

# Comparing dependent combination rules under the belief classifier fusion framework

Asma Trabelsi · Zied Elouedi · Eric Lefevre

Received: date / Accepted: date

**Abstract** Data fusion, within the evidence theory framework, consists of obtaining a unique belief function by the combination of several belief functions induced from various information sources. Considerable attention has been paid to combination rules dealing with beliefs induced from non-distinct information sources. The most popular fusion rule is the cautious conjunctive rule, proposed by Denœux. This latter has the empty set, called also the conflict, as an absorbing element. In fact, the mass assigned to the conflict tends toward 1 when applying a high number of the cautious conjunctive operator, and consequently, the conflict loses its initial role as an alarm signal, announcing that there is a kind of disagreement between sources. This problem has led to the introduction of the normalized cautious rule which totally ignores the mass assigned to the conflict. An intermediate rule between the cautious conjunctive and the normalized cautious rules, named the cautious Combination With Adaptive Conflict (cautious CWAC), has been proposed to preserve the initial alarm role of the conflict. Despite this diversification, no great effort has been devoted until now to

find out the most convenient combination rule. Thus, in this paper, we suggest to evaluate and compare the cautious conjunctive, the normalized cautious and the cautious CWAC rules in order to pick out the most appropriate one within the classifier fusion framework.

**Keywords** Belief function theory · combination rules · dependent information sources · multiclassifier fusion framework.

## 1 Introduction

Pattern recognition is an exciting field in machine learning which has been widely studied over the past few years to solve classification problems. Among the proposed solutions, some have relied on an advanced method named multiple classifier systems or ensemble classifier (Wolpert, 1992; Kittler, Hatef, Duin, & Matas, 1998; Dietterich, 2000; Reformat & Yager, 2008). The underlying idea behind the use of ensemble classifier is that different classifiers can provide higher classification accuracy about patterns to be classified (Sharkey & Sharkey, 1997; Kuncheva, Skurichina, & Duin, 2002; Kuncheva & Whitaker, 2003). Ensemble classifier can take different forms: A such ensemble can be constructed by learning different machine learning classifiers using all the training examples (Trabelsi, Elouedi, & Lefevre, 2015a, 2015b), by training the same classifier on different subsets of training instances [bagging (Breiman, 1996), boosting (Freund & Schapire, 1997) and cross-validation] or on different subsets of features [Random Subspaces (Ho, 1998; Bertoni, Folgieri, & Valentini, 2005; Kuncheva, Rodríguez, Plumpton, Linden, & Johnston, 2010), Rotation Forests (Rodríguez, Kuncheva, & Alonso, 2006)] and also by introducing randomness into a learning procedure [e.g., using different initial weights for the Nearest Neighbor classifier, picking different splits for the

---

Asma Trabelsi  
Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia  
Tel.: +21627109150  
EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), University Artois, 62400, Béthune, France  
E-mail: trabelsyasma@gmail.com

Zied Elouedi  
Larodec, Institut Supérieur de Gestion de Tunis, Université de Tunis, Tunisia  
Tel.: +21698354307  
E-mail: zied.elouedi@gmx.fr

Eric Lefevre  
EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), University Artois, 62400, Béthune, France  
Tel.: +330321632300  
E-mail: eric.lefevre@univ-artois.fr

decision tree classifier (Dietterich, 2000; Geurts, Ernst, & Wehenkel, 2006), etc]. Up to now, various combination approaches have been proposed to combine diverse classifiers such as voting-based systems (Ruta & Gabrys, 2005), plurality (Hansen & Salamon, 1990), fuzzy integrals (Cho & Kim, 1995; Pizzi & Pedrycz, 2010), classifier local accuracy (Woods, Kegelmeyer, & Bowyer, 1997), Bayesian theory (L. Xu, Krzyzak, & Suen, 1992; Kuncheva, Skurichina, & Duin, 2002), multilayered perceptrons (Huang, Liu, & Suen, 1995) and belief function theory (L. Xu, Krzyzak, & Suen, 1992; P. Xu, Davoine, & Denœux, 2014; P. Xu, Davoine, Zha, & Denœux, 2016; Mercier, Cron, Denœux, & Masson, 2005; Le, Huynh, Shimazu, & Nakamori, 2007; Quost, Denœux, & Masson, 2007, 2008; Bi, Guan, & Bell, 2008; Reformat & Yager, 2008). The belief function theory, originated from the work of Dempster (1967) and Shafer (1976), is regarded as a convenient method for representing and managing different kinds of imperfect data and has proved to be an efficient tool for merging. Thanks to its combination rules, a member of classifiers such that their outputs are expressed in terms of mass functions (Al-Ani & Deriche, 2002) can be combined. The basic rule of combination is Dempster's rule also called orthogonal sum. This latter requires a normalization task in order to maintain the basic properties of the belief functions (Lefevre, Colot, & Vannoorenberghe, 2002; Yen, 1990). Zadeh (1986) has proved that this normalization can involve counterintuitive behaviors in the case where the initial conditions of the Dempster rule are not respected (e.g., where the information sources to be combined are not reliable). Thus, several combination rules have been developed to address this problem such as Yager's rule (1987), Dubois and Parde's rule (1988), Smets's rule (1990a) and Murphy's rule (2000). However, these rules can only be used when the combined information sources are independent. A source is presumed to be independent on another one when its belief knowledge does not affect the belief of the other one. Quite recently, considerable attention has been paid to the combination of beliefs coming from dependent sources of information such as the cautious rule and its unnormalized version (Denœux, 2006), the bold rule (Denœux, 2008), the cautious Combination With Adaptive Conflict rule (cautious CWAC) (Boubaker, Elouedi, & Lefevre, 2013), the t-norms-based rules (2008) and Cattaneo's rule (2003), etc. An information concerning the independence of sources or not guides the choice of the type of combination rules to be used. Nevertheless, the selection of the most convenient rule among several alternatives is crucial, but the criteria to be based on have not been deeply reported until now. Thus, in this paper, we suggest to evaluate and compare combination rules within the classifier fusion framework to find out the most adequate rule. More concretely, we evaluate and compare combination rules dealing with non-independent information sources, in particular

the cautious conjunctive rule, the normalized cautious rule and the cautious CWAC rule. This choice is justified by the fact that these alternative rules mainly differ according to the way of managing the mass assigned to the empty set called also conflict. Indeed, the cautious conjunctive rule, which is the first proposed rule to deal with non-independent information sources, does not keep the initial alarm role of the conflict owing to the use of the conjunctive operator that has the empty set as an absorbing element (Lefevre & Elouedi, 2013). In view of this problem, the normalized cautious rule, obtained by the use of the Dempster operator, has been proposed to ignore completely the value of the conflict yielded by combining pieces of evidence. On the other hand, the cautious Combination With Adaptive Conflict rule (cautious CWAC), which is defined by an adaptive weighting between the cautious conjunctive and the normalized cautious rules, gives the conflict its initial role as an alarm signal indicating that there is a sort of disagreement between sources (Boubaker, Elouedi, & Lefevre, 2013). Therefore, in this investigation, we propose to compare the cautious CWAC rule with the cautious conjunctive and the normalized cautious rules according to some proposed assessment criteria. The remaining of this paper is organized as follows: We provide in Sect 2 a brief overview of the fundamental concepts of the belief function theory. We outline various combination rules in Sect 3. We present the procedure of combining multiple classifiers within the belief function framework in Sect 4. Sect 5 is dedicated to highlighting our comparative approach. Our experiments and results are given in Sect 6. The conclusion is reported in Sect 7.

## 2 Fundamental concepts of the belief function theory

The belief function theory is regarded as a practical and efficient framework for representing and managing uncertain knowledge. This section briefly recalls some basic concepts of this theory.

### 2.1 Information representation

Let  $\Theta$  be a finite non-empty set of  $N$  elementary events related to a given problem. These events are assumed to be exhaustive and mutually exclusive. Such set  $\Theta$ , called the frame of discernment, is defined as:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_N\} \quad (1)$$

The agent beliefs on  $2^\Theta$  are represented by the so-called basic belief assignment (bba), named originally by Shafer (1976) basic probability assignment (bpa). The bba is de-

defined as follows:

$$\begin{aligned} m : 2^\Theta &\rightarrow [0, 1] \\ m(\emptyset) = 0 \text{ and } \sum_{A \subseteq \Theta} m(A) &= 1 \end{aligned} \quad (2)$$

The amount  $m(A)$ , called basic belief mass (bbm), states the part of belief that exactly supports the event  $A$ . All the subsets  $A$  of  $\Theta$  such that  $m(A)$  is strictly positive are referred to the focal elements of  $m$ .

In Shafer's original work (1976), the mass assigned to the empty set should have a null value (i.e., impossible proposition). Such bba is called a normalized basic belief assignment. However, Smets (1990b) allows unnormalized bba and interprets  $m(\emptyset)$  as the amount of conflict between the pieces of evidence or as the part of belief given when the true value is not included in the frame of discernment.

As well, it is possible to transform any unnormalized bba, denoted by  $m'$ , into a normalized one, denoted by  $m$ . The normalization process is defined by:

$$m(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ k \cdot m'(A) & \text{otherwise} \end{cases} \quad (3)$$

where  $k^{-1} = 1 - m(\emptyset)$  is called the normalization factor.

From the bba  $m$ , a belief function  $\text{bel}$ , a plausibility function  $\text{pl}$  and a communality function  $q$  are defined, respectively, as:

$$\text{bel}(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad (4)$$

$$\text{pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (5)$$

and

$$q(A) = \sum_{B \supseteq A} m(B) \quad (6)$$

The quantity  $\text{bel}(A)$  represents the total mass of belief that supports the event  $A$ , while  $\text{pl}(A)$  quantifies the maximum possible support that could be given to a subset  $A$  of the frame of discernment  $\Theta$ . On the other hand, the communality  $q(A)$  stands for the sum of masses allocated to the supersets of  $A$ . It is important to note that  $m$ ,  $\text{bel}$ ,  $\text{pl}$  and  $q$  are regarded as various expressions of the same information (Denceux, 1999).

## 2.2 Special bbas

In this subsection, we will focus on some special cases of bbas used for expressing particular situations of uncertainty. A vacuous bba is a normalized bba that has  $\Theta$  as its unique focal element (Shafer, 1976). Such bba, representing the state of total ignorance, is defined as follows:

$$m(\Theta) = 1 \text{ and } m(A) = 0 \forall A \neq \Theta \quad (7)$$

A categorical bba is a normalized bba that has a unique focal element  $A$  different from the frame of discernment  $\Theta$ . It is defined as follows:

$$\begin{aligned} m(A) &= 1, \text{ for some } A \subset \Theta \\ m(B) &= 0, \text{ for } B \subseteq \Theta, B \neq A \end{aligned} \quad (8)$$

A certain bba is a particular case of the categorical belief function where its unique focal element is a singleton. This bba represents the state of total certainty. It is defined as follows:

$$\begin{aligned} m(A) &= 1 \text{ for some } A \subset \Theta \text{ and } |A|=1 \\ m(B) &= 0 \text{ for all } B \subseteq \Theta \text{ and } B \neq A \end{aligned} \quad (9)$$

A bba is called simple support function (ssf) if it has at most two focal elements: the frame of discernment  $\Theta$  and a strict subset of  $\Theta$  called the focus of the ssf (Smets, 1995). A simple support function is defined as follows (Smets, 1995):

$$m(X) = \begin{cases} w & \text{if } X = \Theta \\ 1 - w & \text{if } X = A \text{ for some } A \subseteq \Theta \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

In simple terms, such ssf can be written as  $A^w$  where  $A$  is the focus and  $w \in [0, 1]$ .

A bba is named dogmatic only if the frame of discernment  $\Theta$  is not a focal element ( $m(\Theta) = 0$ ); otherwise it is called non-dogmatic ( $m(\Theta) > 0$ ).

## 2.3 Discounting

The combination of bbas induced from not fully reliable information sources may lead to a non-informative bba. However, a method of discounting seems to be imperative before combining sources. Let  $m$  be a bba induced from an information source with a reliability rate  $(1 - \alpha)$  (Shafer, 1976). The discounted bba  $m^\alpha$  is obtained as follows:

$$\begin{aligned} m^\alpha(A) &= (1 - \alpha)m(A) \forall A \subset \Theta. \\ m^\alpha(\Theta) &= \alpha + (1 - \alpha)m(\Theta) \end{aligned} \quad (11)$$

where the coefficient  $\alpha \in [0, 1]$  is called the discounted factor.

## 2.4 Distance between two bbas

The distance measures of evidence represent the degree of dissimilarity between pieces of evidence. One of the well-known distances is the one proposed by Jousselme. The Jousselme et al. et al. (2001) distance between two bbas  $m_1$  and  $m_2$  is found as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D(m_1 - m_2)} \quad (12)$$

where  $D$  is the Jaccard similarity measure defined by:

$$D(A, B) = \begin{cases} 1 & \text{if } A=B=\emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^\Theta \end{cases} \quad (13)$$

## 2.5 Pignistic transformation

To make a decision within the belief function framework, Smets (1988) proposes to transform beliefs into a probability measure called the pignistic probability and denoted by BetP. It is defined as follows:

$$\text{BetP}(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)} \quad \forall A \in \Theta \quad (14)$$

## 3 Combining information sources

Data fusion consists in obtaining a relevant information by the combination of several imperfect data (uncertain, imprecise and incomplete) (Lefevre, Colot, & Vannoorenberghe, 2002). The belief function framework provides numerous tools for ensuring fusion. Some of them treat the problem of independent information sources, while others assume sources of information combined to be dependent. Two information sources are dependent in the case where the knowledge of the belief induced by one source affects the belief of the other source; otherwise these sources are independent (Denoeux & Masson, 2012).

### 3.1 Combining independent pieces of evidence

As mentioned earlier, there exist several combination rules assuming items of evidence combined to be independent. In this subsection, we emphasize on the conjunctive rule (Smets, 1998), the Dempster's rule (1967) and the CWAC rule (Lefevre & Elouedi, 2013).

The conjunctive rule, proposed by Smets (1998), is used to combine two bbas provided by reliable and distinct information sources. The resulting bba, denoted  $m_1 \odot m_2$ , is

defined by:

$$(m_1 \odot m_2)(A) = \sum_{B, C \subseteq \Theta: B \cap C = A} m_1(B) \cdot m_2(C) \quad (15)$$

The mass assigned to the empty set ( $m_1 \odot m_2(\emptyset)$ ) quantifies the degree of disagreement between the two combined sources. In the case where the number of information sources increases, the conflict tends toward 1 and consequently loses its initial role as an alarm signal reflecting the existence of disagreement between sources.

The Dempster rule, based on the orthogonal sum, is a normalized version of the conjunctive rule where the mass of the empty set must be reallocated over all focal elements in the case where  $m_1 \odot m_2(\emptyset) \neq 0$  thanks to a normalization factor, denoted  $k$  (Shafer, 1976). This rule, assuming pieces of evidence combined to be reliable and distinct, is defined as follows:

$$(m_1 \oplus m_2)(A) = k(m_1 \odot m_2)(A) \quad (16)$$

and

$$(m_1 \oplus m_2)(\emptyset) = 0 \quad (17)$$

where

$$k^{-1} = 1 - (m_1 \odot m_2)(\emptyset) \quad (18)$$

The conjunctive and the Dempster rules are commutative, associative but not idempotent which make them unusable when sources are non-distinct.

Authors, in Lefevre and Elouedi (2013), have proposed the CWAC combination rule which is defined by an adaptive weighting between the conjunctive and the Dempster rules. This adaptive weighting offers an effective way to obtain the same behavior as the conjunctive rule when the bbas are contradictory and the same behavior as the Dempster rule when the bbas are similar. Assume we have  $M$  bbas, denoted as  $m_1, \dots, m_M$ , the result of their combination using the CWAC operator is noted as  $m_{\odot \oplus}$  and is defined as follows:

$$m_{\odot \oplus}(A) = D m_{\odot}(A) + (1 - D) m_{\oplus}(A) \quad (19)$$

and

$$m_{\odot \oplus}(\emptyset) = 1 \text{ when } m_{\odot}(\emptyset) = 1 \quad (20)$$

with  $m_{\odot}(A) = (\odot_i m_i)(A)$ ,  $m_{\oplus}(A) = (\oplus_i m_i)(A)$  and  $D = \max_{i,j} [d(m_i, m_j)]$  is the Jousselme distance between  $m_i$  and  $m_j \forall i, j \in [1, M]$ .

The CWAC rule is commutative, not associative and non-idempotent.

### 3.2 Combining dependent pieces of evidence

By against of the conjunctive operator, the cautious conjunctive rule, proposed by Denœux and denoted by  $\bigotimes$ , is used to combine pieces of evidence induced from reliable dependent information sources based on the conjunctive canonical decomposition stated by Smets (1995). Let  $m_1$  and  $m_2$  be two non-dogmatic bbas, the result of their combination, denoted by  $m_1 \bigotimes m_2$ , is given as follows (Denœux, 2006):

$$m_1 \bigotimes m_2(A) = \bigodot_{A \subseteq \Theta} A^{w_1(A) \wedge w_2(A)} \quad (21)$$

where  $w_1(A) \wedge w_2(A)$  corresponds to the weight function of a bba  $m_1 \bigotimes m_2$  and  $\wedge$  represents the minimum operator. The weights  $w(A)$  for every  $A \subset \Theta$  can be obtained from the commonalities as follows:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|}-1}. \quad (22)$$

It is important to note that information sources may be dependent without being identical (e.g. experts may observe different but correlated variables). As the cautious conjunctive rule inherits the absorbing effect of the conjunctive rule, the conflict may lose its initial alarm role when we apply a significant number of cautious conjunctive combinations.

The normalized version of the cautious conjunctive rule, denoted by  $\bigotimes^*$ , is inspired from the behavior of the Dempster operator in order to overcome the value of the conflictual mass (Denœux, 2006). It is defined as:

$$m_1 \bigotimes^* m_2 = \bigoplus_{\emptyset \neq A \subset \Theta} A^{w_1(A) \wedge w_2(A)} \quad (23)$$

We thus have:

$$m_1 \bigotimes^* m_2(A) = k \cdot m_1 \bigotimes m_2(A) \quad (24)$$

with  $k^{-1} = 1 - m_1 \bigotimes m_2(\emptyset)$  and  $m_1 \bigotimes^* m_2(\emptyset) = 0$ .

This rule has the same properties as its unnormalized version counterpart: it is commutative, associative and idempotent.

An intermediate rule between the cautious conjunctive and the normalized cautious rules has been developed by (Boubaker, Elouedi, & Lefevre, 2013) under the name of cautious CWAC rule and denoted by  $m_{\odot}$ . This proposed rule is defined by an adaptive weighting between the cautious conjunctive rule and its unnormalized version in order to manage the conflict induced from the combination of several information sources. As a consequence, this latter not only ensures the combination of dependent pieces of evidence but also keeps the conflict as an alarm signal announcing that there is a kind of disagreement between sources. Assuming there are  $M$  bbas, denoted as  $m_1, \dots, m_i, \dots,$

,  $m_M$ , the result of the cautious CWAC operator is set to:

$$\forall A \subseteq \Theta, m_{\odot}(\emptyset) \neq 1;$$

$$m_{\odot}(A) = D m_{\bigotimes}(A) + (1 - D) m_{\bigotimes^*}(A) \quad (25)$$

with  $D = \max_{i,j} [d(m_i, m_j)]$  is the Jousselme distance measure between  $m_i$  and  $m_j \forall i, j \in [1, M]$ .

The cautious CWAC rule is commutative but not associative. It is idempotent only in case where  $D$  is equal to 0.

## 4 The combination of multiple classifiers within the belief function framework

An increasing interest in the classification process has relied on the so-called ensemble classifier systems (Dietterich, 2000; Kittler, Hatef, Duin, & Matas, 1998; Wolpert, 1992). These kinds of systems are based on the paradigm of combining the outputs of a member of individual classifiers in order to achieve the best classification accuracy. The belief function framework has proved to be an efficient tool for ensuring classifier fusion (Al-Ani & Deriche, 2002).

Authors, in Xu et al. (1992), have shown that the procedure of merging the outputs of numerous classifiers, when using the framework of belief functions, comprises two distinct levels. The first one consists of building mass functions from classifiers' outputs, whereas the second one is dedicated to combining these mass functions through a combination rule  $R$  to finally give an improved information decision  $D_F$  (see Fig 1).

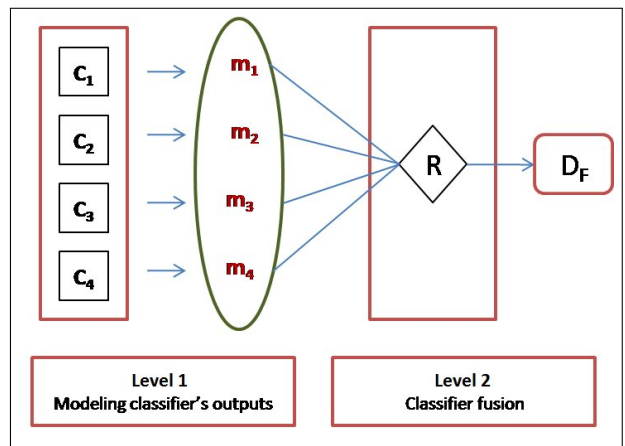


Fig. 1 Classifier fusion within the belief function framework

#### 4.1 Modeling classifiers' outputs

One of the most important key points within the belief function framework concerns the assignment of masses. Therefore, several assignments have been discussed based on the performance of classifiers in terms of recognition rate (correct answer), substitution rate (wrong answer) and rejection rate which are computed from a confusion matrix. Let  $C = \{C_1, \dots, C_M\}$  be a set of  $M$  classifiers, and let  $\Theta = \{\theta_1, \dots, \theta_K\}$  be a set of  $K$  class labels. Each classifier takes as input a pattern test  $x$  and outputs either a class label  $\theta_k$  or a rejection class  $\theta_{k+1}$ . The confusion matrix  $M_i$  relative to a classifier  $C_i$  is given by:

$$M_i = \begin{pmatrix} n_{11}^i & \cdots & n_{1K}^i \\ n_{21}^i & \cdots & n_{2K}^i \\ \vdots & & \vdots \\ n_{K1}^i & \cdots & n_{KK}^i \\ n_{(K+1)1}^i & \cdots & n_{(K+1)K}^i \end{pmatrix}$$

where each row  $k$  corresponds to the predicted class  $\theta_k$ , each column  $l$  corresponds to the actual class  $\theta_l$ , an element  $n_{kl}^i$  represents the number of test patterns of class  $\theta_l$  which have been classified by  $C_i$  in class  $\theta_k$  and  $n^i$  denotes the total number of patterns classified by the classifier  $C_i$ . The performance rates of  $C_i$  are computed as follows:

- The recognition rate:

$$R_i = \frac{\sum_{k=1}^K n_{kk}^i}{n^i} \quad (26)$$

- The substitution rate:

$$S_i = \frac{\sum_{k=1}^K \sum_{l=1; l \neq k}^K n_{kl}^i}{n^i} \quad (27)$$

- The rejection rate:

$$T_i = 1 - (R_i + S_i) \quad (28)$$

There exist several assignments based on the performance rates of classifiers (Mercier, Cron, Denœux, & Masson, 2005) such as the Xu's assignment (1992), the reliability assignment (Johansson, Boström, & Karlsson, 2008), the calibration method (P. Xu, Davoine, & Denœux, 2014). In this paper, we propose to model each classifier output using the reliability assignment. Suppose that the classifier  $C_i$  outputs a class label  $\theta_k$ , the basic belief assignment is computed as follows:

$$m_i(\theta_k) = R_i \quad (29)$$

$$m_i(\Theta) = 1 - R_i$$

The output results associated with  $M$  classification methods are different; by the way, we will have  $M$  number of mass functions for each test pattern. In the case where the

outputs of all classifiers are converted into bbas, we move on to the next step which consists of the classifier fusion through some combination rules provided by the belief function framework.

#### 4.2 Classifier fusion

This level concerns the problem related to the question "How to aggregate the results from a set of classifiers such a way that we obtain the best result?". As there exist several combination rules within the belief function framework, the choice of the most efficient one becomes really a challenging task and it still requires non-trivial effort. In a previous work (Trabelsi, Elouedi, & Lefevre, 2015a), we have proved that the CWAC rule is the well-suited rule in the case of independent classifiers. In this paper, we propose to carry out a comparative study which allows to select the best fusion rule among the cautious conjunctive, the normalized cautious and the cautious CWAC ones.

### 5 Comparative approach

Let us remind that our purpose is to compare the cautious CWAC rule with the cautious conjunctive and the normalized cautious ones. To achieve our goal, we suggest to resort on the distance and the PCC criteria. The distance criterion, corresponding to the Jousselme distance between the testing instance's bba and its real class, is used to compare the cautious CWAC rule with the cautious conjunctive one, whereas the PCC criterion, representing the percent of the correctly classified instances, is used to compare the cautious CWAC rule with the normalized cautious one.

- For comparing the cautious CWAC rule with the cautious conjunctive one, we should proceed as follows:

1. The real class  $\theta_j$  of all test instances must be transformed into a mass function as follows:

$$m_r(\{\theta_j\}) = 1 \quad (30)$$

$$m_r(A) = 0 \quad \forall A \subseteq \Theta \text{ and } A \neq \{\theta_j\}$$

2. Subsequently, we compute for each instance the Jousselme distance between the mass function relative to its real class ( $m_r$ ) and the mass function obtained by combining bbas induced from  $M$  classifiers.
  3. At last, we sum the Jousselme distances produced by all test instances to get a total one.
  4. The most adequate fusion rule is the one that has the minimum value of total distance.
- The comparison of the cautious CWAC rule with the normalized cautious one requires the use of 3 variables  $n_1$ ,  $n_2$  and  $n_3$  which, respectively represent the number

of well-classified, misclassified and rejected instances. Whence, for each combination rule, we suggest the following stages:

1. At first, we have set tolerance threshold values to be  $S = \{0.1, 0.2, \dots, 1\}$ . Then, for any test pattern, we verify the mass assigned to the empty set  $m(\emptyset)$ . In the case where  $m(\emptyset)$  is greater than the threshold  $s$ , the classifier rejects instance in place of misclassifying it and consequently, the number of rejected instances  $n_3$  must be incremented. On the contrary, if  $m(\emptyset)$  is less than the threshold  $s$ , we calculate the pignistic probability (BetP) to make a decision about the class to be chosen. Besides, the actual class will be compared to the real one: We increment  $n_1$  if there are similar and  $n_2$  if they are different.
2. Once we have calculated our well-classified, misclassified and rejected instances, we calculate then the *PCC* for each threshold  $s \in S$  as follows:

$$PCC = \frac{n_1}{n_1 + n_2} * 100 \quad (31)$$

3. Finally, we compute the average of all *PCCs* to obtain a mean *PCC*.
4. The best combination rule is the one that has the highest value of mean *PCC*.

Note that the use of two evaluation criteria is justified by the fact that the mass assigned to the empty set is almost high for the cautious conjunctive rule. In such a case, applying the *PCC* criterion may cause a high number of rejected instances leading to erroneous results. In addition, the mass attributed to the empty set, when using the normalized cautious rule, is null, and consequently, the Jousselme distance, which mostly tends toward 0, may adversely affect the results.

## 6 Implementation and simulation

In this section, we detail the different experimentation tests that we have employed to compare the cautious CWAC rule with both conjunctive and normalized cautious rules under the classifier fusion framework.

### 6.1 Experimental settings

To ensure the implementation of our approach, we have developed our program with MATLAB V7.13 (R2013a). We have further performed our experiments on several real databases obtained from the UCI repository (P. Murphy & Aha, 1996). A brief description of these databases is presented in Table 1 where # Instances, #Attributes and #Classes denote, respectively, the total number of instances, the number of attributes and the number of classes. Although

the classes number of all chosen databases is equal to either two or three, our alternative combination rules can also support a class number greater than three.

**Table 1** Description of databases

Databases	#Instances	#Attributes	#Classes
Nursery	12960	8	5
Chess	3196	36	2
Tic-Tac-Toa	958	9	2
Diabetes	768	2	2
Car Evaluation	1728	6	4
Hepatitis	155	19	2
Iris	150	4	3
Parkinson's	195	23	2
Wine	178	13	3

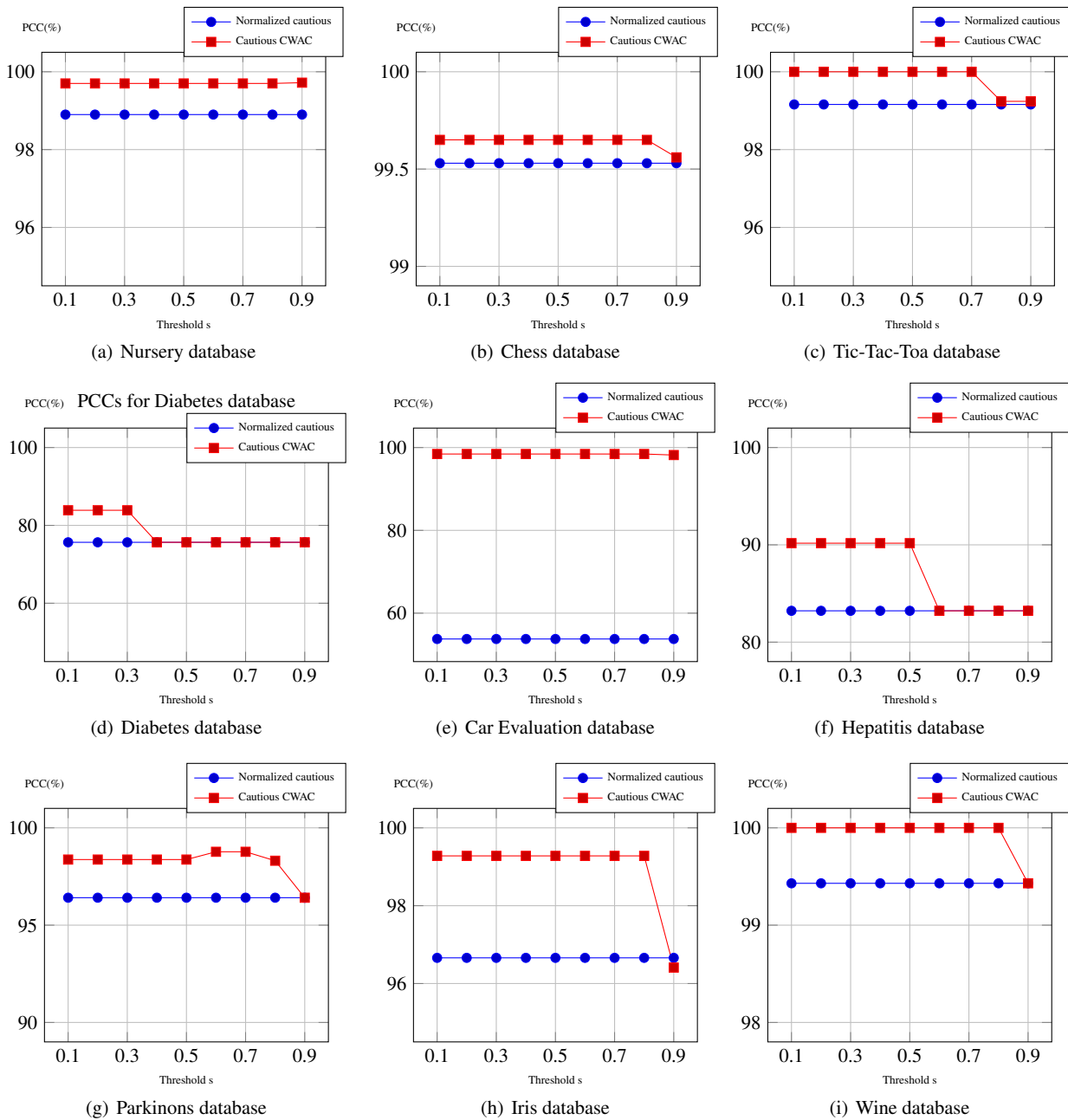
We have carried out experiments using four popular machine learning algorithms implemented in Weka (Hall et al., 2009); Naive Bayes (NB), *k*-Nearest Neighbors (*k*-NN), Decision Tree (DT) and Neural Network (NN) algorithms. We have run these Weka classifiers on Matlab based on the principle of the leave-one-out cross validation method which divides a data set with  $N$  instances into  $N-1$  parts for training and the remaining instance for testing. This process is repeated  $N$  times where each instance is used once as a test set. As a result, we get  $N$  test patterns with their predicted class labels from each classifier.

### 6.2 Experimental results

Within the framework of classifier fusion, several factors including the number of classifiers, the performance of individual classifiers and the dependency between classifiers may affect the combination rule behavior. The influence of classifier dependencies on the performance of various combination rules has been the interest of numerous researchers (Quost, Masson, & Denœux, 2011). As a matter of fact, several correlation measures have been proposed in the literature. One of the most commonly used measures is the one proposed by Goebel and Yan (2004). This latter, denoted by  $\rho$ , allows to quantify the dependency between a member ( $r \geq 2$ ) of classifiers. Based on the outputs of the individual classifiers, the correlation coefficient is computed as follows:

$$\rho = \frac{m.N^f}{N - N^f - N^t + m.N^f} \quad (32)$$

where  $m$  is the number of classifiers,  $N$  is the number of test patterns,  $N^t$  is the number of test patterns for which all classifiers had a right answer, and  $N^f$  is the number of test patterns for which all classifiers had a wrong answer. In



**Fig. 2** PCC results of the normalized cautious and the cautious CWAC rules

the following, we will study the influence of classifier dependencies on the relative performance of the cautious conjunctive, the normalized cautious and the cautious CWAC rules. Thus, we propose to carry out different experimental tests where classifier dependency degrees will differ from one experimentation to another. The results carried out from the different tests will be presented and analyzed in order to pick out the most appropriate fusion rule.

### 6.2.1 Experimentation 1

As a first experimentation, we look forward to combine a set of classifiers with a low correlation through the cautious conjunctive, the normalized cautious and the cautious CWAC rules. Thus, we propose to combine heterogeneous sets of classifiers which are able to make different errors on a new pattern to be classified. Exactly, we have chosen to combine the 1-NN, the NB, the NN and the DT classifiers. The correlation coefficient between these classifiers is presented in Table 2. As seen in Table 2, the degree of



**Table 2** Correlation coefficient between 1-NN, NB, NN and DT classifiers

Databases	$\rho$ -correlation
Nursery	0.005
Chess	0.01
Tic-Tac-Toa	0
Diabetes	0.535
Car Evaluation	0.05
Hepatitis	0.505
Iris	0.2667
Parkinsons	0.1
Wine	0

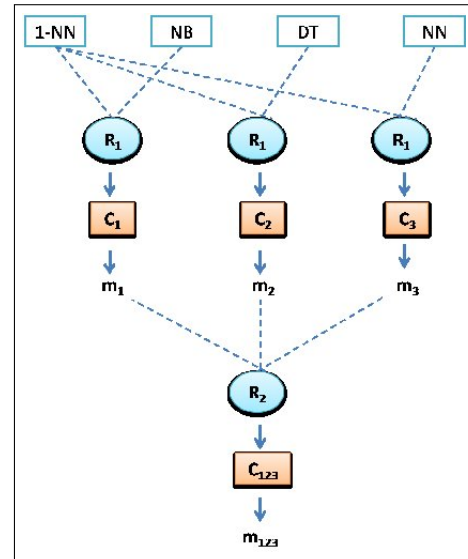
correlation between the 1-NN, the NB, the NN and the DT classifiers is almost low for the different databases (e.g., the  $\rho$ -correlations are equal to 0 for Iris database and 0.1 for Parkinson's database). Accordingly, we will study the behavior of the three alternative operators, where classifiers to be combined have low correlation, in the purpose of finding out the most convenient one. Let us lead off by comparing the cautious conjunctive rule with the cautious CWAC one in terms of the distance criterion. The total distance results are given in Table 3 where we can notice that the total distance relative to the cautious CWAC rule is lower than the one obtained by applying the cautious conjunctive rule for all the databases. For instance, the total distance relative to the Car Evaluation database by the use of the cautious conjunctive operator is equal to 876.39, whereas the total distance obtained by applying the cautious CWAC operator is equal to 852.50. This interpretation is also available for all the other databases. Accordingly, we can say that the cautious CWAC rule is more efficient than the cautious conjunctive one in terms of the distance criterion. From this conclusion, we move on to the comparison of the cautious CWAC rule with the normalized cautious rule in terms of the PCC criterion. Figure 2 represents the mean  $PCC$  values for both the normalized cautious and the cautious CWAC rules relative to all the mentioned databases. This latter indicates that the cautious CWAC rule has the highest value of mean  $PCC$  compared with the normalized cautious one for all the databases. These results may be explained by the fact that the mean number of the rejected instances relative to the cautious CWAC rule is greater than 0, while that relative to the normalized cautious one is equal to 0. Accordingly, we can conclude that the cautious CWAC rule is more efficient than the normalized cautious rule in terms of the  $PCC$  criterion.

### 6.2.2 Experimentation 2

We recall that we aim to study the behavior of the cautious conjunctive, the normalized cautious and the cautious CWAC rules with different classifier dependency degrees. Thus, in this experimentation, we tend to increase the corre-

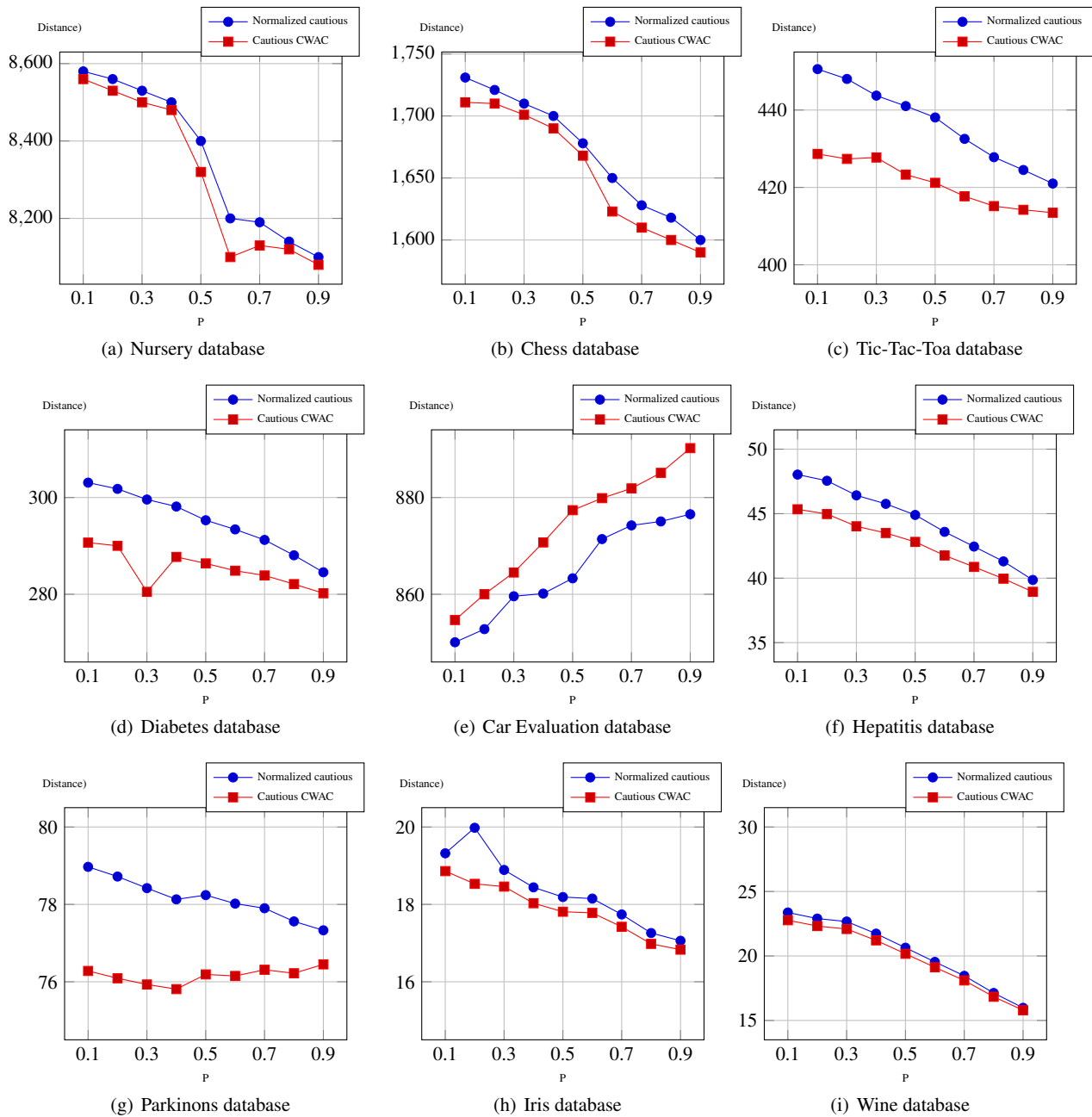
**Table 3** Distance results

Databases	Cautious conjunctive	Cautious CWAC
Nursery	8624.19	8481.42
Chess	1731.05	1713.93
Tic-Tac-Toa	280.17	252.50
Diabetes	334.87	312.45
Car Evaluation	876.39	852.50
Hepatitis	54.90	51.56
Iris	15.09	14.78
Parkinsons	60.92	55.31
Wine	24.47	23.83

**Fig. 3** Two fusion rules to combine classifiers

lation between the 1-NN, the NB, the NN and the DT classifiers. To achieve our goal, we propose to construct three correlated classifiers, denoted by  $C_1$ ,  $C_2$ ,  $C_3$ , obtained respectively by combining 1-NN classifier with NB classifier, 1-NN classifier with DT classifier and 1-NN and NN classifiers. Subsequently, the outputs obtained by  $C_1$ ,  $C_2$  and  $C_3$  will be combined in a same way to obtain a consensus result (see Figure 3 where  $R_1$  corresponds to the conjunctive rule and  $R_2$  will be either the cautious conjunctive rule, the normalized cautious rule or the cautious CWAC rule).

The correlation coefficient between  $C_1$ ,  $C_2$  and  $C_3$  is given in Table 4 for the different mentioned databases. It is clear from the results in Table 4 that the  $\rho$ -correlation coefficient remains high for the different databases (e.g., for the Parkinsons we have 1 as  $\rho$ -correlation value). As a result, we can conclude that  $C_1$ ,  $C_2$  and  $C_3$  are highly correlated. From this conclusion, we move on to analyze the performance of the three alternative combination rules. The total distance and the mean PCC results are presented in Table 5. From this latter, we can notice that the total distance of the cautious conjunctive rule is greater than that relative to the cautious CWAC one for the different databases (e.g.



**Fig. 4** Distances for both cautious conjunctive and cautious CWAC rules

for the database Diabetes, the total distance of the cautious conjunctive and the cautious CWAC rules are respectively 334.87 and 295.23). Accordingly, it can be said that the cautious CWAC rule is better than the cautious conjunctive one in term of the distance criterion. Besides, we can note that the mean PCC of the cautious CWAC rule is greater than that relative to the normalized cautious one for the different databases. For instance, for Hepatitis database, we have 83.22% as normalized cautious mean PCC and 85.79% as cautious CWAC mean PCC. So far, we can assume that this interpretation is also available for the other databases. From

this, there is no doubt that the cautious CWAC rule is the best fusion rule compared with the normalized cautious according to the PCC criterion.

### 6.2.3 Experimentation 3

In the previous experimentation, we have generated a set of correlated classifiers based on the output of multiple distinct classifiers (the NB, the DT the NN and the 1-NN classifiers). However, there exist other ways for building highly dependent classifiers from a unique learning algorithm. One of

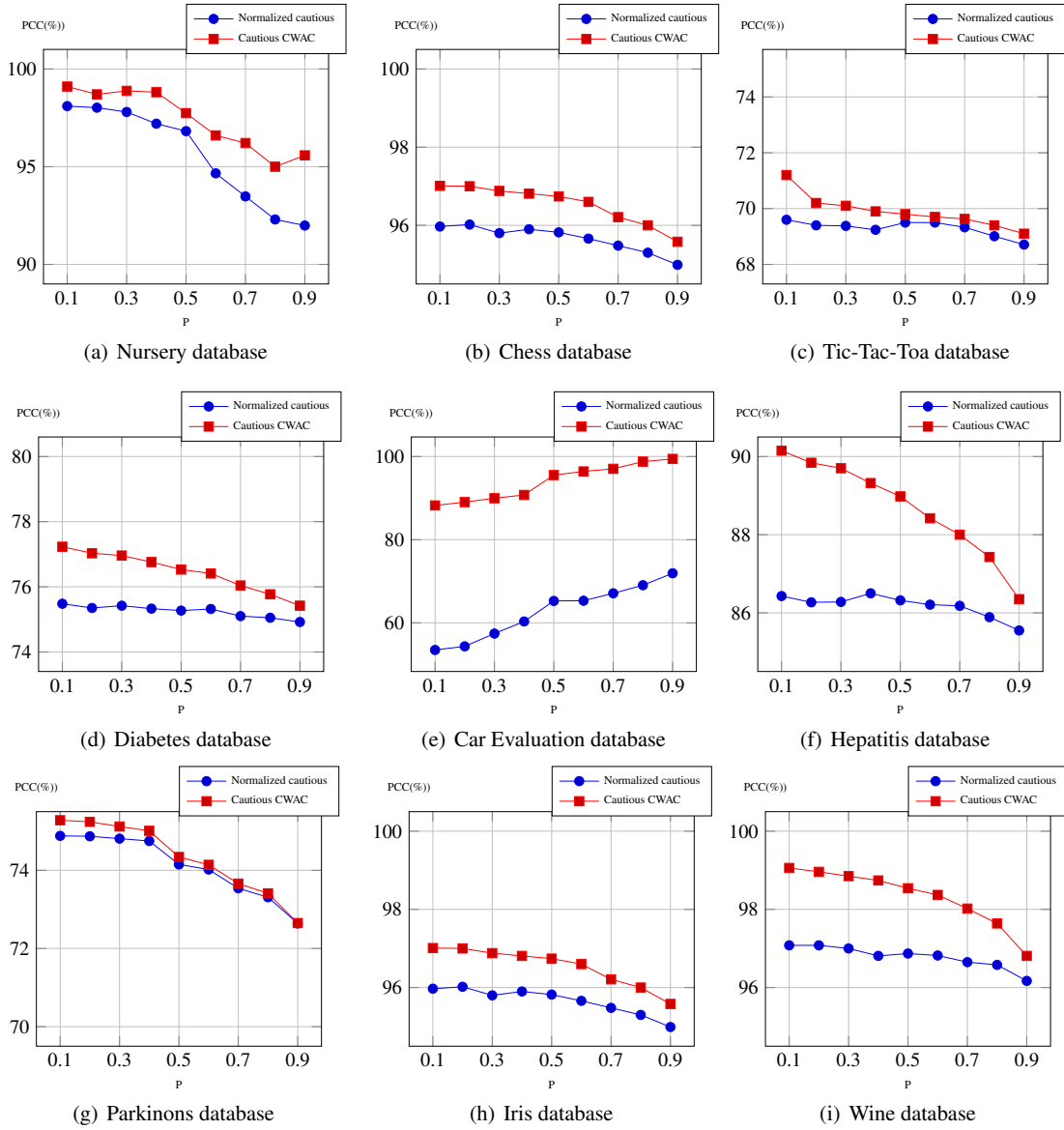


Fig. 5 PCCs for both cautious conjunctive and cautious CWAC rules

Table 4 Correlation coefficient between the classifiers  $C_1, C_2$  and  $C_3$

Databases	$\rho$ -correlation
Nursery	0.725
Chess	0.938
Tic-Tac-Toa	0.723
Diabetes	0.900
Car Evaluation	0.821
Hepatitis	0.950
Iris	0.981
Parkinsons	1
Wine	0.974

the proposed solutions consists of training the same learning algorithm with different learning subsets. Thus, in this experimentation, we propose to randomly divide the train-

ing set into four training partitions of equal size denoted  $P_1, P_2, P_3$  and  $P_4$ . These partitions have common instances which are dependent on the rate  $P$ . Besides, we suggest to learn four Naive Bayes classifiers  $NB_1, NB_2, NB_3$  and  $NB_4$  from respectively  $P_1, P_2, P_3$  and  $P_4$ . Subsequently, we should quantify the degree of dependency between the naive Bayes classifiers for the different values of  $P$  before any combination level takes place. As partitions will be randomly generated, the learning procedure should be executed several times (100 times) and consequently the 100 estimate results will be averaged to yield an overall estimate result. If classifiers are highly correlated, we move on to combine them through the cautious conjunctive, the normalized cautious and the cautious CWAC rules. Table 6 presents the  $\rho$ -

**Table 5** Combination results of the classifiers  $C_1, C_2$  and  $C_3$ 

Databases	Total distance criterion		Mean PCC criterion	
	Cautious conjunctive	Cautious CWAC	Normalized cautious	Cautious CWAC
Nursery	8624.19	<b>8312.07</b>	98.90 %	<b>99.63 %</b>
Chess	1731.055	<b>1653.72</b>	99.52 %	<b>99.53 %</b>
Tic-Tac-Toa	280.17	<b>220.31</b>	99.16 %	<b>99.69 %</b>
Diabetes	334.87	<b>295.23</b>	75.65 %	<b>77.44 %</b>
Car Evaluation	876.39	<b>842.03</b>	53.76%	<b>88.53%</b>
Hepatitis	54.90	<b>49.07</b>	83.22%	<b>85.79%</b>
Iris	15.51	<b>14.52</b>	96.66%	<b>98.75%</b>
Parkinsons	60.92	<b>46.99</b>	96.41%	<b>97.53%</b>
Wine	24.47	<b>19.71</b>	99.43%	<b>99.88%</b>

**Table 6**  $\rho$ -correlation for the different databases

P	Nursery	Chess	Tic-Tac-Toa	Diabetes	Car Evaluation	Hepatitis	Iris	Parkinsons	Wine
0.1	0.823	0.930	0.679	0.794	0.870	0.760	0.917	0.827	0.886
0.2	0.835	0.936	0.696	0.804	0.890	0.769	0.920	0.831	0.890
0.3	0.848	0.942	0.720	0.815	0.896	0.784	0.923	0.840	0.894
0.4	0.852	0.946	0.738	0.827	0.903	0.798	0.927	0.851	0.901
0.5	0.863	0.951	0.752	0.843	0.916	0.813	0.930	0.867	0.910
0.6	0.871	0.954	0.782	0.858	0.924	0.836	0.933	0.877	0.919
0.7	0.873	0.957	0.814	0.877	0.936	0.858	0.943	0.895	0.932
0.8	0.884	0.986	0.847	0.900	0.941	0.880	0.950	0.912	0.945
0.9	0.892	0.991	0.887	0.927	0.950	0.918	0.959	0.941	0.965
Mean	0.860	0.957	0.768	0.849	0.914	0.824	0.934	0.871	0.913

correlation measure for the different databases in term of  $P$  value. This latter clearly show that the degree of dependency between the Naive Bayes classifiers remains high for all the mentioned databases. For instance, it equals 0.934 for Iris database and it is equal to 0.849 for Diabetes database. Therefore, it can be said that the  $NB_1, NB_2, NB_3$  and  $NB_4$  are highly correlated and consequently we try to combine them using the cautious conjunctive, the normalized cautious and the cautious CWAC rules. The total distance result for the different databases is given in Figure 4, where we can note that the total distance of the cautious conjunctive rule is almost greater than that relative to the cautious CWAC rule for the different databases. So, we can deduce the efficiency of the cautious CWAC rule compared with the cautious conjunctive one in term of the distance criterion. Let's move on now to compare the cautious CWAC rule with the normalized cautious rule in term of the PCC criterion. The mean PCC of the different databases is given in Figure 5 where we can notice that the cautious CWAC rule has the highest values of mean PCC compared with the normalized cautious rule for the different databases. So, it can be said that the cautious CWAC rule is better than the normalized cautious one in the context of the PCC criterion.

## 7 Conclusion

In this paper, we have introduced classifier fusion approach adapted to the framework of belief functions using dependent combination rules; the cautious conjunctive, the normalized cautious and the cautious CWAC combination rules. Our objective is to pick out the most appropriate fusion rule among these three ones. To this end, we have proposed a comparative approach that aims to compare the cautious CWAC rule with the cautious conjunctive and the normalized cautious ones in terms of two evaluation criteria, namely the Jousselme distance and the PCC. Experimental results have shown the efficiency of the cautious CWAC rule compared with the cautious conjunctive one according to the distance criterion. Furthermore, we have proved that mostly the cautious CWAC rule is more efficient than the normalized cautious one in term of the PCC criterion. In sum, we can deduce that the cautious CWAC rule is the most adequate fusion rule compared with the two others. As future works, we look forward to using the correlation coefficient as a new assessment criterion to enhance our approach. We also intend to compare the cautious CWAC rule with other existing dependent rules.

## References

Al-Ani, A., & Deriche, M. (2002). A new technique for

- combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, 333–361.
- Bertoni, A., Folgieri, R., & Valentini, G. (2005). Biomolecular cancer prediction with random subspace ensembles of support vector machines. *Neurocomputing*, 63, 535–539.
- Bi, Y., Guan, J., & Bell, D. (2008). The combination of multiple classifiers using an evidential reasoning approach. *Artificial Intelligence*, 172(15), 1731–1751.
- Boubaker, J., Elouedi, Z., & Lefevre, E. (2013). Conflict management with dependent information sources in the belief function framework. In *the 14th International Symposium of Computational Intelligence and Informatics (CINTI)* (Vol. 52, pp. 393–398). IEEE.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.
- Cattaneo, M. E. G. V. (2003). Combining Belief Functions Issued from Dependent Sources. In *the 3rd International Symposium on Imprecise Probabilities and Their Applications (ISIPTA)* (Vol. 3).
- Cho, S.-B., & Kim, J. H. (1995). Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Transactions on Systems, Man and Cybernetics*, 25(2), 380–384.
- Dempster, A. P. (1967). Upper and Lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38, 325–339.
- Denceux, T. (1999). Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20(1), 79–111.
- Denceux, T. (2006). The cautious rule of combination for belief functions and some extension. *9th International Conference on Information Fusion (FUSION'2006)*, 1–8.
- Denceux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2), 234–264.
- Denoeux, T., & Masson, M.-H. (2012). Belief Functions: Theory and Applications. In *the 2nd International Conference on Belief Functions* (Vol. 164). Springer Science & Business Media.
- Dietterich, T. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), 139–157.
- Dubois, D., & Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3), 244–264.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10–18.
- Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*(10), 993–1001.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Huang, Y. S., Liu, K., & Suen, C. Y. (1995). The combination of multiple classifiers by a neural network approach. *International Journal of Pattern Recognition and Artificial Intelligence*, 9(03), 579–597.
- Johansson, R., Bostrom, H., & Karlsson, A. (2008). A Study on Class-Specifically Discounted Belief for Ensemble Classifiers. In *the IEEE international conference on multisensor fusion and integration for intelligent systems* (pp. 614–619).
- Jousselme, A., Grenier, D., & Bosse, E. (2001). A new distance between two bodies of evidence. *Information Fusion*, 2(2), 91–101.
- Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226–239.
- Kuncheva, L., Rodriguez, J., Plumpton, C., Linden, D., & Johnston, S. (2010). Random subspace ensembles for fMRI classification. *IEEE Transactions on Medical Imaging*, 29(2), 531–542.
- Kuncheva, L., Skurichina, M., & Duin, R. P. (2002). An experimental study on diversity for bagging and boosting with linear classifiers. *Information Fusion*, 3(4), 245–258.
- Kuncheva, L., & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2), 181–207.
- Le, C. A., Huynh, V.-N., Shimazu, A., & Nakamori, Y. (2007). Combining classifiers for word sense disambiguation based on Dempster–Shafer theory and OWA operators. *Data & Knowledge Engineering*, 63(2), 381–396.
- Lefevre, E., Colot, O., & Vannoorenberghe, P. (2002). Belief function combination and conflict management. *Information Fusion*, 3(2), 149–162.
- Lefevre, E., & Elouedi, Z. (2013). How to Preserve the Conflict As an Alarm in the Combination of Belief Functions? *Decision Support Systems*, 56, 326–333.
- Mercier, D., Cron, G., Denceux, T., & Masson, M. (2005).

- Fusion of multi-level decision systems using the Transferable Belief Model. In *the 8th International Conference on Information Fusion (fusion'2005)* (Vol. 2, p. 655-658).
- Murphy, C. K. (2000). Combining belief functions when evidence conflicts. *Decision Support Systems*, 29(1), 1–9.
- Murphy, P., & Aha, D. (1996). UCI repository databases. <http://www.ics.uci.edu/mllear>.
- Pizzi, N. J., & Pedrycz, W. (2010). Aggregating multiple classification results using fuzzy integration and stochastic feature selection. *International Journal of Approximate Reasoning*, 51(8), 883–894.
- Quost, B., Denœux, T., & Masson, M.-H. (2007). Pairwise classifier combination using belief functions. *Pattern Recognition Letters*, 28(5), 644–653.
- Quost, B., Denœux, T., & Masson, M.-H. (2008). Adapting a combination rule to non-independent information sources. In *the 12th Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)* (pp. 448–455).
- Quost, B., Masson, M.-H., & Denœux, T. (2011). Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules. *International Journal of Approximate Reasoning*, 52(3), 353–374.
- Reformat, M., & Yager, R. R. (2008). Building ensemble classifiers using belief functions and OWA operators. *Soft Computing*, 12(6), 543–558.
- Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619–1630.
- Ruta, D., & Gabrys, B. (2005). Classifier selection for majority voting. *Information Fusion*, 6(1), 63–81.
- Samet, A., Lefèvre, E., & Ben Yahia, S. (2016). Evidential data mining: precise support and confidence. *Journal of Intelligent Information Systems*. doi: 10.1007/s10844-016-0396-5
- Shafer, G. (1976). *A mathematical theory of evidence* (Vol. 1). Princeton university press Princeton.
- Sharkey, A. J., & Sharkey, N. E. (1997). Combining diverse neural nets. *The Knowledge Engineering Review*, 12(03), 231–247.
- Smets, P. (1988). The Transferable Belief Model for quantified belief representation. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1, 267–301.
- Smets, P. (1990a). The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 447–458.
- Smets, P. (1990b). The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 447–458.
- Smets, P. (1995). The canonical decomposition of a weighted belief. In *the 14th International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 95, pp. 1896–1901).
- Smets, P. (1998). The application of the Transferable Belief Model to diagnostic problems. *International Journal of Intelligent Systems*, 13, 127–157.
- Trabelsi, A., Elouedi, Z., & Lefevre, E. (2015a). Belief function combination: comparative study in classifier fusion framework. In *the 1st International Symposium on Advanced Intelligent Systems and Informatics (AISI)* (Vol. 407, pp. 425–435).
- Trabelsi, A., Elouedi, Z., & Lefevre, E. (2015b). Classifier fusion within the belief function framework using dependent combination rules. In *the 22nd International Symposium on Methodologies for Intelligent Systems (ISMIS)* (Vol. 9384, pp. 133–138).
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
- Woods, K., Kegelmeyer, W. P., Jr., & Bowyer, K. (1997). Combination of Multiple Classifiers Using Local Accuracy Estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 405–410.
- Xu, L., Krzyzak, A., & Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 22(3), 418–435.
- Xu, P., Davoine, F., & Denœux, T. (2014). Evidential logistic regression for binary SVM classifier calibration. In *the 3rd International Conference on Belief Functions (BELIEF)* (pp. 49–57). Springer.
- Xu, P., Davoine, F., Zha, H., & Denœux, T. (2016). Evidential calibration of binary SVM classifiers. *International Journal of Approximate Reasoning*, 72, 55–70.
- Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. *Information sciences*, 41(2), 93–137.
- Yen, J. (1990). Generalizing the Dempster-Shafer Theory to Fuzzy Sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3), 559–570.
- Zadeh, L. A. (1986). A Simple View of the Dempster-Shafer Theory of Evidence and Its Implication for the Rule of Combination. *AI Mag*, 7(2), 85–90.