Classifier Fusion With Contextual Reliability Evaluation

Zhunga Liu, Quan Pan, Jean Dezert, Jun-Wei Han, and You He

***Abstract*—Classifier fusion is an efficient strategy to improve the classification performance for the complex pattern recognition problem. In practice, the multiple classifiers to combine can have different reliabilities and the proper reliability evaluation plays an important role in the fusion process for getting the best classification performance. We propose a new method for classifier fusion with contextual reliability evaluation (CF-CRE) based on inner reliability and relative reliability concepts. The inner reliability, represented by a matrix, characterizes the probability of the object belonging to one class when it is classified to another class. The elements of this matrix are estimated from the K-nearest neighbors of the object. A cautious discounting rule is developed under belief functions framework to revise the classification result according to the inner reliability. The relative reliability is evaluated based on a new incompatibility measure which allows to reduce the level of conflict between the classifiers by applying the classical evidence discounting rule to each classifier before their combination. The inner reliability and relative reliability capture different aspects of the classification reliability. The discounted classification results are combined with Dempster–Shafer's rule for the final class decision making support. The performance of CF-CRE have been evaluated and compared with those of main classical fusion methods using real data sets. The experimental results show that CF-CRE can produce substantially higher accuracy than other fusion methods in general. Moreover, CF-CRE is robust to the changes of the number of nearest neighbors chosen for estimating the reliability matrix, which is appealing for the applications.**

***Index Terms*—Belief functions (BFs), classifier fusion, evidence theory, pattern classification, reliability evaluation.**

I. Introduction

THE FUSION of multiple classifiers provides an efficient way to achieve the best possible classification performance. The multiple classifiers usually can offer (more or less) complementary information about the pattern to classify, and the fusion procedure is expected to reduce the error rate and enhance the robustness of classification compared with any individual classifier. The complementarity (diversity) among classifiers can be achieved by extracting different features, by employing different classifiers, as well as by randomly selecting different training data sets [1]. A typical way is to implement a single classifier on different subsets of attributes or on different training data sets, such as boosting [2], bagging [3], random subspace [4], [5], random forests [6], and so on. For combining multiple classifiers, it requires not only the diversity but also the comparability [7].

How to efficiently combine the multiple classifiers remains an interesting topic. Many fusion methods have been developed for making a class decision from the individual classifiers. For instance, an interesting classifier combination method is introduced in [8] based on the signal strength for ensemble learning systems, and it can effectively combine the individual vote provided by different classifiers. Some research works have been dedicated to the ensemble of multiple neural networks [9]–[11]. Particularly, an effective neural network ensemble approach is presented in [9] to improve the generalization performance of neural networks by selecting and combining the diverse individuals from a pool of neural network classifiers. The decision tree is also integrated with neural network in [12] for improving the comprehensibility as well as the generalization ability of classifier. The selection of appropriate fusion strategy mainly depends on the formats of classifier output. If the output of the classifier consists only of a label value (i.e., a hard-decision classifier), the simple majority voting (MV) method is often recommended. If the classifier can generate soft membership measures, like probability value, fuzzy memberships or belief functions (BFs), many fusion strategies can be used, such as the linear combination way (average, sum, etc.) [13], Bayesian combination [14], fuzzy rules [15], and evidential reasoning technique [1], [16], [17]. The soft classification result generally offers more useful information than a single hard label, and the fusion of soft outputs of different classifiers is expected to improve significantly the classification performance [18].

BFs theory [19]–[22] also known as Dempster–Shafer theory (DST) or evidence theory, provides a very efficient theoretical framework for representing uncertain information and for fusing distinct sources of evidences [23]–[25]. BFs allows the object to be associated with not only the singleton classes but also any sets of classes according to a basic belief

Manuscript received October 30, 2016; revised April 10, 2017 and May 28, 2017; accepted May 28, 2017. Date of publication June 8, 2017; date of current version April 13, 2018. This work was supported by the National Natural Science Foundation of China under Grant 61672431 and Grant 61403310. This paper was recommended by Associate Editor S. Yang. *(Corresponding author: Zhunga Liu.)*

Z. Liu is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: liuzhunga@nwpu.edu.cn).

Q. Pan and J.-W. Han are with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China.

J. Dezert is with ONERA—The French Aerospace Laboratory, F-91761 Palaiseau, France.

Y. He is with the Department of Electronic Engineering, Tsinghua University, Beijing, China.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the authors.

Digital Object Identifier 10.1109/TCYB.2017.2710205

assignment (BBA), and we adopt it here for the ensemble of multiple classifiers. The BFs have been already successfully applied in the information fusion [1], [17], [26], data classification [27]–[32] and clustering [33], decision making support [34], [35], and so on.

In this paper, we focus on the combination of multiple classifiers, which are trained on different attribute sets, based on BFs theory. In the fusion process, the classifiers may have different reliabilities (weights), and the reliability evaluation plays an important role to improve the classification accuracy. So we want to study how to efficiently estimate the reliability of each classifier. The reliability of each classifier is usually determined based on the overall classification performance (e.g., accuracy) in the training set [7], [36], [37], and many methods have been introduced in [7] to calculate the classifier reliabilities (weights) using accuracy rate. In the calculation of training accuracy, the different misclassification cases, e.g., the pattern truly belonging to class No. i (or No. j) but misclassified to class No. j (or No. i), are indiscriminately considered as errors. So the reliability is generally associated with the overall performance of classifier in these methods, and all the elements in (soft) output of classifier are treated by the common reliability.

In many applications, the reliabilities of classification results obtained by one classifier are related with the objects to classify. For example, the object close to the center of a class can be easily classified, and its classification result will be quite reliable. Whereas, the object lying in overlapping borders of several classes could be hard to classify, and its classification result is considered with low reliability. In the soft classification result of one object, the different elements (i.e., the different probabilities assigned to each class) may also have different reliabilities, because the difference between the output value and the expected value (truth) usually is not the same for the different elements. However, the traditional evaluation methods do not specifically reveal the reliability degree of each element in the output of classifier, and the classification accuracy of the individual classifier cannot be improved according to the reliability. It seems interesting to find an appropriate probability redistribution way through refined reliability evaluation for making the classification result closer to the truth. Moreover, high conflict sometimes may occur among the classification results obtained from various classifiers for one object, and this is very harmful for the fusion. The conflict cannot be characterized by the classifier accuracy as done in traditional way. So how to properly reduce the conflict degree is also very important for acquiring the good fusion result.

In this paper, we propose a new method for classifier fusion with contextual reliability evaluation (CF-CRE) to increase the classification accuracy as far as possible. We consider two aspects of reliability for each piece of classification result of the object, i.e., the inner reliability and the relative reliability. Our main contributions lie in the following points.

1) The inner reliability represented by a matrix $\mathbf{R}_{c\times c}$ (c being the number of classes in the data set) is introduced to characterize the conditional probability of the object belonging to class No. i ($i = 1, \dots, c$) when it is classified to class No. j ($j = 1, \dots, c$) by the given classifier. This matrix provides much more refined reliability knowledge than the training accuracy value. We propose to estimate this matrix by exploiting the classification information from the close neighbors of the object taking into account the distance influence.
2) A cautious discounting rule is developed to correct the soft classifier output according to the inner reliability matrix under the BFs framework. It cautiously transfers the masses of belief from the singleton classes to the set of associated classes in order to efficiently decrease the risk of error by appropriately modeling the partial imprecision, which can be refined in the fusion procedure.
3) The relative reliability evaluation is proposed to reduce the bad impact of conflicting information among the classifiers for improving the classification performance. A new incompatibility measure is employed to determine the relative reliability of the classification result of each classifier. The incompatibility measure can well characterize the level of conflict between classifiers, and this measure is also tolerant of the difference of BBA's in some degree, which is very helpful to preserve the complementary information among classifiers for getting good fusion results. The classification knowledge provided by each classifier will be more or less discounted by the classical discounting rule [20] with some mass committed to the total ignorance depending on the relative reliability. With this approach, one can efficiently diminish the influence of the classifier having a low relative reliability in the fusion procedure.

The inner reliability and the relative reliability characterize somehow the confidence we give in the classification result from complementary standpoints. So the joint use of them is expected to provide a finer evaluation capability than using them individually. The multiple discounted classification results from different classifiers will be globally combined using Dempster–Shafer (DS) rule[1] for the final classification of object. This paper is an extended version of preliminary ideas and works [42] which only focused on the inner reliability evaluation in classifier fusion. In this paper, the relative reliability evaluation of classifiers is additionally included for dealing with the high conflicting information to further improve the global classification performance. The experimental test of this new method has been done on more real data sets by comparing with other classical methods to draw more general conclusions.

This paper is organized as follows. After a brief introduction of the BFs in Section II, we present the methods for contextual reliability evaluation in details in Section III. Simulations results are presented in Section IV to evaluate the performance of this new method for different data sets. Section V concludes this paper and gives research perspectives.

[1] Here we denote the fusion rule of DST as DS's rule rather than Dempster's rule, because this rule originally proposed by Dempster gained its popularity thanks to Shafer in the mid of seventies with his mathematical theory of evidence [20].

II. Basics of Belief Functions Theory

A. Definitions and Dempster–Shafer's Rule of Combination

The BFs theory has been introduced by Shafer in his mathematical theory of evidence, and it is often called DST or evidence theory (evidential reasoning) [19]–[22]. In DST, we work with a frame of discernment as $\Omega = \{\omega_i, i = 1, 2, \ldots, c\}$ consisting of c exclusive and exhaustive hypotheses (classes) $\omega_i, i = 1, \ldots, c$. A BBA, also called a mass of beliefs, can be defined over the power-set of Ω denoted by 2^Ω, which is the set of all the subsets of Ω. For example, if the frame of discernment is $\Omega = \{\omega_1, \omega_2, \omega_3\}$, then its power-set is $2^\Omega = \{\emptyset, \omega_1, \omega_2, \omega_3, \omega_1 \cup \omega_2, \omega_1 \cup \omega_3, \omega_2 \cup \omega_3, \Omega\}$. A BBA is mathematically defined as a mapping $m(.)$ from 2^Ω to $[0, 1]$, which satisfies the following conditions:

$$m(\emptyset) = 0 \text{ and } \sum_{A \in 2^\Omega} m(A) = 1. \tag{1}$$

With a BBA $m(.)$, one can allow one object to belong to different elements (singleton elements and sets of elements) in 2^Ω with different masses of belief, and all the elements A of 2^Ω such that $m(A) > 0$ are called the focal elements of the BBA $m(.)$. $m(A)$ represents the support degree of the object associated with class ω_i. In pattern classification problem, if A is a set of classes (e.g., $A = \omega_i \cup \omega_j$), $m(A)$ can be used to characterize the imprecision (partial ignorance) degree among the class ω_i and ω_j in classification of the object. $m(\Omega)$ denotes the total ignorance degree, and it usually plays a particular neutral role in the fusion process because it characterizes the vacuous belief source of evidence.

The lower and upper bounds of imprecise probability associated with a BBA, respectively, correspond to the BF Bel(.) and the plausibility function Pl(.) defined in [20] and $\forall A \subseteq \Omega$ by

$$\mathrm{Bel}(A) = \sum_{B \in 2^\Omega | B \subseteq A} m(B) \tag{2}$$

$$\mathrm{Pl}(A) = \sum_{B \in 2^\Omega | A \cap B \neq \emptyset} m(B). \tag{3}$$

In a multiclassifier system, the output of each classifier can be considered as an evidence represented by a BBA. The well-known DS rule is still widely applied for combining multiple BBA's mainly because of its commutative and associative properties, which makes it relatively easy to implement. DS rule offers a compromise between the specificity and complexity for the combination of BBA's. The DS combination of two distinct sources of evidence characterized by the BBA's $m_1(.)$ and $m_2(.)$ over 2^Ω is denoted $\mathbf{m} = \mathbf{m}_1 \oplus \mathbf{m}_2$, and it is mathematically defined (assuming the denominator is not equal to zero) by $m(\emptyset) = 0$, and $\forall A \neq \emptyset \in 2^\Omega$ by

$$m(A) = \frac{\sum_{B,C \in 2^\Omega | B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B,C \in 2^\Omega | B \cap C = \emptyset} m_1(B) m_2(C)}. \tag{4}$$

DS rule is applicable only when the denominator is strictly positive. So the whole conflicting mass must be smaller than 1 as $\sum_{B,C \in 2^\Omega | B \cap C = \emptyset} m_1(B) m_2(C) < 1$. In DS formula (4), the whole conflicting mass is redistributed back to all the focal elements because of the normalization. However, this conflicting mass redistribution way can generate unreasonable results, specially in the high conflicting cases, but also in some special low conflicting cases [38] as well. So a number of alternative combination rules have been developed to overcome the limitations of DS rule, such as Yager's rule, Dubois–Prade (DP) rule, and proportional conflict redistribution (PCR) rules [38]. These modified rules are unfortunately less attractive from the implementation standpoint because they are not associative. Some of them can degrade the specificity of the result (Yager's or DP rules), and other (typically the PCR6 rule) requires much higher complexity to get better performances which is not always tractable so far for high dimension classification problems.

B. Classical Discounting Technique

In the combination of multiple sources of evidence corresponding to different classifiers, each source may have different reliabilities (or weights). The classical Shafer's discounting method was introduced in [20] to deal with the not completely reliable source of evidence. Shafer's method discounts the partial mass of belief in a BBA to the total ignorance according to the reliability (weight) factor.

Let us consider a BBA denoted by $m(.)$ defined over the frame of discernment Ω, and its reliability factor $\alpha \in [0, 1]$. Then the discounted BBA ${}^\alpha\mathbf{m}$ is given by

$$\begin{cases} {}^\alpha m(A) = \alpha \cdot m(A), A \subset \Omega, A \neq \Omega \\ {}^\alpha m(\Omega) = 1 - \alpha + \alpha \cdot m(\Omega). \end{cases} \tag{5}$$

One can see from (5) that the original mass value of each focal element in $m(.)$ has been proportionally redistributed to the total ignorant element Ω based on the weighting (reliability) factor α. When the source is completely reliable, we take $\alpha = 1$ and one gets ${}^\alpha m(.) = m(.)$ (i.e., no effective discounting is done). If $\alpha = 0$, it indicates that this source is totally unreliable, and in this case one takes ${}^\alpha m(\Omega) = 1$, which corresponds exactly to the vacuous BBA $m_v(\Omega) = 1$ that plays a neutral role in the fusion.[2]

In the classical Shafer's discounting method [20], the reliability of one source of evidence is described by a single number, and the mass values of different focal elements are discounted with the same reliability number. A contextual discounting operation considered as a general extension of the classical discounting has been developed in [39]. It allows to take into account the refined reliability knowledge, which is represented by a vector of discounting rates characterizing the reliability of source associated with different hypotheses (contexts). The contextual discounting operation is suitable for handling the cases where the reliability knowledge about the source of information depending on the truth of the object to classify is available.

[2]It means that any BBA remains unchanged when combined with the vacuous BBA.

C. Brief Recall of Belief-Based Pattern Classification

BFs provides an efficient tool for representing and fusion of uncertain information based on the power-set, and it is has already well applied in the pattern classification problem. An evidential K-nearest neighbor (EK-NN) [27] was proposed by Denœux based on DST to well characterize the uncertainty and ignorance. In EK-NN, one can obtain a piece of classification result represented by a simple BBA according to each selected neighbor. Thus, K pieces of classification results denoted by K BBA's can be acquired, and the combination result of the K BBA's by DS rule is used for the classification of the object. An evidential neuron network was later developed also by Denœux [31] to reduce the computation burden with respect to EK-NN. In our previous works, several credal classifiers [28], [29] were presented based on BFs for dealing with different cases, and the additional meta-class (disjunction of several singleton classes) was introduced to capture the partial imprecise information.

Many methods [10], [16], [17], [40], [41] have been also introduced particularly for the fusion of multiple classifiers based on BFs. Rogova [10] introduced a combination method for dealing with multiple neural network classifiers based on DST, and the evidence represented by a simple BBA was obtained based on the distance between the classifier outputs and the reference vectors (i.e., mean vector of a set of classifier outputs). A combination method for multiple classification results was developed by Al-Ani and Deriche [41] using DST, and an interesting evidence estimation technique was introduced based on the gradient descent learning algorithm to minimize the mean squared error between the combined output and the ground truth for a given training data set. In [40], several combination strategies (i.e., MV, Bayesian formalism, and DS model) were introduced, and the conditional probability of the object belonging to different classes was derived based on the confusion matrix. For DS model, each evidence is represented by dichotomous mass functions including three focal elements (e.g., A, $\bar{A}$, and ignorance element Ω), and it was defined according to the overall performance of classifiers. In [17], a class-indifferent method was proposed for multiclassifier fusion using DS rule, and the classifier decisions were modeled by triplet and quartet evidential structures. These aforementioned methods generally focus on how to efficiently construct the evidence (BBA's), and the DS rule working with the independent sources of information is always employed for classifiers combination. In [1], an optimal combination scheme was presented based on a parameterized family of t-norms for an ensemble of multiple classifiers that provides the partly dependent information, and the parameter can be optimized to achieve the minimum error criterion.

In the fusion of multiple classifiers, the reliability of each classifier may be different, and the proper reliability evaluation is very helpful for improving the fusion performance. In this paper, it will be deeply studied how to efficiently evaluate the reliability for classifier fusion based on BFs theory.

III. Contextual Reliability Evaluation

In this section, a very refined reliability evaluation method will be presented for classifier fusion under BFs framework. We focus on the combination of multiple classifiers trained on different attribute sets. The class of the object $\mathbf{y}$ to classify is assumed to belong to the frame of discernment $\Omega = \{\omega_1, \dots, \omega_c\}$. We consider N classifiers, $C_1, \dots, C_N$ trained on N different attribute spaces $\mathbb{S}_1, \dots, \mathbb{S}_N$. Each classifier $C_n, n = 1, \dots, N$ provides as output a probabilistic mass function (pmf) denoted $\boldsymbol{\mu}_n = [\mu_n(1), \dots, \mu_n(c)]$ based on the attribute knowledge of object in $\mathbb{S}_n$. The value $\mu_n(i)$ represents the probability of the object belonging to the class ω_i estimated by the classifier. By convention, the real class (unknown) of the objet to classify say $\mathbf{y}$ is denoted by $c(\mathbf{y})$, and its estimated class declared by a classifier C_n is denoted $\hat{c}_n(\mathbf{y})$. The classification performance can be improved in taking into account the quality of the classifier, which can be captured by the refined contextual reliability evaluation of the output (pmf) $\boldsymbol{\mu}_n$ of each classifier. Then the output $\boldsymbol{\mu}_n$ will be modified accordingly before entering the classifier fusion process to make the final class decision.

The contextual reliability evaluation consists of two steps.

1) The inner reliability is estimated using the training data close to the object.
2) The relative reliability is determined according to the compatibilities of $\boldsymbol{\mu}_n$ with respect to the other classification results.

A. Inner Reliability Evaluation

In a c-class problem, the classification result of an object $\mathbf{y}$ by classifier C_n in the attribute space $\mathbb{S}_n$ is given as $\boldsymbol{\mu}_n$. The inner reliability of $\boldsymbol{\mu}_n$ is denoted by a matrix $\mathbf{R}_{c\times c}$,[3] and this matrix expresses the conditional probability of the object $\mathbf{y}$ potentially belonging to class $\omega_i, i = 1, \dots, c$ when it is classified to class $\omega_j, j = 1, \dots, c$ by classifier C_n, i.e., $r_{ji} \triangleq P(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j)$. Obviously, if this reliability matrix $\mathbf{R}_{c\times c}$ can be well estimated, the accuracy of the classification result $\boldsymbol{\mu}_n$ could be efficiently improved taking into account this important knowledge. Now we will show how to estimate this reliability matrix.

Because the knowledge about the true class of the object is unavailable in the classification task, we will attempt to estimate this reliability matrix using the training information. In the training data space, the patterns in the nearby neighborhood of the object $\mathbf{y}$ generally have the close attribute values with the object. Thus, the given classifier is expected to produce the similar performance on the object and on its close neighbors. Meanwhile, the ground truth of the class of the training patterns is always known. So the training data lying in the neighborhood of the object will be employed here for the inner reliability evaluation.

The K-nearest neighbors (K-NNs, training patterns) of $\mathbf{y}$ are found at first in the attribute space $\mathbb{S}_n$. The selected neighbors denoted $\mathbf{x}_k$, $k = 1 \dots, K$ will be classified by the given

[3] For notation convenience, the classifier index n is omitted in the sequel.

base classifier C_n,[4] and the classification result $\hat{c}_n(\mathbf{x}_k)$ provided by C_n is represented by the vector $\mathbf{P}_k = [P_k(1), \ldots, P_k(c)]$, where $P_k(i) \triangleq P(\hat{c}(\mathbf{x}_k) = \omega_i)$ is the estimated probability of $\mathbf{x}_k$ classified to ω_i, for $i = 1, \ldots, c$.

If a neighbor $\mathbf{x}_k$ with the real class label ω_i [i.e., $c(\mathbf{x}_k) = \omega_i$] is classified by the base classifier into class ω_j [i.e., $\hat{c}(\mathbf{x}_k) = \omega_j$] with the corresponding probability $P_k(j)$, it indicates that the conditional probability of $\mathbf{x}_k$ classified to ω_j is $P_k(j)$ knowing $\mathbf{x}_k$ truly lies in ω_i as $P_k(j) \triangleq P(\hat{c}(\mathbf{x}_k) = \omega_j | c(\mathbf{x}_k) = \omega_i)$. Because $\mathbf{x}_k$ is a close neighbor of the object $\mathbf{y}$, the given classifier C_n likely produces the similar performance on $\mathbf{x}_k$ and $\mathbf{y}$. We can estimate the conditional probability of the object $\mathbf{y}$ classified to ω_j if its real class label is ω_i, i.e., $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i)$, according to $P(\hat{c}(\mathbf{x}_k) = \omega_j | c(\mathbf{x}_k) = \omega_i)$.

Moreover, there may be multiple patterns with the real class label ω_i in the K selected neighbors, and all of them will be employed to estimate of the conditional probability $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i)$. The bigger number of neighbors from ω_i leads to the bigger value of conditional probability $P(\cdot | c(\mathbf{y}) = \omega_i)$. Meanwhile, the distance[5] between the object $\mathbf{y}$ and the neighbor $\mathbf{x}_k$ must be additionally taken into account. If $\mathbf{y}$ is far from $\mathbf{x}_k$, then $\mathbf{x}_k$ is considered with a small influence (weight) on the estimation of $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i)$. The bigger distance, the smaller weight of the neighbor. The weighted sums of the conditional probabilities of the neighbors $\mathbf{x}_k$ belonging to class ω_i but classified to ω_j (denoted by β_{ij}) is computed by

$$\begin{aligned}\beta_{ij} &= \sum_{\mathbf{x}_k} P\big(\hat{c}(\mathbf{x}_k) = \omega_j | c(\mathbf{x}_k) = \omega_i\big) \cdot \delta_k \\ &= \sum_{\mathbf{x}_k | c(\mathbf{x}_k) = \omega_i} P_k(j) \cdot \delta_k \end{aligned} \tag{6}$$

with

$$\delta_k = e^{-\gamma \cdot d_k} \tag{7}$$

$$d_k = \frac{d(\mathbf{y}, \mathbf{x}_k)}{\min\limits_{k \in [1,K]} d(\mathbf{y}, \mathbf{x}_k)} \tag{8}$$

where δ_k denotes the distance weights, and γ is a tuning parameter used to control the influence of distance, and d_k is the relative distance of the object to the neighbor $\mathbf{x}_k$ with respect to the minimum distance to the nearest neighbors. Because the exponential function usually can well characterize the distance influence in K-NN-based classifier with good performance [27], it is also employed in (7) to calculate the distance weights.

β_{ij} can be interpreted as the weighting factor of the hypothesis that the object is really from class ω_i but classified to ω_j. The conditional probability $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i)$ should be proportional to β_{ij} as $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i) \propto \beta_{ij}$, and it is defined by $P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i) = \rho\beta_{ij}$ (ρ being a positive proportional coefficient). Then the reliability matrix $\mathbf{R}$ expressed by the probability $r_{ji} \triangleq P(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j)$ can be easily derived according to Bayes rule. One gets

$$\begin{aligned} r_{ji} &= P\big(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j\big) \\ &= \frac{P\big(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i\big) P(c(\mathbf{y}) = \omega_i)}{\sum_{l=1}^{c} P(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_l) P(c(\mathbf{y}) = \omega_l)}. \end{aligned} \tag{9}$$

Without extra knowledge, *a priori* probability $P(c(\mathbf{y}) = \omega_l), l = 1, \ldots, c$ is usually assumed uniformly distributed. Therefore, the probability $P(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j)$ can be obtained by

$$\begin{aligned} r_{ji} &= \frac{P\big(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_i\big)}{\sum_{l=1}^{c} P\big(\hat{c}(\mathbf{y}) = \omega_j | c(\mathbf{y}) = \omega_l\big)} \\ &= \frac{\rho\beta_{ij}}{\rho \sum_{l=1}^{c} \beta_{lj}} = \frac{\beta_{ij}}{\sum_{l=1}^{c} \beta_{lj}}. \end{aligned} \tag{10}$$

Then the reliability matrix $\mathbf{R}$ is determined, and it will be used to modify the classification result $\boldsymbol{\mu}_n$ to make it closer to the potential truth. We recall that the matrix is estimated according to several neighbors of the object to classify, and therefore we must not be completely confident about the estimation of this matrix for revising the classifier result of the object $\mathbf{y}$. If we directly calculate the marginal probabilities of the object belonging to each class using this estimated matrix $\mathbf{R}$, it may bring high risk of error. That is why we propose a very cautious discounting method to transfer the classification knowledge to the associated partial imprecision (e.g., $\omega_i \cup \omega_j$) rather than to the specific class (e.g., ω_i). Such imprecision of classification can be reduced (eliminated) through the later combining procedure. More specifically, the contribution of belief from the classifier output $\mu_n(j)$ and the reliability value $r_{ji} \triangleq P(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j)$ is transferred by

$$\begin{aligned} m_{n1}\big(\omega_i \cup \omega_j\big) &= r_{ji} \cdot \mu_n(j) \\ &= P\big(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j\big) \cdot \mu_n(j) \end{aligned} \tag{11}$$

$\omega_i \cup \omega_j$ represents the imprecision between ω_i and ω_j, and it plays a neutral role in the classification between ω_i and ω_j.

Another contribution of belief on $\omega_j \cup \omega_i$, $j \neq i$ is also obtained from $\mu_n(i)$ by considering

$$\begin{aligned} m_{n2}(\omega_j \cup \omega_i) &= r_{ij} \cdot \mu_n(i) \\ &= P\big(c(\mathbf{y}) = \omega_j | \hat{c}(\mathbf{y}) = \omega_i\big) \cdot \mu_n(i). \end{aligned} \tag{12}$$

So that the discounted BBA derived from $\boldsymbol{\mu}_n$ is given for $i = 1, \ldots, c$ and $j = 1, \ldots, c$.

If $i \neq j$, one gets

$$\begin{aligned} m_n\big(\omega_i \cup \omega_j\big) &= m_{n1}\big(\omega_i \cup \omega_j\big) + m_{n2}\big(\omega_j \cup \omega_i\big) \\ &= P\big(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_j\big) \cdot \mu_n(j) \\ &\quad + P\big(c(\mathbf{y}) = \omega_j | \hat{c}(\mathbf{y}) = \omega_i\big) \cdot \mu_n(i). \end{aligned} \tag{13}$$

If $i = j$, one gets

$$\begin{aligned} m_n(\omega_i) &= r_{ii} \cdot \mu_n(i) \\ &= P\big(c(\mathbf{y}) = \omega_i | \hat{c}(\mathbf{y}) = \omega_i\big) \cdot \mu_n(i). \end{aligned} \tag{14}$$

[4]The base classifier can be selected according to the actual application, like Artificial neural network, Bayesian classifier, etc. The classifier can work with probabilistic framework or BFs framework. In this paper, we just consider the belief-based classifier (e.g., evidential neural network) with the output including just the singleton classes and the total ignorant class denoted by Ω as the focal elements. The evidential neural network [31] classifier and Bayes classifier [45] working with probability framework are, respectively, employed as base classifier in our sequel simulations.

[5]The Euclidean distance is used here.

One can see that some imprecision (i.e., $\omega_i \cup \omega_j$) has arisen due to the cautious discounting operation, but such imprecision can be specified through the combination with other classifiers.

If the probabilities of the K neighbors committed to ω_i are all zeros. In this case, the probability of the object $\mathbf{y}$ belonging to ω_i, i.e., $\mu_n(i)$, will be discounted to total ignorance by taking

$$m_n(\Omega) = 1 - \sum_{A \subset \Omega} m_n(A) \tag{15}$$

$m_n(\Omega)$ captures the total ignorant information about the classification done by the classifier C_n, and it plays a neutral role in the combination with the (modified) output of other classifiers. In fact, $m_n(\Omega)$ will always be redistributed to other more specific focal elements in the classifier fusion process based on the conjunctive rule of combination.

The inner reliability of classification result is derived from the classifier performance in the close neighborhoods of object, and it is represented by a matrix characterizing the conditional probability of the object belonging to one class but assigned to another class by the classifier. In fact, this matrix expresses the probabilities of the different misclassification occurrences, and it provides much more refined reliability knowledge for the classification result of the object than the training accuracy (being a single value) of classifier. The new cautious discounting operation is proposed to correct the classification result according to the inner reliability matrix, and the beliefs of the singleton classes (e.g., ω_i) are cautiously redistributed to the associated class set defined by the disjunction of two classes (e.g., $\omega_i \cup \omega_j$). This cautious discounting can reduce the risk of error by modeling the partial imprecision that will be refined in the fusion procedure. It offers an efficient way to improve the classification result. Because the traditional classifier reliability generally depends on the training accuracy which mainly reflects the overall performance of classifier, it is impossible to improve the individual classifier accuracy using such reliability. This is the main reason why the proposed inner reliability evaluation method outperforms the traditional ones.

B. Relative Reliability Evaluation

In the reliability evaluation, the underlying principle is also often followed that the source of information (i.e., output of classifier) quite different (conflict) from the others usually is not very reliable in the information fusion systems. In the multisource information fusion, the high conflict is usually harmful for the DS fusion procedure. The relative reliability evaluation, which can be done by the comparisons of the classification results to be combined, is used to reduce the degree of conflict between the classifiers for achieving the best possible fusion result.[6]

How to efficiently measure the difference of the classification results represented by BBA's plays the crucial role in the relative reliability evaluation. There have been various measurements developed to characterize the dissimilarity of evidences (i.e., BBA), like the J distance [43], conflicting beliefs, etc. Particulary, both conflicting beliefs and distance between betting commitments of beliefs are used to characterize the conflict of evidence in [44]. In our previous work, we also presented a dissimilarity measure taking into account both distance and conflict [34]. These previous methods generally provide the strict measurements, since they are usually sensitive to any small difference among the evidences. When these strict measurements are employed in the reliability evaluation, they generally tend to decrease the difference between evidences, and thus to reduce the conflict between the sources.

In fact, the difference characterizes the complementarity of sources of information to some degree, and the complementary information is helpful to obtain specific result through the fusion procedure. For example, let us consider two pieces of classification results after the cautious discounting, and they are represented by two BBA's as $m_1(\omega_1) = 0.2$, $m_1(\omega_1 \cup \omega_2) = 0.8$, and $m_2(\omega_2 \cup \omega_3) = 1$. Obviously, the two BBA's are different, but they are complementary and compatible. This is because the object is considered quite likely in the class ω_1 or ω_2 according to $\mathbf{m}_1$, whereas it indicates that the objects belongs to class ω_2 or ω_3 in $\mathbf{m}_2$. So both BBA's strongly support the potential class ω_2. In their fusion results, i.e., $\mathbf{m} = \mathbf{m}_1 \oplus \mathbf{m}_2$, it is given by $m(\omega_2) = 1$ using DS rule, and this specific result is consistent with our intuition.

The new measurement should be tolerant of the difference in some degree. In this paper, we propose a new incompatibility measure to evaluate the relative reliability, and the complementary information will be considered compatible.

In the previous cautious discounting step, the belief of the singleton class has been partly discounted to the associate imprecise classes (e.g., $\omega_i \cup \omega_j$). The plausibility function Pl(.) defined by (3) is usually interpreted as the upper probability of the class of the object. It corresponds to all the masses of classes that are compatible with the class of the object and it is used here in the definition of the incompatibility measure.

For the cautious discounted BBA's to be combined, the plausibility function Pl(.) of the singleton classes can be obtained from each BBA. Then the potential class the object very likely lies in can be predicted according to Pl(.). The class having the maximum Pl(.) value is the most possible, and the other classes with the relatively close value (with respect to a given threshold) are also possible. So the set of the possible true classes derived from the classification result by classifier C_i, $i = 1, \ldots, N$ can be defined by

$$\Phi_i = \left\{ A \middle| \frac{\mathrm{Pl}_i(A)}{\max\limits_{B \in \Omega} \mathrm{Pl}_i(B)} > \lambda \right\} \tag{16}$$

where $\lambda \in (0, 1)$ is a small positive threshold.

If one has $\Phi_i \cap \Phi_j \neq \emptyset$, it means that the classifiers C_i and C_j support the common potential classes, and the classification results obtained by the two classifiers will be considered as compatible. In such case, the incompatibility value is zero. If one has $\Phi_i \cap \Phi_j = \emptyset$, it means that the two classifiers support distinct classes with respect to the object $\mathbf{y}$, and their outputs

[6]In some applications, part of attributes could be polluted by noise. The inner reliability mainly evaluates the classifier performances with the given attribute values. Then the relative reliability, as the complement of inner reliability, can be used to capture the credibility of the observations (i.e., attribute), and it works with the assumption that the truth lies in the majority, and the classifier highly conflicting with the others will be considered having a serious pollution and therefore with a small relative reliability.

is considered as incompatible. The incompatibility degree is simply defined by the conjunctive product of the maximum plausibility values of incompatible classes in each BBA, and it represents the conflict degree of two pieces of classification results on the class they mainly support. Let us consider a pair of BBA's, i.e., $\mathbf{m}_i$ and $\mathbf{m}_j$, over the same frame of discernment Ω, and the corresponding plausibility functions can be obtained as $\mathbf{Pl}_i$ and $\mathbf{Pl}_j$. We propose to calculate the incompatibility degree of $\mathbf{m}_i$ and $\mathbf{m}_j$ by

$$\kappa(i,j) = \begin{cases} \sqrt{\max_{A\in\Omega} \mathrm{Pl}_i(A) \max_{B\in\Omega} \mathrm{Pl}_j(B)}, & \Phi_i \cap \Phi_j = \emptyset \\ 0, & \Phi_i \cap \Phi_j \neq \emptyset \end{cases} \tag{17}$$

where A and B are distinct singleton classes that naturally satisfies $|A| = 1, |B| = 1, A \cap B = \emptyset$, and $\kappa(i,j) \in (0,1]$. If $A \cap B \neq \emptyset$, it means hat $\mathbf{m}_i$ and $\mathbf{m}_j$ are compatible, and then one must have $\Phi_i \cap \Phi_j \neq \emptyset$ and $\kappa(i,j) = 0$.

The compatibility degree of $\mathbf{m}_i$ (corresponding to classifier C_i) with respect to the other classifiers is defined by

$$\tilde{\alpha}_i = \sum_{j|j\neq i}(1 - \kappa(i,j)). \tag{18}$$

If the output of a classifier say $\mathbf{m}_i$ is very compatible with that of other classifiers, $\mathbf{m}_i$ is usually considered quite reliable. However, if $\mathbf{m}_i$ highly conflicts with the others, it will be given a small relative reliability factor to reduce its influence in the fusion.

The relative reliability of $\mathbf{m}_i$ can be calculated based on the compatibility as

$$\alpha_i = \frac{\tilde{\alpha}_i}{\max_j \tilde{\alpha}_j}. \tag{19}$$

Because the relative reliability α_i generally captures the credibility of the whole piece of classification result, Shafer's classical discounting rule will be used on the previous modified classification result $\mathbf{m}$, and the discounted information will be transferred to the total ignorance denoted by the framework of discernment, i.e., Ω. The discounted BBA with the relative reliability α_i (i.e., ${}^{\alpha_i}\mathbf{m}_i$) is then given by

$$\begin{cases} {}^{\alpha_i}m_i(A) = \alpha_i m_i(A), \ \forall A \subset \Omega \\ {}^{\alpha_i}m_i(\Omega) = \alpha_i m_i(\Omega) + 1 - \alpha_i. \end{cases} \tag{20}$$

It is generally considered that the ensemble classifier can produce good performance if the individual classifiers have high accuracy and high diversity. In order to produce the high accuracy, the individuals should strongly support a common (true) class, and it indicates the individuals are compatible. For the high diversity, the distances of the individual outputs should be large, but it must be under the condition that the individuals strongly support the common (true) class. It is worth noting that the proposed incompatibility measure is quite different from the previous strict dissimilarity/distance measures. In this paper, if the classifiers strongly support one (or several) common class, their incompatibility measure is zero. So the new incompatibility measure is tolerant of some difference of BBA's in a certain degree, and this can well preserve the complementary information of classifiers, which is very important for producing the good fusion result. This is the main advantage of the new incompatibility measure. The incompatibility value will be positive only when the individuals strongly support the distinct classes. In such case, the relative reliability of individuals having high incompatibility with the others is considered low. Then the classical discounting rule is operated according to the relative reliability. By doing this, it can efficiently reduce the harmful conflicting information among classifiers for achieving the best possible performance of ensemble classifier.

The inner reliability matrix and the relative reliability value generally characterize the credibility of the classification results from different aspects. The inner reliability matrix represents the refined prior knowledge about the classifier performance on the neighborhoods of the object, and it is used to correct the classifier output associated with the object. The inner reliability of each classifier is independently evaluated. Whereas, the relative reliability reflects the importance of different classification results according to their conflict level, and it is employed to control the influence of each classifier in the fusion process. The relative reliability of one classifier is determined based on its degree of incompatibility with respect to the others. So the inner reliability and relative reliability are complementary, and both comprehensively capture our confidence in the classifiers to combine.

The two discounting steps including cautious discounting and classical discounting can efficiently reduce the conflict degree of the classification results. The popular DS rule defined by (4) requiring relatively small computation burden is often used to handle the low conflicting cases, and it will be employed here to combine the discounted classification results, i.e., ${}^{\alpha_i}\mathbf{m}_i, i = 1, \ldots, c$, from different classifiers. Since DS rule is associative, the BBA's can be combined one by one, and the combination order has no influence on the results.

In the final fusion results, some beliefs may remain in the (partial) imprecise focal element (imprecise classes) due to the discounting procedure. So the plausibility functions Pl(.) taking into account all the beliefs of the associated classes is used here for decision making support, and the object is considered belonging to the class receiving the biggest plausibility value, e.g., $\omega_g, \mathrm{Pl}(\omega_g) = \max_j \mathrm{Pl}(\omega_j)$.

1) Guideline for Parameters Tuning: In this new CF-CRE method, the parameter γ involved in (7) and the plausibility threshold λ involved in (16) should be tuned in the real applications. γ is used to penalize the influence of the neighbors in the determination of the inner reliability according to the distance between the object and its neighbor. The bigger γ value, the smaller influence of the neighbor (through its distance to the object) for the reliability evaluation. After many tests conducted on various real data set, this method generally produces good classification performances with $\gamma \in [5, 20]$ in practice, and we recommend to take $\gamma = 10$ as default value. $\lambda \in [0, 1]$ can be considered as the tolerance threshold for the difference of BBA's (i.e., classification results). The smaller λ value will put more classes in potential true class set, so that different classification results will have more likely the common potential true class to support. Hence, more classification results may be considered compatible with a smaller λ

value, and this generally leads to a bigger relative reliability (up to one) for these compatible results. Note that a too small λ value cannot efficiently reduce the conflict degree among classifiers, whereas a bigger λ value will decrease the tolerance degree for the information difference in the incompatibility measure, and it will yield a smaller relative reliability for the classifier incompatible with the other classifiers. A too big λ value will penalize too much the incompatible classifier in the fusion process. As default value, we take $\lambda = 0.5$ based on our heuristics. In applications, the tuning parameters γ and λ can be optimized by cross validation in the training data space, and the optimized value corresponding to the highest accuracy can be chosen.

IV. Experimental Applications

The classification performance of this new CF-CRE method will be evaluated by comparisons with several other fusion methods including (weighted) averaging rule, (weighted) MV rule and (weighted) DS combination rule. The weighting factors of classifiers are usually determined based on the individual classification accuracy η [7], and three often used formulas (21)–(23) for the calculation of classifier weight are briefly introduced here. The simple weight determination is given by (21), and the normalized individual accuracy of each classifier is considered as the weight of this classifier. The best-worst weighted way is defined by (22). The weight of classifier with the maximum (or minimum) accuracy is considered as one (or zero), and the weights of the other classifiers are rated linearly between the two extremes. In (23), the logarithmic function is employed to calculate the classifier weight. It just assigns the positive weight to the classifier with the accuracy bigger than 0.5, and the other classifiers with the accuracy smaller than 0.5 will be removed in the fusion. These three formulas (21)–(23) are commonly applied in practice [7], and they will be also employed here for comparisons

$$w_n = \frac{\eta_n}{\sum_i \eta_i} \tag{21}$$

$$w_n = \frac{\eta_n - \eta_l}{\eta_u - \eta_l} \tag{22}$$

$$w_n = \log \frac{\eta_n}{1 - \eta_n} \tag{23}$$

where $\eta_u \triangleq \max_n \eta_n$, $\eta_l \triangleq \min_n \eta_n$, $\eta_n \triangleq (n_c/T)$, and where n_c is the number of patterns correctly classified, and T is the number of total patterns. η_n denotes the individual accuracy of the classifier C_n on the whole data set.

Moreover, we also consider the local accuracy $\hat{\eta}_n$, which is calculated according to neighborhoods of objects. The local accuracy of classifier C_n with respect to the object $\mathbf{y}$ can be determined based on the classification results of the K-NNs of $\mathbf{y}$ in training data space as $\hat{\eta}_n = (\hat{n}_c/K)$, where K is the total number of the selected neighbors, and $\hat{n}_c$ is the number of correctly classified neighbors. The local accuracy can well reflect the reliability of classifier in close region of object. The local weighting factor $\hat{w}_n$ can be calculated similarly following (21)–(23) by replacing η_n with $\hat{\eta}_n$.

TABLE I
Description of the Used Fusion Methods

Name	Calculation
AF	$\mathbf{p} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{p}_n.$
WAF	$\mathbf{p} = \sum_{n=1}^{N} w_n \mathbf{p}_n.$
LWAF	$\mathbf{p} = \sum_{n=1}^{N} \hat{w}_n \mathbf{p}_n.$
MV	$\mathbf{l} = \sum_{n=1}^{N} \mathbf{l}_n.$
WMV	$\mathbf{l} = \sum_{n=1}^{N} w_n \mathbf{l}_n.$
LWMV	$\mathbf{l} = \sum_{n=1}^{N} \hat{w}_n \mathbf{l}_n.$
DS	$\mathbf{m} = \mathbf{m}_1 \oplus \ldots \oplus \mathbf{m}_N.$
WDS	$\mathbf{m} = {}^{\alpha_1}\mathbf{m}_1 \oplus \ldots \oplus {}^{\alpha_n}\mathbf{m}_N.$
LWDS	$\mathbf{m} = {}^{\hat{\alpha}_1}\mathbf{m}_1 \oplus \ldots \oplus {}^{\hat{\alpha}_n}\mathbf{m}_N.$

Nine related fusion methods have been evaluated in this paper, and it includes average fusion (AF), weighted AF (WAF), local WAF (LWAF), MV, weighted MV (WMV), local WMV (LWMV), DS fusion, weighted DS (WDS) fusion, and local WDS (LWDS) fusion. The brief description of these methods is shown in Table I.

In Table I, $\mathbf{p}_n$ denotes output of classifier C_n, and $\mathbf{l}_n$ means the hard decision making result according to $\mathbf{p}_n$. For example, if $\mathbf{p}_n = [0.7, 0.2, 0.1]$ for a 3-class problem, then one will have $H(\mathbf{p}_n) = [1, 0, 0]$ because the class ω_1 corresponding to the first dimension of $\mathbf{p}_n$ gets the biggest probability value. For DS fusion, $\alpha_n = (w_n / \max_i w_i)$, and $\hat{\alpha}_n = (\hat{w}_n / \max_i \hat{w}_i)$. w_n $(n = 1, \ldots, N)$ is normalized to make the sum of fusion result equal to one. ${}^{\alpha_n}\mathbf{m}_n$ denotes the BBA $\mathbf{m}_n$ is classically discounted using (5) with the reliability factor α_n.

In order to show the effectiveness of the new incompatibility measure, our previous strict dissimilarity measure DismP [34] and the widely used J distance [43], which can well characterize the distance measure between a pair of BBA's, will be, respectively, employed to evaluate the relative reliability of classifications results for comparisons.

The base classifier can be selected according to the actual applications. In this paper, the naive Bayesian classifier [45] and Evidential neural network (ENN) [31] classifier are employed as the base classifiers.[7] ENN working with BFs produces the classification results consisting of the singleton focal elements and the total ignorant focal elements, i.e., Ω. The mass of belief $m(\Omega)$ is usually very small, and it is proportionally distributed to other singleton focal elements here. Thus, the output of ENN classifier can be transferred to probability values, which enters the proposed reliability evaluation procedure. The base classifier(s) will be, respectively, trained using different subsets of attributes, and the multiple classification results obtained by different classifiers will be fused

[7] Any other classical classifiers can be also be used here as base classifier, and the selection of proper base classifier mainly depends on the actual application, which is out of scope of this paper.

TABLE II
BASIC INFORMATION OF THE USED DATA SETS

Data	Class	Attribute	Instance
SPECTF Heart (SH)	2	44	267
Bupa (Bu)	2	6	345
Texture (Te)	11	40	5500
Vehicle (Ve)	4	18	946
Wbdc (Wb)	2	30	569
Ionosphere (Io)	2	33	351
Movement-libras (ML)	15	90	360
Wine quality (WQ)	7	11	4898
Sonar (So)	2	60	208
Page-blocks (PB)	5	10	5472
Segment (Se)	7	19	2310
Magic (Ma)	2	10	19020
Satimage (Sa)	7	36	6435
Tae(Ta)	3	5	151
Connectionist-bench(CB)	11	10	990

for classifying the objects. The predicted class of the object should receive the maximum support degree. For the methods working with probability framework (typically the Naive Bayesian classifier, or with the averaging fusion rule), the support degree is represented by the probability value committed to each class. For classification methods working with the BFs framework (like ENN classifier with DS fusion method, or the Bayesian classifier with our proposed fusion method), the support degree is characterized by a plausibility function Pl(.).

Fifteen real data sets from UCI repository (http://archive.ics.uci.edu/ml) have been used in this paper to evaluate the performance of our CF-CRE method, and to compare it with respect to other nine fusion methods. The basic knowledge of the used data sets are shown by Table II. The patterns in these data sets contain multiple attributes. For each data set, the whole set of attributes will be randomly divided into n distinct subsets,[8] and each subset of attributes will be, respectively, used to train the base classifier. For example, SPECTF Heart data set has 44 attributes that can be divided into four distinct subsets, and each subset contains 11 attributes. The base classifier will be, respectively, learned based on each subset of attributes. In the first test, only ENN is used as the base classifier, whereas only naive Bayesian classifier is considered as base classifier in the second test.

The k-fold cross validation is often used for the classification performance evaluation, but k remains a free parameter [46]. We use the simplest twofold cross validation here, since the training and test sets are large, and each sample can be, respectively, used for training and testing on each fold. In the K-NNs selection, we have tested the classification performance with the K value ranging from 5 to 20 for the local weighted fusion methods, and for our proposed CF-CRE method. The three derivations of weights according to (21)–(23) have been tested and the best results are reported. In the CF-CRE method, the parameters $\gamma \in [5, 20]$ and $\lambda \in [0, 1]$ can be optimized using the training data, and optimized values corresponding to the highest accuracy is adopted. The average classification results (mean accuracy value) with $K \in [5, 20]$ for different methods are reported in Tables III and IV. For the conciseness of notations, the classifier fusion method with inner reliability evaluation and the relative reliability evaluation using DismP, J distance, and the new incompatibility measure is, respectively, denoted by CRE_D, CRE_J, and CRE_I in Tables III and IV. The accuracy curves with the different K values in different methods are shown in Figs. 1 and 2. It is worth noting that the x-axis represents the number of K, and the y-axis corresponds to the classification accuracy in Figs. 1 and 2.

In Tables III and IV, the n value is the number of classifiers, and each classifier corresponds to a subset of attributes. η_l and η_u represent, respectively, the lower and upper bounds of the classification accuracy of these individual classifiers that are combined, and the only maximum accuracy value is labeled in boldface type.

The analysis of the results of Tables III and IV shows that the fusion approaches usually improve the classification accuracy with respect to each classifier taken individually. This demonstrates the advantage and interest of combining classifiers. Meanwhile, one can see that the CRE_J, CRE_D, and CRE_I (i.e., CF-CRE) methods usually produce much higher classification accuracy than other traditional fusion methods thanks to the new refined contextual reliability evaluation strategy, which includes inner reliability (with cautious discounting operation) and relative reliability (with classical discounting operation). The inner reliability denoted by a matrix provides very important and refined prior knowledge on the local classifier performance with respect to the object based on its K-NNs, and it exactly characterizes the conditional probability of the object belonging to one class when it is assigned to another class by the given classifier. The cautious discounting is operated to correct the classification result of the object according to the inner reliability matrix, and this step can improve the classification accuracy. The relative reliability evaluation based on the incompatibility/dissimilarity measures among classifiers is used to reduce the high conflict (if necessary) among the classifiers for obtaining reasonable fusion results. However, with other weighted fusion methods, the weighting factors are mainly calculated based on the overall classification accuracy, and it cannot well capture the refined reliability knowledge of classifier with respect to each object. That is why the fusion methods with our contextual reliability evaluation outperforms the traditional weighted fusion methods. We also find that the CF-CRE method provides the highest classification accuracy in most cases. This is because the new incompatibility measure employed for the relative reliability evaluation in CF-CRE is tolerant of the difference of BBA's in a certain degree, and it can well preserve the complementary information among classifiers, which is helpful for achieving the best possible fusion performance.

According to Figs. 1 and 2, we observe that the classification performance of CF-CRE is not very sensitive to the K value. This is because the influence of the distance from the object to its neighbors is taken into account efficiently. The farther distance will yield the smaller weight (influence) of this neighbor in the inner reliability evaluation. With CF-CRE, the

[8]There is no overlapping attributes in different subsets.

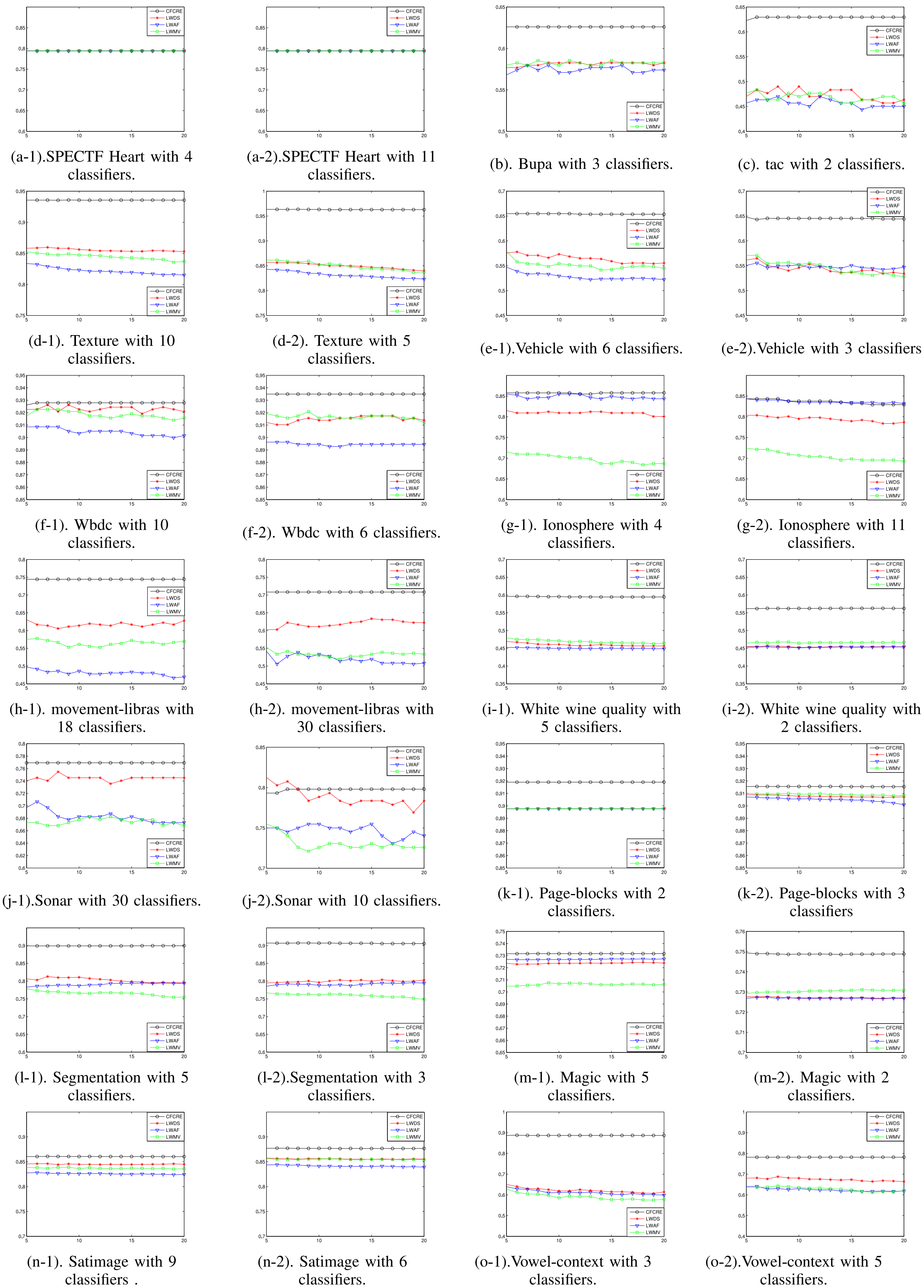
TABLE III
CLASSIFICATION RESULTS OF DIFFERENT METHODS WITH ENN CLASSIFIER (IN %)

Data	n	[η_l, η_u]	MV	WMV	LWMV	AF	WAF	LWAF	DS	WDS	LWDS	CRE$_J$	CRE$_D$	CRE$_I$
SH	4	[79.40, 79.40]	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40
SH	11	[79.40, 79.40]	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40	79.40
Bu	3	[56.82, 57.97]	57.97	57.97	58.23	57.97	56.24	57.45	57.69	57.69	58.12	62.84	**62.77**	62.61
Ta	2	[37.06, 44.39]	43.05	44.39	46.84	49.67	44.38	45.69	45.05	45.72	47.35	59.62	59.62	**62.92**
Te	10	[49.87, 68.44]	77.00	77.47	84.47	82.09	76.84	82.21	81.93	81.35	85.55	93.21	93.35	**93.56**
Te	5	[61.13, 71.35]	76.51	76.60	84.97	80.53	77.87	83.17	79.35	79.84	84.95	95.23	95.55	**96.31**
Ve	6	[39.24, 49.29]	51.42	51.42	55.13	55.67	49.88	52.86	54.96	54.73	56.45	64.65	65.17	**65.42**
Ve	3	[38.18, 50.00]	48.94	48.82	54.67	53.07	53.43	54.79	53.31	52.48	54.44	**66.37**	65.84	64.52
Wb	10	[69.60, 91.56]	91.39	91.21	91.84	91.03	90.68	90.45	91.38	91.03	92.30	92.51	92.51	**92.78**
Wb	6	[65.90, 88.75]	89.27	89.63	91.66	91.21	89.28	89.46	91.03	90.86	91.46	93.49	93.49	93.49
Io	4	[60.98, 85.76]	66.70	67.56	69.82	74.09	84.34	84.81	78.65	79.50	80.92	85.44	85.26	**85.73**
Io	10	[56.70, 81.19]	63.27	63.27	70.48	73.54	81.77	83.62	78.09	78.37	79.44	82.36	80.08	**83.62**
ML	18	[24.44, 39.17]	55.00	55.00	56.51	62.50	42.50	48.09	64.72	61.67	61.72	72.17	71.33	**74.44**
ML	30	[22.50, 37.50]	52.50	53.33	53.35	63.89	48.06	51.91	69.72	62.78	61.93	69.44	68.89	**70.83**
WQ	5	[44.73, 49.86]	45.24	45.24	46.91	45.00	44.90	45.00	44.83	44.86	46.01	59.23	58.89	**59.51**
WQ	2	[44.86, 45.69]	45.61	45.06	46.63	45.10	45.14	45.33	45.12	45.10	45.41	56.24	56.24	56.24
So	30	[51.92, 71.15]	64.42	64.90	67.46	74.52	66.35	68.30	74.04	74.04	74.43	75.64	75.16	**76.92**
So	10	[52.88, 74.52]	72.60	72.12	73.11	77.40	72.60	74.67	78.37	77.40	78.85	78.81	79.25	**79.75**
PB	2	[89.77, 89.77]	89.77	89.77	89.77	89.77	89.77	89.77	89.77	89.77	89.77	91.91	91.91	91.91
PB	3	[89.77, 91.34]	89.77	89.77	90.92	89.77	89.77	90.48	89.77	89.77	90.77	91.07	91.26	**91.55**
Se	6	[30.65, 67.75]	64.42	63.64	76.57	74.46	76.80	79.07	80.82	77.49	80.29	89.29	88.65	**89.93**
Se	3	[47.53,68.44]	68.01	68.31	75.98	77.45	79.65	79.12	84.98	82.68	79.98	89.33	89.13	**90.65**
Ma	5	[64.84, 72.38]	67.01	67.01	70.62	71.61	72.53	72.70	72.19	72.27	72.37	72.61	71.28	**73.16**
Ma	2	[64.84, 72.29]	64.84	72.29	73.04	72.31	72.37	72.69	72.39	72.36	72.72	73.98	73.78	**74.88**
Sa	9	[73.82, 78.51]	78.48	78.42	83.70	81.21	78.28	82.58	82.28	82.49	84.50	85.92	85.02	**86.05**
Sa	6	[74.55, 81.80]	80.72	81.52	85.48	84.13	81.55	84.13	83.82	84.12	85.56	87.85	**88.13**	87.68
CB	3	[30.71, 53.54]	45.86	57.27	59.10	61.01	57.07	61.21	62.53	58.28	62.21	87.76	87.10	**88.69**
CB	5	[24.44, 55.96]	50.20	58.08	62.83	69.90	58.79	62.47	74.95	61.92	67.46	74.34	70.76	**78.18**

TABLE IV
CLASSIFICATION RESULTS OF DIFFERENT METHODS WITH BAYESIAN CLASSIFIER (IN %)

Data	n	[η_l, η_u]	MV	WMV	LWMV	AF	WAF	LWAF	DS	WDS	LWDS	CRE$_J$	CRE$_D$	CRE$_I$
SH	4	[62.94, 66.31]	60.33	66.69	71.89	67.82	63.32	67.97	71.93	70.43	72.70	73.98	73.24	**75.48**
SH	11	[56.95, 68.93]	71.18	71.18	78.83	73.43	75.66	79.58	77.16	76.04	82.80	82.40	82.43	**82.98**
Bu	3	[50.15, 56.53]	56.24	56.24	58.59	60.29	55.08	56.89	60.58	59.13	58.93	**63.17**	62.18	62.89
Ta	2	[42.39, 50.33]	39.08	44.41	52.07	49.66	52.98	50.99	50.32	49.65	51.89	62.38	63.19	**64.25**
Te	5	[56.62, 72.84]	71.80	73.18	80.28	74.40	67.42	77.25	77.45	75.36	81.48	93.02	**93.80**	93.30
Te	10	[46.05, 66.53]	69.45	70.87	78.26	74.35	66.47	75.09	77.45	75.53	80.12	88.01	**89.41**	89.02
Ve	3	[39.72, 51.89]	46.69	49.88	60.77	49.53	52.96	60.42	47.28	50.71	61.55	65.43	65.66	**67.24**
Ve	6	[36.76, 43.38]	46.57	48.35	56.09	47.28	44.33	50.85	45.51	46.34	59.43	64.67	64.72	**65.25**
Wb	6	[71.88, 94.02]	92.62	92.26	92.45	91.21	91.21	91.50	93.32	93.49	93.79	93.49	93.32	**95.10**
Wb	15	[62.39, 93.14]	88.93	90.33	91.75	91.74	91.74	91.94	93.15	92.44	93.53	93.44	93.62	**93.70**
Io	4	[65.82, 87.74]	74.09	77.78	79.87	72.37	87.17	88.90	82.61	87.45	85.51	86.57	86.39	**88.17**
Io	10	[64.40, 88.60]	71.52	74.08	80.78	75.50	90.02	90.41	81.19	87.18	85.65	89.87	89.87	**91.32**
ML	30	[18.61, 32.50]	40.83	40.00	44.13	50.83	36.67	41.42	69.44	49.72	52.66	66.22	65.44	**70.28**
ML	18	[20.83, 33.06]	44.17	41.67	46.06	49.17	35.00	39.86	69.44	46.94	51.04	69.33	69.43	**72.22**
WQ	2	[41.40, 46.82]	46.57	46.82	48.06	47.73	46.63	48.31	48.00	48.24	49.01	57.81	57.81	57.81
WQ	5	[42.83, 48.67]	46.69	46.94	48.66	47.02	47.12	48.04	46.75	47.08	48.93	59.55	59.29	**60.04**
So	10	[53.37, 66.35]	71.63	71.15	74.70	70.19	59.13	64.18	69.23	66.83	73.20	79.05	79.27	**79.81**
So	30	[52.88, 73.56]	74.04	72.12	74.67	73.08	60.10	60.28	71.15	71.63	74.52	79.13	78.37	**80.29**
PB	5	[87.19, 91.94]	90.83	91.25	91.80	91.78	91.05	91.93	93.06	92.64	92.62	92.88	93.13	**93.76**
PB	2	[83.81, 93.42]	93.44	93.42	93.84	93.31	93.57	94.28	93.59	93.75	94.22	94.86	94.86	94.86
Se	3	[52.73, 80.09]	65.93	77.10	83.33	71.08	65.50	73.41	80.39	79.22	83.50	89.67	89.84	**90.93**
Se	6	[27.32, 61.60]	62.29	62.73	80.12	71.73	74.98	78.11	79.40	80.61	79.59	86.85	87.02	**87.13**
Ma	2	[70.25, 75.65]	73.29	75.65	75.31	72.96	71.87	73.45	72.96	73.00	74.16	77.11	77.11	77.11
Ma	5	[64.59, 73.70]	71.93	71.93	73.16	73.43	72.13	73.04	73.43	73.63	74.21	74.20	74.28	**74.69**
Sa	6	[74.51, 78.20]	81.54	81.07	85.59	81.82	77.93	83.21	80.34	81.55	85.71	86.34	86.55	**87.98**
Sa	9	[74.48, 79.15]	81.86	81.60	85.77	82.39	78.71	83.07	80.34	81.66	85.47	86.90	86.92	**87.19**
CB	5	[19.49, 55.35]	38.79	55.96	57.86	63.84	57.07	58.81	71.62	59.29	62.42	72.73	69.09	**78.18**
CB	3	[24.44, 56.87]	45.25	58.99	62.52	66.57	60.20	62.55	71.62	61.72	65.47	89.80	90.10	**92.46**

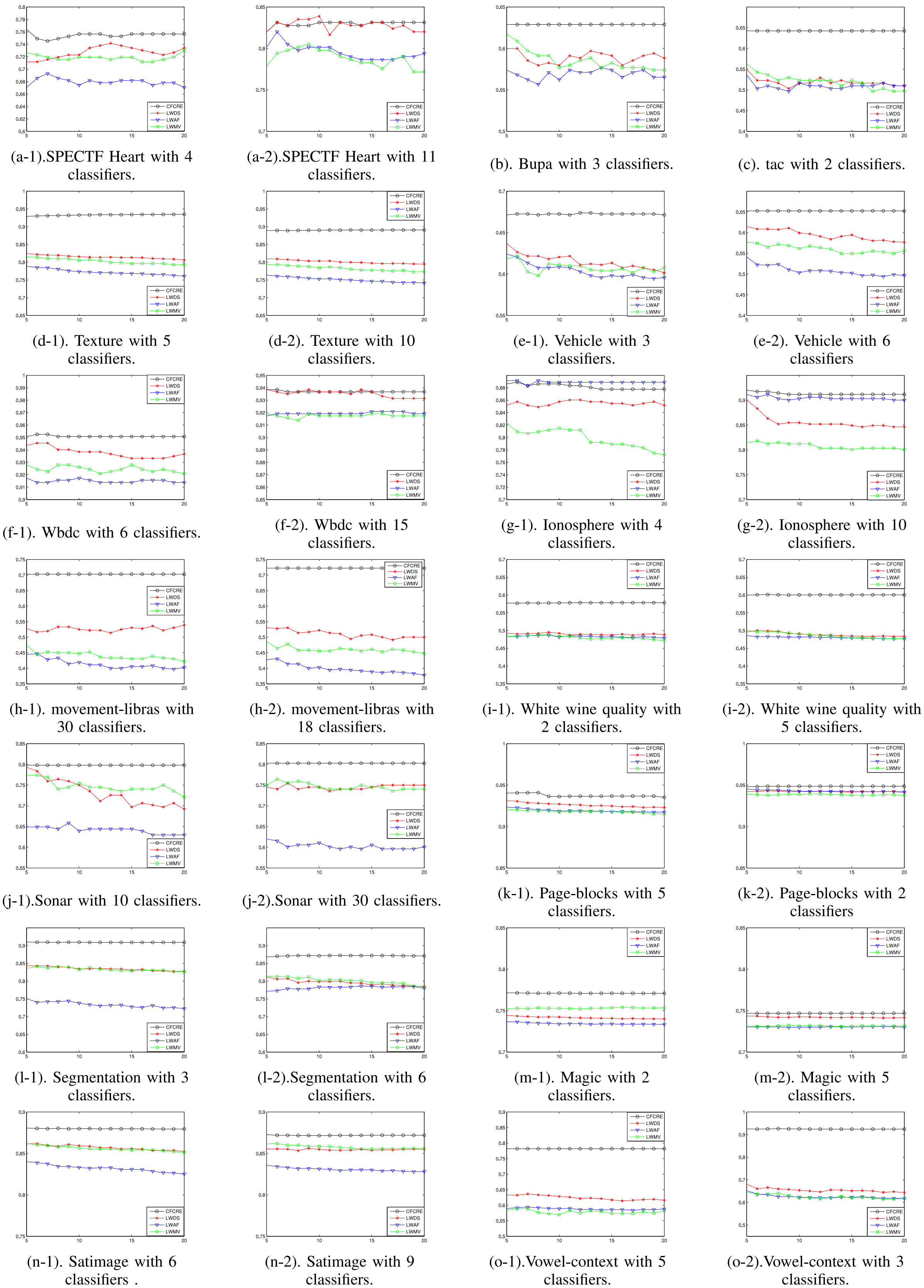
neighbors which are quite far from the object will have only very little influence on the classification of the object. Our results show that the CF-CRE method is robust with respect to the K value, which is a good property for the real applications. So the small number of K can be recommended for applications to reduce the computation burden.



(a-1).SPECTF Heart with 4 classifiers.
(a-2).SPECTF Heart with 11 classifiers.
(b). Bupa with 3 classifiers.
(c). tac with 2 classifiers.
(d-1). Texture with 10 classifiers.
(d-2). Texture with 5 classifiers.
(e-1).Vehicle with 6 classifiers.
(e-2).Vehicle with 3 classifiers
(f-1). Wbdc with 10 classifiers.
(f-2). Wbdc with 6 classifiers.
(g-1). Ionosphere with 4 classifiers.
(g-2). Ionosphere with 11 classifiers.
(h-1). movement-libras with 18 classifiers.
(h-2). movement-libras with 30 classifiers.
(i-1). White wine quality with 5 classifiers.
(i-2). White wine quality with 2 classifiers.
(j-1).Sonar with 30 classifiers.
(j-2).Sonar with 10 classifiers.
(k-1). Page-blocks with 2 classifiers.
(k-2). Page-blocks with 3 classifiers
(l-1). Segmentation with 5 classifiers.
(l-2).Segmentation with 3 classifiers.
(m-1). Magic with 5 classifiers.
(m-2). Magic with 2 classifiers.
(n-1). Satimage with 9 classifiers .
(n-2). Satimage with 6 classifiers.
(o-1).Vowel-context with 3 classifiers.
(o-2).Vowel-context with 5 classifiers.

Fig. 1. Classification results of different fusion methods based on ENN classifier.

In the fusion of classifiers, it is often encountered that the classification results from different classifiers are in high conflict and sometimes in total contradiction, and this quite likely causes unreasonable fusion result (i.e., error). The proposed fusion method CF-CRE is able to well deal with the conflicting case, and that is why CF-CRE can efficiently improve



(a-1).SPECTF Heart with 4 classifiers.
(a-2).SPECTF Heart with 11 classifiers.
(b). Bupa with 3 classifiers.
(c). tac with 2 classifiers.
(d-1). Texture with 5 classifiers.
(d-2). Texture with 10 classifiers.
(e-1). Vehicle with 3 classifiers.
(e-2). Vehicle with 6 classifiers
(f-1). Wbdc with 6 classifiers.
(f-2). Wbdc with 15 classifiers.
(g-1). Ionosphere with 4 classifiers.
(g-2). Ionosphere with 10 classifiers.
(h-1). movement-libras with 30 classifiers.
(h-2). movement-libras with 18 classifiers.
(i-1). White wine quality with 2 classifiers.
(i-2). White wine quality with 5 classifiers.
(j-1).Sonar with 10 classifiers.
(j-2).Sonar with 30 classifiers.
(k-1). Page-blocks with 5 classifiers.
(k-2). Page-blocks with 2 classifiers
(l-1). Segmentation with 3 classifiers.
(l-2).Segmentation with 6 classifiers.
(m-1). Magic with 2 classifiers.
(m-2). Magic with 5 classifiers.
(n-1). Satimage with 6 classifiers .
(n-2). Satimage with 9 classifiers.
(o-1).Vowel-context with 5 classifiers.
(o-2).Vowel-context with 3 classifiers.

Fig. 2. Classification results of different fusion methods based on Bayesian classifier.

the classification accuracy as shown in the experiment. In CF-CRE, the cautious discounting rule is employed to cautiously transfer the masses of belief from the singleton classes to the associated class set (i.e., disjunction of two classes) based on the inner reliability. In fact, this step is to make the classification results from different classifiers closer to the common truth for improving the accuracy. So it can reduce the conflict degree among classifiers by properly modeling the partial imprecision, which will be refined through the combination with other classifiers. After that, the relative reliability of each classifier is evaluated according to its incompatibility with the others, and the classifier highly conflicting with the others is committed with a small relative reliability value. Shafer's discounting rule is applied with the relative reliability to proportionally discount the masses of belief from other focal elements to the total ignorance, which plays a neutral role in fusion process. Hence, it can efficiently decrease the influence of the classifier highly conflicting with the others in the fusion. So the new CF-CRE method generally produces high classification accuracy by properly handling the conflicting information.

V. Conclusion

We have proposed a new CF-CRE including the inner reliability and the relative reliability. The inner reliability denoted by a matrix is introduced to express the conditional probability of the object belonging to each class knowing the classification result given by the classifier. This matrix provides much more refined reliability knowledge than training accuracy that is often used to reflect the overall classifier performance in general, and it is estimated by deeply exploring the classification information of the K-NNs of the object. Then a new cautious discounting rule is developed to prudently redistribute the partial probability (or belief degree) of each class (e.g., A) to the associated imprecise classes (e.g., $A \cup B$) according to the inner reliability matrix under the BFs framework. This modification of classification result can reduce the error risk by modeling the partial imprecision that can be specified in the combination with other classifiers. The relative reliability is mainly used to reduce the harmful conflict among classifiers, and it is calculated based on a new incompatibility measure. This measure can well capture the high conflicting information, and it is also tolerant of the BBA difference in a certain degree for preserving the important complementarity of classifiers, which is very useful for achieving the good fusion performance. The classification result quite incompatible with the others is committed with a small relative reliability value. Shafer's discounting rule is then applied with the relative reliability to efficiently decrease the influence of the classification result highly conflicting with the others in the fusion. The effectiveness of CF-CRE has been clearly demonstrated using various real data sets by the comparisons with other related methods. The experimental results show that CF-CRE can significantly improve the classification accuracy thanks to the refined contextual reliability evaluation. Furthermore, this new method is also robust to the changes of choice of the K value, which is very convenient in practice for the applications. The CF-CRE method will be further tested with some other base classifiers like support vector machine in our future work.

References

[1] B. Quost, M.-H. Masson, and T. Denœux, "Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules," *Int. J. Approx. Reason.*, vol. 52, no. 3, pp. 353–374, 2011.

[2] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[3] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[4] Z. W. Yu, L. Li, J. M. Liu, and G. Q. Han, "Hybrid adaptive classifier ensemble," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 177–190, Feb. 2015.

[5] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[6] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[7] F. Moreno-Seco, J. M. Inesta, P. J. P. de Leon, and L. Micó, *Comparison of Classifier Fusion Methods for Classification in Pattern Recognition Tasks*, D.-Y. Yeung, J. T. Kwok, A. Fred, F. Roli, and D. de Ridder, Eds. Berlin, Germany: Springer-Verlag, 2006, pp. 705–713.

[8] H. He and Y. Cao, "SSC: A classifier combination method based on signal strength," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1100–1117, Jul. 2012.

[9] J. Yang, X. Zeng, S. Zhong, and S. Wu, "Effective neural network ensemble approach for improving generalization performance," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 878–887, Jun. 2013.

[10] G. Rogova, "Combining the results of several neural network classifiers," *Neural Netw.*, vol. 7, no. 5, pp. 777–781, 1994.

[11] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, nos. 1–2, pp. 239–263, 2002.

[12] Z.-H. Zhou and Y. Jiang, "NeC4.5: Neural ensemble based C4.5," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 6, pp. 770–773, Jun. 2004.

[13] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 281–286, Feb. 2002.

[14] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognit.*, vol. 34, no. 2, pp. 299–314, 2001.

[15] N. J. Pizzi and W. Pedrycz, "Aggregating multiple classification results using fuzzy integration and stochastic feature selection," *Int. J. Approx. Reason.*, vol. 51, no. 8, pp. 883–894, 2010.

[16] D. A. Bell, J. W. Guan, and Y. Bi, "On combining classifier mass functions for text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 10, pp. 1307–1319, Oct. 2005.

[17] Y. X. Bi, J. W. Guan, and D. Bell, "The combination of multiple classifiers using an evidential reasoning approach," *Artif. Intell.*, vol. 172, no. 15, pp. 1731–1751, 2008.

[18] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Comput. Inf. Syst.*, vol. 7, no. 1, pp. 1–10, Jan. 2000.

[19] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artif. Intell.*, vol. 205, pp. 1–29, Dec. 2013.

[20] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[21] F. Smarandache and J. Dezert, Eds., *Advances and Applications of DSmT for Information Fusion*, vol. 4. Rehoboth, NM, USA: Amer. Res. Press, 2015. [Online]. Available: http://www.onera.fr/staff/jean-dezert?page=2

[22] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 447–458, May 1990.

[23] A.-L. Jousselme, C. Liu, D. Grenier, and E. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 5, pp. 890–903, Sep. 2006.

[24] T. Denœux, Z. Younes, and F. Abdallah, "Representing uncertainty on set-valued variables using belief functions," *Artif. Intell.*, vol. 174, nos. 7–8, pp. 479–499, 2010.

[25] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowl. Based Syst.*, vol. 94, pp. 114–123, Feb. 2016.

[26] S. Destercke, P. Buche, and B. Charnomordic, "Evaluating data reliability: An evidential answer with application to a Web-enabled data warehouse," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 92–105, Jan. 2013.

[27] T. Denœux, "A k-nearest neighbor classification rule based on Dempster–Shafer Theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.

[28] Z.-G. Liu, Q. Pan, J. Dezert, and G. Mercier, "Credal classification rule for uncertain data based on belief functions," *Pattern Recognit.*, vol. 47, no. 7, pp. 2532–2541, 2014.

[29] Z.-G. Liu, Q. Pan, G. Mercier, and J. Dezert, "A new incomplete pattern classification method based on evidential reasoning," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 635–646, Apr. 2015.

[30] T. Denœuxm and P. Smets, "Classification using belief functions: Relationship between case-based and model-based approaches," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 6, pp. 1395–1406, Dec. 2006.

[31] T. Denœux, "A neural network classifier based on Dempster–Shafer theory," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 2, pp. 131–150, Mar. 2000.

[32] Z.-G. Liu, Q. Pan, J. Dezert, and G. Mercier, "Hybrid classification system for uncertain data," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2016.2622247.

[33] T. Denœux, "Maximum likelihood estimation from uncertain data in the belief function framework," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 119–130, Jan. 2013.

[34] Z.-G. Liu, J. Dezert, Q. Pan, and G. Mercier, "Combination of sources of evidence with different discounting factors based on a new dissimilarity measure," *Decis. Support Syst.*, vol. 52, no. 1, pp. 133–141, 2011.

[35] S. Huang, X. Su, Y. Hu, S. Mahadevan, and Y. Deng, "A new decision-making method by incomplete preferences based on evidence distance," *Knowl. Based Syst.*, vol. 56, pp. 264–272, Jan. 2014.

[36] R. P. W. Duin, "The combining classifier: To train or not to train?" in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, vol. 2. Quebec City, QC, Canada, 2002, pp. 765–770.

[37] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2004.

[38] J. Dezert and A. Tchamova, "On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule," *Int. J. Intell. Syst.*, vol. 29, no. 3, pp. 223–252, Mar. 2014.

[39] D. Mercier, B. Quost, and T. Denœux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Inf. Fusion*, vol. 9, no. 2, pp. 246–258, 2008.

[40] L. Xu, A. Krzyzak, and C. Y. Suen, "Several methods for combining multiple classifiers and their applications in handwritten character recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 418–435, Jun. 1992.

[41] A. Al-Ani and M. Deriche, "A new technique for combining multiple classifiers using the Dempster–Shafer theory of evidence," *J. Artif. Intell. Res.*, vol. 17, pp. 333–361, 2002.

[42] Z.-G. Liu, Q. Pan, and J. Dezert, "Classifier fusion based on cautious discounting of beliefs," in *Proc. 19th Int. Conf. Inf. Fusion*, Heidelberg, Germany, 2016, pp. 363–370.

[43] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *Int. J. Approx. Reason.*, vol. 53, no. 2, pp. 118–145, 2012.

[44] W. Liu, "Analyzing the degree of conflict among belief functions," *Artif. Intell.*, vol. 170, no. 11, pp. 909–924, 2006.

[45] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2003.

[46] S. Geisser, *Predictive Inference: An Introduction*. New York, NY, USA: Chapman & Hall, 1993.

Zhunga Liu was born in China, in 1984. He received the bachelor's, master's, and Ph.D. degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 2007, 2010, and 2013, respectively.

He is an Associate Professor with the School of Automation, NPU. His current research interests include belief function theory and its application in pattern recognition.

Quan Pan was born in China, in 1961. He received the bachelor's degree from the Huazhong University of Science and Technology, Wuhan, China, in 1991, and the master's and Doctoral degrees from Northwestern Polytechnical University (NPU), Xi'an, China, in 1997.

He has been a Professor, since 1998, and the Dean of the School of Automation, NPU, since 2009. His current research interests include information fusion and pattern recognition.

Jean Dezert was born in France, in 1962. He received the electrical engineering and Ph.D. degrees from the University of Paris XI, Orsay, France, in 1985 and 1990, respectively.

Since 1993, he has been a Senior Research Scientist with ONERA—The French Aerospace Laboratory, Palaiseau, France. His current research interests include decision-making support and belief function theory.

Jun-Wei Han was born in China, in 1977. He received the bachelor's, master's, and Ph.D. degrees from Northwestern Polytechnical university (NPU), Xi'an, China, in 1999, 2001, and 2003, respectively.

He has been a Professor with the School of Automation, NPU, since 2010. His current research interest includes pattern recognition.

You He was born in China, in 1956. He received the Ph.D. degree from Tsinghua University, Beijing, China, in 1997.

He is currently a Professor with the Department of Electronic Engineering, Tsinghua University. His current research interest includes information fusion.

Dr. He was elected as an Chinese Academy of Engineering Academician in 2013.