracy. FBR-DCA

is, also, characterized by its lightweight in terms of running time in comparison to RC-DCA. Another characteristic of our FBR-DCA approach, when comparing it to the standard DCA version when applying PCA, is that it holds the semantics of the initial attributes. Adding to this, our fuzzy-rough DCA model, FBR-DCA, can effectively select features with no need for user-supplied information.

7 Conclusion and Future Works

In this paper, we have proposed a new hybrid DCA classification model based on fuzzy rough set theory. Our model aims to select the convenient set of features and to perform their signal categorization using the Fuzzy-Rough QuickReduct algorithm. Our proposed solution, FBR-DCA, ensures a more rigorous data pre-processing, for the DCA, when dealing with databases with real-valued attributes. Results show that FBR-DCA is capable of performing better its classification task than the standard DCA, the crisp rough RC-DCA model and other classifiers.

References

1. 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)
2. Lutz, M., Schuler, G.: Immature, semi-mature and fully mature dendritic cells: which signals induce tolerance or immunity? Trends in Immunology 23, 445–449 (2002)
3. Feng, G., Greensmith, J., Oates, R., Aickelin, U.: Pca 4 dca: The application of principal component analysis to the dendritic cell algorithm. In: Annual Workshop on Computational Intelligence (2009)
4. Chelly, Z., Elouedi, Z.: RC-DCA: A new feature selection and signal categorization technique for the dendritic cell algorithm based on rough set theory. In: Coello Coello, C.A., Greensmith, J., Krasnogor, N., Liò, P., Nicosia, G., Pavone, M. (eds.) ICARIS 2012. LNCS, vol. 7597, pp. 152–165. Springer, Heidelberg (2012)
5. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11, 341–356 (1982)
6. Jensen, R., Shen, Q.: Fuzzy-rough sets for descriptive dimensionality reduction. In: IEEE International Conference on Fuzzy Systems, pp. 29–34 (2002)
7. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. Kluwer Academic Publishers, Dordrecht (1992)
8. Greensmith, J., Aickelin, U., Twycross, J.: Articulation and clarification of the dendritic cell algorithm. In: Bersini, H., Carneiro, J. (eds.) ICARIS 2006. LNCS, vol. 4163, pp. 404–417. Springer, Heidelberg (2006)
9. Jensen, R., Shen, Q.: New approaches to fuzzy-rough feature selection. IEEE Transactions on Fuzzy Systems 17, 824–838 (2009)
10. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007), `http://mlearn.ics.uci.edu/mlrepository.html`
11. Jensen, R., Shen, Q.: Fuzzy-rough sets assisted attribute selection. IEEE Transactions on Fuzzy Systems 15, 73–89 (2007)