An ensemble classifier through rough set reducts for handling data with evidential attributes

Asma Trabelsi[1,2], Zied Elouedi[1], Eric Lefevre[2]

[1] *Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC*
41, Avenue de la Liberté, Tunis, 2000, Tunisia
[2] *Univ. Artois, UR 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A),*
Technoparc Futura, Béthune, F-62400, France.

Abstract

Ensemble classifier is a well-known method that has been used to solve several machine learning problems. To have reliable results, one should ensure the build of a good ensemble. In order to do so, researchers have proposed some heuristics like Random Subspace Ensemble (RSM), Rough set based ensemble, etc. The drawback of these mentioned approaches is their disability to handle uncertain data especially when uncertainty is represented by the evidence theory. The aim of this paper is to adapt both RSM and Rough set based ensemble in order to let them working in the context of evidential data. Three ensemble classifier approaches based on the rough set theory have been proposed and have been compared with each other. For the comparison purpose, we have relied on Ensemble Enhanced Evidential *k* Nearest Neighbor (EE*k*-NN) classifier, real world datasets from the UCI repository as well as synthetic databases.

Keywords:
Ensemble classifier, evidential data, rough set theory, uncertain data, reducts

1. Introduction

Recent years have witnessed the revival of artificial intelligence (AI) in academic and industry circles. One of the popular applications of AI is Machine Learning (ML), which has seen spectacular developments, and continues to find applicability in a wide range of domains, such as Information search, Information retrieval [1], speech recognition [2], and personal assistants on mobile phones [3], etc.

The ever-growing availability of data imposes an increasing demand for new and more powerful learning techniques. Therefore, in the last decade, we have seen the growth of several learning techniques, such as the ensemble learning [4, 5]. This approach allows a significant improvement in the machine learning results by combining several models. Thus, this approach offers better predictive/classification performance compared to a single model [5]. That is why ensemble methods ranked first in many prestigious machine learning competitions, such as the Netflix Competition, KDD 2009, and Kaggle.

The construction of an ensemble system requires two substantial steps. The first one is to select a set of individual classifiers while the second one consists of combining the output predictions of

these classifiers. The choice of the set of individual classifiers as well as the fusion operator could influence the ensemble performance.

The diversity between the base classifiers has been defended as a successful mean to construct a performing ensemble system and could be ensured in different manners. One among the diversity techniques that has been proven to be efficient and effective [6] is to vary the input feature space.

The process of generating diverse feature subsets with good predicting potential is still undergoing research study. The Random Subspace Method (RSM) is often used in the literature [7], but its major limitation is to randomly partition the original input set. So, random selection may potentially increase the risk of selecting irrelevant and redundant features.

Despite the promising results of RSM based ensemble classifier, several other techniques and frameworks have been developed to enhance the prediction performance [8]. Among them, we highlight the rough set theory that has been successfully used to reduce the set of features of any dataset [9]. This theory has been successfully applied in machine learning, particularly for feature reduction. It consists of finding the minimal feature subsets allowing the same discrimination as the initial set. The reduced set is called a reduct.

Ensemble classifiers through rough set reducts have been applied in a wide range of practical problems, such as text classification [10], biomedical classification [11], tumor classification [12], web services classification [13], etc.

It is important to note that real world data are almost imperfect (i.e. incomplete or/ and uncertain). This imperfection is due to multiple external factors such as obstacles, interference, missing information, etc. Several studies have been made to handle data imperfection. From these lines of research, we mention evidential theory [14], the probability theory [15] and the fuzzy set theory [16].

The evidence theory is commonly used to represent data imperfection thanks to its flexibility and capability to represent all kinds of imperfection including total and partial ignorance [17]. The data represented with the evidence framework are called evidential data [18].

In spite of their great importance, rough set ensemble classifiers has not been applied on data with evidential features [19, 20]. To unlock this research field, we propose a rough set based ensemble for processing such kind of data. More particularly, we treat the case of imperfect features represented within the evidence framework. Our approach consists of three main levels: reduct generation, reduct selection for training individual classifiers and classifier fusion.

To sum up, the idea of this paper is to propose new ensemble classifier approaches to cope with uncertain data, more precisely when the uncertainty affects the values of attributes and is represented with the evidence theory. To the best of our knowledge, we are among the first to develop ensemble rough set classifiers in the context of evidential data. Three new approaches based on the rough set theory have been proposed and compared with each other. They will also be compared with individual EEk-NN classifier and the RSM technique. The choice of the rough set theory is

justified by the fact that it has given promising results in a which context [10, 11, 12, 13].

The remaining of this paper is organized as follows. In Section 2, we provide a brief description of the main concepts underlying the evidence theory, we define the concept of data with evidential features and we highlight the fundamental concepts of the rough set theory. We present, in Section 3, our ensemble classifier through rough set reducts to cope with evidential data, particularly data with evidential features. The experimentation settings and results are given in Section 4. We draw conclusions and some future work directions in Section 5.

2. Preliminary knowledge

In this section, we present the preliminary knowledge of two well known mathematical theories, namely the evidence theory and the rough set theory.

2.1. The evidence theory

The evidence theory, also referred to as belief function theory, is regarded as a very effective and efficient way for representing and managing uncertainty [21, 22]. This theory is extensively used for handling several real-world applications, including image processing [23], business decision [24], multi-sensor fusion [25, 26], pattern recognition [27], medical diagnosis [28, 29], classification [30, 31, 32], clustering [33] and target tracking [34]. In what follows, we provide a brief overview of its fundamental concepts as interpreted by the Transferable Belief Model (TBM) [35]. We also point out other basic concepts including the special belief functions, the belief function *bel* and the decision making process. We at last present the Dempster rule for combining distinct information sources.

2.1.1. Frame of discernment

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_c\}$ denotes the frame of discernment including a finite non empty set of c elementary hypotheses that are assumed to be exhaustive and mutually exclusive. The power set of Θ, denoted by 2^Θ, is made up of all the subsets of Θ:

$$2^\Theta = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \{\theta_1, \theta_2\}, \ldots, \Theta\} \tag{1}$$

where each element of 2^Θ is called a proposition or an event.

2.1.2. Basic belief assignment

An expert's belief over the subsets of the frame of discernment Θ are represented by the so-called basic belief assignment (bba) denoted by m. It is carried out in the following manner:

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{2}$$

The basic belief mass (bbm), denoted by $m(A)$, implies the degree of belief exactly assigned to the event A. Because of a lack of information, this quantity cannot be distributed to any strict subset of A. It is worth noting that every subset A of 2^Θ having fulfilled $m(A) > 0$ is called a focal element.

With the aim of expressing some particular uncertainty situations, some special bbas have been proposed, including the vacuous, the certain bbas and the simple support mass function. The vacuous bba represents the total ignorance (i.e. $m(\Theta) = 1$). The certain bba is a bba with a singleton as its unique focal element. A bba is called simple support function (ssf) if it has at most two focal elements: the frame of discernment Θ and a strict subset of Θ called the focus of the ssf.

2.1.3. Belief function bel

A belief function *bel*, relative to a bba m, assigns to any subset A of Θ the sum of beliefs exactly committed to every subset of A by m [22]. In other words, it implies the total belief that one commits to A without being also committed to $\overline{A}$. It has to be noted that $m(\emptyset)$ is not included in $bel(A)$, since $\emptyset$ is a subset of both A and $\overline{A}$. The belief function *bel* is defined as follows:

$$\begin{aligned} bel &: 2^{\Theta} \rightarrow [0,1] \\ bel(A) &= \sum_{\emptyset \neq B \subseteq A} m(B) \end{aligned} \tag{3}$$

2.1.4. The Dempster combination rule

The combination of a set of imperfect data is a crucial task owing to its ability to achieve a piece of more accurate information and improve decision making. The evidence theory is considered as a powerful tool to merge imperfect data including uncertain, imprecise and incomplete data. Indeed, several fusion rules have been proposed to aggregate a set of data induced from distinct information sources. The Dempster rule is one among the commonly used rules. Let m_1 and m_2 be two bbas defined in the same frame of discernment Θ. The Dempster rule is set to:

$$\begin{aligned} &(m_1 \oplus m_2)(A) = k(m_1 \textcircled{\cap} m_2)(A) \\ &(m_1 \oplus m_2)(\emptyset) = 0 \end{aligned} \tag{4}$$

where

$$k^{-1} = 1 - (m_1 \textcircled{\cap} m_2)(\emptyset) \tag{5}$$

and

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B).m_2(C) \tag{6}$$

where $\textcircled{\cap}$ represents the conjunctive fusion rule also proposed within the evidence framework.

2.1.5. Data with evidential features

A dataset with evidential features is a dataset composed by M objects O_j (i.e. $j \in \{1,\ldots,M\}$) where each of them is described by N features $A = \{A_1,\ldots,A_N\}$ that are expressed within the evidence theory and a class label $d_j \in d$. Each feature A_k (i.e. $k \in \{1,\ldots,N\}$) has a domain of discrete values denoted by Θ^{A_k}. The advantage of evidential data over the other theories is the ability to represent all kinds of data uncertainty. In fact, it allows us to manage total certainty, partial ignorance and total ignorance. An example of data characterized by evidential attributes is given below.

Example: *Overall, a bank loan officer has to predict the customer profitability levels d={Good, Bad} on the basis of some parameters (features). To put it simply, in this example, Table 1 describes the data knowledge for training, where we relied on three characteristics:*

- *Income with possible values* Θ^{Income}*={No, Low, Average, High}.*
- *Property: This feature reflects whether the loan requested by the client is greater or less than its property value and consequently it takes values into* $\Theta^{Property}$*={Greater, Less}.*
- *Unpaid Credit: This feature provides information about client's unpaid Credit with possible values* $\Theta^{UnpaidCredit}$*={Yes, No}.*

Table 1: Example of data with evidential attributes

O	Income	Property	UnpaidCredit	d
O_1	$m_1^{Income}(\{High\})=1$	$m_1^{Property}(\{Greater\})=0.6\ m_1^{Property}(\{Less\})=0.3\ m_1^{Property}(\Theta^{Property})=0.1$	$m_1^{UnpaidCredit}(\{Yes\})=1$	Good
O_2	$m_2^{Income}(\{Average\})=1$	$m_2^{Property}(\{Greater\})=1$	$m_2^{UnpaidCredit}(\{No\})=1$	Good
O_3	$m_3^{Income}(\{Low\})=1$	$m_3^{Property}(\{Less\})=0.5\ m_3^{Property}(\Theta^{Property})=0.5$	$m_3^{UnpaidCredit}(\{Yes\})=1$	Bad
O_4	$m_4^{Income}(\{No\})=1$	$m_4^{Property}(\{Less\})=1$	$m_4^{UnpaidCredit}(\{No\})=1$	Bad
O_5	$m_5^{Income}(\{Average\})=1$	$m_5^{Property}(\{Greater\})=0.8\ m_5^{Property}(\Theta^{Property})=0.2$	$m_5^{UnpaidCredit}(\{No\})=1$	Good
O_6	$m_6^{Income}(\{High\})=1$	$m_6^{Property}(\{Greater\})=0.2\ m_6^{Property}(\{Less\})=0.7\ m_6^{Property}(\Theta^{Property})=0.1$	$m_6^{UnpaidCredit}(\{Yes\})=1$	Bad

The example of Table 1 presents different kinds of bbas. The objects of both *Income* and $UnpaidCredit$ are represented by certain mass functions. For the feature $Property$, objects O_2 and O_4 are also represented by certain mass functions, objects O_3 and O_5 are represented by simple support mass functions and objects O_1 and O_6 are represented by general mass functions.

2.2. Basic concepts for rough set theory

The rough set theory, proposed by Pawlak [9], is an efficient way for dealing with various machine learning problems such as clustering [36], classification [37], feature selection [38, 39], etc. One of our ultimate goals throughout this paper is to extract the most suitable feature subsets for classifier ensemble. Thus, we mainly relied on rough sets for generating the smallest subsets of relevant features (i.e. called reducts) and selecting the appropriate ones. In practical terms, a data set has to be represented through a Decision Table (DT) which is defined as a pair $DT = (U, A \cup \{d\})$. The universe $U = \{O_1, \ldots, O_M\}$ reflects a non-empty finite set of M objects, $A = \{A_1, \ldots, A_N\}$ is a non-empty finite set of N features with values $V(O_i) = \{V_1(O_i), \ldots, V_N(O_i)\}$ for each object O_i and d corresponds to the decision value (i.e. the decision class). The discernibility function is one among the efficient solutions for reduct extraction. It consists firstly of computing a discernibility matrix DM from a given decision table DT. Each element $DM(O_i, O_j)$ represents the set of all features discerning objects O_i and O_j with different class d. Entries of the discernibility matrix DM are computed as follows:

$$DM(O_i, O_j) = \{A_k \in A | V_k(O_i) \neq V_k(O_j) \text{ and } d_i \neq d_j\} \quad \forall\, i, j = \{1, \ldots, M\} \tag{7}$$

Once the discernibility matrix DM is computed, the discernibility function can be defined as follows:

$$f(DM) = \wedge\{\vee(DM(O_i, O_j)) | \forall O_i, O_j \in U, DM(O_i, O_j) \neq \emptyset\} \tag{8}$$

The discernibility function has to be converted from a conjunctive normal form into a disjunctive normal form for picking out all possible reducts.

3. A rough set based ensemble for data with evidential features

We propose to develop a rough set based ensemble classifier for dealing with data with evidential features. It aims to construct an ensemble of evidential classifiers (i.e. classifiers trained from data with evidential features) through rough set reducts. The general structure of our proposed framework is depicted in Figure 1 where EC_i (i.e. $i \in \{1,\ldots,L\}$) corresponds to the i^{th} selected Evidential Classifer. As depicted in this figure, our proposed framework consists of three main steps: reduct generation from data with evidential features, reduct selection to ameliorate the prediction performance and classifier combination. In the following, we describe in further detail each of these steps.

3.1. A framework for generating reducts from data with evidential features

Reduct computation has been proven as an NP-hard problem which has led to the introduction of several heuristics. The Rosetta software is one among the most effective approaches for reducts generation. It includes a set of algorithms for extracting multiple reducts such as the SAVGenetic Reducer [40], a genetic algorithm for picking out approximate reducts. As our main purpose is to address data with evidential features, we propose to extend the SAVGenetic algorithm for processing such kind of data. By analogy to the standard version, our evidential SAVGenetic reducer starts by calculating a discernibility matrix Λ' [41]. Assume that U=$\{O_1,\ldots,O_M\}$ is a given data set with M objects. Each object O_i ($i \in \{1,\ldots,M\}$) is described by N evidential features $A = \{A_1,\ldots,A_N\}$ and a certain class label $d_i \in d$. Note that each attribute A_k (i.e. $k \in \{1,\ldots,N\}$) has a domain of discrete values denoted by Θ^{A_k}. The entries of the discernibility matrix Λ' are set to:

$$\Lambda'(O_i, O_j) = \{A_k \in A | bel(O_i \neq O_j | A_k) > T \text{ and } d_i \neq d_j\} \tag{9}$$

where T refers to a tolerance threshold (i.e. T is set to 0.1 with the aim of maximizing the search space) and:

$$bel(O_i \neq O_j | A_k) = \sum_{E_i, E_j, \forall a_i \in E_i, a_j \in E_j, a_i \neq a_j} m_i^k(E_i) m_j^k(E_j) \tag{10}$$

with E_i and E_j are subsets of Θ^{A_k}. The idea behind Equation 10 is to link between the events: attribute A_k of objects O_i and O_j having different values as well as the distance between two mass functions. In particular, two mass functions may well be identical, while we would have absolutely no idea of whether their attribute values are equal or not. Even worse, we may have strong evidence of equal values that would give a positive distance (say, two probabilities slightly different, but with an important weight on the same value), and no evidence at all of equality that would give a null distance (say, two vacuous mass functions). This is obviously an undesirable property.

The process of extracting reducts through a discernibility matrix is regarded as a set cover problem. It consists of finding the minimal hitting sets from the non empty sets of the obtained discernability matrix. Since the minimal hitting set is an NP-hard problem, we relied on the genetic algorithm for picking out approximate hitting sets, meaning approximate reducts. Let us denote by A the set containing the elements of the discernibility matrix Λ' and let us denote by ζ' the set

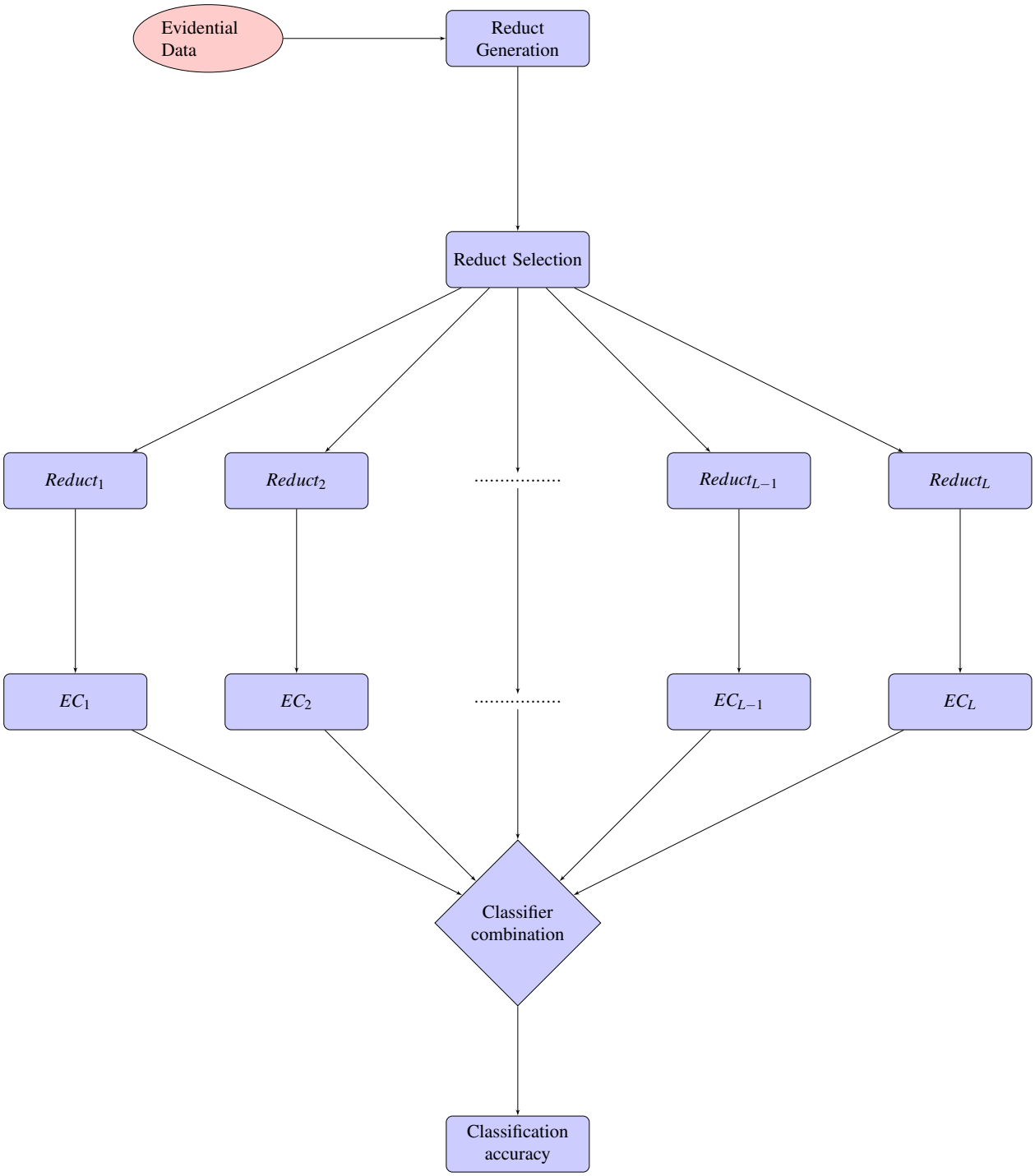

Figure 1: Ensemble evidential classifier through rough set reducts

containing the non empty sets of Λ', the fitness function corresponds to our genetic algorithm for each candidate solution $B \in 2^N$ is set to [42]:

$$f(B) = (1-\alpha) \times \frac{|A| - |B|}{|B|} + \alpha \times min\{\varepsilon, \frac{|[F \in \zeta' | F \cap B \neq \emptyset]|}{|\zeta'|}\} \tag{11}$$

The fitness function $f(B)$ rewards not only subsets that are hitting sets (i.e. meaning subsets having a non empty intersection with all elements of the discernability matrix) but also subsets with shortest size. Herein, $\alpha \in [0,1]$ refers to the adaptive weighting between the two parts and ε reflects the minimal hitting set fraction. In this paper, we set ε with the aim of extracting minimal sets and we set α to 0.5 to balance the two parts.

3.2. Reduct selection for ensemble learning

We could very well wind up with hundreds or even thousands of reducts. The process of constructing ensemble systems with all generated classifiers is extremely costly, especially for high dimensional databases. An alternative solution consists of selecting the most appropriate reducts for ensemble learning. We present, in what follows, three approaches enabling us to pick out the most suitable reducts for an ensemble of evidential classifiers.

3.2.1. Select Diverse Reducts (DR)

One of the main keys for constructing a successful rough set ensemble is to ensure a good diversity between the chosen reducts. Getting inspiration from [43], we propose a new heuristic for selecting diverse reducts from the pool of generated ones. Our algorithm (see Algorithm 1) starts by picking out the reduct with the minimum cost, meaning the smallest one. Then, it computes the diversity between the chosen reduct and the remaining ones using Algorithm 2. The reduct which has to be chosen is the one with the highest diversity degree. As in [43], the diversity measure is the inverse of the average similarity between a candidate reduct R_j and the L' chosen ones (i.e *RED_Chosen*) as follows:

$$Div_j = 1 - \frac{1}{L'} \sum_{R_i \in RED\_Chosen} \frac{|R_j \cap R_i|}{|R_j \cup R_i|} \tag{12}$$

The most diverse reduct R_j will then be chosen for constructing the ensemble system and it will be removed from the current reduct set *RED*. This process will be repeated until at most L reducts are selected or the reduct pool *RED* is empty.

3.2.2. Accuracy-Diversity Assessment Function for reduct selection (AD-AF)

The study conducted by Opitz [44] has demonstrated that both the accuracy of individual classifiers and the diversity of base classifiers may improve the performance of an ensemble system. So, a good classifier ensemble has to be constructed on the basis of accurate individual classifiers that are diverse as much as possible. As already stated above, an ensemble system with rough set reducts is considered as a valid alternative for ensuring good diversity between the base classifiers. Herein, we propose to construct a good rough set based ensemble classifier by making a trade-off between the diversity and the accuracy of each individual classifier. More precisely, we relied on the assessment function, proposed by Opitz [44], to extract the reducts enabling the construction

Algorithm 1 Select diverse reducts

input: A pool of reducts *RED*, L is the maximum Number of chosen reducts
output: L' diverse reducts
RED_Chosen $\leftarrow \emptyset$
R_1=min$_{R \in RED}\, cost(R)$
$RED\_Chosen \leftarrow \{RED\_Chosen, R_1\}$
L'=1
RED= *RED* - R_1;
While $L' <$ L **or** $isEmpty(RED) = false$ **Do**
$Div \longleftarrow ReductDiversity(RED\_Chosen, RED)$ {%Computed through Algorithm 2}
R_best= *arg* max$_{R_j \in RED}$ Div_j
$RED\_Chosen \leftarrow \{RED\_Chosen, R\_best\}$
RED= *RED* - *R_best*;
$L' = L' + 1$
end while

Algorithm 2 ReductDiversity(*RED_Chosen*,*RED*)

input: Candidate reducts *RED* and selected reducts *RED_Chosen*
output: Diversity between reducts *Div*
for j= 1 **to** $|RED|$
Sim_j=0;
for each $R_i \in RED\_Chosen$
Sim_j= Sim_j + $\frac{|RED_j \cap R_i|}{|RED_j \cup R_i|}$
end for
Div_j = 1- $\frac{Sim_j}{|RED\_Chosen|}$
end for

of accurate individual classifiers with the greatest possible level of diversity. Opitz's assessment function for a classifier f is set to:

$$Fitness(f, Ens\_Cls) = Acc(f, Ens\_Cls) + \omega \times Div(f, Ens\_Cls) \quad (13)$$

where Ens_Cls states the current ensemble of classifiers, $Acc(f, Ens\_Cls)$ reflects the average accuracy of the base classifiers, $Div(f, Ens\_Cls)$ represents the diversity between base classifiers and ω corresponds to the parameter that balances *Accuracy* and *Diversity*.

Several classifier diversity measures exist in the literature. Authors in [45] have distinguished pairwise and non-pairwise diversity measures. The choice of the most convenient one remains a tricky problem. In this paper, we relied on the disagreement measure, which is a pairwise one, for computing classifier diversity. The parameter ω has to be adjusted automatically for maximizing the fitness function value. More concretely, we keep the value of ω when *Fitness* is increasing, while we increase it if *Accuracy* is stable and *Diversity* is decreasing and we decrease it if *Accuracy* is decreasing and *Diversity* is stable. The value of ω will be set to 1 as the initial value and the changing amount of ω will be set to 10% based on its current value. Our method differs from Opitz's approach in the extent to which it takes into consideration the accuracy between reducts in addition to the diversity and the accuracy of individual classifiers. The diversity between reducts is used to reduce the search space of reducts to improve search efficiency. It will be computed at each iteration, using Equation 12, and the candidate reducts with a diversity measure smaller than a threshold S (i.e. in the experimentation parts, we have set S to 0.7 to enable the selection of the most diverse reducts) will then be removed from the search space of reducts. Our proposed framework is detailed in Figure 2 and Algorithm 3.

We start by retrieving the reduct R_1 with the lowest cost and constructing the first evidential Classifier EC_1 (i.e. $RED\_Choosen = \{R_1\}$ and $Ens\_Cls = \{EC_1\}$).

The diversity Div_j between the current selected reducts $RED\_Choosen$ and each reduct $R_j \in RED$ is the calculated. The reducts R_j with a diversity measure smaller then a threshold S will be removed from the reduct pool RED. Each candidate reduct $R_j \in RED$ will be evaluated using Equation 13. The reduct R_k enabling the highest fitness function will be selected for constructing our ensemble learning (i.e. $Ens\_Cls = \{Ens\_Cls, EC_k\}$). This process has to be repeated until at most a number L of reducts is reached or the current reduct pool RED is empty.

3.2.3. Ensemble Accuracy Assessment Function for reduct selection (EA-AF)

The wrapper approach, using the classifier accuracy as feature selection criterion, has been successfully used for solving several pattern recognition problems. In fact, it allows us to pick out the feature subset that achieves the greatest classification accuracy. Herein, we follow the same process as the previously presented approach but we relied on the ensemble accuracy as a fitness function for extracting the most appropriate reducts for an ensemble of evidential classifiers. The fitness function is set to:

$$Fitness(f_j, Ens\_Cls) = \max_{R_k \in RED} (EnsAcc(f_k, Ens\_Cls)) \quad (14)$$

where $EnsAcc(f_k, EnsCls)$ reflects the ensemble accuracy of the already chosen classifiers RED_Chosen and the candidate classifier f_k.

One important element which has to be highlighted is the maximum number L of selected reducts,

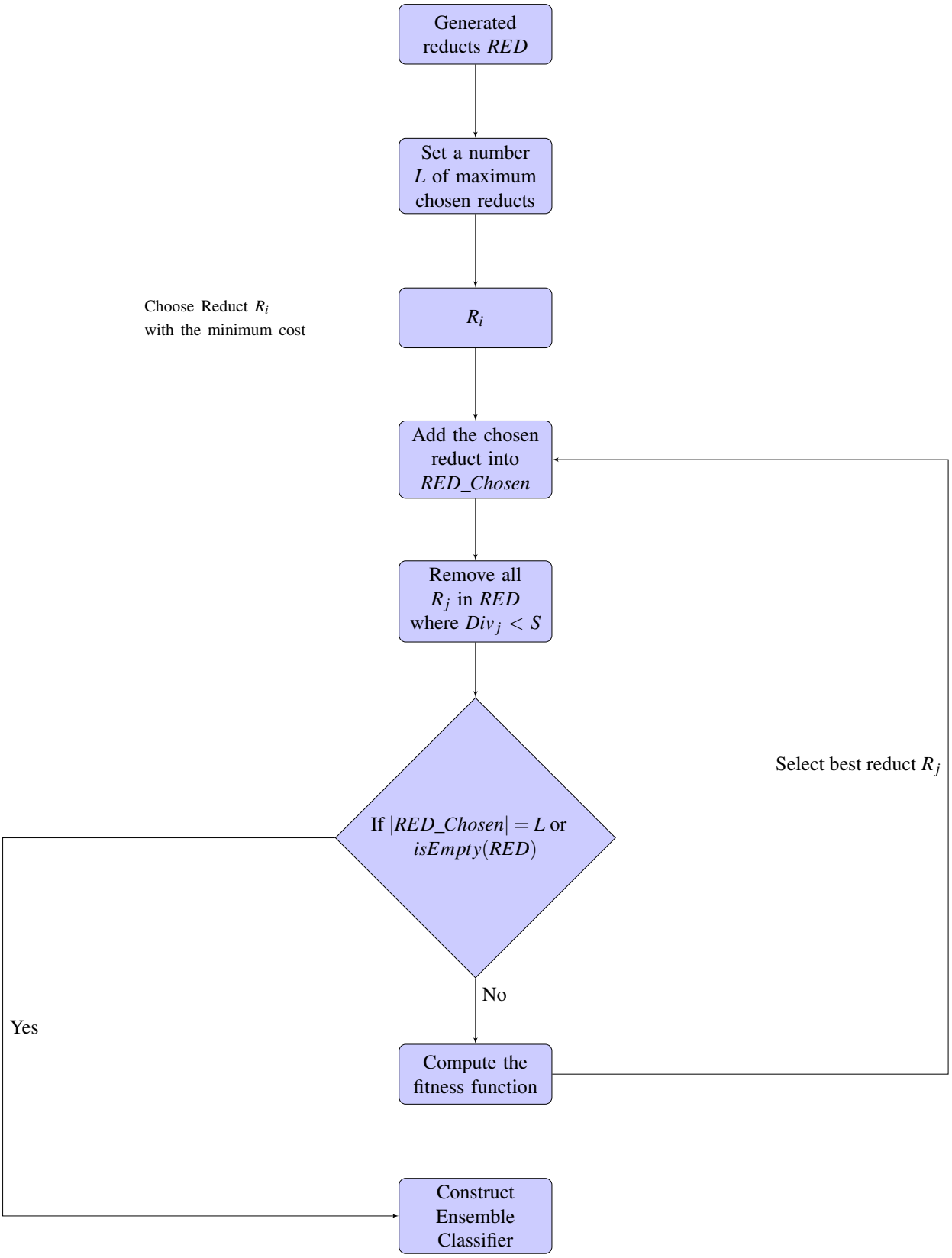

Figure 2: Reduct selection for ensemble learning

Algorithm 3 Accuracy-Diversity assessment function for reduct selection

input: A pool of reducts RED, L is the maximum Number of chosen reducts
output: Chosen L' diverse reducts
$RED\_Chosen \leftarrow \emptyset$
$Ens\_Cls \leftarrow \emptyset$
R_1=min$_{R \in RED}\, cost(R)$
$RED\_Chosen \leftarrow \{RED\_Chosen, R_1\}$
L'=1
RED= RED - R_1;
$Ens\_Cls \leftarrow \{Ens\_Cls, f_1\}$
Repeat
$Div \longleftarrow ReductDiversity(RED\_Chosen, RED)$ {% Computed through Algorithm 2}
$RED\_New \longleftarrow R_j \in RED$ with $Div_j > S$
$RED = RED\_New$
Choose a new reduct R_j from RED satisfying:
$Fitness(f_j, Ens\_Cls) = \max_{R_k \in RED}(Fitness(f_k, Ens\_Cls))$ {% See Equation 20}
$Ens\_Cls \leftarrow \{Ens\_Cls, f_j\}$
$RED\_Chosen \leftarrow \{RED\_Chosen, R_j\}$
RED=RED-R_j
L'=L'+1
until $L' = L$ **or** $isEmpty(RED) = true$

meaning selected classifiers. According to a study conducted in [44], ensembles of 25 classifiers are sufficient to improve the ensemble performance. In the remaining of this paper, we set L to 25.

3.2.4. Time complexity

It is important to point out the complexity of our different proposed approaches. Let us denote by N the total number of reducts, $N^{'}$ the number of candidate reducts, L the maximum number of chosen reducts and $L^{'}$ the real number of chosen reducts, the time complexity of each approach is given computed as:

- **DR:** For the approach Select Diverse Reducts, we repeat the process L times and for each time, we compare reducts with the L' chosen ones. So the process takes $O(L * L^{'} * N)$.
- **AD-AF:** A part of the *DR* approach is included in the *AD-AF* one. We only need to compute the accuracy induced by all candidate reducts. The time complexity will be computed as $O(L * (L^{'} * (N + N^{'})))$
- **EA-AF:** This third approach is the more complex one in terms of time complexity which is computed as $O(L * (L^{'} * (N + (\frac{N' * (N'+1)}{2}))))$

3.3. Multiple classifier integration

As previously mentioned, each individual evidential classifier yields decisions in terms of belief functions. Numerous combination operators have been proposed within the evidence theory to aggregate the classifiers outputs. In this paper, we relied on the Dempster combination rule presented

Table 2: Description of databases

Databases	Heart	Japanese	Vote records	Hepatitis	Wine	Thoracic Surgery	SD	SD_0.15	SD_0.30	SD_0.50
Total instances	270	690	435	155	178	470	700	700	700	700
Total attributes	13	15	16	19	13	17	9	9	9	9
Missing values	No	No	Yes	Yes	No	No	No	No	No	No
Number of classes	3	2	2	2	3	2	4	4	4	4

in subsection 2.1.4. This choice is justified by the fact that we combine individual classifiers trained with diverse reducts.

4. Experimentation settings and results

This section is devoted to examining the performance of the three proposed reduct selection approaches for ensemble learning. In what follows, we detail our experimentation settings and results.

4.1. Experimentation settings

We have relied on some numerical and mixed real world databases acquired from the UCI machine learning databases [46], where some of them contain missing values. We have also generated some Synthetic datasets using Python and Scikit-Learn, a Machine Learning Library. The advantage of this library is that it accepts various parameters allowing to control the looks and feels of datasets. One Artificial dataset has been generated using Python. Then, we have added noise to the attributes of this generated data in order to create newest datasets. Different noise intensities have been explored like 0.15, 0.30 and 0.5. These values represent the fraction of samples whose attributes have been assigned randomly. Consequently, we have obtained four synthetic databases, the first one with no noise is called SD, the second one with a 0.15 as noise intensity is called SD_0.15, the third one with 0.30 as noise intensity is called SD_0.30 and the last one with 0.50 as noise intensity is called SD_0.50. Table 2 provides a description of the used databases.

Once data are collected and generated, we have modeled data thanks to the evidence theory. In a practical point of view, missing values have to be imputed and continuous variables have usually to be discretized into bins. However, the uncertainty introduced by missing values imputation and continuous variables discretization have to be addressed. Herein, we propose to generate databases with evidential features from the mentioned ones. That is, the missing values will be represented by vacuous bbas and symbolic attributes have to be expressed through certain bbas. With regards to continuous variables, they have been transformed into beliefs using the Evidential c-Means approach (EcM) [47]. The EcM approach enables to associate for each instance feature a bba representing its membership to each cluster. In our case, we have set c to 3 for reducing the time complexity. The EcM algorithm starts by creating the user requested number of clusters for each feature. It then estimates, for each feature, the distance between each instance and each cluster's center and generates a bba using the distance value. Afterwards, it tries to minimize an objective function and computes recursively the cluster's center until no more minimization is possible [47, 48].

The choice of the base classifier is also a crucial task. To the best of our knowledge, machine learning classifiers that handle data with evidential features are limited. Note that we have introduced, in previous work, a novel classifier for handing such a kind of data called Enhanced

Evidential k Nearest Neighbor (EEk-NN) [49]. For the evaluation process, we relied on the (EEk-NN) as a base classifier. A key issue that has to be addressed is the parameter k which reflects the number of neighbors. For our tests, we evaluate three k values which correspond to 3, 5 and 7 respectively. To do so, we were based on some standard information retrieval measures, notably the Percentage of Correctly Classification (PCC), the recall and the precision.

4.2. Experimentation results

We conduct some experiments to compare our proposed approach for reduct selection and to pick out the best one among them. Our comparison is based on the ensemble size, the ensemble reduct diversity and the performance measure. We also make comparisons using an hypothesis test.

4.2.1. Comparison in terms of the ensemble size

The comparison results between our three heuristics for reduct selection in terms of the ensemble size are given in Table 3 where *GR* represents the total number of generated reducts. From these tables, we can see that a huge number of reducts may be produced for a given database (e.g. we have 8191 reducts for the Hepatitis database and 975 reducts for the Thoracic Surger database). So, it is crucial to select the most appropriate reducts for an ensemble of classifiers. As previously mentioned, we have proposed three approaches for reduct selection, namely the *DR*, the *AD-AF* and *EA-AF* methods. The experimentation results have proven that both *AD-AF* and *EA-AF* methods have yielded smallest ensembles compared with the *DR* approach. Let's take the Hepatitis database with k equals 3 as an example, the ensemble size achieved by the *DR*, the *AD-AF* and the *EA-AF* are equal to 25, 5 and 3 respectively. From this point of view, we can deduce the efficiency of the *AD-AF* and *EA-AF* approaches for generating ensemble EEk-NN with reduced size.

4.2.2. Comparison in terms of the Reduct diversity

We have mainly relied on the Jaccard distance J_δ for measuring the reduct diversity. This measure highly depends on the number of reducts. In fact, the maximum diversity is yielded when there is an empty intersection of the generated reducts. It is set to:

$$J_\delta = \frac{|R_1 \cup R_2 \cup \ldots \cup R_{M'}| - |R_1 \cap R_2 \cap \ldots \cap R_{M'}|}{|R_1 \cup R_2 \cup \ldots \cup R_{M'}|} \tag{15}$$

The obtained results are given in Table 4 where we can remark that the *DR* method has achieved in most cases the most diverse reducts comparatively with the *AD-AF* and *EA-AF* approaches. This can be explained by the specific feature of the Jaccard measure. In fact, it promotes the ensemble constructed with the largest number of reducts. The results still show that the *AD-AF* and *EA-AF* methods are able to provide sets of reducts with higher diversity compared to *DR* on some datasets (e.g., Hepatitis).

4.2.3. Classification performance Comparison

The evidence theory has not only the advantage to manage and represent uncertainty but it also proposes a set of combination rules to merge evidence acquired from several information sources, notably the evidential outputs of an ensemble of classifiers. The Dempster operator is a well used rule in the context of classifier fusion within the evidence theory [5]. From this, we use the Dempster rule for combining the selected individual classifiers obtained by the *DR*, the *AD-AF* and the *EA-AF* techniques. Following a 5-folds cross-validation process, we carry out a comparative study

Table 3: Comparison in terms of ensemble size

(a) Ensemble size for Heart database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	127	25	**3**	**3**
k=5	127	25	**2**	3
k=7	127	25	4	**2**

(b) Ensemble size for Japanese database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	511	25	3	**2**
k=5	511	25	**3**	**3**
k=7	511	25	**3**	**3**

(c) Ensemble size for Vote Records database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	136	25	**3**	**2**
k=5	136	25	**3**	**2**
k=7	136	25	**3**	**3**

(d) Ensemble size for Hepatitis database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	8191	25	5	**3**
k=5	8191	25	**4**	**4**
k=7	8191	25	**3**	**3**

(e) Ensemble size for Thoracic Surgery database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	975	25	4	**3**
k=5	975	25	**4**	**4**
k=7	975	25	**4**	**4**

(f) Ensemble size for Wine database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	1824	25	4	**3**
k=5	1824	25	4	**3**
k=7	1824	25	**3**	**3**

(g) Ensemble size for SD

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	153	25	4	**3**
k=5	153	25	4	**3**
k=7	153	25	4	**3**

(h) Ensemble size for SD_0.15 database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	153	25	4	**3**
k=5	153	25	4	**3**
k=7	153	25	4	**3**

(i) Ensemble size for SD_0.30 database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	153	25	4	**3**
k=5	153	25	4	**3**
k=7	153	25	4	**3**

(j) Ensemble size for SD_0.50 database

	GR	*DR*	*AD-AF*	*EA-AF*
k=3	153	25	4	**3**
k=5	153	25	4	**3**
k=7	153	25	4	**3**

Table 4: Comparison in terms of reduct diversity

(a) Reduct diversity for Heart database

	DR	*AD-AF*	*EA-AF*
k=3	**0.6**	0.5	0.5
k=5	**0.6**	0.5	0.54
k=7	0.6	0.42	**0.86**

(b) Reduct diversity for Japanese database

	DR	*AD-AF*	*EA-AF*
k=3	**0.69**	0.67	**0.69**
k=5	**0.69**	0.65	0.63
k=7	**0.69**	**0.69**	0.64

(c) Reduct diversity for Vote Records database

	DR	*AD-AF*	*EA-AF*
k=3	**0.5**	0.49	0.47
k=5	**0.5**	**0.5**	0.48
k=7	**0.5**	0.48	0.50

(d) Reduct diversity for Hepatitis database

	DR	*AD-AF*	*EA-AF*
k=3	0.54	**0.58**	**0.58**
k=5	0.54	**0.58**	0.46
k=7	0.54	**0.58**	0.54

(e) Reduct diversity for Thoracic Surger database

	DR	*AD-AF*	*EA-AF*
k=3	**0.94**	0.93	**0.94**
k=5	**0.94**	0.93	0.92
k=7	**0.94**	0.92	0.90

(f) Reduct diversity for Wine Data database

	DR	*AD-AF*	*EA-AF*
k=3	**0.82**	0.75	0.81
k=5	**0.82**	0.79	0.76
k=7	**0.82**	0.81	0.74

(g) Reduct diversity for SD database

	DR	*AD-AF*	*EA-AF*
k=3	**0.79**	0.77	0.74
k=5	**0.79**	0.77	0.74
k=7	**0.79**	0.75	0.73

(h) Reduct diversity for SD_0.15 database

	DR	*AD-AF*	*EA-AF*
k=3	**0.79**	0.78	**0.79**
k=5	**0.79**	0.77	0.78
k=7	**0.79**	0.78	0.77

(i) Reduct diversity for SD_0.30 database

	DR	*AD-AF*	*EA-AF*
k=3	**0.79**	0.77	0.73
k=5	**0.79**	0.76	0.73
k=7	**0.79**	0.77	0.74

(j) Reduct diversity for SD_0.50 database

	DR	*AD-AF*	*EA-AF*
k=3	**0.79**	0.76	0.75
k=5	**0.79**	0.76	0.74
k=7	**0.79**	0.76	0.75

Table 5: Results for Heart database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*AE-AF*
k=3	PCC	60.75	61.48	75.92	75.92	**77.77**
	Recall	66.18	67.90	75.82	75.99	**77.77**
	Precision	56.10	57.49	75.64	75.39	**77.52**
k=5	PCC	71.23	72.96	78.14	**78.88**	**78.88**
	Recall	72.75	73.84	78.16	78.82	**78.93**
	Precision	70.13	71.70	77.59	78.53	**78.59**
k=7	PCC	71.98	72.96	78.88	78.88	**82.86**
	Recall	73.12	74.74	79.32	78.98	**83.34**
	Precision	70.91	71.13	78.27	78.62	**82.43**

Table 6: Results for Japanese database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*AE-AF*
k=3	PCC	45.03	45.21	73.76	76.37	**78.69**
	Recall	55.71	56.22	65.99	67.22	**68.68**
	Precision	70.89	71.38	71.10	72.05	**73.79**
k=5	PCC	51.25	52.17	71.59	78.40	**80.86**
	Recall	62.24	65.83	64.95	69.26	**71.00**
	Precision	55.21	57.33	71.81	74.27	**76.04**
k=7	PCC	54.28	55.94	70.86	79 42	**80.00**
	Recall	62.01	63.72	65.65	69.19	**69.90**
	Precision	59.45	61.05	68.55	73.33	**74.04**

between these methods when based on the PCC, the Recall and the Precision. This approach has also been compared with individual EE*k*-NN classifier and the Random Selected Reduct (*RSR*) that consists of selecting a defined number of reducts to build the ensemble system [50]. The experimentation results for Heart, Japanese, SD, SD_0.15, SD_0.30 and SD_0.50 databases are given from Table 5 to Table 10. From the obtained results, we can remark that the performance of an ensemble system is greatly influenced by the selected reduct approach.

From those tables, we can also remark that both individual EE*k*-NN classifier and ensemble classifiers built from the *RSR* method provide the most poorly performance comparatively with the *DR*, the *AD-AF* and the *EA-AF* approaches for all the tested databases and for all values of *k* but *RSR* still a little bit better than individual EE*k*-NN classifier. We can conclude that ensemble systems stay anyway better than individual classifier and we can attribute the failure of *RSR* with

Table 7: Results for SD database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*AE-AF*
k=3	PCC	65.83	67.90	77.95	82.34	**85.39**
	Recall	55.15	56.93	73.87	77.23	**85.83**
	Precision	56.20	58.43	75.36	81.19	**86.66**
k=5	PCC	67.43	69.03	78.17	84.52	**85.39**
	Recall	57.46	59.89	75.78	79.38	**85.83**
	Precision	60.12	60.67	76.13	83.24	**86.74**
k=7	PCC	70.21	71.80	79.43	84. 21	**85.9**
	Recall	60.13	60.07	76.18	80.33	**86.43**
	Precision	61.72	62.86	77.15	84.81	**86.94**

Table 8: Results for SD_15 database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*EA-AF*
k=3	PCC	59.76	60.18	70.02	73.19	**76.89**
	Recall	53.20	54.67	68.07	74.15	**76.31**
	Precision	54.10	55.34	70.24	76.12	**78.38**
k=5	PCC	63.67	64.50	73.89	76.56	**78.30**
	Recall	55.01	55.16	70.11	76.13	**79.43**
	Precision	56.17	57.78	73.24	76.34	**77.50**
k=7	PCC	55.99	66.78	74.50	77.89	**79.6**
	Recall	56.63	57.05	71.67	77.43	**78.90**
	Precision	58.04	58.09	72.45	77.13	**80.88**

Table 9: Results for SD_30 database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*EA-AF*
k=3	PCC	57.15	58.67	60.01	68.04	**69.6**
	Recall	52.71	53.22	67.62	68.16	**69.2**
	Precision	53.24	54.90	65.67	67.18	**69.98**
k=5	PCC	58.35	59.14	61.45	70.08	**71.4**
	Recall	53.52	54.63	65.34	67.66	**69.80**
	Precision	55.73	56.82	67.89	69.90	**72.13**
k=7	PCC	60.04	60.89	63.56	71.41	**72.2**
	Recall	54.18	55.14	66.02	67.31	**68.43**
	Precision	56.74	57.20	69.33	73.50	**75.03**

Table 10: Results for SD_50 database

		EE*k*-NN	*RSR*	*DR*	*AD-AF*	*EA-AF*
k=3	PCC	54.17	55.24	57.34	58.51	**59.3**
	Recall	50.25	51.73	59.80	61.71	**64.59**
	Precision	55.10	52.97	60.81	63.23	**58.21**
k=5	PCC	55.19	56.12	58.26	60.41	**61.80**
	Recall	52.71	53.24	59.85	60.05	**60.59**
	Precision	56.15	57.21	62.43	64.88	**65.19**
k=7	PCC	60.99	61.07	62.15	62.98	**63.15**
	Recall	51.17	52.20	59.43	62.76	**67.21**
	Precision	53.20	54.34	59.91	60.97	**61.89**

respect to *DR*, the *AD-AF* and the *EA-AF* to the existence of redundant features as part of the *RSR* selected reducts. To cope with a redundancy problem, we have firstly proposed the *DR* approach allowing the selection of diverse reducts from the original pool. The experimentation results have proven the impact of diversity when constructing an ensemble system. In fact, the PCC, the Recall and the Precision results achieved by the *DR* approach are greater than those obtained by the *RSR* method for almost all cases. Since the diversity between reducts is not sufficient for achieving the best performance, we have proposed the *AD-AF* method that takes into consideration in addition to the diversity between the selected reducts the diversity and the accuracy of base classifiers. The obtained results have proven the effectiveness of this method over the *DR* approach for almost all databases. Thus, we can deduce the effectiveness of the *AD-AF* technique over the *DR* in terms of the ensemble size as well as the classification performance. At last, we showed that the *EA-AF* has yielded the best classification performance for these databases and for the different values of k. We have obtained the same interpretation for all databases.

In this paper, we have also studied the effect of noise on our proposed approach. Different noise levels have been considered 0, 0.15, 0.30 and 0.50. The experimentation results are given from Table 7 to Table 10 and from Figure 3 to Figure 6. From these tables, we can remark that the more uncertainty we have, the lower accuracy we obtain but the *EA-AF* approach still yield the best results compared with the other approaches. For that reason, we have shown the precision-recall curves only for *EA-AF* results. These curves show the tradeoff between the Precision and the Recall for different thresholds. We can remark that in case of no and low uncertainty, we have a high area under the curves representing high recall and high precision and consequently low false positive and low false negative rate. For high uncertainty (Figure 5 and Figure 6), we can remark that curves approach the less accurate area.

4.2.4. Results based on hypothesis test

As presented earlier, a big part of machine learning is to select the best machine learning approach. That means that we have to select the model maximizing the accuracy. Researchers today ask this key question "Is the difference in skill between two machine learning models real, or due to a statistical chance?". In this part of the paper, we use statistical hypothesis testing to address this question. There exist several statistical tests. To cite a few, we mention the McNemar's test and 5×2 Cross-Validation, Paired t-test, etc. These statistical tests are widely used for comparing machine learning approaches. The idea is to compute a $p-value$ and to compare it with a variable $alpha$ to decide if we have to reject null hypothesis or not (i.e. null hypothesis means that there is no difference between two models).To sum up:

- if $p-value >= \alpha$, we fail to reject null hypothesis
- if $p-value < \alpha$, we reject null hypothesis

In this paper, we have chosen to compare the different machine learning approaches based on the Paired t-test Test Statistic. We have ten pairs of models to check: {EEk-NN, *RSR*}, {EEk-NN, $AD-AF$}, {EEk-NN, $EA-AF$}, {*RSR*, *DR*}, {*RSR*, $AD-AF$}, {*RSR*, $EA-AF$}, {EEk-NN, RSR}, {*DR*,$AD-AF$}, {*DR*, $EA-AF$},{$AD-AF$, $EA-AF$}. For the statistical test, we set α to 0.05, the default value used in the majority of scientific articles.

The achieved results have mentioned that the $p-value$ for this present case study is between 0.001 and 0.023 for all pairs of models, all databases and all values of k, As the $p-value$ is smaller

than the considered significance level (i.e., 0.05), we can reject the null hypothesis. Therefore, the results statistically provide convincing evidence that the EEk-NN, the DR, the $AD-AF$ and the $EA-AF$ approaches perform differently. On average, the accuracy for the $EA-AF$ approaches is better than the DR and the $AD-AF$ approaches.

5. Conclusion and future works

In this paper, we have proposed a novel framework for selecting a successfully ensemble of classifiers for addressing data with evidential features. Our framework consists firstly of generating all possible reducts and then selecting the most suitable ones for training individual classifiers. Three approaches have been proposed for selecting the best reducts, namely the Diversity Reduct method (DR), the Accuracy-Diversity Assessment Function method (AD-AF) and the Ensemble Accuracy Assessment Function method (EA-AF). These mentioned approaches have been compared in terms of ensemble size, reduct diversity, classification performance and the difference between the approaches using a statistical test. The achieved results have shown that the EA-AF approach has yielded the best results in terms of the performance criterion using the Dempster combination rule to merge the outputs of the classifiers. Unfortunately, we cannot demonstrate that our method brings better or similar results to further other methods as we are almost the first authors providing this framework. The comparison could not be limited to ensemble classifiers, but generally to classification methods. One limitation of the proposed approach is that bba calculation for big databases requires too many computation resources, as clustering can be numerically hard. As a future work, we look forward to reducing the approach's complexity. We also aim to study the impact of some combination rules in the ensemble performance as well as to use other interesting evaluation criteria like the statistical test.

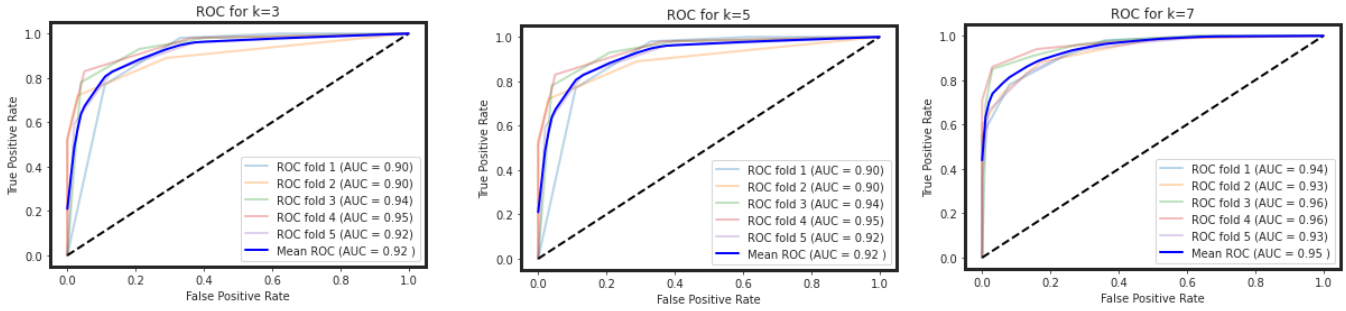

Figure 3: The precision-recall curves for the SD dataset

Compliance with Ethical Standards

Conflict of interest: The authors declare that they have no conflict of interest.
Ethical approval: This article does not contain any studies with animals performed by any of the authors.

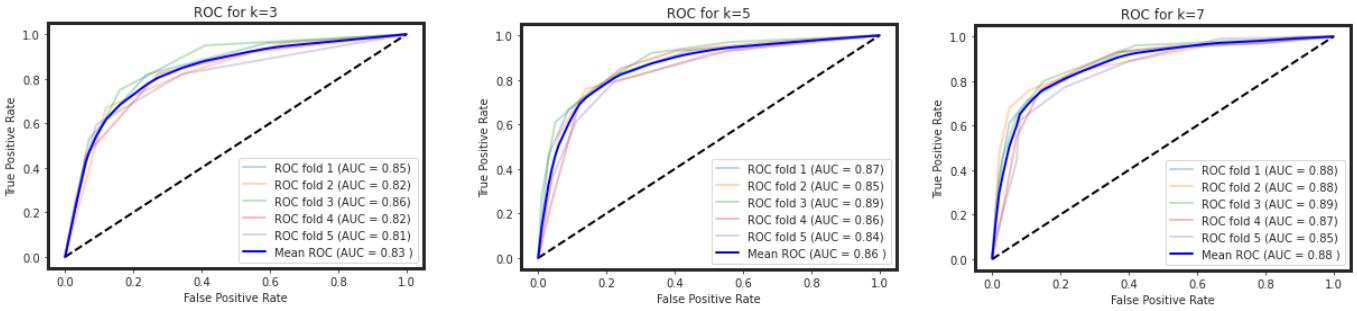

Figure 4: The precision-recall curves for the SD_0.15 dataset

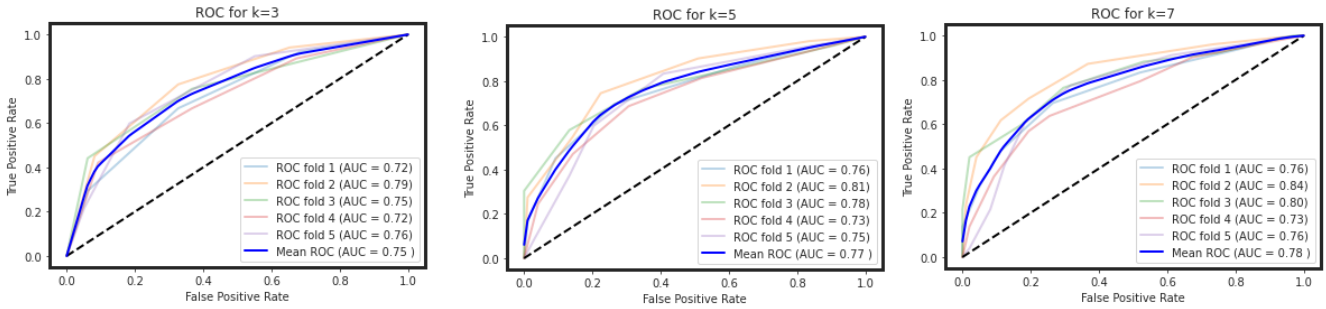

Figure 5: The precision-recall curves for the SD_0.30 dataset

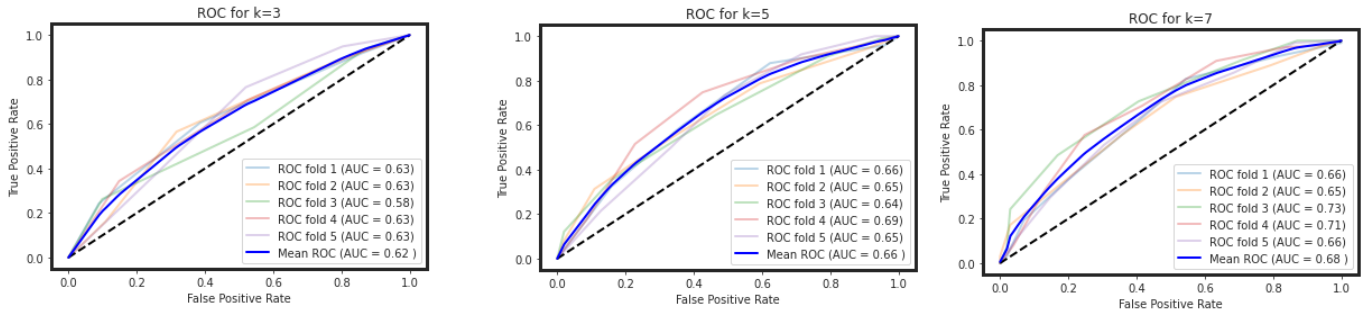

Figure 6: The precision-recall curves for the SD_0.50 dataset

References

[1] H. Chen, Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms, Journal of the American society for Information Science 46 (3) (1995) 194–216.

[2] J. Padmanabhan, M. J. Johnson Premkumar, Machine learning in automatic speech recognition: A survey, IETE Technical Review 32 (4) (2015) 240–251.

[3] N. G. Canbek, M. E. Mutlu, On the track of artificial intelligence: Learning with intelligent personal assistants, Journal of Human Sciences 13 (1) (2016) 592–601.

[4] M. Reformat, R. R. Yager, Building ensemble classifiers using belief functions and owa operators, Soft Computing 12 (6) (2008) 543–558.

[5] B. Quost, M.-H. Masson, T. Denœux, Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules, International Journal of Approximate Reasoning 52 (3) (2011) 353–374.

[6] R. Bryll, R. Gutierrez-Osuna, F. Quek, Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets, Pattern recognition 36 (6) (2003) 1291–1302.

[7] S. Kotsiantis, D. Kanellopoulos, Combining bagging, boosting and random subspace ensembles for regression problems, International Journal of Innovative Computing, Information and Control 8 (6) (2012) 3953–3961.

[8] S. Saha, C. Murthy, S. K. Pal, Rough set based ensemble classifier forweb page classification, Fundamenta Informaticae 76 (1-2) (2007) 171–187.

[9] Z. Pawlak, Rough sets, International Journal of Computer & Information Sciences 11 (5) (1982) 341–356.

[10] L. Shi, X. Ma, L. Xi, Q. Duan, J. Zhao, Rough set and ensemble learning based semi-supervised algorithm for text classification, Expert Systems with Applications 38 (5) (2011) 6300–6306.

[11] L. Shi, L. Xi, X. Ma, M. Weng, X. Hu, A novel ensemble algorithm for biomedical classification based on ant colony optimization, Applied Soft Computing 11 (8) (2011) 5674–5683.

[12] S.-L. Wang, X. Li, S. Zhang, J. Gui, D.-S. Huang, Tumor classification by combining PNN classifier ensemble with neighborhood rough set based gene reduction, Computers in Biology and Medicine 40 (2) (2010) 179–189.

[13] S. Saha, C. Murthy, S. K. Pal, Classification of web services using tensor space model and rough ensemble classifier, in: Proceedings of International Symposium on Methodologies for Intelligent Systems, Springer, 2008, pp. 508–513.

[14] S. K. Lee, Imprecise and uncertain information in databases: An evidential approach, in: 1992 Eighth International Conference on Data Engineering, IEEE Computer Society, 1992, pp. 614–615.

[15] R. Alizadehsani, M. Roshanzamir, S. Hussain, A. Khosravi, A. Koohestani, M. H. Zangooei, M. Abdar, A. Beykikhoshk, A. Shoeibi, A. Zare, et al., Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020), Annals of Operations Research (2021) 1–42.

[16] N. Arunraj, S. Mandal, J. Maiti, Modeling uncertainty in risk assessment: An integrated approach with fuzzy set theory and monte carlo simulation, Accident Analysis & Prevention 55 (2013) 242–255.

[17] M. Song, C. Sun, D. Cai, S. Hong, H. Li, Classifying vaguely labeled data based on evidential fusion, Information Sciences 583 (2022) 159–173.

[18] A. Samet, E. Lefèvre, S. B. Yahia, Evidential database: a new generalization of databases?, in: Proceedings of International Conference on Belief Functions, Springer, 2014, pp. 105–114.

[19] S. Trabelsi, Z. Elouedi, Learning decision rules from uncertain data using rough sets, in: Proceedings of the 8th international FLINS conference on Computational Intelligence In Decision And Control, World Scientific, 2008, pp. 109–114.

[20] A. Campagner, D. Ciucci, T. Denœux, Belief functions and rough sets: Survey and new insights, International Journal of Approximate Reasoning 143 (2022) 192–215.

[21] A. P. Dempster, Upper and Lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.

[22] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[23] L. Huang, S. Ruan, T. Denœux, Application of belief functions to medical image segmentation: A review, Information Fusion 91 (2023) 737–756.

[24] R. Srivastava, T. Mock, Belief Functions in Business Decisions, Physica-Verlag, Heidelberg, Springer-Verlag Company, 2002.

[25] I. Ullah, J. Youn, Y.-H. Han, Multisensor data fusion based on modified belief entropy in Dempster–Shafer theory for smart environment, IEEE Access 9 (2021) 37813–37822.

[26] Q. Liang, Z. Liu, Z. Chen, A networked method for multi-evidence-based information fusion, Entropy 25 (1) (2023).

[27] T. Denœux, L. M. Zouhal, Handling possibilistic labels in pattern classification using evidential reasoning, Fuzzy Sets and Systems 122 (3) (2001) 409–424.

[28] E. Straszecka, Combining uncertainty and imprecision in models of medical diagnosis, Information Sciences 176 (20) (2006) 3026–3059.

[29] A. Ellouze, O. Kahouli, M. Ksantini, H. Alsaif, A. Aloui, B. Kahouli, Artificial intelligence-based diabetes diagnosis with belief functions theory, Symmetry 14 (10) (2022).

[30] Z. Elouedi, K. Mellouli, P. Smets, Belief decision trees: theoretical foundations, International Journal of Approximate Reasoning 28 (2) (2001) 91–124.

[31] S. Trabelsi, Z. Elouedi, P. Lingras, Classification systems based on rough sets under the belief function framework, International Journal of Approximate Reasoning 52 (9) (2011) 1409–1432.

[32] C. Zhu, B. Qin, F. Xiao, Z. Cao, H. M. Pandey, A fuzzy preference-based Dempster-Shafer evidence theory for decision fusion, Information Sciences 570 (2021) 306–322.

[33] T. Denœux, NN-EVCLUS: Neural network-based evidential clustering, Information Sciences 572 (2021) 297–330.

[34] F. Daum, Multitarget-multisensor tracking: principles and techniques, Aerospace and Electronic Systems Magazine, IEEE 11 (2) (1996) 41–44.

[35] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (2) (1994) 191–234.

[36] D. Parmar, T. Wu, J. Blackhurst, MMR: an algorithm for clustering categorical data using rough set theory, Data & Knowledge Engineering 63 (3) (2007) 879–893.

[37] L. Khoo, S. Tor, L. Zhai, A rough-set-based approach for classification and rule induction, International Journal of Advanced Manufacturing Technology 15 (6) (1999) 438–444.

[38] R. B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern recognition letters 26 (7) (2005) 965–975.

[39] P. Kumar, P. Vadakkepat, L. A. Poh, Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets, Applied Soft Computing 11 (4) (2011) 3429–3440.

[40] M. A. El-Monsef, M. Seddeek, T. Medhat, Classification of sand samples according to radioactivity content by the use of euclidean and rough sets techniques, in: proceedings of the 4th nuclear and particle physics, 2003.

[41] A. Trabelsi, Z. Elouedi, E. Lefevre, Feature selection from partially uncertain data within the belief function framework, in: Proceedings of the 16th International Conferenc Information Processing and Management of Uncertainty in Knowledge-Based IPMU, Part II, 2016, pp. 643–655.

[42] Y. Jin, Multi-objective machine learning, Vol. 16, Springer Science & Business Media, 2006.

[43] E. Debie, K. Shafi, C. Lokan, K. Merrick, Reduct based ensemble of learning classifier system for real-valued classification problems, in: proceedings of the IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), IEEE, 2013, pp. 66–73.

[44] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research 11 (1999) 169–198.

[45] L. Kuncheva, C. J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Machine learning 51 (2) (2003) 181–207.

[46] K. Bache, M. Lichman, UCI repository databases, http://www.ics.uci.edu/mlear (1996).

[47] M.-H. Masson, T. Denoeux, ECM: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (4) (2008) 1384–1397.

[48] A. Samet, E. Lefèvre, S. B. Yahia, Evidential data mining: precise support and confidence, Journal of Intelligent Information Systems 47 (1) (2016) 135–163.

[49] A. Trabelsi, Z. Elouedi, E. Lefèvre, A novel k-nn approach for data with uncertain attribute values, in: proceedings of the 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017, Part I, 2017, pp. 160–170.

[50] A. Bar, P. Sai Prasad, Multiple reducts computation in rough sets with applications to ensemble classification, in: Proceedings of ICETIT 2019, Springer, 2020, pp. 449–461.