Higher Order Fractal Belief Rényi Divergence with Its Applications in Pattern Classification

Yingcheng Huang, Fuyuan Xiao, *Senior Member, IEEE*, Zehong Cao, *Member, IEEE*, Chin-Teng Lin, *Fellow, IEEE*

Abstract—Information can be quantified and expressed by uncertainty, and improving the decision level of uncertain information is vital in modeling and processing uncertain information. Dempster–Shafer evidence theory can model and process uncertain information effectively. However, the Dempster combination rule may provide counter-intuitive results when dealing with highly conflicting information, leading to a decline in decision level. Thus, measuring conflict is significant in the improvement of decision level. Motivated by this issue, this paper proposes a novel method to measure the discrepancy between bodies of evidence. First, the model of dynamic fractal probability transformation is proposed to effectively obtain more information about the non-specificity of basic belief assignments (BBAs). Then, we propose the higher-order fractal belief Rényi divergence (HOFBReD). HOFBReD can effectively measure the discrepancy between BBAs. Moreover, it is the first belief Rényi divergence that can measure the discrepancy between BBAs with dynamic fractal probability transformation. HoFBReD has several properties in terms of probability transformation as well as measurement. When the dynamic fractal probability transformation ends, HoFBReD is equivalent to measuring the Rényi divergence between the pignistic probability transformations of BBAs. When the BBAs degenerate to the probability distributions, HoFBReD will also degenerate to or be related to several well-known divergences. In addition, based on HoFBReD, a novel multisource information fusion algorithm is proposed. A pattern classification experiment with real-world datasets is presented to compare the proposed algorithm with other methods. The experiment results indicate that the proposed algorithm has a higher average pattern recognition accuracy with all datasets than other methods. The proposed discrepancy measurement method and multisource information algorithm contribute to the improvement of decision level.

Index Terms—Dempster–Shafer evidence theory, Rényi divergence, probability transformation, Multisource information fusion, Pattern classification.

1 Introduction

Uncertain information is ubiquitous in the real world, and how to deal with uncertain information and improve the decision level is one of the hot issues in recent years [1]–[3]. A variety of theories have been presented to model and process uncertain information, including the extended probability theory [4], [5], fuzzy theory [6], Z-number [7], quantum evidence theory [8], [9], random permutation set [10], [11] and entropy-based approach [12]. These methods have been widely applied in many fields including COVID-19 waves superposition [13], graph clustering [14], dynamic modeling of multi-agent learning [15], [16], driving task recordings [17], decision making [18]–[20], outranking relations [21], network analysis [22], engineering systems design [23], fault analysis [24], rapid source localization [25]. In this paper, Dempster-Shafer evidence theory is focused on improving the decision level. As the extension of the probability theory [26], Dempster-Shafer evidence theory is presented by Dempster [27] and extended by Shafer [28], which can model and process uncertain information effectively.

However, highly conflicting information is often hard to handle in Dempster-Shafer evidence theory, which has declined the decision level of it [29]. Measuring the discrepancy between bodies of evidence is essential to the effectiveness of a combination of evidence [30], [31]. For measuring the discrepancy between bodies of evidence, scholars have proposed a number of methods [32]. Belief divergence is a well-known theory that can measure the discrepancy between basic belief assignments (BBAs) in Dempster-Shafer evidence theory [33], [34]. Xiao proposed the belief Jensen–Shannon divergence on the basis of Shannon entropy [35]. After BJS divergence is proposed, $\mathcal{B}$ divergence and $\mathcal{RB}$ divergence that can take the correlation between BBAs into account are proposed [36]. Although these methods can measure the discrepancy between BBAs in some cases, they may provide counter-intuitive and invalid results. For instance, consider a numerical example:

Example 1.1. Let a frame of discernment be $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Define BBAs m_1, m_2, m_3 and m_4 in Ω as follows:

$$
\begin{aligned}
&m_1 : m_1(\{\omega_1\}) = \frac{1}{2}; m_1(\{\omega_2\}) = \frac{1}{2};\\
&m_2 : m_2(\{\omega_1\omega_2\}) = 1;\\
&m_3 : m_3(\{\omega_1\omega_2\omega_3\}) = \frac{1}{2}; m_3(\{\omega_2\omega_3\omega_4\}) = \frac{1}{2};\\
&m_4 : m_4(\{\omega_1\omega_2\omega_3\omega_4\}) = 1.
\end{aligned}
$$

As shown in Table 1, classical discrepancy measurement

- *(Corresponding author: Fuyuan Xiao.)*
  Y. Huang and F. Xiao are with the School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China. (e-mail: xiaofuyuan@cqu.edu.cn; doctorxiaofy@hotmail.com)
- *Zehong Cao is with STEM, Mawson Lakes Campus, University of South Australia, Adelaide, SA 5095, Australia (e-mail: zhcaonctu@gmail.com).*
- *Chin-Teng Lin is with the Australia AI Institute, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: Chin-Teng.Lin@uts.edu.au).*

TABLE 1
Different discrepancy measurement methods in Example 1.1.

BBAs	$\mathcal{K}$	BJS	$\mathcal{B}$	$\mathcal{RB}$	Bχ^2	d$_{BBA}$
m_1 and m_2	0	1	0.2688	0.5185	1	0.5000
m_2 and m_3	0	1	-0.0118	0.1088	1	0.5401
m_3 and m_4	0	1	0.1957	0.4424	1	0

methods were invalid in Example 1.1. The whole conflicting factor, the BJS divergence, cannot distinguish the discrepancy between the four evidence well. The $\mathcal{B}$ divergence between m_2 and m_3 is a negative number. The $\mathcal{RB}$ divergence between m_1 and m_2 is larger than that between m_2 and m_3. This is not intuitive because m_2 only supports $\omega_1\omega_2$, and it is noticeable that m_1 supports ω_1 and ω_2, and m_3 supports $\omega_1\omega_2\omega_3$ and $\omega_2\omega_3\omega_4$. From the perspective of intuitiveness, the divergence between m_1 and m_2 should be smaller than that between m_2 and m_3. The belief $\mathcal{X}^2$ divergence [37] is not able to significantly distinguish the discrepancy between the four evidence either. Jousselme et al.'s evidence distance [38] also showed the limited measurement of the discrepancy between m_3 and m_4. The major problem of low discrepancy across different evidence is caused by discord and non-specificity which are two essential properties of BBA [26]. Especially for, non-specificity, it can differ between BBAs and probability distributions, representing the distribution's uncertainty. For instance, we assume from m_1 to m_4, the non-specificity gradually increases from an intuitive point of view. However, the results in Table 1 demonstrated the counterfactual outcomes, due to ineffective information received from the non-specificity of BBAs where m_1, m_2, m_3 and m_4, leading to the invalid discrepancy measurement.

To handle this problem, it is required to develop an effective method to obtain more valuable discrepancy knowledge about the non-specificity of BBAs. Some previous work discussed how to deal with the information about the non-specificity of BBAs in several fields. For example, probability transformation transforms BBAs into probability distributions [39], which can eliminate the non-specificity of BBAs. Deng's fractal-based belief entropy [40] analyzed the process of probability transformation in entropy measurement, showing that the process of probability transformation contains information about non-specificity. However, the existing methods still cannot deal with the situation in Example 1.1, and how to develop an effective method to obtain more valuable discrepancy knowledge about the non-specificity of BBAs is still an open issue. This motivates this work to develop a novel belief divergence method.

This paper proposes a mathematical model called dynamic fractal probability transformation, which can effectively obtain more information about the non-specificity of BBAs with robust representation. A novel belief divergence with dynamic fractal probability transformation, namely higher order fractal belief Rényi divergence (HoFBReD), is proposed. HoFBReD is the first belief Rényi divergence that can measure the discrepancy between BBAs with dynamic fractal probability transformation. HoFBReD has several properties in terms of probability transformation as well as measurement. As the dynamic fractal probability transformation ends, HoFBReD is equivalent to measuring the Rényi divergence between the pignistic probability transformations of BBAs. When the BBAs degenerate to the probability distributions, HoFBReD will also degenerate to or be related to several well-known divergences, such as Kullback-Leibler divergence as α changes. The advantages of HoFBReD are illustrated by several numerical examples. In addition, based on HoFBReD, a novel multisource information fusion algorithm is proposed. A pattern classification experiment with real-world datasets is presented to compare the proposed algorithm with other methods. The results of the pattern classification experiment with real-world datasets indicate that the proposed algorithm has a higher pattern recognition accuracy compared to other methods.

The main contributions of this article are as follows:

- We propose a higher order fractal belief Rényi divergence with dynamic fractal probability transformation. It is the first belief Rényi divergence that can measure the discrepancy between BBAs with dynamic fractal probability transformation and extends the order to a 'higher' order. Besides, the proposed belief divergence can effectively obtain more information about the non-specificity of BBAs during discrepancy measurement with robust representation, which is superior to other methods.
- A novel multisource information fusion algorithm is proposed. A pattern classification experiment with real-world datasets is presented, and the results show that the proposed method has a higher pattern recognition accuracy than other methods. The novel algorithm contributes to the improvement of the decision level.

This paper is organized as follows. Section 2 introduces the preliminaries of Dempster-Shafer evidence theory, divergence theory and probability transformation methods. In section 3, the model of dynamic fractal probability transformation is proposed, which is applied to a novel belief divergence, and its properties are proven. Section 4 illustrates the advantages of the proposed belief divergence by numerical examples. In section 5, a novel multisource information fusion algorithm is proposed. Section 6 presents a pattern recognition experiment with real-world datasets. Section 7 concludes the study.

2 Preliminaries

In this section, a brief introduction is given to the Dempster-Shafer (D-S) evidence theory, several well-known divergences and probability transformation methods.

2.1 Dempster–Shafer evidence theory

Dempster–Shafer evidence theory, also known as evidence theory, is presented by Dempster and developed by Shafer [27], [28]. D-S evidence theory is extended from probability

theory and fuzzy set theory. It has versatile applications, including software risk evaluation [41], multisource quantum information fusion [42], rescuer assignments [43], evidential reasoning [44]–[46], output control [47], clustering [48] and heuristic representation learning [49].

Definition 1. *Frame of discernment*

A discernment framework is defined as [27], [28]

$$\Omega = \{\omega_1, \omega_2, \dots \omega_N\}, \tag{1}$$

which consists of N mutually exclusive and collectively exhaustive status of a certain pattern.

The power set of Ω is defined as

$$2^{\Omega} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \cdots, \{\omega_N\}, \{\omega_1 \cup \omega_2\}, \cdots, \Omega\}, \tag{2}$$

where the empty set is denoted as $\emptyset$ [50]. There are both singleton sets and multielement sets in 2^{Ω}. The multielement sets (e.g., $\omega_1 \cup \omega_2 \cup \omega_3$) can express the uncertainty and imprecision degree among classes ω_1, ω_2 and ω_3.

Definition 2. *Basic belief assignments*

A basic belief assignment (BBA), which is a mapping $m(\cdot)$ from 2^{Ω} to [0,1], is defined as [27], [28]

$$m : 2^{\Omega} \to [0, 1], \tag{3}$$

and satisfies the following conditions:

$$\sum_{X \in 2^{\Omega}} m(X) = 1, \tag{4}$$

$$m(\emptyset) = 0, \tag{5}$$

where $m(X)$ represents the support degree of the pattern associated with the element X. If $m(X) > 0$, X is called a focal element. In D-S evidence theory, BBAs are also called mass functions of belief. $m(\Omega)$ denotes the total ignorance degree, which describes the vacuous belief source of evidence.

Definition 3. *Dempster's rule of combination*

In multisource information fusion processing, each classification result can be regarded as an evidence source represented by a BBA. The well-known Dempster's rule is applied to combine the multiple BBAs. Let $m_1(\cdot)$ and $m_2(\cdot)$ be two distinct BBAs in a frame of discernment 2^{Ω}. For BBAs $m_1(\cdot)$ and $m_2(\cdot)$, Dempster's rule of combination, represented in the form $m = m_1 \oplus m_2$, is defined as [27]

$$m(X) = (m_1 \oplus m_2)(X) = \frac{\sum\limits_{Y \cap Z = X} m_1(Y) m_2(Z)}{1 - \mathcal{K}}, \tag{6}$$

and

$$\mathcal{K} = \sum_{Y \cap Z = \emptyset} m_1(Y) m_2(Z) < 1, \tag{7}$$

where $X, Y, Z \subseteq \Omega$ and $\mathcal{K}$ is the whole conflicting factor.

Dempster's rule of combination is commutative, conjunctive and associative.

2.2 Divergence theory

Divergence theory is presented to measure the discrepancy between two probability distributions or two BBAs [51], [52]. Specifically, the divergences that can measure the discrepancy between BBAs are called belief divergences.

Definition 4. *Rényi divergence*

For two probability distributions $P = (p_1, p_2, \cdots, p_N)$ and $Q = (q_1, q_2, \cdots, q_N)$, the Rényi divergence between P and Q is defined as [53]

$$D_{\alpha}(P||Q) = \frac{1}{\alpha - 1} \ln \sum_{i=1}^{N} p_i^{\alpha} q_i^{1-\alpha}, \tag{8}$$

where $\alpha \in (0, 1) \cup (1, +\infty)$. When $\alpha > 1$, the rules that $\frac{0}{0} = 0$ and $\frac{p_i^{\alpha}}{0} = \infty$ are adopted.

Definition 5. *Belief Jensen–Shannon divergence*

Recently, the Belief Jensen–Shannon (BJS) divergence was presented by Xiao to measure the discrepancy between BBAs in D-S evidence theory:

$$\begin{aligned} &D_{BJS}(m_1||m_2) \\ &= \frac{1}{2} \sum_{i=1}^{2^N - 1} m_1(E_i) \log_2 \left[\frac{2m_1(E_i)}{m_1(E_i) + m_2(E_i)} \right] \\ &+ \frac{1}{2} \sum_{i=1}^{2^N - 1} m_2(E_i) \log_2 \left[\frac{2m_2(E_i)}{m_1(E_i) + m_2(E_i)} \right], \end{aligned} \tag{9}$$

where m_1 and m_2 are two BBAs in 2^{Ω}, and E_i is any element in 2^{Ω} except $\emptyset$.

After BJS divergence is proposed, scholars have presented several new belief divergences [37], [54], [55].

2.3 Probability transformation methods

Definition 6. *Pignistic probability transformation*

Let a frame of discernment be $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ with a BBA m in Ω. Its pignistic probability transformation, denoted by $BetP(\omega_i)$ can be defined as [39]

$$BetP(\omega_i) = \sum_{\omega_i \in E | E \in 2^{\Omega}} \frac{m(E)}{|E|}, \tag{10}$$

where E is any element in 2^{Ω} except $\emptyset$ here and $|E|$ denotes the cardinality of E.

Definition 7. *Plausibility transformation method*

Let a frame of discernment be $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ with a BBA m in Ω. Its plausibility transformation method, denoted by $PnPl(\omega_i)$, is defined as [56]

$$PnPl(\omega_i) = \frac{Pl(\omega_i)}{\sum\limits_{j=1}^{N} Pl(\omega_j)}, \tag{11}$$

where Pl is defined as the plausibility function:

$$Pl(X) = \sum\nolimits_{Y \cap X \neq \emptyset | X, Y \subseteq 2^{\Omega}} m(Y). \tag{12}$$

The fractal theory can be useful in the expression of the dimension of information [57]. And Deng's information volume of mass function applies the concept of fractal.

Definition 8. *Deng's information volume of mass function*

Let a frame of discernment be $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ with a BBA m in Ω. The Deng's information volume of mass function is defined as follows [58]:

Step 1: Input the BBA m.

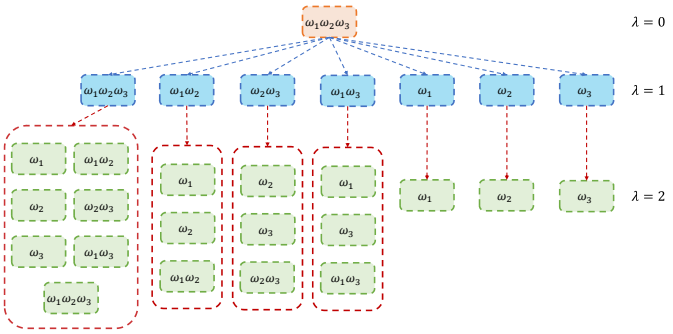


Fig. 1. Two times power splitting of $A = \{\omega_1, \omega_2, \omega_3\}$.

Step 2: According to the distribution of the maximum Deng entropy, the mass functions of the multielement sets are split.

Step 3: Obtain the Deng entropy of the new BBA. Execute Step 4 if the result is less than ϵ. Otherwise, repeat Steps 2-3. ϵ is the error coefficient.

Step 4: Output the information volume of the mass function.

Deng's information volume of mass function can obtain information from the process of splitting the mass functions.

3 A NOVEL BELIEF DIVERGENCE WITH DYNAMIC FRACTAL PROBABILITY TRANSFORMATION

In this section, dynamic fractal probability transformation is presented. Then, a novel belief divergence, namely higher order fractal belief Rényi divergence, is proposed. Its properties are then analyzed.

3.1 Dynamic fractal probability transformation

A mathematical model, namely dynamic fractal probability transformation, is proposed to handle the problems in Example 1.1 by effectively obtaining more information about non-specificity. It is analyzed from the perspective of generating function and power splitting. First, a hypothesis about power splitting is presented.

Hypothesis 1. *Let a frame of discernment be* $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$. *Let* A *be any element in* 2^Ω *except* $\emptyset$ *and* B *be any subset of* A *except* $\emptyset$. *If power splitting of* A *is conducted for* λ *times, the number of all elements except is* $(\lambda+1)^{|A|} - \lambda^{|A|}$, *and the number of* B *is* $\lambda^{|A|-|B|}$.

For example, let A be $\{\omega_1, \omega_2, \omega_3\}$ and conduct power splitting for 2 times to A as shown in Figure 1. The green squares are the elements that are generated after splitting two times. According to hypothesis 1, for green squares, the number of $\{\omega_1\}$, $\{\omega_2\}$ and $\{\omega_3\}$ is 4; The number of $\{\omega_1\omega_2\}$, $\{\omega_1\omega_3\}$ and $\{\omega_2\omega_3\}$ is 2; The number of $\{\omega_1\omega_2\omega_3\}$ is 1, and the number of all elements is 19. The proof is presented as follows.

Proof. Define a variable σ_i and denote ω_i as σ_i. According to the generating function theory, the generating function of the element A is defined as $f(A)$

$$\begin{aligned} f(A) &= (\sigma_1+1)(\sigma_2+1)...(\sigma_{|A|}+1) \\ &= \sigma_1\sigma_2...\sigma_{|A|} + \sum_{cyc}\sigma_2\sigma_3...\sigma_{|A|} + ... + \sum_{cyc}\sigma_1 + 1, \end{aligned} \tag{13}$$

where $\sum_{cyc}$ is defined as the cyclic sum. Each term in the polynomial $f(A)$ represents a certain subset of A, and each term's coefficient represents the subset's number. Suppose the generating function after λ times of power splitting is:

$$f(A,\lambda) = \prod_{i=1}^{|A|}(\sigma_i+\lambda) = (\sigma_1+\lambda)(\sigma_2+\lambda)...(\sigma_{|A|}+\lambda). \tag{14}$$

As follows, mathematical induction is applied to prove Eq. (17). When $\lambda = 1$, according to Eq. (13), Eq. (17) is established. Suppose Eq. (13) is established when $\lambda = s-1$:

$$\begin{aligned} f(A,s-1) &= \prod_{i=1}^{|A|}(\sigma_i+s-1) \\ &= (\sigma_1+s-1)(\sigma_2+s-1)...(\sigma_{|A|}+s-1) \\ &= \sigma_1\sigma_2...\sigma_{|A|} + (s-1)\sum_{cyc}\sigma_2\sigma_3...\sigma_{|A|} + ...+ \\ &\quad (s-1)^{|A|-1}\sum_{cyc}\sigma_1 + (s-1)^{|A|}. \end{aligned} \tag{15}$$

Conduct one extra power splitting on $f(A, s-1)$, that is, let $\sigma_i = \sigma_i + 1$. Then Eq. (15) is changed into

$$\begin{aligned} &(\sigma_1+1)(\sigma_2+1)...(\sigma_{|A|}+1) + (s-1)\sum_{cyc}(\sigma_2+1)(\sigma_3+1)...(\sigma_{|A|}+1) \\ &+ ... + (s-1)^{|A|-1}\sum_{cyc}(\sigma_1+1) + (s-1)^{|A|}. \end{aligned} \tag{16}$$

By merging items of equal time in Eq. (16), we obtain

$$\begin{aligned} &(\sigma_1+1)(\sigma_2+1)...(\sigma_{|A|}+1) + (s-1)\sum_{cyc}(\sigma_2+1)(\sigma_3+1)...(\sigma_{|A|}+1) \\ &+ ... + (s-1)^{|A|-1}\sum_{cyc}(\sigma_1+1) + (s-1)^{|A|} \\ &= \sigma_1\sigma_2...\sigma_{|A|} + s\sum_{cyc}\sigma_2\sigma_3...\sigma_N + ... + s^{|A|-1}\sum_{cyc}\sigma_1 + s^{|A|} \\ &= (\sigma_1+s)(\sigma_2+s)...(\sigma_{|A|}+s) = f(A,s). \end{aligned}$$

Therefore, mathematical induction is established. When power splitting is conducted λ times to A, the generating function is

$$\begin{aligned} f(A,\lambda) &= \prod_{i=1}^{|A|}(\sigma_i+\lambda) = (\sigma_1+\lambda)(\sigma_2+\lambda)...(\sigma_{|A|}+\lambda) \\ &= \sigma_1\sigma_2...\sigma_{|A|} + \lambda\sum_{cyc}\sigma_2\sigma_3...\sigma_{|A|} + ... + \lambda^{|A|-1}\sum_{cyc}\sigma_1 + \lambda^{|A|}. \end{aligned}$$

Obviously, the sum of each coefficient in $f(A,\lambda)$ is $(\lambda+1)^{|A|}$, meaning the number of all elements is $(\lambda+1)^{|A|}$ after power splitting of A is conducted for λ times. Since the number of constant term is $\lambda^{|A|}$, meaning the number of $\emptyset$ is $\lambda^{|A|}$. Hence, the number of all elements except $\emptyset$ is

$$(\lambda+1)^{|A|} - \lambda^{|A|}.$$

Besides, for any A's subset B, the coefficient of B is

$$\lambda^{|A|-|B|}.$$

Based on the analysis above, hypothesis 1 is proven.

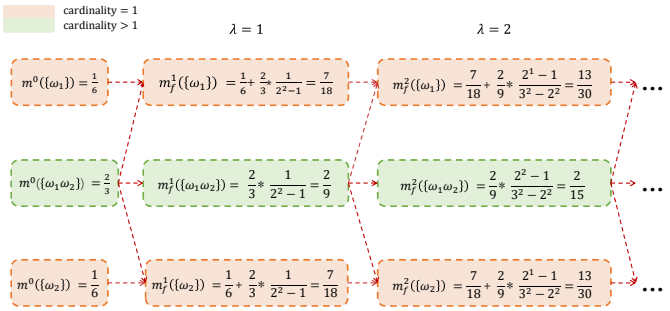


Fig. 2. flowchart of a dynamic fractal probability transformation in HoFBReD

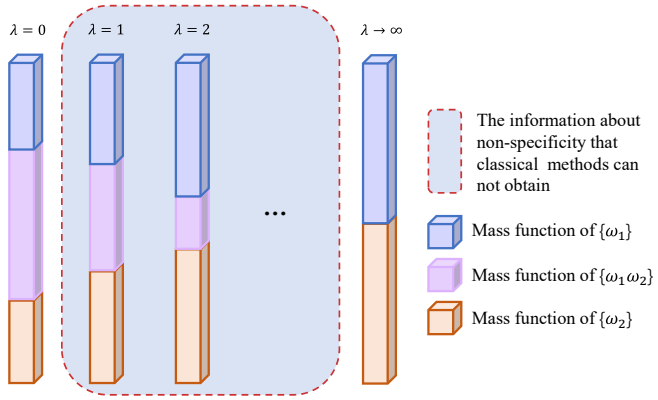


Fig. 3. The proportion of the mass function of $\{\omega_1\}$, $\{\omega_2\}$ and $\{\omega_1\omega_2\}$ in Example 3.1.

Apparently, the power splitting reflects the concept of self-similarity, which is the core of fractal theory [59]. Furthermore, the times of power splitting can be determined dynamically according to real cases. Consequently, a novel probability transformation method called dynamic fractal probability transformation is defined as follows.

Definition 9. *Dynamic fractal probability transformation*

Let a frame of discernment be Ω with a BBA m in Ω, dynamic fractal probability transformation of BBA m is defined as

$$\begin{aligned} m_f^{\lambda}(E_i) &= \sum_{E_i \subseteq E_j} \frac{\lambda^{|E_i|} - (\lambda-1)^{|E_i|}}{(\lambda+1)^{|E_j|} - \lambda^{|E_j|}} m_f^{\lambda-1}(E_j) \\ &= \sum_{E_i \subseteq E_j} \frac{\lambda^{|E_j|-|E_i|}}{(\lambda+1)^{|E_j|} - \lambda^{|E_j|}} m(E_j), \end{aligned} \tag{17}$$

where E_i is the element in 2^{Ω} except $\emptyset$; E_i is any subset of E_j; $|E_i|$ and $|E_j|$ denotes the cardinality of the element E_i and E_j; λ is the times of power splitting. For convenience, we name the BBA m_f^{λ} as the λth order of BBA m, and λ is called the order coefficient. It is obvious that m_f^0 is m itself.

On the basis of hypothesis 1, it is easy to find that in Eq. (17), when the order coefficient is λ, each of the mass function $m(E_j)$ is split into $(\lambda+1)^{|E_j|} - \lambda^{|E_j|}$ parts uniformly. Among these $(\lambda+1)^{|E_j|} - \lambda^{|E_j|}$ parts, totally $\lambda^{|E_j|-|E_i|}$ parts are distributed to $m_f^{\lambda}(E_i)$. In addition, when the order is added 1, each of the mass functions $m_f^{\lambda-1}(E_j)$ is spilled again and distributed to the powerset's mass functions of it. Among the $(\lambda+1)^{|E_j|} - \lambda^{|E_j|}$ parts that are split from $m(E_j)$, there are $\lambda^{|E_i|} - (\lambda-1)^{|E_i|}$ parts that should be distributed to $m(E_i)$ because there are $\lambda^{|E_i|} - (\lambda-1)^{|E_i|}$ elements that contain E_i.

Here, a numerical example illustrates the so-called dynamic fractal probability transformation. For a 2-element discernment framework $\Omega = \{\omega_1, \omega_2\}$, a BBA in Ω is defined as:

$$m : m(\{\omega_1\}) = \frac{1}{6}, \quad m(\{\omega_2\}) = \frac{1}{6}, \quad m(\{\omega_1\omega_2\}) = \frac{2}{3}.$$

Figure 2 shows the flowchart of dynamic fractal probability transformation of m in Example 3.1. As shown in Figure 2, the mass functions of multielement sets $\{\omega_1\omega_2\}$ are split, and distributed respectively to subsets of $\{\omega_1\omega_2\}$ except $\emptyset$, i.e., $\{\omega_1\}$, $\{\omega_2\}$ and $\{\omega_1\omega_2\}$. Additionally, with λ changing, the proportion of the mass function of $\{\omega_1\}$, $\{\omega_2\}$ and $\{\omega_1\omega_2\}$ is shown in Figure 3. Figure 3 shows that with dynamic fractal probability transformation, the mass function of $\omega_1\omega_2$ is gradually split and distributed to mass functions of ω_1 and ω_2. When $\lambda \to \infty$, the dynamic fractal probability transformation ends, and the BBA m is completely transformed into a probability distribution. This will be proved in the following content. Figure 3 also shows that dynamic fractal probability transformation can obtain more information about non-specificity because the classical transformation methods, such as PPT, can only obtain one result. However, dynamic fractal probability transformation can obtain more than one result while splitting the mass functions.

In this sub-section, dynamic fractal probability transformation is proposed. Based on dynamic fractal probability transformation, a novel belief divergence will be presented in the following sub-section.

3.2 A novel belief divergence and its properties

Definition 10. *Higher order fractal belief Rényi divergence*

Let m_1 and m_2 be two BBAs in 2^{Ω}. Higher order fractal belief Rényi divergence (HoFBReD) is defined as

$$D_{FB}^{(\alpha,\lambda)}(m_1||m_2) = \frac{1}{\alpha-1} \ln \sum_{i=1}^{2^N-1} m_{f_1}^{\lambda}(E_i)^{\alpha} m_{f_2}^{\lambda}(E_i)^{1-\alpha}, \tag{18}$$

and

$$\begin{aligned} m_{f_k}^{\lambda}(E_i) &= \sum_{E_i \subseteq E_j} \frac{\lambda^{|E_i|} - (\lambda-1)^{|E_i|}}{(\lambda+1)^{|E_j|} - \lambda^{|E_j|}} m_{f_k}^{\lambda-1}(E_j) \\ &= \sum_{E_i \subseteq E_j} \frac{\lambda^{|E_j|-|E_i|}}{(\lambda+1)^{|E_j|} - \lambda^{|E_j|}} m_k(E_j), \end{aligned} \tag{19}$$

where $k = 1, 2$; E_i is the element in 2^{Ω} except $\emptyset$; E_i is any subset of E_j; $|E_i|$ and $|E_j|$ denotes the cardinality of the element E_i and E_j; λ is the order coefficient here; It is obvious that $m_{f_k}^0$ is m_k itself. Since λ can certainly be larger than 1, which indicates that HoFBReD extends the order to a 'higher' order.

HoFBReD has several properties in terms of probability transformation (Property 1) as well as measurement (Properties 2-9), which are analyzed as follows.

Property 1. *When $\lambda \to \infty$, HoFBReD is equivalent to measuring the Rényi divergence between the pignistic probability transformations of m_1 and m_2:*

$$\lim_{\lambda\to\infty} D_{FB}^{(\alpha,\lambda)}(m_1||m_2) = \frac{1}{\alpha-1}\ln\sum_{i=1}^{N} BetP_1(\omega_i)^{\alpha} BetP_2(\omega_i)^{1-\alpha}.$$

Proof. Denote the pignistic probability transformations of m_1 and m_2 by $BetP_k$, $k = 1, 2$:

$$BetP_k(\omega_i) = \sum_{\omega_i\in E|E\in 2^{\Omega}} \frac{m_k(E)}{|E|}.$$

When the cardinality of E_i in Eq. (19) is 1, E_i is equivalent to ω_i. Hence, when $\lambda \to \infty$ and $|E_i|$=1,

$$\begin{aligned}\lim_{\lambda\to\infty} m_{f_k}^{\lambda}(E_i) &= \lim_{\lambda\to\infty}\sum_{E_i\subseteq E_j}\frac{\lambda^{|E_j|-|E_i|}}{(\lambda+1)^{|E_j|}-\lambda^{|E_j|}} m_k(E_j)\\ &= \lim_{\lambda\to\infty}\sum_{E_i\subseteq E_j}\frac{\lambda^{||E_j|-1|}}{(\lambda+1)^{|E_j|}-\lambda^{|E_j|}} m_k(E_j),\end{aligned}$$

where E_j are subsets of E_i. When $|E_i|$=1, obviously $|E_j| \geq 1$. Thus

$$\begin{aligned}\lim_{\lambda\to\infty} m_{f_k}^{\lambda}(E_i) &= \lim_{\lambda\to\infty}\sum_{E_i\subseteq E_j}\frac{\lambda^{|E_j|-1}}{(\lambda+1)^{|E_j|}-\lambda^{|E_j|}} m_k(E_j)\\ &= \lim_{\lambda\to\infty}\sum_{E_i\subseteq E_j}\frac{1}{\lambda(\frac{1}{\lambda}+1)^{|E_j|}-\lambda} m_k(E_j).\end{aligned}$$

When $\lambda \to \infty$, $\frac{1}{\lambda} \to 0$. Denote $\frac{1}{\lambda}$ by t, $t \to 0$. According to Taylor expansion:

$$(1+t)^{|E_j|} = 1 + |E_j|t + o(t),$$

where $o(t)$ is the equivalent infinitesimal of t. Then

$$\begin{aligned}\lim_{t\to 0} m_{f_k}^{\lambda}(E_i) &= \lim_{t\to 0}\sum_{E_i\subseteq E_j}\frac{1}{\frac{1}{t}(t+1)^{|E_j|}-\frac{1}{t}} m_k(E_j)\\ &= \lim_{t\to 0}\sum_{E_i\subseteq E_j}\frac{1}{\frac{1}{t}(1+|E_j|t+o(t))-\frac{1}{t}} m_k(E_j)\\ &= \lim_{t\to 0}\sum_{E_i\subseteq E_j}\frac{m_k(E_j)}{|E_j|}\\ &= \sum_{E_i\subseteq E_j}\frac{m_k(E_j)}{|E_j|}\\ &= BetP_k(E_i) = BetP_k(\omega_i)\\ &= \sum_{\omega_i\in E|E\in 2^{\Omega}}\frac{m_k(E)}{|E|}.\end{aligned}$$

Therefore, when $\lambda \to \infty$,

$$\begin{aligned}\lim_{\lambda\to\infty} m_{f_k}^{\lambda}(E_i) &= \sum_{\omega_i\in E|E\in 2^{\Omega}}\frac{m_k(E)}{|E|}\\ &= BetP_k(\omega_i),\end{aligned}$$

so

$$\lim_{\lambda\to\infty} D_{FB}^{(\alpha,\lambda)}(m_1||m_2) = \frac{1}{\alpha-1}\ln\sum_{i=1}^{N} BetP_1(\omega_i)^{\alpha} BetP_2(\omega_i)^{1-\alpha}.$$

Proof completed.

Property 2. *For $\alpha \in [0, \infty]$, when m_1 and m_2 degenerate to probability distributions $P = (p_1, p_2, \cdots, p_N)$ and $Q = (q_1, q_2, \cdots, q_N)$, HoFBReD degenerates to Rényi divergence:*

$$D_{FB}^{(\alpha,\lambda)}(m_1||m_2) = D_{\alpha}(m_1||m_2).$$

Proof. Obviously, when m_1 and m_2 degenerate to probability distributions P and Q:

$$\forall E_i \in \Omega, m_1(E_i) = p_i; m_2(E_i) = q_i, \tag{20}$$

and

$$\sum_{i=1}^{N} m_1(E_i) = \sum_{i=1}^{N} m_2(E_i) = 1.$$

It is obvious that

$$\begin{aligned}\forall E_i \in \Omega, m_{f_k}^{\lambda}(E_i) &= \sum_{E_i\subseteq E_j}\lambda^{|E_j|-|E_i|}\frac{m_k(E_j)}{(\lambda+1)^{|E_j|}-\lambda^{|E_j|}}\\ &= m_k(E_i).\end{aligned} \tag{21}$$

Therefore,

$$\begin{aligned}D_{FB}^{(\alpha,\lambda)}(m_1||m_2) &= \frac{1}{\alpha-1}\ln\sum_{i=1}^{N} m_{f_1}^{\lambda}(E_i)^{\alpha} m_{f_2}^{\lambda}(E_i)^{1-\alpha}\\ &= \frac{1}{\alpha-1}\ln\sum_{i=1}^{N} m_1^{\lambda}(E_i)^{\alpha} m_2^{\lambda}(E_i)^{1-\alpha}\\ &= \frac{1}{\alpha-1}\ln\sum_{i=1}^{N} p_i^{\alpha} q_i^{1-\alpha}\\ &= D_{\alpha}(P||Q).\end{aligned}$$

Proof completed. Eq. (20) and Eq. (21) will be directly applied in the following properties.

Property 3. *When $\alpha = 0$, and m_1 and m_2 degenerate to probability distributions P and Q, HoFBReD satisfies:*

$$D_{FB}^{(0,\lambda)}(m_1||m_2) = 0.$$

Proof. When $\alpha = 0$, and m_1 and m_2 degenerate to probability distributions P and Q, it can be calculated as:

$$\begin{aligned}D_{FB}^{(0,\lambda)}(m_1||m_2) &= -\ln\sum_{i=1}^{N} m_{f_2}^{\lambda}(E_i)\\ &= -\ln\sum_{i=1}^{N} m_2(E_i)\\ &= -\ln 1 = 0.\end{aligned}$$

Proof completed.

Property 4. *When $\alpha = \frac{1}{2}$, and m_1 and m_2 degenerate to probability distributions P and Q, HoFBReD is related to Hellinger distance.*

Proof. When $\alpha = \frac{1}{2}$, and m_1 and m_2 degenerate to probabil-

ity distributions P and Q, HoFBReD can be written as

$$\begin{aligned}
&D_{FB}^{(\frac{1}{2},\lambda)}(m_1||m_2)\\
&=-2\ln\sum_{i=1}^{N}\sqrt{m_{f_1}^{\lambda}(E_i)m_{f_2}^{\lambda}(E_i)}\\
&=-2\ln\sum_{i=1}^{N}\sqrt{m_1(E_i)m_2(E_i)}\\
&=-2\ln\sum_{i=1}^{N}\frac{(m_1(E_i)+m_2(E_i))-(\sqrt{m_1(E_i)}-\sqrt{m_2(E_i)})^2}{2}\\
&=-2\ln[1-\sum_{i=1}^{N}\frac{(\sqrt{m_1(E_i)}-\sqrt{m_2(E_i)})^2}{2}]\\
&=-2\ln[1-Hel^2(m_1,m_2)],
\end{aligned}$$

where

$$\begin{aligned}
Hel^2(m_1,m_2)&=\sum_{i=1}^{N}\frac{(\sqrt{m_1(F_i)}-\sqrt{m_2(F_i)})^2}{2}\\
&=\sum_{i=1}^{N}\frac{(\sqrt{p_i}-\sqrt{q_i})^2}{2}.
\end{aligned}$$

Proof completed.

Property 5. *When $\alpha=\frac{1}{2}$, HoFBReD is symmetric:*

$$D_{FB}^{(\frac{1}{2},\lambda)}(m_1||m_2)=D_{FB}^{(\frac{1}{2},\lambda)}(m_2||m_1).$$

Proof. When $\alpha=\frac{1}{2}$:

$$\begin{aligned}
D_{FB}^{(\frac{1}{2},\lambda)}(m_1||m_2)&=-2\ln\sum_{i=1}^{2^N-1}m_{f_1}^{\lambda}(E_i)^{\frac{1}{2}}m_{f_2}^{\lambda}(E_i)^{\frac{1}{2}},\\
D_{FB}^{(\frac{1}{2},\lambda)}(m_2||m_1)&=-2\ln\sum_{i=1}^{2^N-1}m_{f_2}^{\lambda}(E_i)^{\frac{1}{2}}m_{f_1}^{\lambda}(E_i)^{\frac{1}{2}}.
\end{aligned}$$

Thus,

$$D_{FB}^{(\frac{1}{2},\lambda)}(m_1||m_2)=D_{FB}^{(\frac{1}{2},\lambda)}(m_2||m_1).$$

Proof completed.

Property 6. *When $\alpha=1$, and m_1 and m_2 degenerate to probability distributions P and Q, HoFBReD degenerates to the Kullback-Leibler divergence:*

$$D_{FB}^{(1,\lambda)}(m_1||m_2)=D(P||Q).$$

Proof. L'hopital's Rule is applied:

$$\begin{aligned}
&\lim_{\alpha\to1}D_{FB}^{(1,\lambda)}(m_1||m_2)\\
&=\lim_{\alpha\to1}\frac{\ln\sum_{i=1}^{N}m_{f_1}^{\lambda}(E_i)^{\alpha}m_{f_2}^{\lambda}(E_i)^{1-\alpha}}{\alpha-1}\\
&=\lim_{\alpha\to1}\frac{\ln\sum_{i=1}^{N}m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}}{\alpha-1}\\
&=\lim_{\alpha\to1}\frac{\frac{\partial}{\partial\alpha}[\ln\sum_{i=1}^{N}m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}]}{\frac{\partial}{\partial\alpha}(\alpha-1)}\\
&=\lim_{\alpha\to1}\frac{\sum_{i=1}^{N}\ln[m_1(E_i)]m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}}{\sum_{i=1}^{N}m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}}\\
&-\lim_{\alpha\to1}\frac{\ln[m_2(E_i)]m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}}{\sum_{i=1}^{N}m_1(E_i)^{\alpha}m_2(E_i)^{1-\alpha}}\\
&=\sum_{i=1}^{N}m_1(E_i)(\ln m_1(E_i)-\ln m_2(E_i))\\
&=\sum_{i=1}^{N}p_i(\ln p_i-\ln q_i)\\
&=D(P||Q),
\end{aligned}$$

where the Kullback-Leibler divergence is defined as

$$D(P||Q)=\sum_{i=1}^{N}p_i\ln\frac{p_i}{q_i}.$$

Proof completed.

Property 7. *When $\alpha=2$, and m_1 and m_2 degenerate to probability distributions P and Q, HoFBReD is related to the $\mathcal{X}^2$ divergence:*

$$D_{FB}^{(2,\lambda)}(m_1||m_2)=\ln(\mathcal{X}^2(m_1||m_2)+1).$$

Proof. When $\alpha=2$, and m_1 and m_2 degenerate to probability distributions P and Q, HoFBReD can be written as follows

$$\begin{aligned}
&D_{FB}^{(2,\lambda)}(m_1||m_2)\\
&=\ln\sum_{i=1}^{2^N-1}\frac{m_{f_1}^{\lambda}(F_i)^2}{m_{f_2}^{\lambda}(F_i)}\\
&=\ln\sum_{i=1}^{N}\frac{m_1(F_i)^2}{m_2(F_i)}\\
&=\ln\sum_{i=1}^{N}[2m_1(F_i)-m_2(F_i)+\frac{(m_1(F_i)-m_2(F_i))^2}{m_2(F_i)}]\\
&=\ln[\mathcal{X}^2(m_1||m_2)+1],
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{X}^2(m_1||m_2)&=\sum_{i=1}^{N}\frac{(m_1(F_i)-m_2(F_i))^2}{m_2(F_i)}\\
&=\sum_{i=1}^{N}\frac{(p_i-q_i)^2}{q_i}.
\end{aligned}$$

Proof completed.

Property 8. *For $\alpha\in(0,1)\cup(1,+\infty)$, the following equation is satisfied:*

$$D_{FB}^{(\alpha,\lambda)}(m_1||m_2)=\frac{\alpha}{1-\alpha}D_{FB}^{(1-\alpha,\lambda)}(m_2||m_1).$$

Proof. For $\alpha \in (0,1) \cup (1,+\infty)$, there is

$$\begin{aligned}
&\frac{\alpha}{1-\alpha} D_{FB}^{(1-\alpha,\lambda)}(m_2||m_1) \\
&= \frac{\alpha}{1-\alpha}\frac{1}{-\alpha} \ln \sum_{i=1}^{2^N-1} m_{f_2}^{\lambda}(E_i)^{1-\alpha} m_{f_1}^{\lambda}(E_i)^{\alpha} \\
&= \frac{1}{\alpha-1} \ln \sum_{i=1}^{2^N-1} m_{f_1}^{\lambda}(E_i)^{\alpha} m_{f_2}^{\lambda}(E_i)^{1-\alpha} \\
&= D_{FB}^{(\alpha,\lambda)}(m_1||m_2).
\end{aligned}$$

Proof completed.

Property 9. *When m_1=m_2, the value of HoFBReD is 0:*

$$D_{FB}^{\alpha}(m_1||m_2) = 0.$$

Proof. When m_1=m_2:

$$\begin{aligned}
D_{FB}^{(\alpha,\lambda)}(m_1||m_2) &= \frac{1}{\alpha-1} \ln \sum_{i=1}^{2^N-1} m_{f_1}^{\lambda}(E_i) \\
&= \frac{1}{\alpha-1} \ln \sum_{i=1}^{2^N-1} m_{f_2}^{\lambda}(E_i) \\
&= \frac{1}{\alpha-1} \ln 1 \\
&= 0.
\end{aligned}$$

Proof completed.

The properties above show that HoFBReD can be related to pignistic probability transformations when $\lambda \to \infty$ and have some critical properties in measurement. The following section will further show the advantages of HoFBReD by numerical examples.

4 NUMERICAL EXAMPLES

In this section, several specific numerical examples are presented to demonstrate the advantages of HoFBReD. When $\alpha = 1$, Rényi divergence will have degenerated into the well-known Kullback-Leibler divergence. Moreover, in the following examples, $\alpha = 1$ will be applied to the HoFBReD to show its advantages better.

Example 4.1. Let a frame of discernment be $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ with powerset 2^{Ω}, and let m_1 and m_2 be two BBAs:

$$\begin{aligned}
&m_1 : m_1(\{\omega_1\omega_2...\omega_{N-1}\}) = \frac{1}{3}, \quad m_1(\{\omega_2\omega_3...\omega_N\}) = \frac{2}{3}; \\
&m_2 : m_2(\{\omega_1\omega_2...\omega_{N-1}\}) = \frac{2}{3}, \quad m_2(\{\omega_2\omega_3...\omega_N\}) = \frac{1}{3}.
\end{aligned}$$

Example 4.1 indicates the advantage of HoFBReD that it can obtain more uncertain information than any other belief divergences. When N varies form 2 to 7, Figure 4(a) shows the performance of BJS, $\mathcal{B}$ and $\mathcal{RB}$ divergences between m_1 and m_2. In comparison, Figure 4(b) shows the performance of HoFBReD with the order coefficient λ varying from 0 to 7 when N varies from 2 to 7. Figure 4(c) shows the performance of HoFBReD with λ varying from 0 to 20 when N varies from 2 to 7.

As shown in Figure 4(a), when the values of N change from 2 to 7, the values of BJS divergence stay constant at 0.0817, and the values of $\mathcal{B}$ and $\mathcal{RB}$ divergences gradually decrease from 0.2858 to 0.1167. In Figure 4(b), when $\lambda = 0$, HoFBReD stays constant at 0.2310, which has exactly the same change trend as BJS divergence in Figure 4(a). Furthermore, when λ varies from 1 to 7, the values of HoFBReD gradually decrease with N varying from 2 to 7, which also has the same change trend as $\mathcal{B}$ divergence and $\mathcal{RB}$ divergence in Figure 4(a). Hence, when the values of λ change, the performance of HoFBReD is similar to the BJS, $\mathcal{B}$ and $\mathcal{RB}$ divergences. Besides, apparently, there are more curves of HoFBReD than other methods. From the analysis above, it can be summarized that HoFBReD can effectively obtain more information about the non-specificity of BBAs with robust representation, which is superior to other methods.

According to Eq. (19), the values of λ represent the times of splitting in the dynamic fractal probability transformation. So it can be seen that there are multiple curves of HoFBReD in Figure 4(b) and 4(c). In Figure 4(b), there are 8 curves of HoFBReD because λ varies from 0 to 7. In Figure 4(c), when the value of N is fixed, the curves of HoFBReD with λ changing present the dynamic fractal probability transformation as well.

From Example 4.1, it is illustrated that HoFBReD obtains more information about non-specificity compared to the other belief divergences due to its dynamic fractal probability transformation.

From both Figure 4(b) and 4(c), it is apparent that the values of HoFBReD decrease as λ increases. It shows that the discrepancy between BBAs m_1 and m_2 gets gradually smaller as the dynamic fractal probability transformation continues. This is intuitive because the non-specificity of BBAs decreases as the dynamic fractal probability transformation continues.

Example 4.2. Let a frame of discernment be $\Omega = \{\omega_1, \omega_2\}$ and let m_1 and m_2 be two BBAs in Ω:

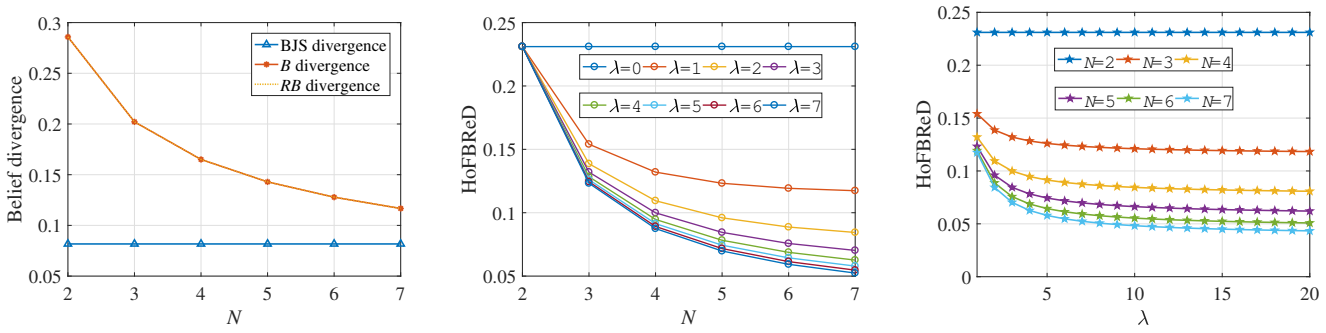
$$\begin{aligned}
&m_1 : m_1(\{\omega_1\}) = \frac{1}{2} - \frac{x}{2}, \quad m_1(\{\omega_2\}) = \frac{1}{2} - \frac{x}{2}, \quad m_1(\{\omega_1\omega_2\}) = x; \\
&m_2 : m_2(\{\omega_1\}) = \frac{1}{2} - \frac{y}{2}, \quad m_2(\{\omega_2\}) = \frac{1}{2} - \frac{y}{2}, \quad m_2(\{\omega_1\omega_2\}) = y.
\end{aligned}$$

As shown in Figure 5, the HoFBReD with different λ provides different performance. When the difference between the absolute values of x and y is small, the value of HoFBReD is relatively small. On the other hand, when the difference between the absolute values of x and y becomes more prominent, the value of HoFBReD becomes relatively large, which is intuitive. The maximum value of HoFBReD becomes smaller as λ increases, which also illustrates the conclusion in Example 4.1 that the discrepancy of BBAs m_1 and m_2 becomes gradually smaller as the dynamic fractal probability transformation process goes on. It illustrates that HoFBReD can effectively measure the discrepancy between BBAs with robust representation.

Example 4.3. Considering the condition of Example 1.1:

$$\begin{aligned}
&m_1 : m_1(\{\omega_1\}) = \frac{1}{2}; m_1(\{\omega_2\}) = \frac{1}{2}; \\
&m_2 : m_2(\{\omega_1\omega_2\}) = 1; \\
&m_3 : m_3(\{\omega_1\omega_2\omega_3\}) = \frac{1}{2}; m_3(\{\omega_2\omega_3\omega_4\}) = \frac{1}{2}; \\
&m_4 : m_4(\{\omega_1\omega_2\omega_3\omega_4\}) = 1.
\end{aligned}$$

In Example 1.1, the whole conflicting factor in D-S theory (denoted by $\mathcal{K}$), BJS divergence (denoted by BJS), $\mathcal{B}$ divergence (denoted by $\mathcal{B}$) and $\mathcal{RB}$ divergence (denoted by $\mathcal{RB}$)

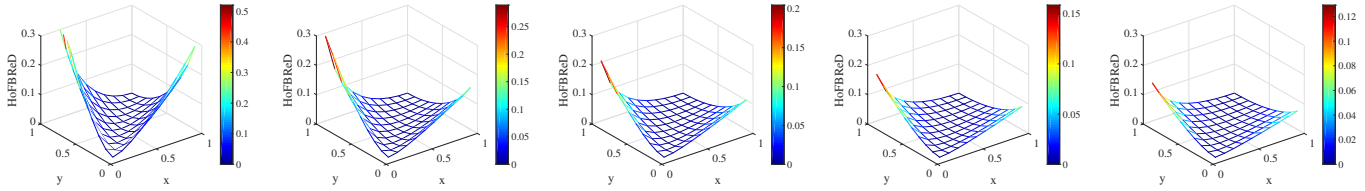


(a) Performance of BJS, $\mathcal{B}$ and $\mathcal{RB}$ divergences between m_1 and m_2.

(b) The performance of HoFBReD with the order coefficient λ varying from 0 to 7 when N varies from 2 to 7.

(c) The performance of HoFBReD with λ varying from 0 to 20 when N varies from 2 to 7.

Fig. 4. Comparison between HoFBReD and other belief divergences in Example 4.1.



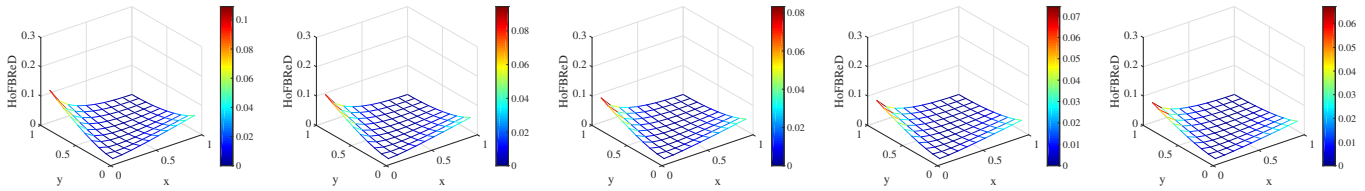
(a) HoFBReD between m_1 and m_2 with $\lambda = 1$.

(b) HoFBReD between m_1 and m_2 with $\lambda = 2$.

(c) HoFBReD between m_1 and m_2 with $\lambda = 3$.

(d) HoFBReD between m_1 and m_2 with $\lambda = 4$.

(e) HoFBReD between m_1 and m_2 with $\lambda = 5$.



(f) HoFBReD between m_1 and m_2 with $\lambda = 6$.

(g) HoFBReD between m_1 and m_2 with $\lambda = 7$.

(h) HoFBReD between m_1 and m_2 with $\lambda = 8$.

(i) HoFBReD between m_1 and m_2 with $\lambda = 9$.

(j) HoFBReD between m_1 and m_2 with $\lambda = 10$.

Fig. 5. HoFBReD between m_1 and m_2 with λ varying from 1 to 10 in Example 4.2.

TABLE 2
Different discrepancy measurement methods in Example 4.3.

BBAs	$\mathcal{K}$	BJS	$\mathcal{B}$	$\mathcal{RB}$	Bχ^2	d$_{BBA}$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
m_1 and m_2	0	1	0.2688	0.5185	1	0.5000	1.0986	0.9163	0.8473
m_2 and m_3	0	1	-0.0118	0.1088	1	0.5401	0.8473	0.6419	0.5664
m_3 and m_4	0	1	0.1957	0.4424	1	0	0.7621	0.5368	0.4553

can not effectively measure the discrepancy between m_1 and m_2, m_2 and m_3 and m_3 and m_4. However, as shown in Table 2, HoFBReD effectively measures the discrepancy between m_1 and m_2, m_2 and m_3 and m_3 and m_4. Example 4.3 shows that the performance of HoFBReD is superior to other methods. The order coefficient λ are selected as 1, 2 and 3 (denoted by $\lambda = 1$, $\lambda = 2$ and $\lambda = 3$). From Example 4.3, it can be concluded that HoFBReD can effectively obtain more information about BBA's non-specificity than other methods with robust representation.

5 A NOVEL MULTISOURCE INFORMATION FUSION ALGORITHM

This section proposes a novel multisource information fusion method, namely higher order fractal belief Rényi divergence-based weighted multisource information fusion (HOFBReD-WMSIF). The HOFBReD-WMSIF can provide good performance in pattern classification problems.

Considering a R-dimensional dataset $\mathcal{X}$ in class $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$. Each dimension of attributes $\mathcal{X}^r$ ($r \in \{1, 2, ..., R\}$) independently trains a classifier $\mathcal{C}_r$. For a certain pattern $x \in \mathcal{X}$, the classifier $\mathcal{C}_r$ will provide a probability distribution or a BBA as the output. The classification result given from an attribute is called a sub-classification result. A discounted combination of the R sub-classification

results improves the final classification accuracy. The discounting indexes are determined by two factors: classifier reliability and the attribute's contribution. These two factors are respectively referred to as the reliability and contribution weights.

The HOFBReD-WMSIF is mainly comprised of three parts: Calculation of classifier reliability, evaluation of attributes' contribution, and the final combination of classifiers with decision-making. For convenience, the pseudo-code of the HOFBReD-WMSIF algorithm is provided in Algorithm 1. The flowchart of the HOFBReD-WMSIF algorithm is given in Figure 6.

Algorithm 1 Higher order fractal belief Rényi divergence-based weighted multisource information fusion

Input: Training set $\mathcal{X}_{training} = \{x_1, x_2, ..., x_L\}$ and testing set $\mathcal{X}_{testing} = \{x_1, x_2, ..., x_M\}$;
Output: Class label $\mathcal{L}(x)$
for $r = 1, r \leq R$ **do**
  Train the rth basic classifiers $\mathcal{C}_r$ based on the rth attribute of the training set $\mathcal{X}_{training}$.
end for
Obtain the optimal discounting vector as classifiers reliability using Eq. (23), Eq. (24) and Eq. (25);
for $r = 1, r \leq R$ **do**
  Calculate the Pearson correlation coefficient using Eq. (26);
end for
for $r = 1, r \leq R$ **do**
  Calculate the contributions of each attribute using Eq. (27);
end for
Calculate the discounting factors δ_r using Eq. (28);
Normalize the discounting factors using Eq. (29) and Eq. (30);
for $i = 1, i \leq M$ **do**
  Obtain the classification results of x_i with the trained classifiers;
  Discount the various classification results using Eq. (31);
  Combine the discounted BBAs using Eq. (32);
  Determine the class label of x_i using Eq. (35).
end for

5.1 Calculation of classifier reliability

In real pattern classification problems, the classifier obtained by global optimization may be unsuitable for a specific query pattern [60], [61]. Thus, the reliability of each classifier is different from the others, which should be reviewed and reset [62]. All patterns in the training set are applied in the calculation of classifier reliability. The reliability is obtained as follows:

Let the reliability of classifier $\mathcal{C}_r$ be β_r, and all the reliability values of R classifiers are expressed by a 1×R vector $\beta = [\beta_1, \beta_2, ..., \beta_R]$, subject to $\sum_{r=1}^{R} \beta_r = 1$, $\beta_r \in [0, 1]$. Let the training set be $\mathcal{X}_{training}$ and the testing set be $\mathcal{X}_{testing}$. If the output of the classifier $\mathcal{C}_r$ is a probability distribution, denote the output on the rth attribute of query pattern x by $p_r^x = [p_r^x(1), ..., p_r^x(N)]$. $p_r^x(i)$ is the possibility of the query pattern x belonging to the class ω_i, and $p_r^x(i) \in [0, 1]$ with $\sum_{i=1}^{N} p_r^x(i) = 1$. If the output of the classifier $\mathcal{C}_r$ is a BBA, denote the output on the rth attribute of query pattern x by $m_r^x = [m_r^x(1), ..., m_r^x(\Omega)]$. $m_r^x(i)$ is defined as the support degree of the query pattern x belonging to the ith element in 2^Ω except $\emptyset$, and $m_r^x(i) \in [0, 1]$ with $\sum_{i=1}^{2^N-1} m_r^x(i) = 1$. For the classifier working with probability distributions, the ground truth of the query pattern's class is denoted by $G^x = [G^x(1), ..., G^x(N)]$. Let the label of query pattern x be $\mathcal{L}(x)$, then

$$G^x(i) = \begin{cases} 1, if \mathcal{L}(x) = \omega_i \\ 0, otherwise. \end{cases} \tag{22}$$

For the classifier working with BBAs, the ground truth vector of the query pattern's class is denoted by $G^x = [G^x(1), ..., G^x(N), 0, ..., 0]$, where there are $2^N - 1 - N$ of 0 added to the G^x vector, which makes G^x a $1 \times (2^N - 1)$ vector. For example, for a 2-class problem, if query pattern x belongs to the first class, $G^x = [1, 0]$ if classifiers work with probability distributions and $G^x = [1, 0, 0]$ classifiers work with BBAs.

Based on the above, for classifiers that work with probability distributions, an optimal discounting vector β is calculated by

$$\beta = arg \min_{\beta} \left(\sum_{l=1}^{L} D_{FB}^{(\alpha,\lambda)}(G^l || \sum_{r=1}^{R} \beta_r p_r^l) \right), \tag{23}$$

where $l \in \{1, 2, ..., L\}$ denotes the index of the training pattern and there are L patterns in the training set $\mathcal{X}_{training}$ in total.

For classifiers that work with BBAs, an optimal discounting vector β is calculated by

$$\beta = arg \min_{\beta} \left(\sum_{l=1}^{L} D_{FB}^{(\alpha,\lambda)}(G^l || \sum_{r=1}^{R} \beta_r m_r^l) \right), \tag{24}$$

$$s.t. \sum_{r=1}^{R} \beta_r = 1. \tag{25}$$

Based on the sequential quadratic programming (SQP) method and the constraints linear constraint Eq. (25), the solutions of classical constrained nonlinear least optimization problem in Eq. (23) and Eq. (24) are calculated.

The optimization problem in Eq. (23) and Eq. (24) is to obtain the reliability vector β that makes the loss of the classifier's error minimized. HOFBReD is applied to measure the discrepancy between the ground truth of the lth query pattern G^l and the weighted sum result $\sum_{r=1}^{R} \beta_r p_r^l$ or $\sum_{r=1}^{R} \beta_r m_r^l$.

5.2 Evaluation of attributes' contribution

In this work, the reliability vector β is obtained by the optimal global process, which may not be suitable for some specific patterns. Besides, if the class labels of the patterns are sensitive to changes in attribute values, this attribute's contribution is more significant. On the other hand, the attribute's contribution is more negligible if the class labels of the patterns are not sensitive to changes in attribute values. Hence, evaluating the attributes' contribution will make the attributes that are influential to the class label play more critical roles in the final decision-making. Since many attribute values are continuous, Pearson coefficient [63] is applied to express the attributes' contribution. To be specific, let the rth attribute be $\mathcal{X}_r$ and the label of each pattern be $\mathcal{L}$. The Pearson correlation coefficient ρ_r is defined as

$$\rho_r = \frac{cov(\mathcal{X}_r, \mathcal{L})}{\sigma \mathcal{X}_r \sigma \mathcal{L}}, \tag{26}$$

where $cov(\mathcal{X}_r, \mathcal{L})$ denotes the covariance between $\mathcal{X}_r$ and $\mathcal{L}$, σ the standard deviation.

The contribution of the rth attribute is defined as

$$\gamma_r = \frac{|\rho_r|}{\sum_{i=1}^{R} |\rho_i|}, \tag{27}$$

where each attribute's contribution is constrained between 0 and 1.

5.3 Final combination of classifiers with decision making

After calculating classifier reliability and evaluating attributes' contribution, the R sub-classification results of the query pattern are globally combined for final decision-making.

Various classifiers provide different prediction results of the testing set $\mathcal{X}_{testing}$ supporting different classes, which may lead to high conflict if the BBAs are directly combined. To avoid this trend, discounting techniques are employed for weighing the impact of different sources of evidence separately. In this work, weights of different sources of evidence are not only relevant to the classifier reliability β but also the attributes' contribution $gamma$, and the discounting factor of the rth output is defined as

$$\delta_r = \beta_r \gamma_r. \tag{28}$$

The discounting factors are then normalized:

$$\hat{\delta_r} = \frac{\delta_r}{\max_j \delta_j}, \tag{29}$$

where

$$\max_j \delta_j = \max\{\delta_1, \delta_2, ..., \delta_R\}. \tag{30}$$

Shafer's discounting method [28] is employed here for the final combination of discounted BBAs. The discounted BBA m_r with discounting factor δ_r is denoted by $m_r^{\delta_r}$ and given by

$$\begin{cases} {m_r}^{\delta_r}(X) = \delta_r m_r(X), & \forall X \subsetneqq \Omega \\ {m_r}^{\delta_r}(\Omega) = \delta_r m_r(\Omega) + 1 - \delta_r, \end{cases} \tag{31}$$

where total ignorance $m_r(\Omega)$ plays a neutral role in the fusion process, which can efficiently manage the conflict among different classifiers.

It is worth noting that if the classifier provides probability distributions as the prediction results of the testing set $\mathcal{X}_{testing}$, it becomes a BBA after being discounted according to Eq. (31). The reason is that a certain probability distribution p_r can be regarded as a BBA m_r whose total ignorance $m_r(\Omega)$ is always zero and $\sum_{i=1}^{N} m_r(\omega_i) = 1$. After discounting, the value of total ignorance is greater than 0 unless $\delta_r = 0$; thus, it becomes a BBA. Thus, in HOFBReD-WMSIF, both classifiers working with probability distributions and BBAs are applicable.

After Shafer's discounting, the conflict is properly managed, and the discounted BBAs can finally be combined directly by Dempster's rule of combination, according to Eq. (6). For a certain query pattern x, the final combination result $\mathbf{m}$ is obtained after applying the Dempster's rule of combination $R-1$ times to fuse the R discounted BBAs:

$$\mathbf{m} = m_1^{\delta_1} \oplus m_2^{\delta_2} \oplus ... \oplus m_R^{\delta_R}. \tag{32}$$

After obtaining the final fusion result $\mathbf{m}$, the label $\mathcal{L}(x)$ of the query pattern x is then determined by $BetP(\mathbf{m})$. The class whose possibility is the highest will be determined as the label $\mathcal{L}(x)$.

$$\mathcal{L}(x) = arg \max_{\omega_i \in \Omega} \left(BetP(\mathbf{m}(\omega_i))\right). \tag{33}$$

On account of the effective discrepancy measurement of the HOFBReD, the HOFBReD-WMSIF algorithm can improve the decision level, which will be illustrated in the next section by a pattern classification experiment.

6 Applications in Pattern Classification

Pattern recognition aims to classify objects of interest into one of a number of categories or classes [64]–[66]. Multi-source information fusion is an effective method that can be applied in pattern classification [56], [67].

To validate that the HOFBReD-WMSIF algorithm is effective, we apply it to a pattern classification experiment. In the experiment, the HOFBReD-WMSIF algorithm is compared with several machine learning methods and conventional information fusion methods. Through cross-validation, the performance of the HOFBReD-WMSIF algorithm is verified to be effective in pattern classification problems using real-world data based on the common criteria of pattern recognition accuracy (in %). According to the result, HOFBReD-WMSIF provides the highest accuracy among all methods in this experiment.

6.1 Compared methods

The performance of HOFBReD-WMSIF is evaluated compared with several machine learning methods and conventional information fusion methods including K nearest neighbor algorithm (KNN), support vector machines (SVM), discriminant analysis (DA), decision tree (DT), evidential K nearest neighbor (EKNN), average fusion method (AF) and the simple majority voting method (MV). In KNN, the principle is that when a new pattern x is predicted, the class label of x is determined according to the class label of the nearest K patterns. In SVM, the basic idea is to solve the separation hyperplane that can correctly divide the training data set and has the most considerable geometric interval. DA determines the patterns to which the grouping and other multivariate variable information belong for discriminant grouping, according to the known information about one grouping variable and corresponding other multivariate variables of a certain number of patterns. The DT is a tree structure, in which each internal node represents a judgment on an attribute, each branch represents the output of a judgment result, and finally, each leaf node represents a classification result. AF calculates the average results of the base classifiers. MV calculates the average complex decision-making results of the base classifiers. For more details about the AF and MV methods, please refer to [68]. In comparison, HOFBReD-WMSIF applies the proposed belief divergence to obtain the classifier reliability. Based on the classifier reliability, the discounting factors are generated, and the discounted BBAs are combined. Belief $\mathcal{X}^2$ divergence [37] and Jousselme et al.'s evidence distance [38] ablation experiments in the same algorithmic

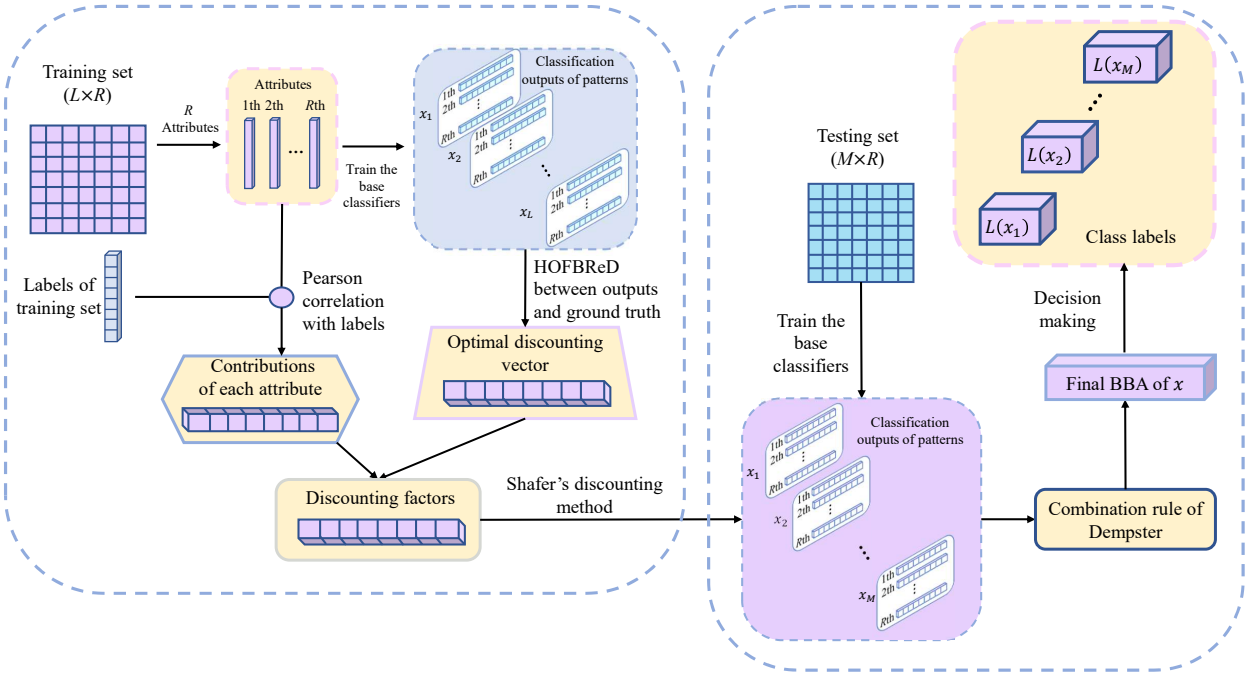


Fig. 6. The flowchart of the HOFBReD-WMSIF algorithm.

TABLE 3
The accuracy of different methods with $\lambda = 2$ with Bayesian classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	90.00±3.62%	69.31±11.19%	84.31±3.88%	89.31±4.58%	89.54±3.44%	84.77±3.90%	75.68±4.06%	**92.51**±2.91%	**92.51**±2.91%	89.77±3.68%
Cle	47.49±7.15%	**53.88**±10.14%	46.48±12.98%	37.72±10.58%	46.49±9.96%	**53.88**±10.14%	**53.88**±6.37%	53.15±6.67%	53.83±5.85%	53.21±11.08%
HCV	50.10±2.30%	48.73±4.45%	47.72±2.38%	49.74±3.55%	50.10±2.30%	50.25±1.49%	50.10±1.49%	**51.69**±1.96%	49.89±2.46%	49.81±1.92%
Ir	95.33±2.98%	96.66±2.35%	**98.00**±2.98%	94.66±3.80%	95.33±2.98%	96.00±2.78%	88.66±9.00%	93.33±2.35%	94.00±1.49%	95.33±1.82%
Trans	76.47±4.42%	73.53±4.88%	**76.87**±3.27%	72.19±5.29%	76.47±4.42%	**76.87**±3.43%	76.20±3.06%	76.47±4.39%	75.54±4.76%	75.80±2.75%
VC	95.48±2.10%	98.06±2.10%	91.93±3.42%	**99.35**±0.88%	94.51±2.44%	73.22±2.44%	87.09±1.61%	98.70±1.76%	98.70±1.76%	98.70±1.34%
Wh	88.86±3.25%	66.81±13.78%	82.50±4.86%	88.86±1.48%	88.63±3.11%	84.31±5.41%	75.90±6.03%	79.31±10.82%	77.04±4.91%	**89.77**±2.12%
AVG	78.97±3.22%	75.08±5.61%	75.79±4.42%	77.46±3.72%	78.66±3.65%	74.90±3.87%	73.27±5.39%	77.88±4.41%	77.36±3.45%	**79.34**±3.27%

TABLE 4
The accuracy of different methods with $\lambda = 3$ with Bayesian classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.93±0.49%	87.55±2.31%	86.94±1.99%	**99.84**±0.34%	98.93±0.49%	90.07±1.32%	78.39±3.91%	92.36±3.01%	92.36±3.01%	92.51±2.74%
Cle	47.49±6.22%	53.57±5.95%	48.16±6.58%	38.05±4.09%	43.11±5.61%	**53.91**±6.04%	**53.91**±6.04%	53.20±5.33%	53.87±4.50%	53.23±5.95%
HCV	48.08±2.77%	**50.90**±3.40%	49.16±1.03%	50.03±2.34%	48.08±2.77%	48.44±4.95%	48.59±3.87%	48.95±1.79%	47.58±2.04%	46.85±1.54%
Ir	96.66±2.35%	97.33±1.49%	**98.00**±1.82%	95.33±2.98%	96.66±2.35%	95.33±3.80%	88.66±6.05%	94.00±3.65%	94.00±3.65%	96.00±1.49%
Trans	75.66±2.60%	74.85±5.20%	**77.53**±2.38%	74.86±2.59%	75.79±2.54%	76.59±3.29%	76.19±3.18%	76.33±1.82%	75.53±1.23%	75.52±3.89%
VC	94.83±3.10%	98.06±0.72%	91.29±4.91%	99.67±0.72%	94.83±2.65%	72.90±4.32%	87.74±2.44%	98.70±1.34%	98.70±1.34%	**99.03**±0.88%
Wh	89.77±4.75%	57.72±26.88%	84.77±7.07%	88.40±3.44%	89.31±5.18%	85.45±5.04%	75.68±5.54%	75.00±2.27%	75.00±1.39%	**90.00**±3.15%
AVG	78.64±2.75%	75.69±4.37%	76.25±4.12%	78.14±2.20%	78.04±2.58%	74.94±4.06%	73.11±4.13%	76.93±2.74%	76.72±2.45%	**78.71**±2.70%

TABLE 5
The accuracy of different methods with $\lambda = 4$ with Bayesian classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.70±0.57%	87.02±3.13%	86.56±2.34%	**99.84**±0.20%	98.70±0.57%	89.77±1.27%	80.07±6.16%	92.29±1.24%	92.29±1.24%	92.21±1.58%
Cle	44.46±6.54%	**53.88**±8.22%	48.49±6.14%	40.40±8.61%	41.75±10.56%	**53.88**±8.22%	**53.88**±8.22%	53.53±7.03%	53.86±6.75%	53.20±8.74%
HCV	48.30±2.64%	**51.98**±1.48%	47.50±2.08%	45.55±4.33%	48.30±2.64%	49.74±2.03%	50.75±2.98%	49.45±3.61%	49.24±3.86%	49.09±1.59%
Ir	96.00±3.65%	**98.00**±2.98%	**98.00**±2.98%	94.00±2.78%	96.00±3.65%	94.00±2.78%	92.00±7.67%	94.66±1.82%	94.66±1.82%	96.00±4.34%
Trans	74.47±3.33%	75.26±2.70%	76.46±1.90%	75.53±1.48%	74.33±3.42%	**76.60**±1.81%	76.19±2.53%	76.33±4.54%	75.40±4.15%	75.26±2.61%
VC	93.87±3.85%	98.38±1.61%	94.19±3.71%	**99.67**±0.72%	93.22±3.85%	72.58±13.44%	86.77±3.85%	98.38±1.14%	98.38±1.14%	98.06±2.10%
Wh	89.31±3.55%	69.31±11.86%	83.86±0.95%	**90.22**±3.64%	88.86±3.96%	85.00±1.48%	74.77±5.41%	72.72±4.88%	75.00±3.21%	89.31±1.29%
AVG	77.87±3.45%	76.26±4.57%	76.44±2.87%	77.89±3.11%	77.31±4.09%	74.51±5.26%	73.49±4.60%	76.77±3.47%	76.97±3.17%	**79.02**±3.18%

TABLE 6
The accuracy of different methods with $\lambda = 5$ with Bayesian classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.93±0.68%	87.25±1.19%	86.87±1.31%	**99.92**±0.17%	98.93±0.68%	90.00±1.91%	81.52±6.94%	92.36±2.00%	92.36±2.00%	92.36±1.34%
Cle	45.14±8.72%	**53.91**±6.50%	50.22±9.47%	40.79±10.40%	44.47±8.43%	**53.91**±6.50%	**53.91**±6.50%	53.88±8.69%	53.88±8.69%	52.89±5.80%
HCV	48.73±1.42%	49.02±3.43%	47.22±1.56%	49.67±3.57%	48.73±1.42%	**50.54**±0.67%	50.32±0.54%	49.60±3.75%	48.88±1.92%	48.23±3.26%
Ir	96.66±2.35%	**98.00**±1.82%	**98.00**±1.82%	96.66±2.35%	96.66±2.35%	95.33±3.80%	90.00±6.23%	92.00±1.82%	92.66±2.78%	95.33±1.82%
Trans	76.07±2.70%	75.80±2.52%	**77.93**±3.23%	73.93±3.18%	76.07±2.94%	76.73±2.78%	76.19±1.74%	76.20±2.68%	75.40±2.75%	75.52±3.40%
VC	95.48±1.76%	98.38±1.14%	86.45±8.65%	**99.35**±0.88%	94.19±3.14%	71.93±8.58%	87.41±3.67%	98.70±1.34%	98.70±1.34%	98.70±1.34%
Wh	**90.00**±3.62%	69.31±11.19%	84.31±3.88%	89.31±4.58%	89.54±3.44%	84.77±3.90%	75.68±4.06%	83.18±8.24%	82.50±10.30%	89.77±3.68%
AVG	78.48±3.47%	76.05±4.25%	76.68±3.50%	78.23±4.18%	78.23±3.38%	75.15±3.89%	72.51±4.79%	77.99±4.07%	77.77±4.25%	**79.21**±2.96%

TABLE 7
The accuracy of different methods with $\lambda = 2$ with KNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.77±0.95%	87.32±3.20%	86.64±2.99%	99.84±0.34%	98.77±0.95%	93.20±3.11%	87.78±9.42%	99.69±0.49%	99.69±0.49%	**99.77**±0.51%
Cle	45.74±8.46%	53.79±12.34%	47.74±10.99%	36.95±11.22%	43.06±6.87%	53.79±12.34%	53.79±12.34%	53.21±4.86%	**53.88**±4.13%	53.79±12.34%
HCV	48.51±2.89%	51.26±2.46%	47.65±2.79%	48.95±4.16%	48.51±2.89%	48.15±4.14%	50.32±3.35%	51.33±2.59%	50.68±3.24%	**53.93**±3.33%
Ir	96.66±3.33%	97.33±2.78%	**98.00**±2.98%	94.66±2.98%	96.66±3.33%	92.00±2.98%	86.00±6.41%	95.33±3.80%	94.66±4.47%	96.00±2.78%
Trans	75.53±2.95%	76.46±1.74%	**77.00**±1.43%	73.52±2.32%	75.67±2.87%	73.40±3.01%	71.93±3.06%	75.26±2.90%	75.26±2.90%	72.58±6.84%
VC	94.83±2.88%	98.06±1.34%	93.87±2.88%	**99.35**±0.88%	93.87±3.49%	75.48±5.51%	80.96±2.88%	98.70±1.34%	98.70±1.34%	98.70±0.72%
Wh	89.09±1.72%	73.63±6.88%	84.31±3.80%	88.40±3.44%	88.86±1.86%	88.40±1.24%	85.45±2.46%	73.18±7.30%	81.13±3.37%	**90.00**±2.46%
AVG	81.79±2.90%	79.53±4.04%	79.67±3.58%	80.51±3.02%	81.18±2.77%	77.84±4.03%	76.17±5.22%	70.77±2.95%	71.90±2.38%	**83.48**±3.66%

TABLE 8
The accuracy of different methods with $\lambda = 3$ with KNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.54±0.31%	87.63±1.28%	87.02±0.97%	**99.69**±0.31%	98.54±0.31%	93.43±1.55%	86.25±7.69%	99.46±0.57%	99.46±0.57%	**99.69**±0.41%
Cle	47.46±5.04%	**53.83**±9.05%	49.46±6.98%	42.77±6.07%	45.11±4.11%	**53.83**±9.05%	**53.83**±9.05%	53.53±8.73%	53.87±9.03%	**53.83**±9.05%
HCV	**51.26**±3.47%	49.96±4.14%	47.00±1.75%	50.75±3.24%	**51.26**±3.47%	48.88±2.57%	49.67±3.40%	49.96±1.85%	50.68±2.41%	48.73±2.77%
Ir	96.00±1.49%	96.66±4.08%	**98.00**±2.98%	94.00±2.78%	96.00±1.49%	94.00±5.47%	87.33±4.34%	94.00±3.65%	92.66±2.78%	93.33±3.33%
Trans	**77.67**±5.53%	73.65±5.03%	77.27±4.67%	75.40±4.82%	**77.67**±5.53%	73.40±3.18%	72.19±3.75%	75.79±4.54%	75.79±4.54%	73.52±7.58%
VC	94.19±4.64%	98.38±2.28%	89.35±5.30%	**99.35**±0.88%	92.90±5.30%	77.09±2.10%	81.29±5.30%	99.35±0.88%	99.35±0.88%	99.03±1.44%
Wh	**90.00**±3.53%	63.18±18.63%	85.45±4.43%	89.77±2.89%	89.77±3.40%	86.13±4.97%	85.00±4.04%	74.54±4.06%	85.90±4.14%	88.86±2.03%
AVG	79.30±3.43%	74.76±6.35%	76.22±3.87%	78.82±3.00%	78.75±3.37%	75.25±4.13%	73.65±5.37%	78.09±3.47%	**79.67**±3.48%	79.57±3.80%

TABLE 9
The accuracy of different methods with $\lambda = 4$ with KNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	89.54±1.86%	64.09±12.87%	83.63±6.35%	88.18±2.35%	89.09±1.29%	86.13±2.94%	85.00±3.15%	99.69±0.31%	99.69±0.31%	**90.00**±2.32%
Cle	45.44±6.79%	**53.88**±5.01%	49.52±7.58%	40.09±10.93%	40.73±5.06%	**53.88**±5.01%	**53.88**±5.01%	53.52±6.25%	53.85±6.50%	**53.88**±5.01%
HCV	50.10±2.27%	49.67±2.37%	47.87±3.96%	49.67±1.52%	50.10±2.27%	48.01±3.40%	48.51±4.43%	51.33±2.92%	**51.62**±2.96%	50.68±1.00%
Ir	96.66±3.33%	96.66±3.33%	**98.00**±1.82%	94.00±4.94%	96.66±3.33%	90.00±5.27%	86.00±3.65%	92.00±1.82%	92.66±2.78%	94.00±2.78%
Trans	77.13±2.88%	65.40±21.07%	**78.34**±4.82%	73.79±3.00%	77.27±2.62%	74.06±2.14%	72.46±2.68%	75.52±4.34%	75.52±4.34%	72.84±7.36%
VC	93.87±1.76%	98.38±1.97%	94.51±3.71%	**99.35**±0.88%	93.22±3.49%	77.41±7.30%	82.90±5.17%	99.03±2.16%	99.03±2.16%	99.03±0.88%
Wh	89.31±4.44%	62.50±19.30%	84.77±2.96%	89.09±3.27%	88.86±3.96%	86.81±3.81%	85.22±2.89%	71.36±4.91%	72.50±11.59%	**90.22**±5.05%
AVG	78.75±3.13%	73.43±7.89%	77.10±3.87%	77.97±3.53%	77.93±3.02%	74.96±4.12%	73.64±4.48%	77.49±3.24%	77.84±4.38%	**80.05**±3.20%

TABLE 10
The accuracy of different methods with $\lambda = 5$ with KNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.93±0.49%	87.55±1.76%	86.64±2.01%	**99.84**±0.20%	98.93±0.49%	92.67±2.50%	89.08±7.17%	99.61±0.27%	99.61±0.27%	99.69±0.17%
Cle	45.45±5.59%	**53.88**±3.32%	47.46±5.38%	38.35±7.95%	43.75±5.53%	**53.88**±3.32%	**53.88**±3.32%	53.53±3.85%	53.87±3.71%	**53.88**±3.32%
HCV	49.74±1.60%	**50.97**±3.34%	50.75±1.43%	49.38±4.70%	49.74±1.60%	50.68±3.19%	51.91±4.29%	50.10±3.17%	51.69±4.18%	49.16±2.24%
Ir	94.66±4.47%	96.66±2.35%	**98.00**±1.82%	94.66±3.80%	94.66±4.47%	91.33±7.67%	84.66±6.91%	92.00±1.82%	92.66±2.78%	93.33±6.23%
Trans	**77.67**±2.54%	65.39±20.47%	77.40±1.89%	73.13±3.59%	77.80±2.26%	74.60±3.12%	72.73±3.06%	75.53±2.23%	75.53±2.23%	75.39±2.30%
VC	94.19±3.34%	97.74±0.88%	92.25±1.34%	**99.35**±0.88%	92.90±4.50%	77.41±3.78%	81.29±3.53%	98.70±1.34%	98.70±1.34%	99.03±0.88%
Wh	89.54±1.86%	64.09±12.87%	83.63±6.35%	88.18±2.35%	89.09±1.29%	86.13±2.94%	85.00±3.15%	68.18±7.83%	73.86±13.25%	**90.00**±2.32%
AVG	78.60±2.84%	73.75±6.43%	76.59±2.89%	77.56±3.35%	78.12±2.88%	75.24±3.79%	74.08±4.49%	76.80±2.93%	77.98±3.96%	**80.07**±2.49%

TABLE 11
The accuracy of different methods with $\lambda = 2$ with EKNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.24±0.69%	87.25±2.13%	86.41±1.97%	99.69±0.49%	98.24±0.69%	93.58±1.95%	88.93±5.04%	**99.69**±0.17%	**99.69**±0.17%	**99.69**±0.49%
Cle	45.78±3.92%	**53.85**±3.65%	49.82±2.87%	44.79±7.13%	41.07±4.95%	**53.85**±3.65%	**53.85**±3.65%	52.81±5.91%	52.81±5.91%	**53.85**±3.65%
HCV	50.03±1.70%	50.54±3.39%	50.68±4.00%	51.98±2.64%	50.03±1.70%	50.83±2.27%	51.84±1.07%	49.67±1.93%	49.67±1.93%	**52.49**±1.95%
Ir	96.66±2.35%	99.33±1.49%	**98.00**±1.82%	94.66±4.47%	96.66±2.35%	94.00±3.65%	86.66±9.71%	94.00±3.65%	93.33±2.35%	94.66±1.82%
Trans	74.20±3.98%	58.47±23.76%	**77.81**±3.97%	72.46±5.02%	74.07±4.02%	70.58±3.62%	72.45±3.09%	75.13±2.66%	75.13±2.66%	75.41±4.18%
VC	95.16±1.97%	98.06±1.34%	91.93±1.97%	99.35±0.88%	94.19±1.83%	75.48±8.41%	81.61±5.65%	98.70±1.34%	98.70±1.34%	**99.03**±0.88%
Wh	89.54±2.71%	69.54±12.43%	85.22±3.40%	88.18±1.29%	89.31±2.61%	84.54±3.27%	84.54±3.27%	74.77±10.76%	62.27±2.82%	**90.45**±2.06%
AVG	78.52±2.47%	73.86±6.88%	77.12±2.86%	78.73±3.13%	77.65±2.59%	74.69±3.83%	74.27±4.50%	77.82±3.78%	75.94±2.47%	**80.80**±2.15%

TABLE 12
The accuracy of different methods with $\lambda = 3$ with EKNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.70±0.99%	87.48±1.58%	86.71±1.15%	**99.84**±0.20%	98.70±0.99%	93.35±1.54%	88.24±6.32%	99.69±0.31%	99.69±0.31%	99.69±0.41%
Cle	43.10±5.88%	**53.85**±4.05%	50.84±1.33%	38.75±8.19%	42.74±7.25%	**53.85**±4.05%	**53.85**±4.05%	52.81±5.91%	52.81±5.91%	**53.85**±4.05%
HCV	50.25±3.60%	47.00±1.36%	50.10±2.57%	51.69±3.26%	50.25±3.60%	51.33±5.02%	50.75±4.71%	49.67±1.93%	49.67±1.93%	**52.27**±4.75%
Ir	96.00±5.47%	**97.33**±4.34%	98.00±3.65%	94.00±6.41%	96.00±5.47%	92.00±2.98%	86.66±2.98%	94.00±1.49%	94.00±2.78%	94.66±4.47%
Trans	**76.87**±2.15%	75.26±2.34%	76.86±1.80%	76.47±2.71%	76.74±2.15%	70.19±4.58%	70.73±3.58%	75.27±2.45%	75.27±2.45%	75.93±2.70%
VC	94.83±2.10%	98.38±1.61%	90.64±4.75%	98.70±2.10%	94.51±1.83%	73.54±6.09%	86.77±3.85%	**98.70**±1.34%	**98.70**±1.34%	**98.70**±1.34%
Wh	88.86±5.10%	58.40±17.24%	84.31±6.83%	89.09±3.46%	88.63±4.75%	84.54±6.89%	75.68±7.85%	71.59±7.66%	62.50±6.52%	**89.54**±4.78%
AVG	78.37±3.61%	73.96±4.64%	76.78±2.89%	78.36±3.76%	78.22±3.72%	74.12±4.45%	72.95±4.76%	77.39±3.02%	76.09±3.05%	**80.66**±3.21%

TABLE 13
The accuracy of different methods with $\lambda = 4$ with EKNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.77±0.62%	87.25±2.29%	86.87±1.58%	**99.84**±0.20%	98.77±0.62%	91.83±1.13%	84.58±6.94%	99.69±0.31%	99.69±0.31%	99.69±0.17%
Cle	46.45±5.60%	**54.20**±2.97%	47.46±3.08%	39.37±7.60%	45.44±7.15%	53.87±3.06%	53.87±3.06%	52.81±5.91%	52.81±5.91%	53.87±3.06%
HCV	49.45±1.94%	50.10±2.18%	49.24±1.76%	51.40±2.56%	49.45±1.94%	49.67±2.80%	50.32±2.56%	49.67±1.93%	49.67±1.93%	**52.56**±0.74%
Ir	95.33±1.82%	99.33±1.49%	**98.00**±2.98%	94±2.78%	95.33±1.82%	90.00±5.27%	88.66±5.05%	94.66±3.80%	94.00±4.34%	94.66±3.80%
Trans	76.61±3.25%	75.14±4.64%	**77.67**±3.42%	71.78±2.15%	76.20±2.84%	71.12±3.55%	73.00±2.65%	75.26±2.54%	75.26±2.54%	73.01±9.29%
VC	94.19±4.35%	98.06±1.76%	93.22±4.17%	**99.35**±0.88%	94.19±4.20%	76.45±8.73%	81.29±5.30%	99.03±1.44%	99.03±1.44%	98.70±1.34%
Wh	88.63±1.39%	63.41±14.27%	84.77±4.07%	87.95±3.73%	88.63±1.39%	85.68±3.27%	85.68±3.27%	82.72±7.29%	62.04±3.64%	**89.54**±1.48%
AVG	78.49±2.71%	75.35±4.23%	76.75±3.01%	77.67±2.84%	78.29±2.85%	74.09±3.97%	73.91±4.12%	79.12±3.33%	76.07±2.65%	**80.29**±2.84%

TABLE 14
The accuracy of different methods with $\lambda = 5$ with EKNN classifier.

Dataset	KNN	SVM	DA	DT	EKNN	AF	MV	Bχ^2-MSIF	d-MSIF	Proposed
Bank	98.62±0.34%	87.48±1.89%	86.48±1.58%	**99.84**±0.34%	98.62±0.34%	93.51±1.70%	82.59±9.12%	99.77±0.20%	99.77±0.20%	99.61±0.53%
Cle	46.49±7.42%	53.54±5.91%	49.15±4.36%	37.41±8.62%	41.09±5.14%	**53.88**±5.94%	**53.88**±5.94%	52.81±5.91%	52.81±5.91%	**53.88**±5.94%
HCV	49.38±0.93%	48.59±1.60%	48.66±1.43%	**52.34**±2.74%	49.38±0.93%	51.40±1.10%	51.62±0.44%	49.67±1.93%	49.67±1.93%	50.18±2.35%
Ir	96.66±3.33%	96.66±5.77%	**97.33**±2.78%	96.00±3.65%	96.66±3.33%	96.00±3.65%	88.00±3.80%	94.00±2.78%	93.33±4.08%	96.00±1.49%
Trans	76.19±2.58%	69.35±15.71%	**77.67**±2.37%	75.92±2.78%	76.19±2.58%	71.25±1.95%	73.51±3.74%	75.93±3.03%	75.93±3.03%	75.79±1.16%
VC	94.19±3.71%	**99.35**±1.44%	92.25±4.17%	99.03±2.16%	92.90±4.20%	74.83±6.41%	87.09±5.46%	98.70±0.72%	98.70±0.72%	98.70±2.10%
Wh	89.77±2.12%	71.81±12.45%	85.00±4.71%	**90.22**±2.73%	89.54±2.46%	84.77±1.90%	84.77±1.90%	79.77±5.23%	64.54±9.98%	88.63±2.78%
AVG	78.76±2.56%	75.25±5.60%	76.65±2.68%	78.68±2.88%	77.77±2.38%	75.09±2.83%	74.49±3.80%	78.66±2.84%	76.39±3.70%	**80.40**±2.05%

framework are also applied to prove the effectiveness of the proposed divergence. These two methods are denoted as Jousselme et al.'s evidence distance multisource information fusion (d-MSIF) and belief $\mathcal{X}^2$ divergence multisource information fusion (B$\mathcal{X}^2$-MSIF) respectively. For the AF and MV methods, if the base classifier works with probability distributions, the final predicted class of x is determined by:

$$\mathcal{L}(x) = arg \max_{\omega_i \in \Omega}(p(\omega_i)). \tag{34}$$

And if the base classifier works with BBAs, the final predicted class of x is determined by:

$$\mathcal{L}(x) = arg \max_{\omega_i \in \Omega}(m(\omega_i)). \tag{35}$$

6.2 Base classifiers

EKNN is a very classical method of generating evidence [69], and by using EKNN, the effectiveness in improving the decision level of applying HoFBReD in HOFBReD-WMSIF can be fairly evaluated and compared with other methods including d-MSIF and $B\mathcal{X}^2$-MSIF. So in this work, the evidential K-nearest neighbor classifier (EKNN), the KNN classifier and the naive Bayesian classifier are utilized as

the base classifiers. KNN and naive Bayesian classifiers provide probability distributions as the output, and EKNN classifiers provide BBAs as the output. These classifiers can be directly applied to HOFBReD-WMSIF because HOFBReD can effectively measure the discrepancy between probability distributions or BBAs.

6.3 Real world datasets

Real-world datasets are employed in this experiment from the UCI repository (http://archive.ics.uci.edu/ml) and KEI-dataset (http://sci2s.ugr.es/keel/datasets.php). In this experiment, the datasets Bank (Bank) with 1310 instances, 4 attributes and binary classes, Cleveland (Cle) with 297 instances, 13 attributes and 5 classes, hepatitis C virus (HCV) with 1385 instances, 28 attributes and binary classes, Iris (Ir) with 150 instances, 4 attributes and 3 classes, transfusion (Trans) with 748 instances, 4 attributes and binary classes, VC (VC) with 310 instances, 6 attributes and binary classes, Wholesale (Wh) with 440 instances, 7 attributes and 3 classes are utilized.

6.4 Evaluation of performance

A 5-fold cross-validation is applied to verify the proposed algorithm's effectiveness. Besides, the order coefficient λ and the parameter α in HOFBReD are determined based on the scenario applications' commands. In HOFBReD-WMSIF, the order coefficient λ is selected to be 2, 3, 4 and 5 here, and the parameter α of HOFBReD is selected to be $\frac{1}{2}$. For each base classifier and order coefficient λ, 5-fold cross-validation of the 8 methods are conducted once on all datasets. After the 5-fold cross-validation, for the various 4-order coefficient λ and 3 base classifiers, the accuracies of the 8 methods are shown from Table 3 to Table 14, where the average accuracies among the datasets are also presented.

In this paper, the average performances of the proposed algorithm are evaluated to prove the effectiveness of the algorithm. From the results above, it is observed that the HOFBReD-WMSIF generally provides generally better results in comparison to other methods. Specifically, from Table 3 to Table 14, the average accuracies of HOFBReD-WMSIF among all datasets are higher than other methods in almost all cases.

For the Bayesian classifier with the order coefficient $\lambda = 2$, the proposed algorithm has the highest average accuracy (79.34%), while the highest accuracy among other methods only reaches 78.97%. For the Bayesian classifier with the order coefficient $\lambda = 3$, the proposed algorithm has the highest average accuracy (78.71%), while the highest accuracy among other methods only reaches 78.64%. For the Bayesian classifier with the order coefficient $\lambda = 4$, the proposed algorithm has the highest average accuracy (79.02%), while the highest accuracy among other methods only reaches 77.89%. For the Bayesian classifier with the order coefficient $\lambda = 5$, the proposed algorithm has the highest average accuracy (79.21%), while the highest accuracy among other methods only reaches 78.48%.

For the KNN classifier with the order coefficient $\lambda = 2$, the proposed algorithm has the highest average accuracy (83.48%), while the highest accuracy among other methods only reaches 81.79%. For the KNN classifier with the order coefficient $\lambda = 3$, the proposed algorithm has the average accuracy (79.57%), while the highest accuracy among other methods reaches 79.67%. For the KNN classifier with the order coefficient $\lambda = 4$, the proposed algorithm has the highest average accuracy (80.05%), while the highest accuracy among other methods only reaches 78.75%. For the KNN classifier with the order coefficient $\lambda = 5$, the proposed algorithm has the highest average accuracy (80.07%), while the highest accuracy among other methods only reaches 78.60%.

For the EKNN classifier with the order coefficient $\lambda = 2$, the proposed algorithm has the highest average accuracy (80.80%), while the highest accuracy among other methods only reaches 78.73%. For the EKNN classifier with the order coefficient $\lambda = 3$, the proposed algorithm has the highest average accuracy (80.66%), while the highest accuracy among other methods only reaches 78.36%. For the EKNN classifier with the order coefficient $\lambda = 4$, the proposed algorithm has the highest average accuracy (80.29%), while the highest accuracy among other methods only reaches 79.12%. For the EKNN classifier with the order coefficient $\lambda = 5$, the proposed algorithm has the highest average accuracy (80.40%), while the highest accuracy among other methods only reaches 78.76%.

In this work, three traditional base classifiers are taken into account, including EKNN, KNN and naive Bayesian classifiers. In addition to these base classifiers, other base classifiers can also be applied in different applications accordingly. It is believed that HOFBReD-WMSIF can be applied to the improved EKNN methods as base classifiers, and gain the similar advantages of it. Therefore, HOFBReD-WMSIF can be effectively deployed to pattern classification problems, and provide high accuracy performance by recognizing effective patterns, such as the significant discrepancy between two probability distributions or BBAs. Based on the measuring of HOFBReD, it can benefit the classifiers' reliability and contribute to the good final fusion result performance in decision level.

7 Conclusion

This paper proposes a novel belief divergence to measure the discrepancy between BBAs in D-S evidence theory effectively. Furthermore, its application in pattern classification and multisource information fusion is demonstrated. The study's main contribution is that the proposed method is the first belief Rényi divergence that can measure the discrepancy between BBAs with dynamic fractal probability transformation and extends the order to a 'higher' order. And the proposed belief divergence can effectively obtain more information about the non-specificity of BBAs during discrepancy measurement with robust representation, which is superior to other methods. Additionally, a novel multisource information fusion algorithm is proposed, and a pattern classification experiment with real-world datasets is presented. The results show that the proposed method has a higher pattern recognition accuracy than other methods. To summarize, the proposed discrepancy measurement method and multisource information fusion algorithm contribute to improving the decision level.

In addition, the proposed HOFBReD-WMSIF also has a potential limitation. When the number of elements in the frame of discernment increases extremely, the time complexity of the HOFBReD-WMSIF spread exponentially, which is difficult for practical applications. In the future, it is worth considering quantum information processing technology [70] to deal with this limitation. And the HoFBReD can be further applied to complexity analysis for biological systems, EEG data analysis and deforestation focus detection in future work. For instance, we expect HoFBReD can identify the challenging patterns of different stages across a clinical cycle to prove the effectiveness of the HOFBReD-WMSIF algorithm, including migraine [71] and depression detections [72] based on EEG data.

ACKNOWLEDGMENTS

The author greatly appreciates the reviewers' suggestions and the editor's encouragement. This research is supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), Natural Science Foundation of Chongqing, China (No. CSTB2022NSCQ-MSX0531), Chongqing Overseas Scholars Innovation Program (No. cx2022024), and Fundamental Research Funds for the Central Universities (No. 2023CDJXY-035). In addition, this study is partially supported by Dr. Cao's Australian Research Council (ARC) DECRA Fellowship DE220100265.

REFERENCES

[1] R. R. Yager, "Multi-criteria decision making with interval criteria satisfactions using the golden rule representative value," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 1023–1031, 2017.

[2] Z.-g. Liu, Y.-m. Fu, Q. Pan, and Z.-w. Zhang, "Orientational distribution learning with hierarchical spatial attention for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8757 – 8772, 2022.

[3] R. R. Yager, "Ordinal scale based uncertainty models for ai," *Information Fusion*, vol. 64, pp. 92–98, 2020.

[4] Q. Wu, Y. Deng, and N. Xiong, "Exponential negation of a probability distribution," *Soft Computing*, vol. 26, no. 3, pp. 2147–2156, 2022.

[5] P. Liu, Y. Li, X. Zhang, and W. Pedrycz, "A multiattribute group decision-making method with probabilistic linguistic information based on an adaptive consensus reaching model and evidential reasoning," *IEEE Transactions on Cybernetics*, vol. 53, no. 3, pp. 1905–1919, 2022.

[6] R. Tao, Z. Liu, R. Cai, and K. H. Cheong, "A dynamic group MCDM model with intuitionistic fuzzy set: Perspective of alternative queuing method," *Information Sciences*, vol. 555, pp. 85–103, 2021.

[7] X. Deng, S. Xue, and W. Jiang, "A novel quantum model of mass function for uncertain information fusion," *Information Fusion*, p. DOI: 10.1016/j.inffus.2022.08.030, 2022.

[8] F. Xiao, "Quantum X-entropy in generalized quantum evidence theory," *Information Sciences*, p. DOI: 10.1016/j.ins.2023.119177, 2023.

[9] ——, "Generalized quantum evidence theory," *Applied Intelligence*, pp. DOI: 10.1007/s10 489–022–04 181–0, 2022.

[10] Y. Deng, "Random permutation set," *International Journal of Computers Communications & Control*, vol. 17, no. 1, p. 4542, 2022.

[11] J. Deng and Y. Deng, "Maximum entropy of random permutation set," *Soft Computing*, pp. 10.1007/s00 500–022–07 351–x, 2022.

[12] Y. Che, Y. Deng, and Y.-H. Yuan, "Maximum-entropy-based decision-making trial and evaluation laboratory and its application in emergency management," *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 7, pp. 1–16, 2022.

[13] J. W. Lai and K. H. Cheong, "Superposition of COVID-19 waves, anticipating a sustained wave, and lessons for the future," *BioEssays*, vol. 42, no. 12, p. 2000178, 2020.

[14] Z. Wang, Z. Li, R. Wang, F. Nie, and X. Li, "Large graph clustering with simultaneous spectral embedding and discretization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4426–4440, 2021.

[15] Z. Wang, C. Mu, S. Hu, C. Chu, and X. Li, "Modelling the dynamics of regret minimization in large agent populations: a master equation approach," in *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI-22)*, 2022, pp. 534–540.

[16] C. Chu, Y. Li, J. Liu, S. Hu, X. Li, and Z. Wang, "A formal model for multiagent q-learning dynamics on regular graphs," in *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI-22)*, 2022, pp. 194–200.

[17] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Scientific Data*, vol. 6, pp. DOI: 10.1038/s41 597–019–0027–4, 2019.

[18] Y. Song, Q. Fu, Y.-F. Wang, and X. Wang, "Divergence-based cross entropy and uncertainty measures of Atanassov's intuitionistic fuzzy sets with their application in decision making," *Applied Soft Computing*, vol. 84, p. 105703, 2019.

[19] M. Zhou, Y.-Q. Zheng, Y.-W. Chen, B.-Y. Cheng, E. Herrera-Viedma, and J. Wu, "A large-scale group consensus reaching approach considering self-confidence with two-tuple linguistic trust/distrust relationship and its application in life cycle sustainability assessment," *Information Fusion*, vol. 94, pp. 181–199, 2023.

[20] H. Liao, Z. Zhang, Z. Xu, and A. Banaitis, "A heterogeneous regret-theory-based method with Choquet integral to multiattribute reverse auction," *IEEE Transactions on Engineering Management*, vol. 69, no. 5, pp. 2248–2259, 2022.

[21] J. Zhan, H. Jiang, and Y. Yao, "Three-way multi-attribute decision-making based on outranking relations," *IEEE Transactions on Fuzzy Systems*, p. 10.1109/TFUZZ.2020.3007423, 2020.

[22] L. Ni, Y.-w. Chen, and O. de Brujin, "Towards understanding socially influenced vaccination decision making: An integrated model of multiple criteria belief modelling and social network analysis," *European Journal of Operational Research*, vol. 293, no. 1, pp. 276–289, 2021.

[23] D. Meng, S. Yang, C. He, H. Wang, Z. Lv, Y. Guo, and P. Nie, "Multidisciplinary design optimization of engineering systems under uncertainty: a review," *International Journal of Structural Integrity*, vol. 13, no. 4, pp. 565–593, 2022.

[24] D. Meng, H. Wang, S. Yang, Z. Lv, Z. Hu, and Z. Wang, "Fault analysis of wind power rolling bearing based on EMD feature extraction," *CMES-Computer Modeling in Engineering & Sciences*, vol. 130, no. 1, pp. 543–558, 2022.

[25] Z. Wang, D. Hou, C. Gao, J. Huang, and Q. Xuan, "A rapid source localization method in the early stage of large-scale network propagation," in *Proceedings of the ACM Web Conference (WWW-22)*, 2022, p. 1372.

[26] Y. Deng, "Uncertainty measure in evidence theory," *SCIENCE CHINA Information Sciences*, vol. 63, no. 11, p. 210201, 2020.

[27] A. Dempster, "Upper and lower probabilities induced by multivalued mapping, a. of mathematical statistics, ed," *AMS-38*, 1967.

[28] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976.

[29] Q. Shang, H. Li, Y. Deng, and K. H. Cheong, "Compound credibility for conflicting evidence combination: an autoencoder-K-Means approach," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, p. 10.1109/TSMC.2021.3130187, 2021.

[30] L. Xiong, X. Su, and H. Qian, "Conflicting evidence combination from the perspective of networks," *Information Sciences*, vol. 580, pp. 408–418, 2021.

[31] Z.-G. Liu, G.-H. Qiu, S.-Y. Wang, T.-C. Li, and Q. Pan, "A new belief-based bidirectional transfer classification method," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 8101–8113, 2021.

[32] D. Han, J. Dezert, and Y. Yang, "Belief interval-based distance measures in the theory of belief functions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 6, pp. 833–850, 2018.

[33] J. Huang, X. Song, F. Xiao, Z. Cao, and C.-T. Lin, "Belief f-divergence for eeg complexity evaluation," *Information Sciences*, vol. 643, p. 119189, 2023.

[34] Y. Huang and F. Xiao, "Higher order belief divergence with its application in pattern classification," *Information Sciences*, vol. 635, pp. 1–24, 2023.

[35] F. Xiao, "Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy," *Information Fusion*, vol. 46, pp. 23–32, 2019.

[36] ——, "A new divergence measure for belief functions in d–s evidence theory for multisensor data fusion," *Information Sciences*, vol. 514, pp. 462–483, 2020.

[37] L. Zhang and F. Xiao, "A novel belief $\chi 2$ divergence for multisource information fusion and its application in pattern classification," *International Journal of Intelligent Systems*, p. DOI: 10.1002/int.22912, 2022.

[38] A.-L. Jousselme, D. Grenier, and Éloi Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.

[39] L. Chen, Y. Deng, and K. H. Cheong, "Probability transformation of mass function: A weighted network method based on the ordered visibility graph," *Engineering Applications of Artificial Intelligence*, vol. 105, p. 104438, 2021.

[40] Q. Zhou and Y. Deng, "Fractal-based belief entropy," *Information Sciences*, vol. 587, pp. 265–282, 2022.

[41] X. Chen and Y. Deng, "An evidential software risk evaluation model," *Mathematics*, vol. 10, no. 13, p. 10.3390/math10132325, 2022.

[42] F. Xiao and W. Pedrycz, "Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. DOI: 10.1109/TPAMI.2022.3167045, 2022.

[43] L. Fei and Y. Wang, "An optimization model for rescuer assignments under an uncertain environment by using Dempster-Shafer theory," *Knowledge-Based Systems*, p. 109680, 2022.

[44] X. Xu, D. Zhang, Y. Bai, L. Chang, and J. Li, "Evidence reasoning rule-based classifier with uncertainty quantification," *Information Sciences*, vol. 516, pp. 192–204, 2020.

[45] S.-W. Tang, Z.-J. Zhou, C.-H. Hu, J.-B. Yang, and Y. Cao, "Perturbation analysis of evidential reasoning rule," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 8, pp. 4895–4910, 2021.

[46] C. Fu, B. Hou, M. Xue, L. Chang, and W. Liu, "Extended belief rule-based system with accurate rule weights and efficient rule activation for diagnosis of thyroid nodules," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 251–263, 2022.

[47] L. Chang, L. Zhang, C. Fu, and Y.-W. Chen, "Transparent digital twin for output control using belief rule base," *IEEE Transactions on Cybernetics*, p. DOI: 10.1109/TCYB.2021.3063285, 2021.

[48] Z.-W. Zhang, Z.-G. Liu, A. Martin, and K. Zhou, "Bsc: Belief shift clustering," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 3, pp. 1748–1760, 2022.

[49] H. Fujita and Y.-C. Ko, "A heuristic representation learning based on evidential memberships: Case study of UCI-SPECTF," *International Journal of Approximate Reasoning*, vol. 120, 2020.

[50] Y. Song and Y. Deng, "Entropic explanation of power set," *International Journal of Computers Communications & Control*, vol. 16, no. 4, p. 4413, 2021.

[51] F. Xiao, J. Wen, and W. Pedrycz, "Generalized divergence-based decision making method with an application to pattern classification," *IEEE Transactions on Knowledge and Data Engineering*, p. DOI: 10.1109/TKDE.2022.3177896, 2022.

[52] F. Xiao, "GEJS: A generalized evidential divergence measure for multisource information fusion," *IEEE Transactions on Systems, Man, and Cybernetics - Systems*, p. DOI: 10.1109/TSMC.2022.3211498, 2022.

[53] T. van Erven and P. Harremos, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[54] C. Zhu, F. Xiao, and Z. Cao, "A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis," *Information Sciences*, p. DOI: 10.1016/j.ins.2022.05.012, 2022.

[55] W. Fan and F. Xiao, "A complex Jensen-Shannon divergence in complex evidence theory with its application in multi-source information fusion," *Engineering Applications of Artificial Intelligence*, p. DOI: 10.1016/j.engappai.2022.105362, 2022.

[56] D. Li, Y. Deng, and K. H. Cheong, "Multisource basic probability assignment fusion based on information quality," *International Journal of Intelligent Systems*, vol. 36, no. 4, pp. 1851–1875, 2021.

[57] T. Wen and K. H. Cheong, "The fractal dimension of complex networks: A review," *Information Fusion*, vol. 73, pp. 87–102, 2021.

[58] Y. Deng, "Information volume of mass function," *International Journal of Computers Communications & Control*, vol. 15, no. 6, p. 3983, 2020.

[59] C. Qiang, Y. Deng, and K. H. Cheong, "Information fractal dimension of mass function," *Fractals*, vol. 30, p. 2250110, 2022.

[60] N. Masuyama, Y. Nojima, C. K. Loo, and H. Ishibuchi, "Multi-label classification via adaptive resonance theory-based clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2022.

[61] Q. Chen, F. Cao, Y. Xing, and J. Liang, "Evaluating classification model against bayes error rate," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2023.

[62] T. T. Mueller, J. C. Paetzold, C. Prabhakar, D. Usynin, D. Rueckert, and G. Kaissis, "Differentially private graph neural networks for whole-graph classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–11, 2022.

[63] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," 2009.

[64] F. Xiao, Z. Cao, and C.-T. Lin, "A complex weighted discounting multisource information fusion with its application in pattern classification," *IEEE Transactions on Knowledge and Data Engineering*, p. DOI: 10.1109/TKDE.2022.3206871, 2022.

[65] W. Yu, H. Wang, G. Li, N. Xiao, and B. Ghanem, "Knowledge-aware global reasoning for situation recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–13, 2023.

[66] S. Li, X. He, W. Song, A. Hao, and H. Qin, "Graph diffusion convolutional network for skeleton based sematic recognition of two-person actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–17, 2023.

[67] G. Costantini, D. Casali, and M. Carota, "A pattern classification method based on a space-variant cnn template," in *2006 10th International Workshop on Cellular Neural Networks and Their Applications*, 2006, pp. 1–5.

[68] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1605–1618, 2018.

[69] T. Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.

[70] H. He and F. Xiao, "A new quantum dempster rule of combination," *arXiv preprint arXiv:2304.14966*, 2023.

[71] Z. Cao, C.-T. Lin, K.-L. Lai, L.-W. Ko, J.-T. King, K.-K. Liao, J.-L. Fuh, and S.-J. Wang, "Extraction of ssveps-based inherent fuzzy entropy using a wearable headband eeg in migraine patients," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 1, pp. 14–27, 2020.

[72] Z. Cao, C.-T. Lin, W. Ding, M.-H. Chen, C.-T. Li, and T.-P. Su, "Identifying ketamine responses in treatment-resistant depression using a wearable forehead eeg," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1668–1679, 2019.