Deep Evidential Remote Sensing Landslide Image Classification With a New Divergence, Multiscale Saliency and an Improved Three-Branched Fusion

Jiaxu Zhang , Qi Cui , and Xiaojian Ma

Abstract—Hitherto, image-level classification on remote sensing landslide images has been paid attention to, but the accuracy of traditional deep learning-based methods still have room for improvement. The evidence theory is found efficient to boost the accuracy of neural networks, however, the present study argues three challenges that hinder the lead-in of this theory in deep landslide image classification. Aiming at the three problems, this study makes three improvements. For the interpretability and decision-invariance losses of three previous divergences, we propose a belief Jensen–Renyi divergence with properties proven. To couple the evidence theory with deep remote sensing landslide image classification, a channelwise multiscale visual saliency fusion is developed. We additionally find that the channelwise fusion is capable to reduce false recognition of networks as compared with original RGB images. To avoid decision failures in evidence-theoretic fusion process, we design an interpretability improved three-branched fusion. Experiments on Bijie Landslide dataset corroborate the synergistic benefits of the three improvements, where the proposal is compared with state-of-the-art image classification backbone networks, remote sensing image scene classifiers, evidence fusion algorithms, and versatile evidence-theoretic deep learning classifiers. We also evaluated the new method with two sort of image degradation, as well as an actual scenario in Luding County, China, whose data is publicly available.

Index Terms—Evidence theory, evidential deep learning, Jensen–Renyi (JR) divergence measure, multisource decision fusion, remote sensing landslide image classification.

I. Introduction

LANDSLIDE is a sort of natural geological hazard that occurs worldwide, primarily caused by rainfalls, crustal movements, volcanic eruption, human activities, and other synthetical factors [1]. Typically, the types of landslides include falls, topples, slides, spreads, and flows [2], to name a few, most of which have severe adverse effects on human life. According to Casagli et al. [3], from at least 2004 to 2016, the global total death toll from 4862 landslide events had approximated 56 000. Beyond any doubt, it has been an urgent request to monitor, detect, and classify landslides with guarantees of a high accuracy.

The landslide hazard can be fast recognized using visual surface morphology analysis through remote sensing technology [4]. Considering the computer-assisted recognition systems, deep learning technique has been increasingly accepted due to its highly performed "black-box" feature extraction, which is driven by data avoiding sophisticated manual design on geomorphologic feature descriptors [5]. From the subtasks of deep learning and computer vision, at least three manners have been investigated to link deep learning with autonomous landslide recognition, i.e.,

1) landslide target detection;
2) pixel-level or object-based segmentation;
3) image-level classification.

The landslide target detection originates from traditional object detection [6] in computer vision, which indicates to box-up potential landslide areas using bounding boxes from images, such as the radar, multispectral, and panchromatic images remotely captured by airborne or satellite sources [7]. Hou et al.[8] improved the YOLOX network to enhance the extraction of landslide features in optical remote sensing. Liu et al.[9] devised a SE-YOLOv7 network to improve detection precision. However, it is still an open question that how to find an optimal network structure for remote sensing landslide detection, especially when the background scenarios are intricate or the target landslide is small [10]. The pixel-level landslide segmentation aims to identify the pixels (or subpixels) belonging to landslide areas using distinguishable visual features, whereas object-based methods group up adjacent pixels to enhance the utility of spatial distribution information [11]. Some studies also refer to segmentation-based scheme as landslide detection or landslide semantic classification [12]. In this realm, Ji et al.[13] associated attention mechanism with convolutional neural networks for optical remote sensing landslide segmentation. Lv et al.[14] also proposed a ShapeFormer for the same task. Except the mentioned works, many state-of-the-art learning techniques, including multiclass classification [15] and few-shot learning [16],

Jiaxu Zhang is with the College of Science, Northeast Forestry University, Harbin 150000, China, and also with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 510275, China (e-mail: zhangjx283@mail2.sysu.edu.cn).

Qi Cui is with the College of Science, Northeast Forestry University, Harbin 150000, China, and also with the College of science, Northeastern University, Shenyang 110000, China (e-mail: cuiqi15998476980@126.com).

Xiaojian Ma is with the College of Science, Northeast Forestry University, Harbin 150000, China (e-mail: mxjzy@nefu.edu.cn).

Code is available online at https://github.com/defzhangaa/deeplandslideDS.

are also found effective. Nevertheless, pixel or object-level detection has a large computational burden, and Wang and Qiao, [17] pointed out in many scenarios, the accurate identification of landslide boundary is challenging. In contrast with pixel-level segmentation, image-level landslide classification is able to differentiate landslide and other nonlandslide scenes according to geological and visual features, such as object colors, textures, and topographic patterns, from the level of image labels. In this group, Defang et al.[18] investigated transfer learning for landslide image classification combining hybrid datasets. Fang et al.[19] presented a framework with generative adversarial networks. Among the three conventional computer vision tasks, from the view of remote sensing, the image-level classification is more fundamental [20], [21], [22], [23]. However, few of the mentioned image-level classification schemes, dominated by deep learning, though, are coupled with the well-known evidence theory to achieve a higher accuracy. The accuracy boosting effect of this theory in neural network sciences, especially in universal deep image-level classification, have been extensively discussed in very recent documents [24], [25], [26]. Nevertheless, the remote sensing images are distinct from natural images, since the former has a higher spatial resolution, and is often disturbed by object rotation and shape distortion [27], [28], [29]. For landslides, their scales and patterns are often undetermined, which can make them be falsely recognized as other objects, such as swamps, woods, and hills [10]. Herein, considering the importance of deep learning-boosted landslide classification, the present study considers the next problem that maintains relatively unstudied: The research community is not clear how and to what extent the evidence theory can boost the accuracy of deep learning in image-level remote sensing landslide classification. In consequence, to further investigate the mentioned issue, this article newly constructs an evidence-theoretic scheme, which will be introduced next. The evidence theory, also known as Dempster–Shafer's theory, is an advanced technique for intelligent decision support [30]. It has been exploited as a plug-and-play module in copious domains, such as machine learning [31], [32], [33], [34], object detection [35], and opinion aggregation [36]. This theory is sensitive to decision uncertainty [37], and specially tailors a "Dempster's combination rule" for multiple evidence fusion [38], [39]. On the basis of its merits, the present study investigates evidence theory-coupled deep learning in remote sensing landslide image classification. However, we still argue three challenges that hinder the lead-in of evidence theory.

1) Belief divergence challenge: Some belief divergences (BDs) may limit evidence-theoretic landslide image classification due to their explainability and decision-invariance losses.

2) Evidence theory involvement challenge: For landslide image classification, how to efficiently form evidence set via deep learning to implement evidence theory is not off-the-shelf.

3) Evidence conflict challenge: The Dempster's combination rule may bring about counterintuitive classification results when "evidence conflict" occurs.

For the first challenge, the present study mainly focuses on two BDs' intrinsic defects. Typically, the "BDs" refer to "the divergences in evidence theory" to distinguish them from the divergences in other theories [40]. By far, dozens of BDs, such as belief Jensen–Shannon divergence [41] and fractal belief Kullback–Leibler divergence [42], have brought about a tectonic change on evidence fusion. However, some BDs have inherent shortcomings [40]. The present study points out the belief Renyi divergence [43] and generalized belief Renyi divergence [44] can lead to explainability and decision-invariance losses on the newly developed evidence-theoretic system (before formally providing a solution, we set two analyses to illustrate the two mentioned losses at first). To address this issue, this article provides a multi-evidence remedy, i.e., the new belief Jensen–Renyi (BJR) divergence, which is more generalized than current solution [45]. With its properties justified, we perfectly settle the two mentioned losses. In addition, the ill-definition of its probabilistic prototype [46] is also addressed.

The second challenge can be unique to landslide image classification, but it will drastically affect the proposal's accuracy. Currently, at least two manners are majorly used to implement evidence theory into deep learning: One is to directly attach evidence-theoretic operations behind deep feature extractions and the other is to combine decisions of networks using evidence fusion. The first manner shows advantages, but its ability to handle multisource information is limited. For the second manner, a diversified evidence set is the prerequisite [47]. For instance, Xu et al.[48] formed an evidence set via multimodals; Tong et al.[49] combined heterogeneous datasets to generate such a set. Nevertheless, at least two modals or datasets are required in the two mentioned methods, limiting their real-world applications. In [50], the evidence set is obtained via multicolor spaces, but their simple color space conversion is adverse to network accuracy. The ideas of the authors in[51] and [52] may be efficient, but they are not designed for image classification. Then, this study presents a new design ethos from landslides that can avoid the listed issues, i.e., using channelwisely fused images by multiscale visual saliency to form the desired evidence set. We show that the geographical visual saliency from landslides is efficient to diversify networks' decisions, and can reduce network's false classification in comparison with original RGB images.

In the third challenge, "evidence conflict" indicates that the multievidence contains high conflict [53]. The evidence conflict is an intrinsic shortcoming of evidence theory, but in landslide image classification, it may result in false recognition. One of the solutions to overcome evidence conflict is to assign weights to the evidence. One can categorize the weighting-based methods into three groups [47]. The first group utilizes "belief entropy" or "BD" to form weight, like in[54]. The second group often incorporates "belief distance (divergence) + belief entropy", such as in [41], [55], [56], and [57]. The third group also shows efficiency, which may choose a "belief distance + belief entropy + impurity" strategy. Zhang et al. [47] investigated this approach and proposed a three-branched fusion model. However, Zhang et al.'s[47] model lacks theoretical explainability on involving
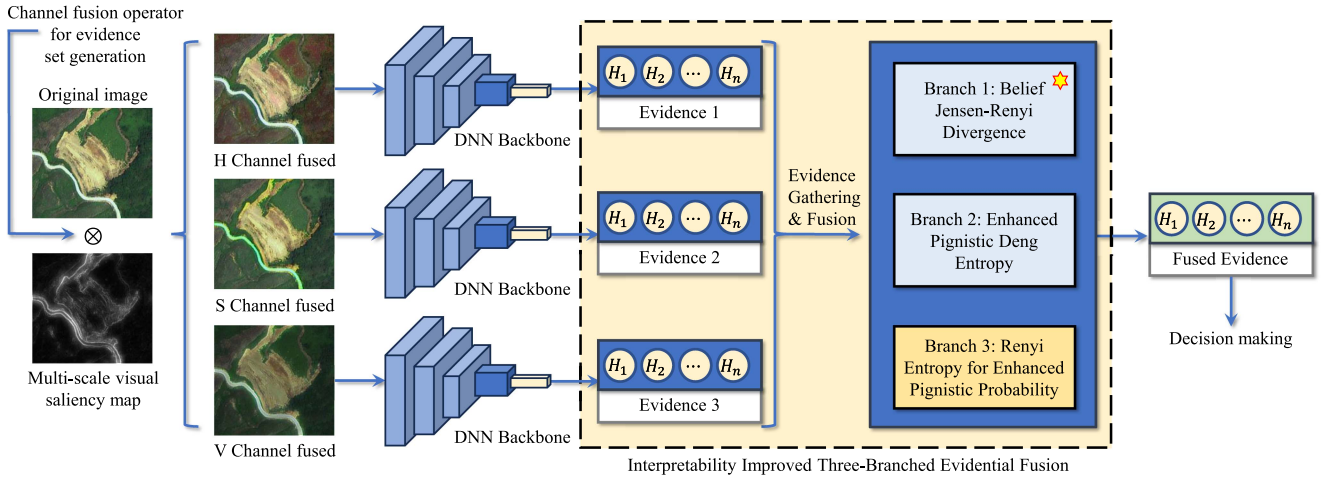
Fig. 1. Proposed deep evidence-theoretic remote sensing landslide image classification workflow with a BJR divergence and a new fusion model called interpretability improved three-branched evidential fusion. n denotes the size of the power set of FoD.

their "impurity" branch, making it fall in the degeneration of model interpretability. Motivated by this problem, we develop a theoretical explainable model, named as, interpretability improved three-branched fusion. The new model discards the original "impurity"-based structure, and is more competitive in landslide image classification.

Fig. 1 briefly displays the proposed method. Since the effect of evidence theory in deep remote sensing landslide image classification is relatively unclear, and when leading in the evidence theory, the three challenges still requires a prompt solution, we conduct the present research. In a nutshell, the contributions of this article are threefold.

1) The present study simultaneously solves three vital challenges, which block the lead-in of evidence theory on deep remote sensing landslide image classification.
2) We first verify that the evidence theory can boost accuracy of at least 11 state-of-the-art image classification backbone networks on classifying remote sensing landslide images.
3) Three improvements are made in terms of the three challenges. A BJR divergence, a channelwise evidence set generation strategy, and a new three-branched evidence fusion are proposed for the first time.

The rest of this article is organized as follows. Section II proposes the BJR divergence with properties analyzed. Section III describes the proposed algorithm. Section IV provides a theoretical comparison study, and in Section V, experiments are displayed. Finally, Section VI concludes this article.

II. THEORETICAL JUSTIFICATIONS: NEW BJR DIVERGENCE

Before the evidence theory is applied in the geotechnical image classification task, the present study starts from the theoretical analysis on explainability and decision-invariance losses, with which the new BJR divergence is proposed to manage the "BD" challenge.

A. Intrinsic Limitations of the Two Belief Renyi Divergences

The major motivation to develop the new divergence is to fix the "BD" challenge of belief Renyi divergence and generalized belief Renyi divergence via a multiBPAs' approach. Namely, the belief Renyi divergence is defined as follows [43]:

$$D_\alpha(m_1||m_2) = \frac{1}{\alpha - 1} \ln \left(\sum_{i=1}^n PBl_{m_1}^\alpha(\theta_i) PBl_{m_2}^{1-\alpha}(\theta_i) \right) \tag{1}$$

where $\alpha \in (0, 1) \cup (1, +\infty)$, $\theta_i \in \Theta$, Θ is the frame of discrimination (FoD) (the set containing all the hypothesis in evidence theory [47]), and $PBl_m(\theta_i) = \text{Bel}(\theta_i) + Pl(\theta_i)/\sum_{\theta_i \in \Theta} \text{Bel}(\theta_i) + Pl(\theta_i)$. The Bel and Pl functions can be found in [43] and [44].

And the generalized belief Renyi divergence is as follows [44]:

$$\text{GRD}_\alpha(m_1||m_2) = \frac{1}{\alpha - 1} \ln \left(\sum_{\Phi_j} \frac{m_1^\alpha(\Phi_j) m_2^{1-\alpha}(\Phi_j)}{2^{|\Phi_j|-1}} \right) \tag{2}$$

where $\alpha \in (0, 1) \cup (1, +\infty)$ and $|\Phi_j|$ is the cardinality of Φ_j.

Note that the two mentioned BDs are symmetric iff. $\alpha = 1/2$. This property is essential to their limitation analysis. To be precise, most of the BDs are symmetric between two pieces of evidence,[1] i.e., suppose SBD is a symmetric belief divergence, m_1 and m_2 are two arbitrary BPAs, then the next relationship holds

$$\text{SBD}(m_1, m_2) = \text{SBD}(m_2, m_1). \tag{3}$$

However, the mentioned two divergences are only symmetric on limited conditions as aforementioned (addressed as limited symmetric hereinafter), with which property the "BD" challenge, the explainability and decision-invariance losses will occur. Since

---

[1]Here, the multiBPAs based BDs, for instance, the generalized evidential Jensen–Shannon divergence $GEJS_\omega$ [55], are excluded.
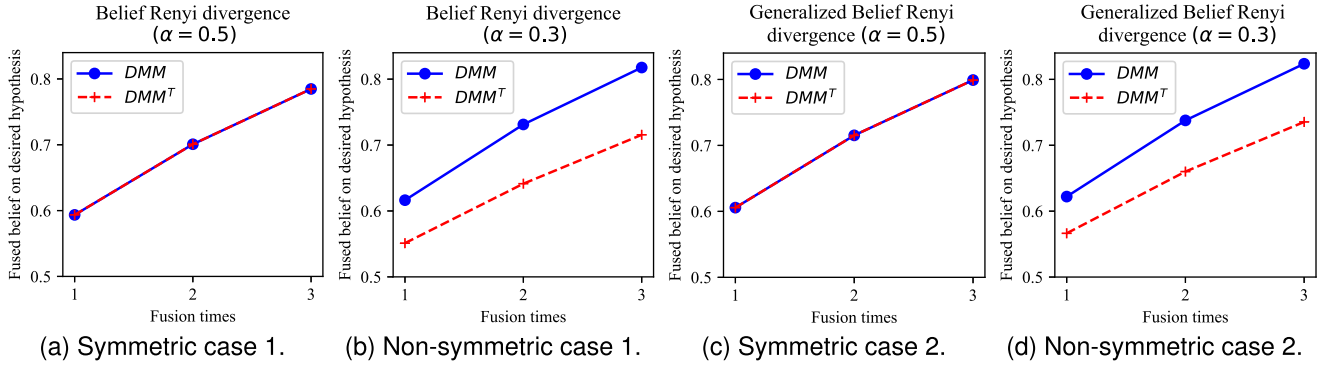
Fig. 2. Fused belief on the preferred hypothesis regarding the fusion times from (a) and (b) belief Renyi divergence [43] and (c) and (d) the generalized Renyi divergence [44]. Note that (a) and (c) are symmetric cases, whereas (b) and (d) are nonsymmetric cases. The abscissa is on behalf of the number of fusion times with Dempster's combination rule [38] when fusing the weighted evidence in [43] and [44]'s proposal. (a) Symmetric case 1. (b) Nonsymmetric case 1. (c) Symmetric case 2. (d) Nonsymmetric case 2.

TABLE I
FOUR SETS OF BPAS

Hypothesis	m_1	m_2	m_3	m_4
$\{\theta_1\}$	0.3337	0.0000	0.6699	0.6996
$\{\theta_2\}$	0.3165	0.9900	0.2374	0.2120
$\{\theta_3\}$	0.2816	0.0100	0.0884	0.0658
$\{\theta_1, \theta_2\}$	0.0307	0.0000	0.0000	0.0000
$\{\theta_1, \theta_3\}$	0.0052	0.0000	0.0000	0.0000
$\{\theta_2, \theta_3\}$	0.0271	0.0000	0.0043	0.0226
$\{\theta_1, \theta_2, \theta_3\}$	0.0052	0.0000	0.0000	0.0000

before this article, few prior studies have *both qualitatively and quantitatively* discussed this topic, without loss of generality the present study provides two short analyses as behind.

Analysis 1 (Explainability loss): This analysis mainly regards the explainability loss. Suppose, m_1 and m_2 are two arbitrary BPAs and a nonsymmetric belief divergence (NBD). Considering the nonsymmetry, it is rational to suppose the following relationship:

$$NBD(m_1, m_2) = 0.2 < NBD(m_2, m_1) = 0.9 \qquad (4)$$

whose semantic *can* be "the dissimilarity between evidence m_2 and m_1 is greater than m_1 and m_2," which is self-contradictory and illogical. This article believes that it is reasonable to regard this nonlogicality as an explainability loss when a NBD is applied.

Analysis 2 (Decision-invariance loss): This analysis focuses on the decision-invariance loss. A divergence measure matrix (DMM) among N sets of BPAs pertaining an arbitrary BD can be characterized as next

$$DMM_{N \times N} = \begin{cases} BD(m_i, m_j), & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \qquad (5)$$

where $i, j \in \{1, \ldots, N\}$. It is easy to derive that for the symmetric BDs, the transposition invariance holds, i.e., $DMM = DMM^T$. And for those who are nonsymmetric, since $BD(m_i, m_j) \neq BD(m_j, m_i)$ if $i \neq j$, there is $DMM \neq$

DMM^T. It means that if the divergence is nonsymmetric, the mentioned transposition invariance will vanish. This article finds that for the symmetric BDs, their fusion results will also keep invariant even if the matrix is transposed. But for the rest, their decision invariance will collapse. The decision invariance-loss will force the system to store the full matrix, which reduces storage efficiency.

We will use an example to illustrate the numerical effect of the mentioned decision-invariance loss. Suppose $m_1, \ldots, m_4$ are four BPAs on $\Theta = \{\theta_1, \theta_2, \theta_3\}$, which are listed in Table I. Then in line with the fusion algorithm in [43] and [44], use the mentioned belief Renyi divergence and generalized belief Renyi divergence to make decisions. For the symmetric case, $\alpha = 0.5$ is adopted, whereas for the nonsymmetric case, $\alpha = 0.3$ is chosen instead. The fused belief on the desired hypothesis $\{\theta_1\}$ regarding DMM and DMM^T have been shown in Fig. 2. It is interesting that in the nonsymmetric case the fusion indeed loses the decision invariance with respect to the matrix transformation on both the divergences, suggesting that the decision system is getting ill-posed, whereas in the opposite case this invariance is kept.

We still remember the belief Renyi divergence [43] and generalized BD [44] are two nonsymmetric divergences when $\alpha \neq 0.5$. For the at least two adverse effects of the nonsymmetry of divergence as we analyzed ahead, this article will try to fix their undesired limited symmetry. Two possible solutions are mainly considered.

i) Only choose $\alpha = 1/2$ when they are applied for information fusion, just as preconditioned in [43] and [44].
ii) To develop a BD that is symmetric on α's whole definition domain, like the method in [45].

But this study concerns that the approach i) may waste the definition domain of α. The method in [45] satisfies approach ii), but their proposal is symmetric within evidence pairs, less generalized than being symmetric among evidence set [55]. Fortunately, this article finds that the classical Jensen–Renyi (JR) divergence in (6) can be a potential candidate, which is symmetric among multiple BPAs. The JR divergence is defined

TABLE II
PROBABILITY DISTRIBUTIONS OVER SAMPLE SPACE Θ

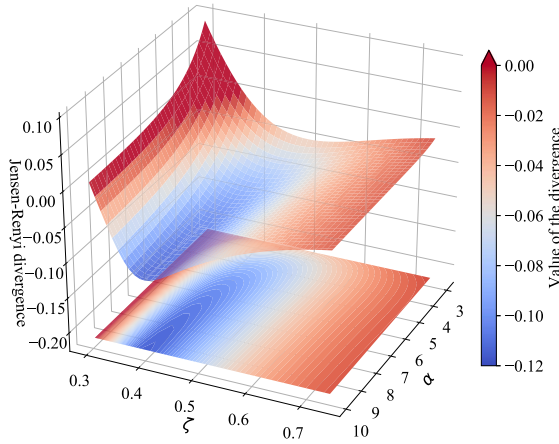| Distribution	θ_1	θ_2	θ_3
P_1	ζ	0.2500	0.7500-ζ
P_2	0.9725	0.0116	0.0159
P_3	0.9868	0.0014	0.0108

Fig. 3.   Surface and contour graphs of the classical JR divergence with an evenly distributed weight vector. In this figure, it is easy to derive that negative divergences indeed have occurred.

in the next form [46]

$$\text{JRD}_{\omega,\alpha}(P_1,\ldots P_N) = R_\alpha\left(\sum_{i=1}^{N} w_i P_i\right) - \sum_{i=1}^{N} w_i R_\alpha(P_i) \quad (6)$$

where $P_1, P_2, \ldots, P_N$ are N sets of discrete probability distributions, and their corresponding weight vector is $\omega = \{w_1, w_2, \ldots, w_N\}$ that satisfies $w_i > 0$, $\sum_{i=1}^{N} w_i = 1$, $\alpha > 0$, $\alpha \neq 1$, and $R_\alpha(\cdot)$ represents the Renyi entropy [46]

$$R_\alpha(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{u} p_i^\alpha\right). \quad (7)$$

Nevertheless, by demonstrating the next example, this study discovers the JR divergence can be negative when $\alpha > 1$, manifesting in certain conditions it may not be well-defined.

Proposition 1 (Ill-definition of JR divergence): When $\alpha > 1$, the traditional JR divergence [46] can be negative.

Proof: An example is exploited to proof this proposition. Suppose $\Theta = \{\theta_1, \theta_2, \theta_3\}$ is the sample space, P_1 (parameterized by ζ), P_2, and P_3 are three discrete probability distributions listed in Table II. To make the results explicit, this study sets $3 \leq \alpha \leq 10$, $0.30 \leq \zeta \leq 0.75$. Fig. 3 shows the corresponding JR divergences.

As can be derived in Fig. 3, negative divergences indeed appear. Thus, now the mentioned proposition is proven. ∎

Up to now, the present study has verified the two belief Jensen divergences may harm the theoretical explainability and decision invariance of fusion framework, which is the "BD" challenge. Besides, we find the classical JR divergence can be

ill-defined. Aiming at these issues, a new BJR divergence should be provided.

B. Proposed BJR Divergence Measure

Next, the BJR divergence is formally defined to address the "BD" challenge. To differentiate other Renyi entropy-like belief entropies [59], the Renyi entropy for the Enhanced Pignistic probability is formalized.

Definition 1 (Renyi entropy for the Enhanced Pignistic probability): Suppose Θ is the FoD, $H_j \in 2^\Theta$, then, the Renyi entropy for the Enhanced Pignistic probability is defined

$$R_\alpha(EBetP) = \frac{1}{1-\alpha} \log\left(\sum_{\Phi_j \in 2^\Theta} EBetP(\Phi_j)^\alpha\right) \quad (8)$$

where $\alpha > 0, \alpha \neq 1$ is the order, $EBetP$ is the Enhanced Pignistic probability that can be found in [58].

As an entropy for belief functions, it is clear that the newly defined Renyi entropy for the Enhanced Pignistic probability is not a disaggregated information measure, which inspects the randomness uncertainty only [60]. With this new entropy, the BJR divergence can be safely defined.

Definition 2 (BJR divergence): Suppose $m_1, m_2,..., m_N$ are N sets of BPAs, $\omega = \{w_1, w_2, \ldots, w_N\}$ is their corresponding weight vector that satisfies $w_i > 0$, $\sum_{i=1}^{N} w_i = 1$. Then, the JR divergence measure is formally defined as follows:

$$BJR_{\omega,\alpha}(m_1,\ldots,m_N)$$
$$= \exp\left\{R_\alpha\left(\sum_{i=1}^{N} w_i EBetP_i\right) - \sum_{i=1}^{N} w_i R_\alpha(EBetP_i)\right\} \quad (9)$$

where R_α is the Renyi entropy for Enhanced Pignistic probability, $\exp(\cdot)$ is the exponential function, $EBetP_i$ is the corresponding Enhanced Pignistic probability of $m_i, i \in \{1, \ldots, N\}$.

It is effortless to derive that the new BJR divergence can manage the limited symmetry of the two belief Renyi divergences, so the next theorem is provided without proof. This theorem points out that the new divergence is rearrangement invariant regarding ω within α's full definition domain, rather than $1/2$ solely, and is symmetric among multiple BPAs, satisfying our expectation.

Theorem 1 (Rearrangement invariance regarding ω): Unlike divergences in (1) and (2), suppose $m_{r_1}, m_{r_2}, ..., m_{r_N}$ is an arbitrary rearrangement of the BPAs $m_1, m_2, \ldots, m_N$, $\omega_r = \{w_{r_1}, w_{r_2}, \ldots, w_{r_N}\}$ is the corresponding rearrangement of ω. Then, for $\forall \alpha > 0, \alpha \neq 1$, the proposed BJR divergence is invariant with rearrangement

$$BJR_{\omega,\alpha}(m_1,\ldots,m_N) = BJR_{\omega_r,\alpha}(m_{r_1},\ldots,m_{r_N}). \quad (10)$$

In addition, the ill-definition of the JR divergence revealed in Proposition 1 has been fixed due to the $\exp(\cdot)$ operator, which is self-explanatory in the next theorem.

Theorem 2 (Nonnegativity): Unlike the divergence in (6), for $\forall \alpha > 0, \alpha \neq 1$, the proposed divergence is nonnegative.
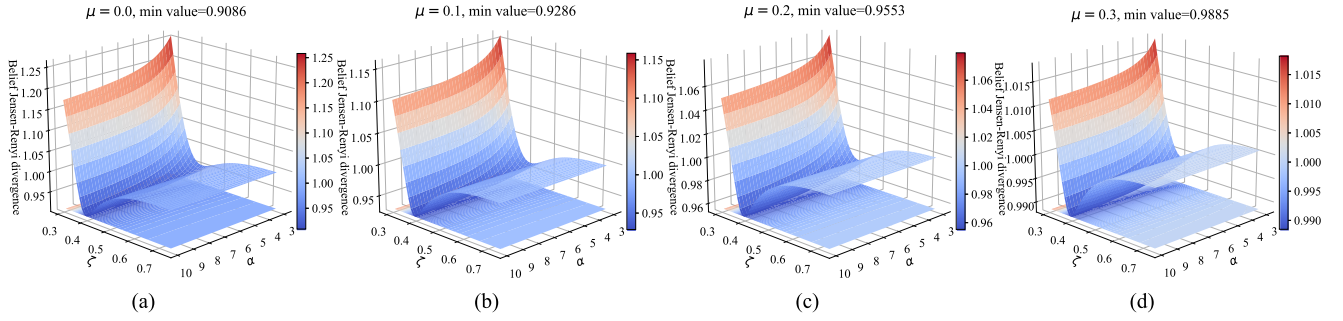
Fig. 4. Proposed BJR divergence in Example 1, where (a)–(d) are the surface graphs with μ =0.0, 0.1, 0.2, and 0.3, respectively.

The theorem of nonnegativity is pivotal and further verified via the next example. For a clear comparison with the probabilistic JR divergence, the data where negative JR divergences are discovered has just been retained.

Example 1 (Verification of nonnegativity): This example converts the three discrete probability distributions in Table II into BPAs with FoD $\Theta = \{\theta_1, \theta_2, \theta_3\}$ unchanged. α and ζ are the same with that table, and the weight vector is $\omega = \{1/3 - \mu, 1/3 + \mu, 1/3\}$, where μ is set as 0.0, 0.1, 0.2, and 0.3, respectively. Results of BJR divergence among the three BPAs have been shown in Fig. 4 .

It is obvious that in Fig. 4, no matter what the parameters μ and α are set, the minimum value of BJR divergence will always be greater than zero. In a significant contrast, the nonnegativity property of our proposal is thoroughly different with the traditional JR divergence, which has been proven in Proposition 1 and exemplified in the proof.

Then, except for the two desired properties, i.e., Theorems 1 and 2, two additional properties are also provided with proofs attached in Appendixes A and B, respectively.

Theorem 3 (Boundness): For $\forall \alpha > 0$, $\alpha \neq 1$, the BJR divergence is both lower and upper bounded.

Theorem 4 (Generalization): The BJR divergence can be considered as the generalization of the generalized evidential Jensen–Shannon ($GEJS_\omega$) divergence [55] on the exponent operator. When $\alpha \to 1$ and using the $\bar{m}$ function to replace the $EBetP$ function, the next relationship holds

$$\lim_{\alpha \to 1} BJR^*_{\omega,\alpha}(m_1, \ldots, m_N) = \exp \{GEJS_\omega(m_1, \ldots, m_N)\} \tag{11}$$

where $BJR^*_{\omega,\alpha}(m_1, \ldots, m_N)$ means replacing the $EBetP$ functions with $\bar{m}$ functions, which are

$$\bar{m}(\Phi_j) = \frac{m(\Phi_j)}{|\Phi_j|} \tag{12}$$

and $|\Phi_j|$ remarks the cardinality of Φ_j, and $\Phi_j \in 2^\Theta$.

This article believes the Theorem 4 is elegant since it uncovers that the new BD is coherently linked with the famous generalization circle of Jensen–Shannon divergence-like BDs, such as belief Jensen–Shannon divergence [41] and generalized Jensen–Shannon divergence [55].

Finally, when the evidence reliability is nonobservable, one can directly assign an average weight to the divergence, i.e., one can set $\omega = \{1/N, \ldots, 1/N\}_{1 \times N}$, where N denotes the number of evidence. This case will be used later, thus, it is useful to address the following corollaries for the proposed BD with an evenly distributed weighting vector.

Corollary 1 (Rearrangement invariance): The proposed BJR divergence is rearrangement invariant regarding N sets of arbitrary evidence with an evenly distributed weighting vector $\omega = \{1/N, \ldots, 1/N\}_{1 \times N}$.

Corollary 2 (Non-negativity, Boundness and Generalization): For $\forall \alpha > 0, \alpha \neq 1$, the BJR divergence with an evenly distributed weighting vector

1) is rearrangement invariant and non-negative;
2) is both lower and upper bounded; and
3) will degenerate to $GEJS_{1/N}$ on the exponent operator when $\alpha \to 1$ and $\bar{m}$ function is applied. $GEJS_{1/N}$ denotes the averagely weighted $GEJS_\omega$ divergence [55].

III. METHODOLOGY: THE DEEP EVIDENCE-THEORETIC REMOTE SENSING LANDSLIDE IMAGE CLASSIFICATION

As described, the BJR divergence is devised to manage "BD" challenge. Next, we will deal with "evidence theory involvement" and "evidence conflict" challenges, respectively. Note that the new BD is applied in Section III-B.

A. Proposed Strategy to Involve Evidence Theory

The accurate recognition of landslide images is very thorny primarily due to their irregular shape patterns, but the shape information can be reflected on multiscale visual saliency maps [27]. The "evidence theory involvement" challenge also indicates a diversified evidence set is vital to lead in the evidence theory [50]. Therefore, we develop a solution with channelwise fusion, which is also the first usage of geological visual saliency for evidence set generation.

The calculation of multiscale visual saliency map can be shortly introduced as next [61]. Assume the input image is I_{ori} which is converted to hue, saturation, and value [62] (HSV) color space, then, down sample I_{ori} to its 2/3 and 1/3 sizes. Denote the downsampled images as $I_{2/3}$ and $I_{1/3}$ respectively, and remark $\eta(\cdot)$ as upsampling an image to its original size. Next, the multiscale visual saliency can be
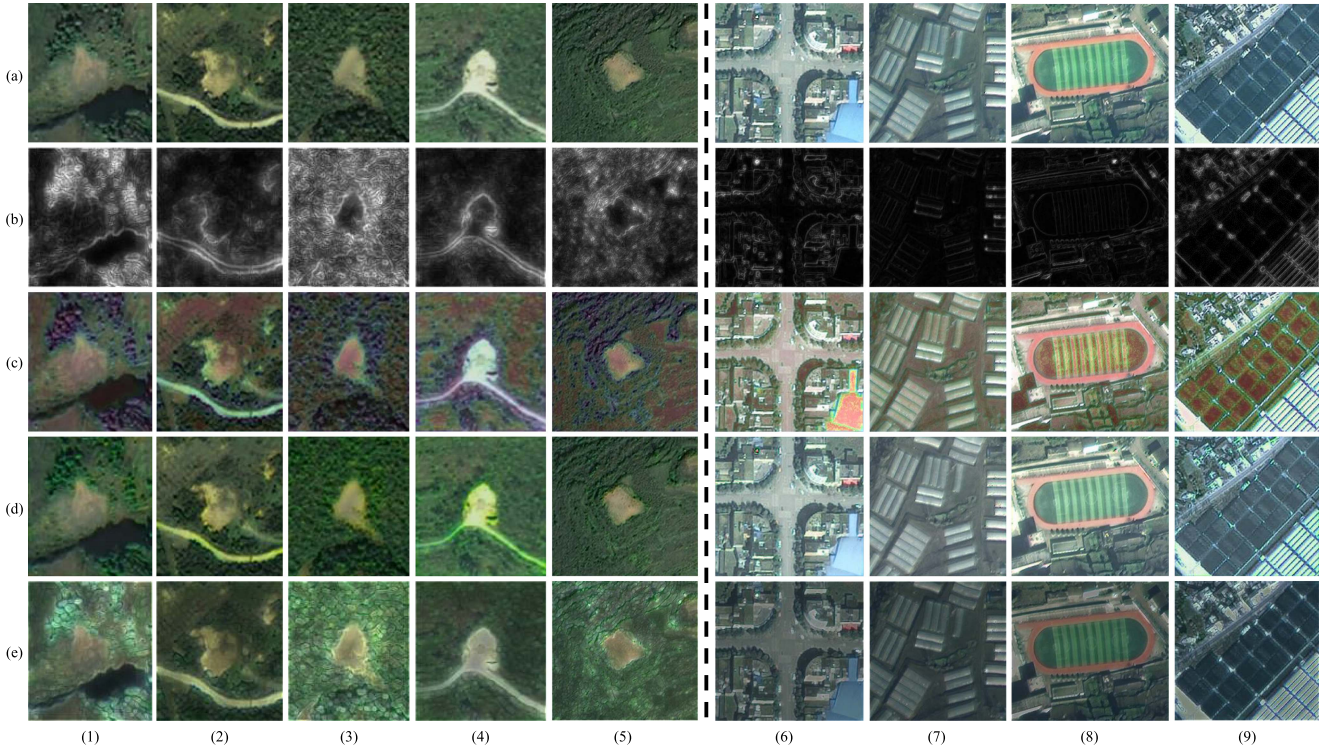
Fig. 5. (a) Original images. (b) Multiscale visual saliency map. (c) H channel fused images. (d) S channel fused images. (e) V channel fused images. Columns (1)–(5) are landslide images, and (6)–(9) are nonlandslide images.

calculated

$$\text{MS}(I_{\text{ori}}) = \sum_{id} w_{id} \eta(\psi(I_{id})) + G(I_{\text{ori}}) \tag{13}$$

where $id \in \{\text{ori}, 2/3, 1/3\}$. ψ is the spectral residual visual saliency operator [63], and $G(\cdot)$ is the Sobel image gradient extractor [64]. w_{id} represents the image entropy [65] of the image I_{id} after a step of normalization.

The multiscale visual saliency in [61] can automatically generate visual attention maps containing geological information from remote sensing images, yelling the importance degree of pixels. With $\text{MS}(I_{\text{ori}})$, this article proposes a channelwise fusion structure to obtain diversified evidence set. Denote $\otimes$ as the Hadamard multiplier to channelwisely fuse the multiscale visual saliency to the remote sensing landslide images

$$F_C = MS(I_{\text{ori}}) \otimes I_{\text{ori}}^C \tag{14}$$

where $C \in \{\text{H,S,V}\}$ represents the image channels.

The major two advantages of HSV space-based channelwise fusion are i) it provides the prerequisite for diversified evidence set, and ii) the channel fused images can experimentally promote accuracy of neural networks than direct classification on original RGB images. For the advantage i), the RGB images, multiscale visual saliency maps and channelwise fused images are displayed in Fig. 5. The colors in H channel fused images are diversified. The saturation of salient regions are changed in S channel fused images. Artificial halos can be observed in V channel fused images. All the diversifications are identical with the expected evidence set, and with such a set one can safely

exploit the evidence fusion. The advantage ii) is because the multiscale visual saliency of landslides in Fig. 5 is commonly higher than nonlandslides, which makes the landslide images more diversified, and the classification improvement is verified in Section V-F. Fusion on spaces other than HSV is discussed in Section V-G.

Since the evidence set is premise of evidential fusion [47], after channelwise fusion, three homogenous networks are trained on each category of channelwisely fused images. The class activation maps in Fig. 6 present that within homogenous network, networks' attentions on pixel of interests are changed or shifted (both the activation location and weighting degree are considered), manifesting the networks are focusing on different locations, or assigning different attention degrees on the same location [66]. The observed phenomenon indicates the desired evidence set is formed [47].

As suggested in [47], the algorithm should appropriately convert evidence set into fusion candidates. Therefore, after network training, the evidence theory is implemented. Suppose the predicted probabilities from networks are P_H, P_S, and P_V for each image category, and FoD is $\Theta = \{\theta_l, \theta_{\text{nl}}\}$ denoting "landslide" and "non-landslide," respectively. Then, the BPAs before normalization are formed via the next approach

$$\hat{m}_C(\Phi) = \begin{cases} P_C(\theta_j), & \text{if } \Phi = \theta_j, j \in \{l, \text{nl}\} \\ \frac{1 - |P_C(\theta_l) - P_C(\theta_{\text{nl}})|}{e}, & \text{if } \Phi = \Theta \end{cases} \tag{15}$$

where $C \in \{\text{H, S, V}\}$ marks the image channel. For the case of ignorance, i.e., $\hat{m}_C(\Theta)$, one can easily derive its lower and
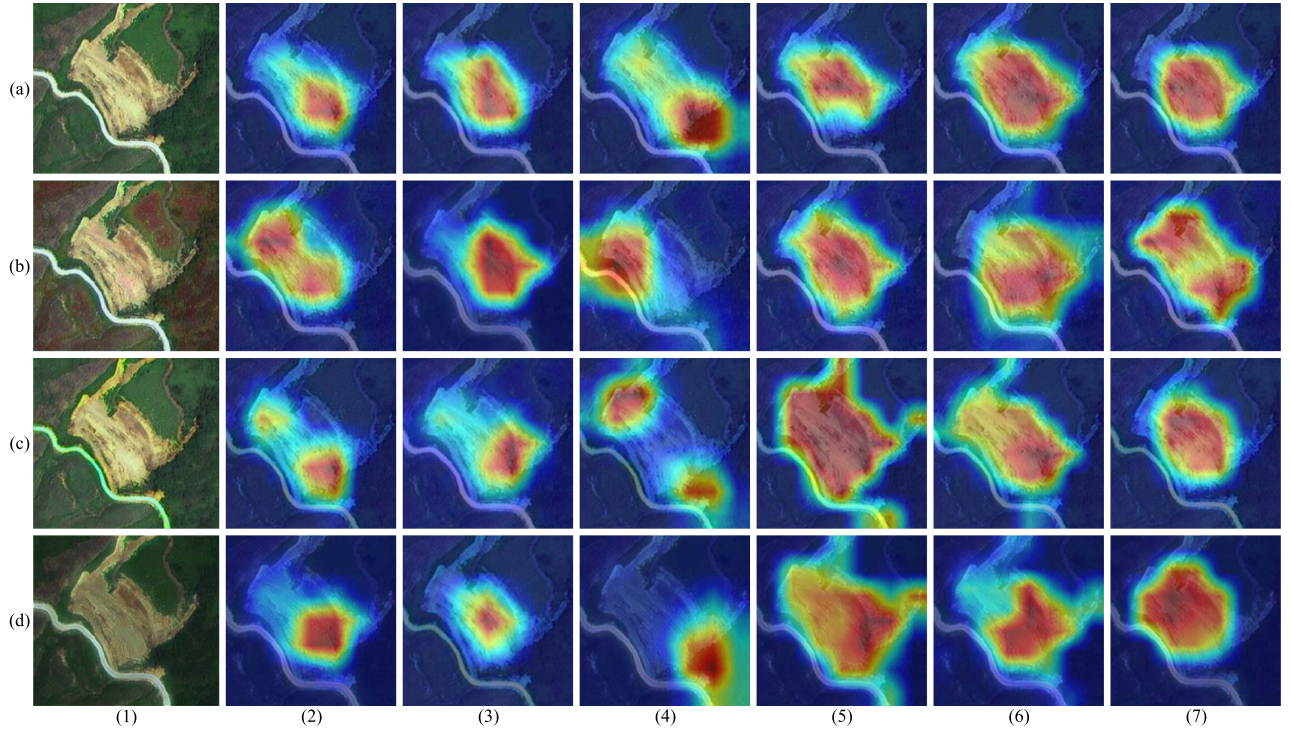
Fig. 6.   Class activation maps via EigenGradCAM algorithm [67]. The rows are on behalf of the images and class activation maps from (a) original image, (b) H channel fused image, (c) S channel fused image, and (d) V channel fused image. The columns refer to (1) exemplified images, the class activation maps generated by (2) the EfficientNet [73], (3) MobileNet V3 [74], (4) ResNet101 [72], (5) Swin Transformer-Base [77], (6) Swin Transformer V2-Base [78], (7) VAN-Base [79]. All these results are calculated from the last layer of the network backbones.

upper bounds

$$0 \leq \hat{m}_C(\Theta) \leq \frac{1}{e} \tag{16}$$

where $C \in \{H, S, V\}$. A step of normalization is required to form BPAs

$$m_C(\Phi) = \frac{\hat{m}_C(\Phi)}{\sum_{\Phi_j \in 2^\Theta} \hat{m}_C(\Phi_j)} \tag{17}$$

where $\Phi \in 2^\Theta$ and $C \in \{H, S, V\}$. It is easy to verify that the generated BPAs satisfy the property in (18)

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{\Phi \in 2^\Theta} m(\Phi) = 1. \end{cases} \tag{18}$$

By verifying (18), it can be determined the computations of evidence theory on landslide classification can be valid [43].

B. New Interpretability Improved Three-Branched Fusion With Proposed BJR Divergence

The final decision on landslides is difficult, which should be made with caution. To tackle the "evidence conflict challenge" in evidence-theoretic decision fusion, after fusion candidate generation, this study proposes the interpretability improved three-branched evidential fusion algorithm.

The motivation on the new algorithm is that in [47], the three-branched fusion is devised, but it lacks an explicit interpretability for their third weighting branch, i.e., the "impurity" branch, since their third branch is just the first-order expansion

of their second branch, indicating that it is difficult to answer why the concept of a third weighting branch is indispensable. To manage the described issue, the new model abandons the "impurity" participated structure, and improves justifiability for introducing a third branch in landslide classification. The new workflow can be roughly divided into five steps, where from Step 2 to Step 4 are the three branches.

Step 1. (Calculate transformed belief functions): Appropriately transformed BPAs can improve decision quality [47]. To start with, suppose N sets of BPAs are obtained (in the present task, $N = 3$), then, they are converted into Enhanced Pignistic probabilities

$$m_j \longrightarrow EBetP_j \tag{19}$$

where $j \in \{1, \ldots, N\}$.

Step 2. (Calculate branch 1: Supporting degree branch) The first branch takes supporting degrees into consideration. Since no prior knowledge on the information weight is provided, the proposed BJR divergence with an averagely distributed weighting vector is exploited for the calculation:

$$\text{Dis}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N | m_j)$$
$$= \text{BJR}_{\frac{1}{N-1}, \alpha}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N) \tag{20}$$

where $\text{Dis}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N | m_j)$ means the average dissimilarity among the N sets of BPAs except m_j, $j \in \{1, \ldots, N\}$. Then, the supporting degree of the jth evidence

before normalization can be obtained

$$\widehat{\text{Sup}}_j = \frac{\text{Dis}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N | m_j)}{\text{BJR}_{\frac{1}{N}, \alpha}(m_1, \ldots, m_N)} \quad (21)$$

where $j \in \{1, \ldots, N\}$.

According to the suggestions by Xiao [55], if $\widehat{\text{Sup}}_j < 1$, it manifests that the average dissimilarity with evidence m_j is greater than the average dissimilarity without m_j, thus, m_j is a conflicted evidence, and a smaller credibility should be assigned. And for the case $\widehat{\text{Sup}}_j > 1$, it means that after providing evidence m_j, the average dissimilarity is decreased, thus, a larger credibility should be assigned. For the case $\widehat{\text{Sup}}_j = 1$, it means that no significant dissimilarity change is obtained. Then, the final supporting degree is gained via this approach

$$\text{Sup}_j = \frac{\widehat{\text{Sup}}_j}{\sum_{s=1}^{N} \widehat{Sup}_s} \quad (22)$$

where $j \in \{1, \ldots, N\}$.

Step 3. (Calculate branch 2: Information volume branch): The information volume measures quantify informational quality of evidence, and Deng entropy has achieved success [47], thus, the Enhanced Pignistic Deng entropy (EPDE) [58] is utilized to measure the uncertainty degree of the jth evidence

$$\text{EPDE}_j = - \sum_{\Phi \in 2^{\Theta}} EBetP_j(\Phi) \log EBetP_j(\Phi)$$
$$+ \sum_{\Phi \in 2^{\Theta}} EBetP_j(\Phi) \log(2^{|\Phi|} - 1) \quad (23)$$

where $j \in \{1, \ldots, N\}$. Then, the information volume of the jth evidence is formed via the next equation

$$\text{IV}_j = \frac{\text{EPDE}_j}{\sum_{s=1}^{N} \text{EPDE}_s} \quad (24)$$

which terminates the Step 3.

Note that as a Deng entropy-like entropy, the EPDE linearly combines the uncertainty measured from randomness and nonspecificity [60]. However, in [68], Deng entropy is argued that its nonspecificity part may expand too large, thus, may cancel the uncertainty measurement from randomness. As a belief entropy analogous to Deng entropy, the EPDE also faces this problem. Keep this motivation in mind, unlike the original method in [47], this study adopted another entropy of belief function in Step 4, which enhances the explainability of three-branched fusion progress.

Step 4. (Calculate branch 3: Improved information volume branch): Following the drawback analysis of the Deng entropy as presented in Step 3, to enhance the uncertainty measurement from randomness, the Renyi entropy for the Enhanced Pignistic probability defined in (8) is exploited

$$\widehat{EIV}_j = R_\alpha(EBetP_j) \quad (25)$$

where $j \in \{1, \ldots, N\}$. Then, a step of normalization is required

$$EIV_j = \frac{\widehat{EIV}_j}{\sum_{s=1}^{N} \widehat{EIV}_s}. \quad (26)$$

According to (8), the proposed belief entropy only measures the uncertainty from randomness, and is parameterized by α, which is more feasible. And unlike its former counterpart in [47], the new approach abandons an "impurity" that can be more beneficial to the fusion model, because the third branch inherently corresponds to the shortcoming of the second branch. Consequently, the third branch can serve as an improvement of the second branch with a better theoretical significance. Note that though the branches can be conflicted, what is determined is a computational crash will not occur, as the three branches are only weights, rather than decisions.

Step 5. (Multiple evidence combination for landslide classification): The evidence combination determines decision belief assignment on landslides [58]. First, fuse the three weight components to attain the evidential weight of jth evidence

$$\hat{W}_j = \text{Sup}_j \times IV_j \times EIV_j \quad (27)$$

for $j \in \{1, \ldots, N\}$. Then, a step of normalization is required

$$W_j = \frac{\hat{W}_j}{\sum_{s=1}^{N} \hat{W}_s} \quad (28)$$

where $j \in \{1, \ldots, N\}$. Next, form the weighted evidence via the next equation

$$WEBetP = \sum_{j=1}^{N} W_j \times EBetP_j \quad (29)$$

Then, fuse the weighted evidence for $N-1$ times, with which the decision evidence function can be obtained

$$DEBetP = \underbrace{WEBetP \oplus WEBetP \oplus \cdots \oplus WEBetP}_{N-1 \text{ times}}$$
$$(30)$$

where $\oplus$ marks the Dempster's combination rule [38]

$$\begin{cases} m(\emptyset) = 0 \\ m(\Phi_j) = \frac{1}{1-K} \sum_{\cap \Phi = \Phi_j} \prod_{1 \leqslant s \leqslant N} m_s(\Phi_j), \Phi_j \in 2^{\Theta} \end{cases} \quad (31)$$

and $K = \sum_{\cap \Phi = \emptyset} \prod_{1 \leqslant s \leqslant N} m_s(\Phi_j)$ is the conflict coefficient. Up to now, the interpretability improved three-branched fusion for landslide image classification has terminated.

Fig. 7 displays the full evidence-theoretic fusion progress. The refined fusion framework can help to gather and fuse the diversified evidence from deep neural networks, thus, the decisions can be integrated more reasonably. For the step of decision-making, the Pignistic transformation [69] is utilized to convert belief functions back into probability functions:

$$BetP(\theta) = \sum_{\Phi \in 2^{\Theta}} DEBetP(\Phi) \frac{|\theta \cap \Phi|}{|\Phi|} \quad (32)$$

where $\theta \in \Theta$. Ultimately, the $BetP$ function carries the final classification on remote sensing landslide images.

IV. THEORETICAL COMPARISON ON EVIDENCE CONFLICT MANAGEMENT

In Section III, an interpretability improved three-branched fusion framework is proposed. However, we still remember the
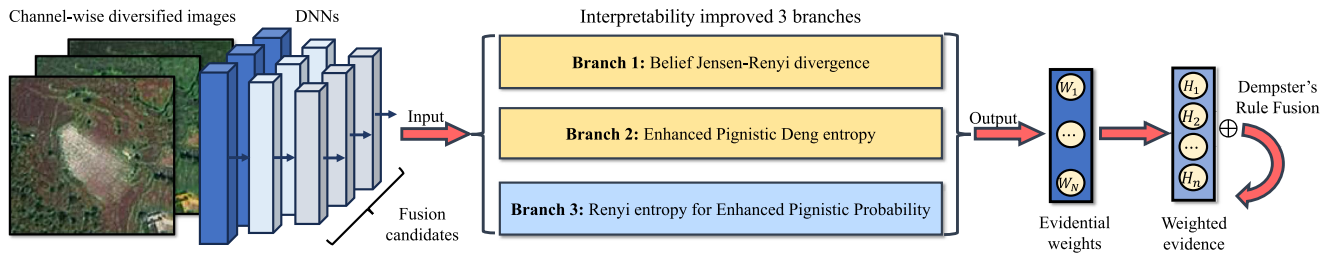
Fig. 7.   Workflow of interpretability improved three-branched evidence fusion. n denotes the size of FoD's power set.

TABLE III
FUSION RESULTS OF EXAMPLE 2

Hypothesis	Branch-free		Single-Branched	Bi-Branched			Tri-Branched			
	Dempster [38]	Murphy [80]	Deng [81]	Hellinger [57]	Renyi [43]	Generalized Renyi [44]	Zhang [47]	Proposed ($\alpha = 0.1$)	Proposed ($\alpha = 0.5$)	Proposed ($\alpha = 1.5$)
$\{\theta_l\}$	0.3684	**0.6272**	**0.8529**	**0.8882**	**0.8945**	**0.8911**	**0.9887**	**0.9642**	**0.9906**	**0.9905**
$\{\theta_{nl}\}$	**0.6303**	0.3705	0.1436	0.1052	0.0970	0.1015	0.0111	0.0356	0.0091	0.0091
$\{\theta_l, \theta_{nl}\}$	0.0013	0.0023	0.0034	0.0067	0.085	0.0074	0.0002	0.0002	0.0003	0.0004
Decision	$\{\theta_{nl}\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$	$\{\theta_l\}$

The involved algorithms include dempster's rule [38], Murphy's method [80], Deng's method [81], belief hellinger distance-based [57], Belief renyi divergence-based [43], generalized belief renyi divergence-based [44], Zhang's method [47], and the proposal with respect to changed order α. The text in bold represents the belief on the preferred hypothesis of each algorithm.
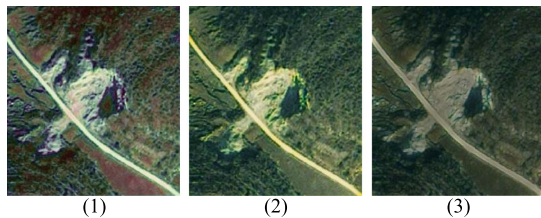

Fig. 8.   In Example 2, (1) H channel fused, (2) S channel fused, and (3) V channel fused remote sensing landslide images.

new fusion model is majorly proposed to cope with evidence conflict [41], thus, like [41], [55], and [56], this study provides a short fusion analysis on evidence conflict, which emerged in remote sensing landslide image classification.

Example 2 (Numerical validation on conflict management): This example is from a real landslide classification problem. Fig. 8 displays the channelwisely fused remote sensing images that contain a landslide occurred near a mountain road. Suppose FoD $\Theta = \{\theta_l, \theta_{nl}\}$, where θ_l and θ_{nl} denote landslide and nonlandslide respectively. m_H, m_S, and m_V are three BPAs converted from networks' predictions on H, S, and V channel fused images. The BPAs have been displayed in Table V, and m_H is noticed to be a piece of conflicting evidence with the highest belief 0.9591 on $\{\theta_{nl}\}$. Next, this example involves seven previous evidential fusion schemes to fuse the evidence. Table III displays the results.

In Table III, it is clear that except Dempster's rule [38], the remaining algorithms have successfully discriminated the desired hypothesis, i.e., $\{\theta_l\}$. In addition, the proposed fusion with $\alpha = 0.5$ reaches the highest belief on the desired hypothesis

with 0.9906, which reflects its efficiency in evidence conflict management [41]. In summary, this result primarily verifies the practicality of the proposed evidential fusion scheme. Even when confronting evidence conflict, the proposal can still keep robust to desired fusion results.

Next, the contribution ratio of each weighting branch and the final weights of evidence are shown in Table IV. One can easily derive that the six involving algorithms have all assigned the lowest weight on the untrusted evidence m_H as desired, but the three-branched schemes, including Zhang's method [47] and the proposal, is averagely even lower, which are 0.0563 and 0.0184, respectively, (and the proposal is even lower than Zhang's method [47]). It suggests that the three-branched schemes can be more sensitive to evidence conflict, because once the third weighting branch is added, via the multiplier in (27) the canceling effect on the conflicted evidence will be reinforced to a higher level. And the proposal is even more sensitive to evidential conflict than the original Zhang's method [47] since an even lower evidential weight (0.0184) on the untrusted m_H has been assigned, which will be more beneficial to conflict-based evidential fusion.

Finally, since the participated weighting-based fusion methods, except Zhang's method in [47], mainly choose to fuse weighted BPAs rather than the weighted $EBetP$ functions with Dempster's combination rule [38], this study also provides the fused belief on the desired hypothesis $\{\theta_l\}$ under two mentioned cases in Fig. 9. It is clear that when using weighted $EBetP$ functions, the model will be more competitive (+0.1109 belief at most). But even to fuse weighted BPAs, except the case when $\alpha = 0.1$, the proposal can still obtain a higher belief on the desired hypothesis than the models that choose to fuse weighted BPAs (the best is generalized Renyi [44] with a fused belief of

TABLE IV
Contribution Ratio of Each Weighting Branch Toward BPAs as Well as the Final Weights in Example 2, Where "↓" Means Smaller is Better

Method	Branch 1			Branch 2			Branch 3			Evidential Weight		
	m_H	m_S	m_V	m_H	m_S	m_V	m_H	m_S	m_V	$m_H \downarrow$	m_S	m_V
Deng [81]	0.1805	0.3774	0.4421	-	-	-	-	-	-	0.1805	0.3774	0.4421
Hellinger [57]	0.2447	0.3573	0.3980	0.1498	0.2543	0.5959	-	-	-	0.1005	0.2492	0.6503
Renyi [43]	0.1952	0.2941	0.5106	0.1498	0.2543	0.5959	-	-	-	0.0716	0.1832	0.7452
Generalized Renyi [44]	0.2244	0.3322	0.4434	0.1498	0.2543	0.5959	-	-	-	0.0879	0.2210	0.6911
Zhang [47]	0.2460	0.3532	0.4007	0.1208	0.3209	0.5583	0.2614	0.3180	0.4205	0.0563	0.2615	0.6821
Proposed ($\alpha = 0.5$)	0.2186	0.3615	0.4199	0.2082	0.2409	0.5509	0.0725	0.2661	0.6614	0.0184	0.1291	0.8525

The branch-free methods in Table III are excluded from this analysis.

TABLE V
BPAs in Example 2

Hypothesis	IS1: m_H	IS2: m_S	IS3: m_V
$\{\theta_l\}$	0.0321	0.8588	0.6144
$\{\theta_{nl}\}$	0.9591	0.0813	0.1798
$\{\theta_l, \theta_{nl}\}$	0.0088	0.0599	0.2058

The "IS" is short for "information source."

TABLE VI
Size of Each Set After Stochastically Splitting 2773 Images

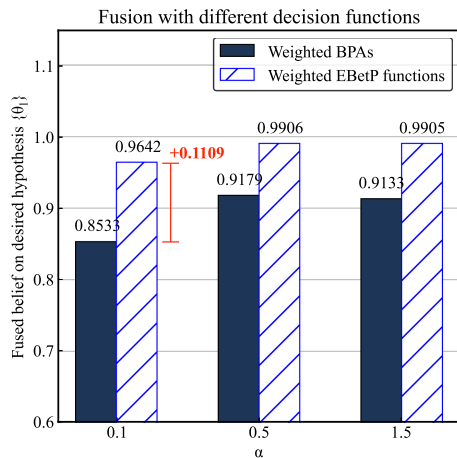| Class	Train	Evaluation	Test
Landslide	493	123	154
Non-landslide	1282	321	400

Fig. 9. Fused belief on desired hypothesis $\{\theta_l\}$ using two different decision functions in the proposal.

0.8911). Consequently, in this case study we present that even when the weighted BPAs are set for fusion can still the proposal be robust to evidence conflict.

V. Experiments

A. Dataset Description

To guarantee a fair comparison, the involved dataset is Bijie Landslide open source dataset [13], which is an optical remote sensing landslide image dataset collected by research groups in Wuhan University. This dataset consists of 2773 images captured by TripleSat satellite in Bijie, China, from May to August 2018, including 770 optical landslide images and 2003 nonlandslides images. The exemplified images have been demonstrated in Fig. 10.

Then, this article utilizes a proportion of "0.64:0.16:0.20" to randomly split the train, evaluation and test datasets for network training. Details have been listed in Table VI.

B. Experimental Setup and Training Details

The following experiments are executed on an Intel I9-11900 K 3.50 GHz CPU, 64.00-GB RAM platform. A NVIDIA GeForce RTX 3090 graphical processing unit is also exploited to accelerate the network training process.

To guarantee a reproducible study, the involved networks are trained for 100 epochs with initial weights pretrained on ImageNet-1 K dataset [70]. Images are center cropped into 224×224 with batch size 32. Optimization details of the participated networks are shown in Table VII. For the "Step" updater, the decay steps are 30, 60, and 90. For the "CA" updater, the minimum learning rate ratio is 0.01. The label smooth value is 0.1. Data argumentation [71] is also employed to solve the issues of data imbalance and small amount of data.

C. Comparison Algorithms

Considering that this article is a *de facto* integrated study, we involve four comparative experiments, where Fig. 11 shows the comparative studies' organization.

Deep image classification backbone networks are essential to image classification task in computer vision-related areas. Thus, in the first experimental study, this article involves eight classical or very recently researched image classification backbone networks (12 networks in total) for the proposal's efficiency verification, which consists of ResNet101 [72], EfficientNet (b0) [73], MobileNetV3 (small) [74], RepVGG (B0) [75], Visual Attention Network (VAN, Tiny and Base) [76], Swin Transformer (Tiny and Base) [77], Swin Transformer V2 (Tiny and Base) [78], and HorNet (Tiny-GF and Base-GF) [79]. In the second comparison, the comparison between the improved three-branched evidential fusion and SOTA evidence fusion algorithms are investigated on the best performed network, which include Dempster's rule [38], Murphy's method [80], Deng's method [81], belief Hellinger distance [57], belief Renyi divergence [43] and its generalization [44], Zhang's method [47] and this proposal.
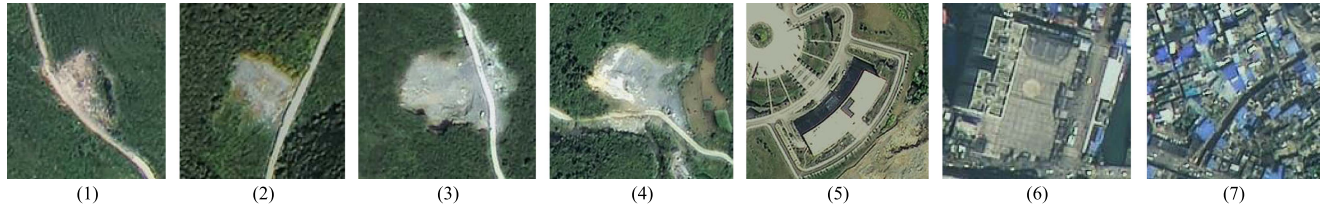
(1) (2) (3) (4) (5) (6) (7)

Fig. 10. Exemplified images of the Bijie Landslide dataset [13], where (1)–(4) are landslide images, and (5)–(7) are nonlandslide images.

TABLE VII
TRAINING DETAILS OF THE BACKBONE NETWORKS, WHERE "CA" IS SHORT FOR COSINE ANNEALING

	ResNet101 [72]	EfficientNet-b0 [73]	MobileNetV3-S [74]	RepVGG-B0 [75]	VAN-T [76]	VAN-B [76]
Loss function	CrossEntropy Loss	CrossEntropy Loss	CrossEntropy Loss	CrossEntropy Loss	LabelSmooth Loss	LabelSmooth Loss
Optimizer	SGD	SGD	RMSprop	SGD	AdamW	AdamW
Learning rate	0.001	0.0125	0.001	0.0125	0.000125	0.000125
Updater	Step	Step	Step	CA	CA	CA
	SwinTrans-T [77]	SwinTrans-B [77]	HorNet-T-GF [79]	HorNet-B-GF [79]	SwinTransV2-T [78]	SwinTransV2-B [78]
Loss function	LabelSmooth Loss	LabelSmooth Loss	LabelSmooth Loss	LabelSmooth Loss	LabelSmooth Loss	LabelSmooth Loss
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW
Learning rate	0.00025	0.00025	0.000125	0.000125	0.00025	0.00025
Updater	CA	CA	CA	CA	CA	CA

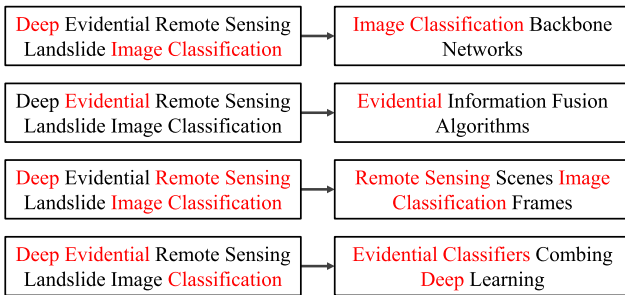The unlisted parameters of the optimizers are set as their default values.

Fig. 11. Comparative studies' organization.

Then, in the third comparative study, this article provides the comparison with eight remote sensing scenes image classification algorithms. Bazi et al. [82] employed ViT-32 for this vision task. In [83], a SCCovNet is devised for end-to-end remote sensing scenes image classification with skip connections and covariance pooling. In [84], the network VGG-VD-16 is applied. In [85], a multiscale feature fusion covariance network named MF ^{2} CNet with octave convolution is constructed. Tang et al. [86] endowed EMTCAL, a remote sensing scene classification algorithm with multiscale transformer and cross-level attention learning. A homo-heterogenous transformer learning (HHTL)-based remote sensing scene image classification framework, is demonstrated in [87]. And Huang et al. [50] proposed ECMS, a multicolor spaces-based remote sensing image classification with an optimizable BPA function discounting weights. This paper choose their best performed network, the GSANet [88] in comparison study.

As for the fourth study, we fairly compare the proposal with three recently prevalent evidence-theoretic deep learning

classifiers. In [90], an evidence-theoretic deep-learning algorithm with expected utility theory was developed. In [89], dynamic evidence fusion is connected with trusted multiview classification. Debaque et al.[91] presented a deep evidence-theoretic frame for accurate sheep classification. Then, this study fairly compares them in the deep image-level classification task on remote sensing landslide images.

D. Evaluation Criteria

In experiments, four metrics (overall accuracy, precision, true positive rate, and F-measure score) are involved to guarantee a fair comparison [50]

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \qquad (33)$$

$$Pre = \frac{TP}{TP + FP} \qquad (34)$$

$$TPR = \frac{TP}{TP + FN} \qquad (35)$$

$$F - measure = \frac{2 \times Pre \times TPR}{Pre + TPR} \qquad (36)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative.

E. Comparative Analyses

1) Efficiency Analysis on Network Structures: One of this study's goals is to demonstrate the efficiency of evidence theory-coupled deep learning in remote sensing landslide image classification. Thus, to begin with, this article compares solely using a single image classification backbone network with the proposal employing the evidence theory (both channelwise fusion and
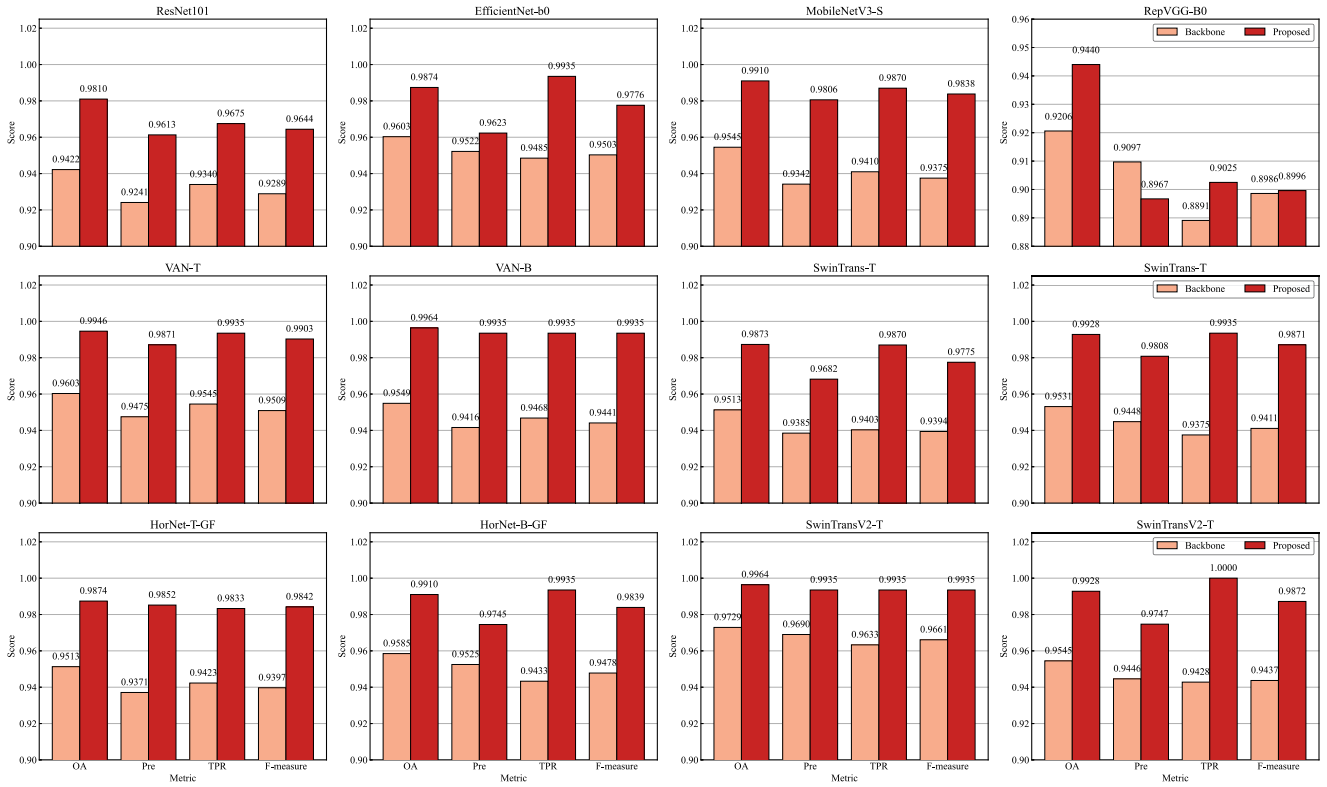
Fig. 12. Results of efficiency analysis. "Backbone" means only a single SOTA backbone network is applied for landslide image classification.

interpretability improved three-branched fusion are included). The evaluation results in Fig. 12 are revealing in two ways, which are discussed as behind.

1) The leading-in of evidence theory has proven efficient with almost all the involved neural networks. Except the slightly decreased Pre score on RepVGG-B0, the proposed frame significantly improves the OA, Pre, TPR, and F-measure scores on the remaining 11 neural networks (with +4.15%, +5.19%, +5.72%, and +4.94% improvement at most, respectively). The dominating reason is the evidence theory can combine multiple meaningful features and information, and consequently make comprehensive decisions, whereas a single network does not have this property. And it also reflects that the evidence theory *seems* not to have an explicit preference on a particular network depth or structure. Namely, the tested networks' scale varies drastically, and they may be structured on CNN or ViT, or their combinations, but we find most of them equally have a tendency to be improved.

2) Under the four given criteria, the Swin Transformer V2-Tiny and VAN-Base achieve the highest Acc, TPR, and F-measure with 0.9963, 0.9935, and 0.9935, respectively. For Swin Transformer V2-Tiny, it is because its scaled cosine attention improves the learning of diversified features from landslides. As for VAN-Base, its channel adaptability is improved, which is compatible with the proposed channelwise fusion. In the following parts, this article chooses the Swin Transformer V2-Tiny as the best backbone network for further investigations.

2) Comparison With Previous Evidence Fusion Algorithms: Since except for the theoretical comparison in Section IV, one should also investigate the improved three-branched fusion in remote sensing image classification, the present article sets the comparison between the improved model and seven state-of-the-art evidence fusion algorithms, and the results have been shown in Table VIII. The parameter analysis on α is also included.

The data in Table VIII reveals that the proposed fashion yields the highest OA, Pre, TPR, and F-measure scores with 0.9964, 0.9935, 0.9935, and 0.9935, when $\alpha = 0.1$ and 0.5. This is because in contrast with earlier proposals, the proposed BD introduces a more general mathematical form that can handle multiple evidence. The impact is the apriori knowledge from landslide visual features can be reasonably fused with a global interplay. In consequence, the decision accuracy is further improved. Note that when $\alpha = 1.5$, the performance of the proposal drops to a lower level, suggesting that when applying this scheme in real world landslide image classification problems, it is better to finetune α to guarantee the algorithm's performance.

3) Comparison With Deep Remote Sensing Image Scene Classification Algorithms: In this section, the present study turns to the comparison with state-of-the-art (evidential) remote sensing image scene classification schemes.

To start with, the algorithms' performances are displayed in Table IX. A positive finding is the new method with the best performed backbone network ranks first under OA, Pre, TPR, and F-measure with 0.9964, 0.9935, 0.9935, and 0.9935, respectively. The result from ECMS-GSANet is also satisfying.

TABLE VIII
COMPARISON STUDY BETWEEN THE BJR DIVERGENCE-INVOLVED FRAME AND STATE-OF-THE-ART EVIDENTIAL SCHEMES

Metric	Branch-free		Single-Branched	Bi-Branched			Tri-Branched			
	Dempster [38]	Murphy [80]	Deng [81]	Hellinger [57]	Renyi [43]	Generalized Renyi [44]	Zhang [47]	Proposed ($\alpha = 0.1$)	Proposed ($\alpha = 0.5$)	Proposed ($\alpha = 1.5$)
OA	0.9928	0.9928	0.9946	0.9928	0.9946	0.9910	0.9946	**0.9964**	**0.9964**	0.9838
Pre	0.9807	0.9870	0.9935	0.9870	0.9935	0.9806	0.9870	**0.9935**	**0.9935**	0.9801
TPR	0.9935	0.9870	0.9870	0.9870	0.9870	0.9870	0.9935	**0.9935**	**0.9935**	0.9610
F-measure	0.9871	0.9870	0.9902	0.9870	0.9902	0.9838	0.9902	**0.9935**	**0.9935**	0.9705

The backbone network is Swin Transformer V2-Tiny.

TABLE IX
COMPARISON BETWEEN THE PROPOSAL AND SOTA (EVIDENTIAL) REMOTE SENSING SCENES IMAGE CLASSIFICATION ALGORITHM

Method	Evidential	OA	Pre			TPR			F-measure		
		Total	L	NL	Total	L	NL	Total	L	NL	Total
ViT-32 [82]		0.9528	0.9324	0.9606	0.9324	0.8961	0.9750	0.8961	0.9139	0.9677	0.9139
SCCovNet-AlexNet [83]		0.9531	0.9388	0.9607	0.9465	0.8961	0.9775	0.9356	0.9169	0.9690	0.9408
SCCovNet-VGG16 [83]		0.9549	0.9324	0.9606	0.9497	0.8961	0.9790	0.9368	0.9319	0.9677	0.9430
VGG-VD-16 [84]		0.9516	0.9296	0.9582	0.9320	0.8571	0.9750	0.8896	0.8919	0.9665	0.9103
MF^{2}CNet [85]		0.9350	0.8688	0.9619	0.8688	0.9026	0.9475	0.9025	0.8854	0.9547	0.8853
EMTCAL [86]		0.9747	0.9430	0.9874	0.9430	0.9675	0.9775	0.9675	0.9551	0.9824	0.9551
HHTL [87]		0.9819	0.9737	0.9851	0.9736	0.9610	0.9900	0.9610	0.9673	0.9875	0.9673
ECMS-GSANet [50]	✓	0.9910	0.9745	0.9975	0.9745	0.9935	0.9900	0.9935	0.9839	0.9937	0.9839
Proposed (RepVGG-B0 backbone)	✓	0.9440	0.8182	0.9600	0.8967	0.9007	0.9320	0.9025	0.8575	0.9458	0.8966
Proposed (SwinTransV2-T backbone)	✓	**0.9964**	**0.9935**	**0.9975**	**0.9935**	**0.9935**	**0.9975**	**0.9935**	**0.9935**	**0.9975**	**0.9935**

"L" and "NL" remark landslide and nonlandslide, respectively.
The bold texts indicates the highest score under the metrics.
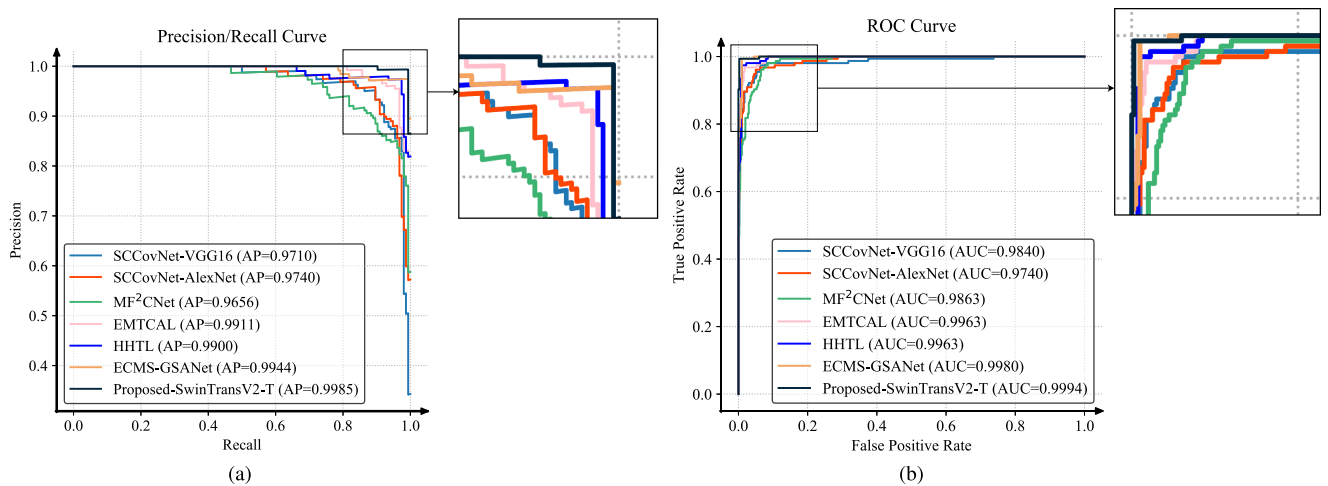

Fig. 13.   (a) Precision/Recall and (b) ROC curves over 554 testing samples (2773 in total) among remote sensing image classification algorithms.

Nonetheless, its simple channel conversion is adverse to network accuracy. That is, although it also uses evidence theory to combine multisource information, its fusion candidates have already made more false classifications than the proposal. The lower competitiveness of the rest methods accounts for 1) they are generalist networks, whose structures are less optimized for the complex landslide features, and 2) they rely to a great extent on a single image, which is less comprehensive than exploiting integrated visual patterns and features.

Then, we go a step further. Since the Precision/Recall curve and receiver operating characteristic (ROC) curve are two more objective performance indicators [92], the present article involves them in Fig. 13. In the Precision/Recall curves, it is apparent that the proposal with best performed backbone network significantly dominates this metric with the highest average precision score 0.9985, demonstrating its superiority. Then, in the figure of ROC curves, since both the proposal and ECMS-GSANet yield satisfying performance, they seem
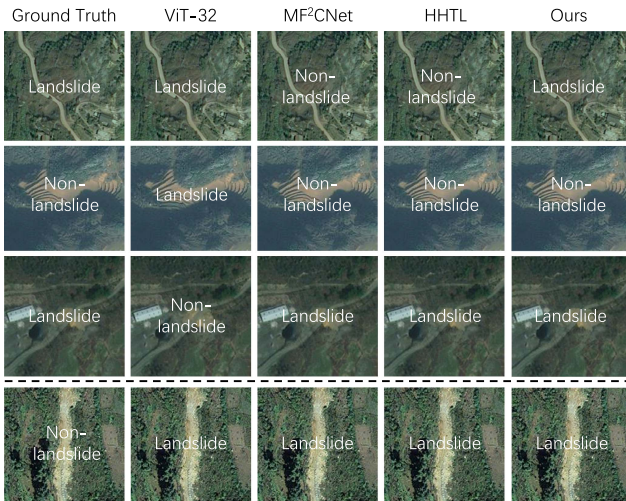
Fig. 14. Classification results on difficult samples.

TABLE X
COMPARISON WITH DEEP EVIDENCE-THEORETIC CLASSIFIERS

Method	OA	Kappa
EDLS [89]	0.8520	0.6298
ETMC [91]	0.9621	0.9042
EDLU [90] ($\gamma = 0.4$)	0.9765	0.9417
EDLU [90] ($\gamma = 0.6$)	0.9765	0.9417
EDLU [90] ($\gamma = 0.8$)	0.9783	0.9462
Proposed (SwinTransV2-T)	**0.9964**	**0.9910**

The bold texts indicates the highest score under the metrics.

to simultaneously dominate this metric. However, because only when an algorithm ranks the first in ROC curves can it dominate the Precision/Recall curves [93], and the proposal has gained the highest area under curve score with 0.9994, this proposal is still more competitive than the second best performed algorithm.

Eventually, the present study sets Fig. 14 to compare the classification on difficult samples. The images in the figure were randomly selected from the samples that are confusing even for humans. A positive finding is from the first three lines, our approach stands out as compared with ViT-32, MF ^{2} CNet, and HHTL. However, our method also fails when dealing with extremely difficult samples, as presented in the last line. The khaki scene in the image center closely resembles the color and shape pattern of real landslides, making it very challenging for the recognition of artificial intelligence classifiers.

4) Comparison With Evidence-Theoretic Deep Learning Classifiers: In this section, the present study further compares the proposal with three versatile deep learning classifiers combining evidence theory, which are evidence deep learning with utility theory (EDLU) [90], evidence deep learning for sheep classification (EDLS) [89], and evidential trusted multi-view classification, (ETMC)[91], respectively. The backbone network of EDLU and the proposal is Swin Transformer V2-Tiny, where Table X exhibits the classification results.

Through data analysis, it is clear that the EDLS achieves inferior performance. The core reason is its ResNet18 backbone

TABLE XI
INVOLVING CONDITIONS IN ABLATION STUDY 2

Weighting strategy	Impurity branch	Renyi entropy branch
Two branch	✗	✗
Zhang's branch [47]	✓	✗
Proposed	✗	✓

networks are too shallow to capture useful visual features for landslides. And the proposal has a significant potential to surpass both ETMC and EDLU. For ETMC, it seems because its fusion rule is akin to traditional Demspter's fusion rule, which is not robust against evidence conflict when it has to face confusing apriori knowledge. Then, the underlying reason for EDLU is it lacks a decision fusion module, which hinders the evidence theory's accuracy boost effect when integrating the landslide information, such as colors and shape patterns. The proposal can avoid the three analyzed disadvantages, thus, a better performance can be obtained.

F. Ablation Studies

1) Ablation Study on Each Fusion Stage: First, we present the effects of each fusion stage. Fig. 15 displays the confusion matrix before and after each fusion (the results of channelwise fusion are listed in accordance with their fused channel, i.e., the H, S, or V channel). The results strongly support the efficiency of our proposal, since in contrast with the RGB images, the number of miss classifications significantly drops on all the H, S, and V channel fused images. Namely, we find the HSV space conversion plus channelwise fusion result in three better feature spaces than traditional RGB space. This encouraging phenomenon can be observed on 11 backbone networks (still excluded the RepVGG-B0), which is also the first study to discover that the proposed HSV channelwise fusion can enhance the neural networks on remote sensing landslide images. Then, with the new three-branched fusion, final results with the least miss classification can be achieved. In summary, both the channelwise fusion and the interpretability improved three-branched fusion are found effective.

2) Ablation Study on Improved Three-Branched Fusion: The second study is about the refined three-branched fusion, where the conditions are presented in Table XI. Fig. 17 displays comparison results on RepVGG-B0 and Swin Transformer V2-Tiny backbone networks.

The Fig. 17 uncovers that if the third weighting branch is not involved, the OA scores on the participated two networks both shrink to a lower level. Then, if we adopt the third branch, the proposed fashion achieves the same performance with the current best three-branched fashion, i.e., Zhang's fashion. But we still remember that Zhang's model falls in the dilemma of numerical efficiency and model explainability, which is avoided by the proposal. Consequently, this article considers that the proposed scheme is better not only from the aspect of model explainability, but also from its guarantee on computational effectiveness.
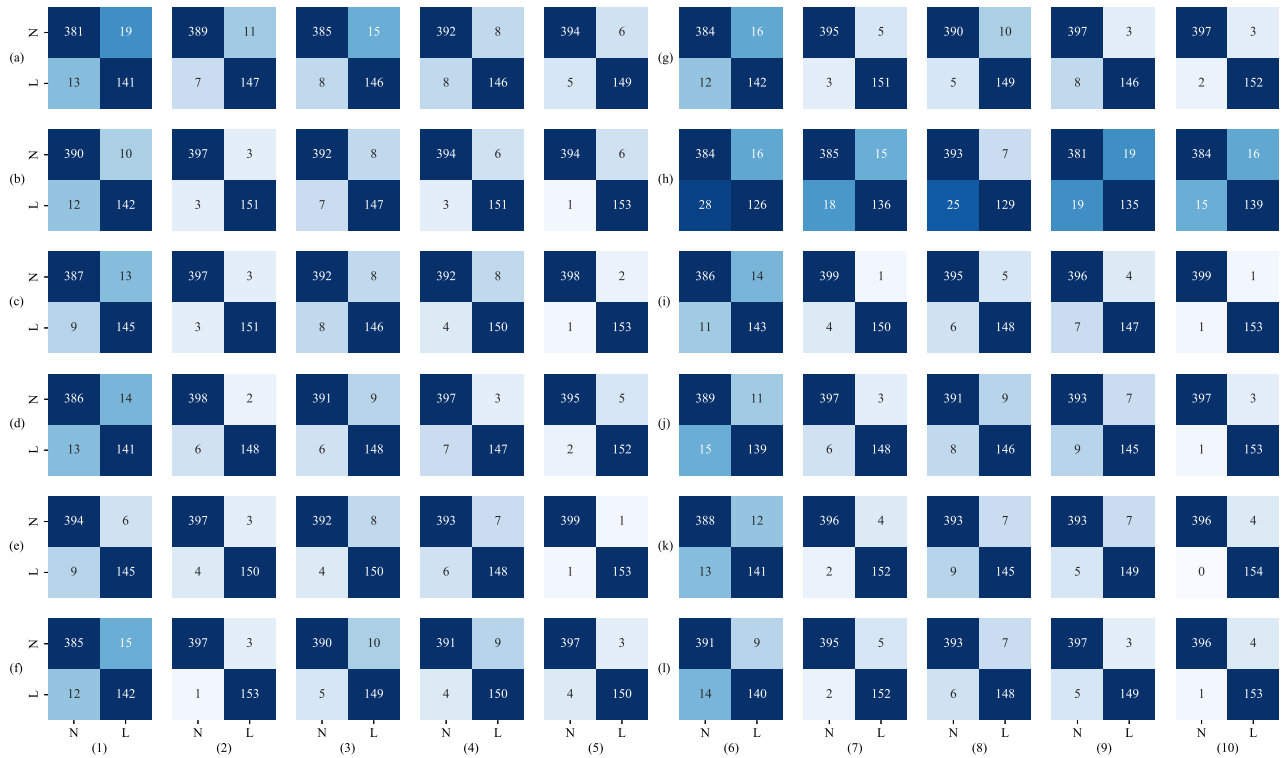
Fig. 15. Classification confusion matrix. Rows are results from (a) ResNet101 [72], (b) MobileNetV3-S [74], (c) VAN-T [76], (d) SwinTransformer-T [77], (e) SwinTransformer V2-T [78], (f) HorNet-T-GF [79], (g) EfficientNet [73], (h) RepVGG-B0 [75], (i) VAN-B [76], (j) SwinTransformer-B [77], (k) SwinTransformer V2-B [78] and (l) HorNet-B-GF [79], respectively. The columns (1) and (6) are the classification results from the original images, (2) and (7) are from the H channel fused images, (3) and (8) are from the S channel fused images, (4) and (9) are from V channel fused images, (5) and (10) are from the proposed fusion flow (including the channelwise fusion and the new divergence-based interpretability improved three-branched fusion).

G. Sensitivity Analyses

1) Sensitivity on Fused Color Spaces: In the proposed channelwise fusion, this article chooses a HSV space conversion strategy. However, the fusion in other color spaces, such as RGB, XYZ, and Lab, has not been verified yet. Therefore, this article also tests the model performance when the fusion is applied in spaces rather than HSV.

Fig. 16 demonstrates the OA score when the channelwise fusion is applied in RGB, XYZ, Lab, HSV, and YIQ spaces. A key conclusion is that except EfficientNet-b0 and RepVGG-B0, the HSV-space fused images tend to reach a better classification. That may be explained that the HSV space is naturally more suitable for feature representation, which is in line with literature [50]. Thus, by appending a HSV transformation, the neural networks can automatically extract better features on landslides than features in the remaining color spaces. This finding also answers why it is the HSV space that is chosen as the basis for our later operations.

2) Sensitivity on Evidence Conflict Management: This section investigates the sensitivity of evidential weight assignment regarding α, which determines the sensitivity of evidence conflict management. This test is performed on Swin Transformer V2-Tiny network, and results are demonstrated in Fig. 18.

An essential outcome is with the increase of α, the evidential weights gradually approximate long-tailed distributions. More weights are assigned nearing 0 and 1, the two boundary values, which indicates the sensitivity of evidence conflict gets reinforced. In accordance with [58], a evidence fusion system with more sensitivity to evidence conflict might be more helpful to manage the counterintuitive results of Dempster's combination rule. Therefore, we can learn with the rising of α, the management of evidence conflict also varies dramatically.

3) Sensitivity to Image Noise Degradation: The image degradation will galactically influence deep learning on remote sensing image classification [95]. This article stimulates two types of sensor-affected image degradation in remote sensing, i.e., addictive and multiplicative Gaussian noise, and compares the proposal with single backbone network. Fig. 19 displays the exemplified noisy landslide images affected by stochastic zero mean Gaussian signal with a variance of 0.1.

The comparison has been presented in Fig. 20. One can learn that with the increase of noise variance, both the single network and the proposal shows a performance degeneration. Nevertheless, the proposal can still produce a better classification on OA score than single networks, which suggests a robustness against noise-caused image degradation. This phenomenon can be explained that in contrast with a single network, the evidence theory is more adept in organically capturing and unifying several information sources. Consequently, a more robust decision-making is observed. It also provides a new evidence about the evidence theory's effectiveness in remote sensing landslide image classification.
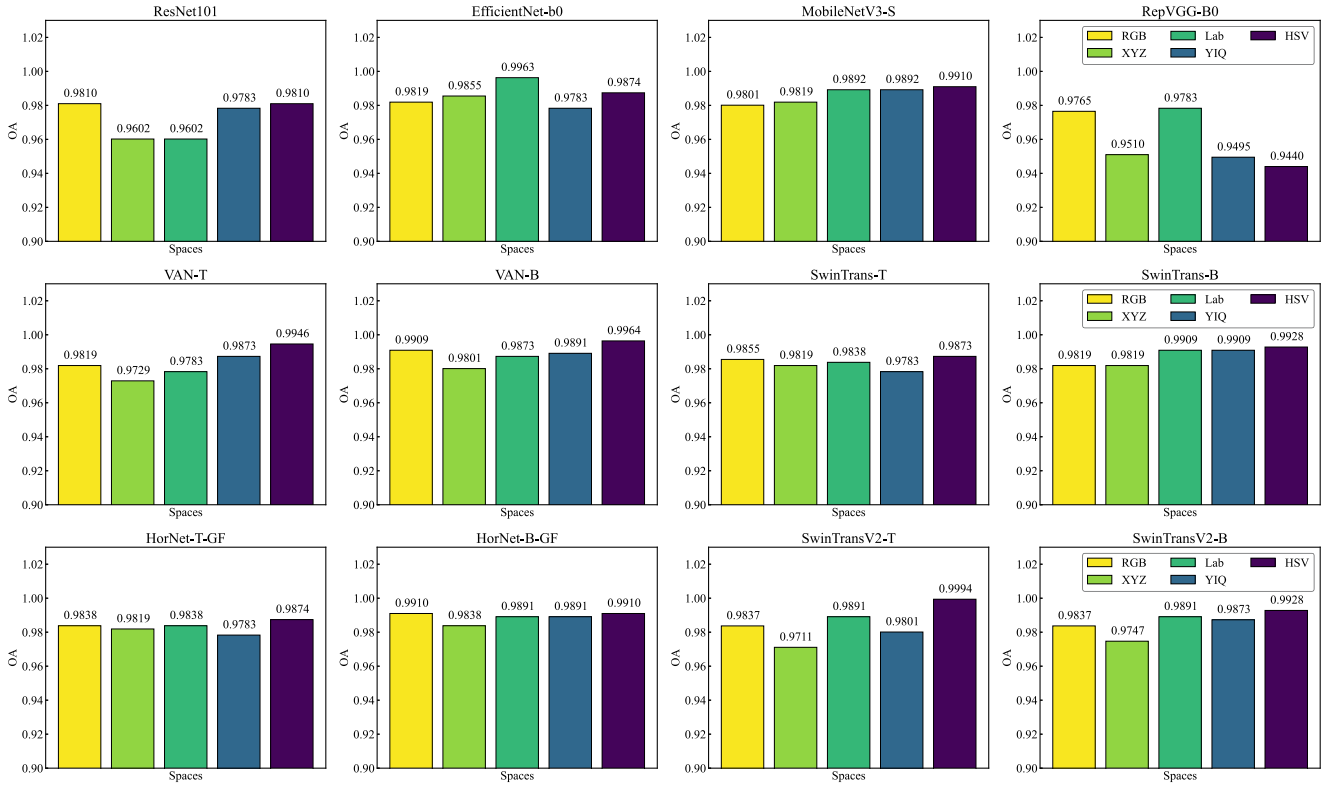
Fig. 16. Classification results in Sensitivity analysis 1 when the proposed channelwise fusion is applied on different color spaces. Here the RGB, XYZ, Lab, YIQ, and HSV spaces are investigated.
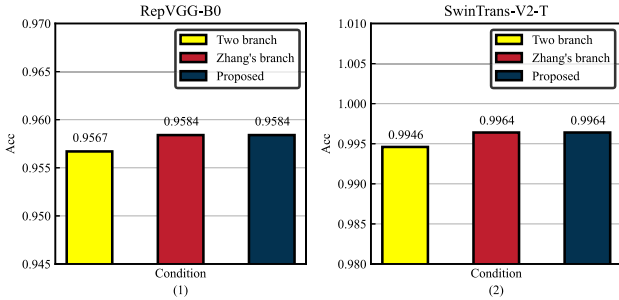

Fig. 17. OA score under three conditions in Ablation study 2. (1) and (2) are results from RepVGG-B0 and SwinTransV2-T, respectively.
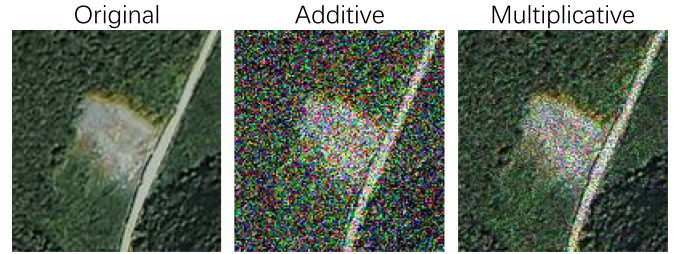

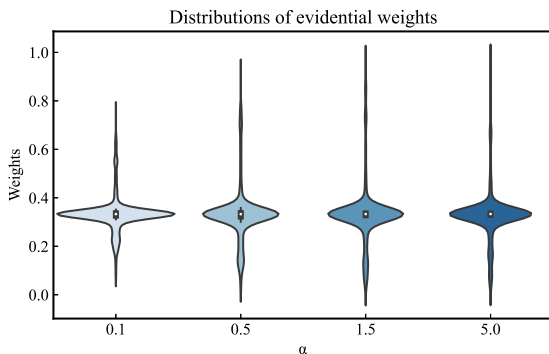Fig. 19. Exemplified landslide images degenerated by additive and multiplicative Gaussian noise.

Fig. 18. Distribution analysis of evidential weights regarding α.

H. Discussions

1) Computational Complexity Analysis: The propose of this section is to discuss time complexity. Since the proposal does not specify a particular network, suppose an input image with size $H \times W$, then, the time complexity on multiscale saliency is $O(F)$, where $F = H \times W \times n_F^2 \log(n_F)$, $n_F = \exp(2, \lceil \log_2(\min\{H, W\}) \rceil)$ [96]. Assume the inference time complexity of network is M, the number of information sources is N, and the Dempster's rule's complexity is E, which has an exponential relation with FoD's cardinality (fixed as three in this study). Table XII displays the results.

With the data from Table XII, it is clear that the proposal ranks first as compared with the remaining algorithms, which reflects the proposal's efficiency. The underlying cause is that the proposal chooses an improved three-branched structure to
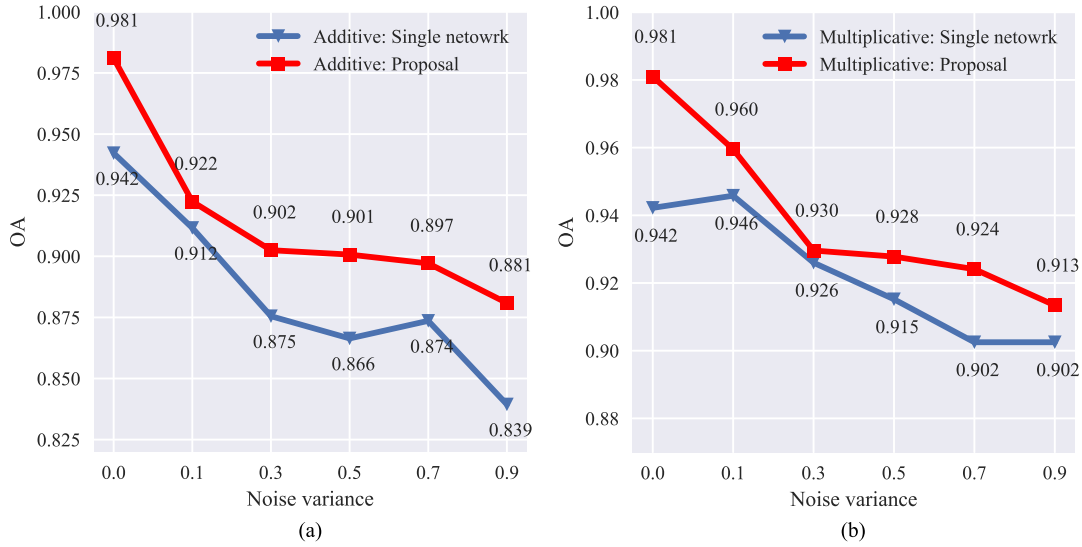
Fig. 20. Classification results with (a) additive and (b) multiplicative Gaussian noise degenerated images. (a) Additive Gaussian noise. (b) Multiplicative Gaussian noise.

TABLE XII
COMPUTATIONAL COMPLEXITY COMPARISON WITH COMPETING EVIDENTIAL
FUSION ALGORITHMS

Method	Complexity	Ranking on OA
Single network only	$O(F + M)$	8
Dempster	$O(F + M + D \times N)$	4
Murphy	$O(F + M + N)$	6
Deng	$O(F + M + N + D \times N)$	2
Hellinger	$O(F + M + N^2 + D \times N)$	4
Renyi	$O(F + M + N^2 + D \times N)$	2
Generalized Renyi	$O(F + M + N^2 + D \times N)$	7
Zhang	$O(F + M + N^2 + D \times N)$	2
Proposed	$O(F + M + N^2 + D \times N)$	**1**

The bold text means the ranking first under OA score.

Fig. 21. Satellite terrain image of Luding from Google Earth. The data collectors did not point out where these landslides locate.
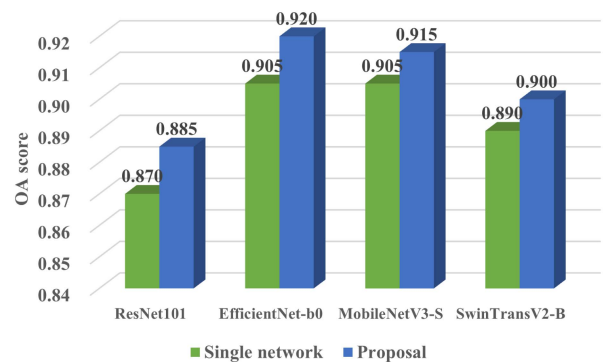
Fig. 22. Comparison on the real landslides at Luding, Southwest China, in September 2022.

evaluate the information sources, which is more comprehensive than its former counterparts. In addition, from the ranking of OA score, the law "the faster, the better" is not strictly followed. Namely, the simplest algorithm participated in this study faces the heaviest risk on performance degradation. This finding suggests that our algorithm is still far from being too complex to face generalization ability loss.

2) Real Scenario Verification: By far, the proposed method has only been verified on Bijie Landslide dataset. Next, we execute a verification on real scenario in Luding, China, to further evaluate the proposal.

The Luding County is located in Sichuan Province, Southwest China. As a conclusion in Fig. 21, due to its complex terrain and fragile ecological environment, landslides are frequent in this area. Fu et al. [97] published a series of real landslide scenarios in Luding, which includes 200 optical remote sensing images on earthquake-included landslides captured by GF-6 satellite on September 5, 2022. Although they did not point out the exact regions that the landslides are located at, one can still use this real image set as unseen samples to test the models' generalization ability. In consequence, we apply

four image classification networks that are coupled with evidence theory and pretrained on Bijie dataset to make predictions on the Luding image set, where the results are shown in Fig. 22.

The finding significantly stands out from Fig. 22 is the evidence theory-coupled methods outperform the single networks. The dominating reason is when facing unseen instances, the evidence-theoretic fusion can efficiently exploit its nature in dealing with "uncertain" and "ignorant" apriori information, which brings about a more robust decision-making. In contrast, the original single networks is less competitive in processing uncertainty in dataset shift, thus, they can be less accurate. The authors notice that the work in [98] can support our analysis.

3) Proposal's Shortcoming and Future Works: An algorithm can hardly be competitive in *every* aspect. Even though the proposal has satisfying decision accuracy, it is still limited to pure vision-based classification. Some environmental factors, such as the profile curvature and slope angle, are not involved. Besides, yet the current model has a high complexity, and has not been pruned for in-orbit embedded systems. Future works can focus on the combination of geological landslide discriminators, as well as the evidence theory-boosted fast recognition on in-space platforms.

VI. CONCLUSION

In the task of remote sensing landslide image classification, the accuracy of conventional deep learning still has room for improvement. Aiming at this issue, the present study introduces the evidence theory to enhance deep learning classifiers. We address three insurmountable challenges that hinder the implementation of evidence theory: the "evidence conflict," "BD," and "evidence theory involvement" challenges, and makes three improvements. To tackle the BD challenge, the BJR divergence is proposed. Next, for the evidence theory involvement challenge, we newly design a channelwise fusion strategy with multiscale visual saliency. Its reduction effect on false classification is also witnessed when compared with RGB landslide images. Eventually, to address the challenge of evidence conflict, an interpretability improved three-branched fusion is meticulously devised, which successfully refines the model explicability of its former counterpart.

To comprehensively evaluate the proposal, in theoretical comparison, we verify its robustness against evidence conflict; in experimental studies, we uncover its suitability on state-of-the-art image classification backbone networks. The proposal also tends to yield better classification as compared with remote sensing scenes classifiers, versatile deep evidence-theoretic classifiers and evidential information fusion schemes. We also painstakingly discussed the proposal's sensitivity on two types of image degradation, as well as its performance under difference color spaces. Eventually, the proposal is verified in a real landslide scenario in Luding County, China, whose data is publicly available.

Despite effectiveness, the present article has pointed out that the proposal is still restricted to pure vision-based landslide classification, and the current model has not been pruned for in-orbit platforms. In future works, the authors will focus attention on the evidence theory-boosted landslide classification combining multiple topographic discriminators, as well as their fast and accurate identifications on in-orbit instruments.

APPENDIX A
PROOF OF THEOREM 3

Before the formal proof, two lemmas should be clarified.

Lemma 1: When $\alpha \in (0, 1)$, the JR divergence is upper bounded [46].

Lemma 2: When $\alpha \in (1, +\infty)$, the Renyi entropy for a probability P is always upper bounded by the traditional Shannon entropy

$$R_\alpha(P) \leq H(P) \tag{37}$$

where $H(P)$ is the classical Shannon entropy for the probability distribution P [99].

With the lemmas this theorem can be proven safely.

Proof: It is clear that the inequality $\text{BJR}_{\omega,\alpha} \geq 0$ holds. By using Lemma 1, it is rational to suppose when $\alpha \in (0, 1)$, the upper bound of the JR divergence is M, then, it is easy to derive the next relationship

$$\text{BJR}_{\omega,\alpha} \leq e^M. \tag{38}$$

As for the case $\alpha \in (1, +\infty)$, considering the Lemma 2, we can derive the next inequality

$$\text{BJR}_{\omega,\alpha}(m_1, \ldots, m_N)$$

$$= \exp\left\{ R_\alpha\left(\sum_{i=1}^N w_i EBetP_i\right) - \sum_{i=1}^N w_i R_\alpha(EBetP_i) \right\}$$

$$\leq \exp\left\{ R_\alpha\left(\sum_{i=1}^N w_i EBetP_i\right) \right\}$$

$$\leq \exp\left\{ H\left(\sum_{i=1}^N w_i EBetP_i\right) \right\} \tag{39}$$

where H represents the traditional Shannon entropy. Since when P is a n-dimensional discrete uniform distribution, the Shannon entropy can achieve its maximum value, we have

$$H(P) \leq \log n \tag{40}$$

Thus, the BJR divergence is upper bounded by $e^{\log n}$.

In summary, for $\alpha \in (0, 1) \cup (1, +\infty)$, the BJR divergence has an upper bound $M' = \max\{e^M, e^{\log n}\}$. ∎

APPENDIX B
PROOF OF THEOREM 4

Proof: For $\alpha \to 1$, the Renyi entropy will degenerate to Shannon's entropy [100], i.e.,

$$\lim_{\alpha \to 1} R_\alpha(P) = H(P) \tag{41}$$

where H represents Shannon entropy. Then, if we use $\bar{m}$ function to replace the $EBetP$ function, it is easy to derive

$$
\begin{aligned}
&\lim_{\alpha \to 1} \exp\left\{ R_\alpha \left(\sum_{i=1}^{N} w_i \bar{m}_i \right) - \sum_{i=1}^{N} w_i R_\alpha(\bar{m}_i) \right\} \\
&= \exp\left\{ H\left(\sum_{i=1}^{N} w_i \bar{m}_i \right) - \sum_{i=1}^{N} w_i H(\bar{m}_i) \right\} \\
&= \exp\left\{ \text{GEJS}_\omega \right\}.
\end{aligned}
\tag{42}
$$

Thus, this theorem holds. ∎

References

[1] C. Ye, R. Wei, Y. Ge, Y. Li, J. M. Junior, and J. Li, "GIS-based spatial prediction of landslide using road factors and random forest for Sichuan-Tibet highway," *J. Mountain Sci.*, vol. 19, pp. 461–476, 2022.

[2] Y. Zhou, H. Wang, R. Yang, G. Yao, Q. Xu, and X. Zhang, "A novel weakly supervised remote sensing landslide semantic segmentation method: Combining CAM and cycleGAN algorithms," *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3650.

[3] N. Casagli, E. Intrieri, V. Tofani, G. Gigli, and F. Raspini, "Landslide detection, monitoring and prediction with remote-sensing techniques," *Nature Rev. Earth Environ.*, vol. 4, no. 1, pp. 51–64, 2023.

[4] A. Mohan, A. K. Singh, B. Kumar, and R. Dwivedi, "Review on remote sensing methods for landslide detection using machine and deep learning," *Trans. Emerg. Telecommun. Tech.*, vol. 32, 2021, Art. no. e3998.

[5] J. Chen, L. Li, C. Xu, Y. Huang, and Z. Luo, "Freely accessible inventory and spatial distribution of large-scale landslides in Xianyang City Shaanxi Province, China," *Earthq. Res. Adv.*, vol. 3, no. 3, 2023, Art. no. 100217.

[6] Y. Zhang et al., "RoI fusion strategy with self-attention mechanism for object detection in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5990–6006, 2023.

[7] Q. Xu et al., "Remote sensing for landslide investigations: A progress report from China," *Eng. Geol.*, vol. 321, 2023, Art. no. 107156.

[8] H. Hou, M. Chen, Y. Tie, and W. Li, "A Universal landslide detection method in optical remote sensing images based on improved YOLOX," *Remote Sens.*, vol. 14, no. 19, 2022, Art. no. 4939.

[9] Q. Liu, T. Wu, Y. Deng, and Z. Liu, "SE-YOLOv7 landslide detection algorithm based on attention mechanism and improved loss function," *Land*, vol. 12, no. 8, 2023, Art. no. 1522.

[10] C. Zhong et al., "Landslide mapping with remote sensing: Challenges and opportunities," *Int. J. Remote Sen.*, vol. 41, no. 4, pp. 1555–1581, 2020.

[11] O. Ghorbanzadeh, H. Shahabi, A. Crivellari, S. Homayouni, T. Blaschke, and P. Ghamisi, "Landslide detection using deep learning and object–based image analysis," *Landslides*, vol. 19, pp. 929–939, 2022.

[12] X. Liu et al., "Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 61, 2023, Art. no. 4500314.

[13] S. Ji, D. Yu, C. Shen, W. Li, and Q. Xu, "Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks," *Landslides*, vol. 17, pp. 1337–1352, 2020.

[14] P. Lv, L. Ma, Q. Li, and F. Du, "ShapeFormer: A shape-enhanced vision transformer model for optical remote sensing image landslide detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2681–2689, 2023.

[15] A. A. Reyes, S. Paheding, A. Rajaneesh, K. S. Sajinkumar, and T. Oommen, "C-PLES: Contextual progressive layer expansion with self-attention for multi-class landslide segmentation on Mars using multimodal satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2023, 354–364.

[16] D. T.-Anh et al., "Integrative few-shot classification and segmentation for landslide detection," *IEEE Access*, vol. 10, pp. 120200–120212, 2022.

[17] L. Wang and Y. Qiao, "Application of optimized CNN algorithm in landslide boundary detection," *J. Meas. Eng.*, vol. 11, no. 4, pp. 1–14, 2023.

[18] L. Defang, J. Li, and F. Fan, "Classification of landslides on the southeastern Tibet plateau based on transfer learning and limited labelled datasets," *Remote Sens. Lett.*, vol. 12, no. 3, pp. 286–295, 2021.

[19] B. Fang, G. Chen, L. Pan, R. Kou, and L. Wang, "GAN-based Siamese framework for landslide inventory mapping using bi-temporal optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 391–395, Mar. 2021.

[20] S. K. Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, "Multimodal fusion transformer for remote sensing image classification," *IEEE Trans. Geosci. Remote*, vol. 61, 2023, Art. no. 5515620.

[21] X. Wu, D. Hong, and J. Chanussot, "Convolutional neural networks for multimodal remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5517010.

[22] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514415.

[23] J. Yao, D. Hong, H. Wang, H. Liu, and J. Chanussot, "UCSL: Towards unsupervised common subspace learning for cross-modal image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514212.

[24] M. Xie, Z. Han, C. Zhang, Y. Bai, and Q. Hu, "Exploring and exploiting uncertainty for incomplete multi-view classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 19873–19882.

[25] Z. Xu, X. Yue, Y. Lv, W. Liu, and Z. Li, "Trusted fine-grained image classification through hierarchical evidence fusion," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 10657–10665.

[26] Z. Zhang, H. Wang, J. Geng, W. Jiang, X. Deng, and W. Miao, "An information fusion method based on deep learning and fuzzy discount-weighting for target intention recognition," *Eng. Appl. Artif. Intell.*, vol. 109, 2022, Art. no. 104610.

[27] M. Mishra, K. N. P. Raju, and P. C. Pandey, "Need for pan- India compatibility of geospatial databases in terms of map projections and parameters," *Geocarto Int.*, vol. 38, no. 1, 2023, Art. no. 2215206.

[28] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513412.

[29] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-net in U-Net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 364–76, 2022.

[30] F. Belmahdi, M. Lazri, F. Ouallouche, K. Labadi, R. Absi, and S. Ameur, "Application of Dempster-Shafer theory for optimization of precipitation classification and estimation results from remote sensing data using machine learning," *Remote Sens. Appl.*, vol. 29, 2023, Art. no. 100906.

[31] L. Jiao, H. Yang, Z. Liu, and Q. Pan, "Interpretable fuzzy clustering using unsupervised fuzzy decision trees," *Inf. Sci.*, vol. 61, pp. 540–563, 2022.

[32] Z. Liu, Y. Bi, and P. Liu, "An evidence theory-based large group FMEA framework incorporating bounded confidence and its application in supercritical water gasification system," *Appl. Soft Comput*, vol. 129, 2022, Art. no. 109580.

[33] L. Jiao, F. Wang, Z. Liu, and Q. Pan, "TECM: Transfer learning-based evidential c-means clustering," *Knowl.-Based Syst.*, vol. 257, 2022, Art. no. 109937.

[34] L. Jiao, T. Denœux, Z. Liu, and Q. Pan, "EGMM: An evidential version of the Gaussian mixture model for clustering," *Appl. Soft Comput.*, vol. 129, 2022, Art. no. 109619.

[35] P. Liu, G. Yu, Z. Wang, B. Zhou, and P. Chen, "Object classification based on enhanced evidence theory: Radar–vision fusion approach for roadside application," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 5006412.

[36] W. Liu, X. Yue, Y. Chen, and T. Denoeux, "Trusted multi-view deep learning with opinion aggregation," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 7, pp. 7585–7593.

[37] X. Deng, S. Xue, and W. Jiang, "A novel quantum model of mass function for uncertain information fusion," *Inf. Fusion*, vol. 89, pp. 619–631, 2023.

[38] G. Shafer, "Dempster's rule of combination," *Int. J. Approx. Reason.*, vol. 79, pp. 26–40, 2016.

[39] L. Ma, W. Yao, X. Dai, and R. Jia, "A new evidence weight combination and probability allocation method in multi-sensor data fusion," *Sensors*, vol. 23, no. 2, Art. no. 722.

[40] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *Int. J. Approx. Reasoning*, vol. 53, pp. 118–145, 2012.

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

[42] J. Zeng and F. Xiao, "Belief Kullback-Leibler divergence," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 106027.

[43] C. Zhu and F. Xiao, "A belief Rényi divergence for multi-source information fusion and its application in pattern recognition," *Appl. Intell.*, vol. 53, no. 8, pp. 8941–8958, 2022.

[44] C. Zhu, F. Xiao, and Z. Cao, "A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis," *Inf. Sci.*, vol. 605, pp. 225–243, 2022.

[45] Z. Chen and R. Cai, "A novel divergence measure of mass function for conflict management," *Int. J. Intell. Syst.*, vol. 37, pp. 3709–3735, 2022.

[46] Y. He, A. B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1211–1220, May 2003.

[47] J. Zhang, J. Zhang, J. Chen, S. Wang, and L. Wang, "Rail surface defect detection through bimodal RSDINet and three-branched evidential fusion," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 2508714.

[48] S. Xu, Y. Chen, C. Ma, and X. Yue, "Deep evidential fusion network for medical image classification," *Int. J. Approx. Reasoning*, vol. 150, pp. 188–198, 2022.

[49] Z. Tong, P. Xu, and T. Denoeux, "Fusion of Evidential CNN Classifiers for Image Classification," in *Proc. 6th Int. Conf. Belief Functions*, 2021, pp. 168–176.

[50] L. Huang, W. Zhao, A. W. Liew, and Y. You, "An evidential combination method with multi-color spaces for remote sensing image scene classification," *Inf. Fusion*, vol. 93, pp. 209–226, 2023.

[51] L. Huang, T. Denoeux, P. Vera, and S. Ruan, "Evidence fusion with contextual discounting for multi-modality medical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv. Lecture Notes Comput. Sci.*, 2022, vol. 13435, pp. 401–411.

[52] Z. Diao, H. Jiang, X.-H. Han, Y.-D. Yao, and T. Shi, "EFNet: Evidence fusion network for tumor segmentation from PET-CT volumes," *Phys. Med. Biol.*, vol. 66, 2021, Art. no. 205005.

[53] J. Zeng and F. Xiao, "A fractal belief KL divergence for decision fusion," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 106027.

[54] Z. Hua and X. Jing, "An improved belief Hellinger divergence for Dempster-Shafer theory and its application in multi-source information fusion," *Appl. Intell. 53*, pp. 17965–17984, 2023.

[55] F. Xiao, "GEJS: A. generalized evidential divergence measure for multisource information fusion," *IEEE Trans. Syst. Man. Cybern.: Syst.*, vol. 53, no. 4, pp. 2246–2258, Apr. 2023.

[56] L. Zhang and F. Xiao, "A novel belief χ^2 divergence for multisource information fusion and its application in pattern classification," *Int. J. Intell. Syst.*, vol. 37, pp. 7968–7991, 2022.

[57] C. Zhu and F. Xiao, "A belief Hellinger distance for D–S evidence theory and its application in pattern recognition," *Eng. Appl. Artif. Intell.*, vol. 106, 2021, Art. no. 104452.

[58] J. Zhang, X. Ma, A. Wang, and Y. Lin, "An enhanced pignistic transformation-based fusion scheme with applications in image segmentation," *IEEE Access*, vol. 11, pp. 19892–19913, 2023.

[59] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy, "Entropy for evaluation of Dempster-Shafer belief function models," *Int. J. Approx. Reasoning*, vol. 151, pp. 164–181, 2022.

[60] Q. Zhou and Y. Deng, "Fractal-based belief entropy," *Inf. Sci.*, vol. 587, pp. 265–282, 2022.

[61] C. Xu, C. Yin, D. Wang, and W. Han, "Fast ship detection combining visual saliency and a cascade CNN in SAR images," *IET Radar Sonar Navigation*, vol. 18, no. 14, pp. 1879–1887, 2020.

[62] S. R. Prakash and P. N. Singh, "Background region based face orientation prediction through HSV skin color model and K-means clustering," *Int. J. Inf. Techol.*, vol. 15, no. 3, pp. 1275–1288, 2023.

[63] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[64] A. Shakya, M. Biswas, and M. Pal, "Fusion and classification of SAR and optical data using multi-image color components with differential gradients," *Remote Sens.*, vol. 15, no. 1, 2023, Art. no. 274.

[65] F. Qi and T. Wang, "A Semantic segmentation method based on image entropy weighted spatio-temporal fusion for blade attachment recognition of marine current turbines," *J. Mar. Sci. Eng.*, vol. 11, no. 4, p. 691, 2023.

[66] B. N. Patro, M. Lunayach, and V. P. Namboodiri, "Uncertainty class activation map (U-CAM) using gradient certainty method," *IEEE Trans. Image Proces.*, vol. 30, pp. 1910–1924, 2021.

[67] H. A. Hsia et al., "CLIPCAM: A simple baseline for zero-shot text-guided object and action localization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 4453–4457.

[68] S. M. García and J. Abellán, "Critique of modified Deng entropies under the evidence theory," *Chaos Solution Fractals*, vol. 140, 2020, Art. no. 110112.

[69] P. Smets, "Decision making in the TBM: The necessity of the pignistic transformation," *Int. J. Approx. Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.

[70] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[71] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[72] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[73] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[74] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.

[75] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets Great Again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13733–13742.

[76] M. H. Guo, C. Z. Lu, Z. N. Liu, M. M. Cheng, and S. M. Hu, "Visual attention network," *Comput. Vis. Media*, vol. 9, no. 4, pp. 733–752, 2022.

[77] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[78] Z. Liu et al., "Swin transformer V2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12009–12019.

[79] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S. N. Lim, and J. Lu, "HorNet: Efficient high-order spatial interactions with recursive gated convolutions," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 10353–10366, 2022.

[80] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decis. Support Syst.*, vol. 29, no. 1, pp. 1–9, 2000.

[81] Y. Deng, W. Shi, Z. Zhu, and Q. Liu, "Combining belief functions based on distance of evidence," *Decis. Support Syst.*, vol. 38, no. 3, pp. 489–493, 2004.

[82] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.-Basel*, vol. 13, no. 3, 2021, Art. no. 516.

[83] N. He, L. Fang, S. Li, J. Plaza, and A. Plaza, "Skip-connected covariance network for remote sensing scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1461–1474, May 2020.

[84] R. S. A. Kareem et al., "Multilabel land cover aerial image classification using convolutional neural networks," *Arab. J. Geosci.*, vol. 14, no. 17, 2021, Art. no. 1681.

[85] L. Bai, Q. Liu, C. Li, Z. Ye, M. Hui, and X. Jia, "Remote sensing image scene classification using multiscale feature fusion covariance network with octave convolution," *IEEE Trans. Geosci. Remote*, vol. 60, 2022, Art. no. 5620214.

[86] X. Tang, M. Li, J. Ma, X. Zhang, F. Liu, and L. Jiao, "EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote*, vol. 60, 2022, Art. no. 5626915.

[87] J. Ma, M. Li, X. Tang, X. Zhang, F. Liu, and L. Jiao, "Homo-heterogenous transformer learning framework for RS scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2223–2239, 2022.

[88] W. Chen, S. Ouyang, W. Tong, X. Li, X. Zheng, and L. Wang, "GCSANet: A global context spatial attention deep learning network for remote sensing scene classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1150–1162, 2022.

[89] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Trans. Pattern Anal.*, vol. 45, no. 2, pp. 2551–2566, Feb. 2023.

[90] Z. Tong, P. Xu, and T. Denœux, "An evidential classifier based on Dempster-Shafer theory and deep learning," *Neurocomputing*, vol. 450, pp. 275–293, 2021.

[91] B. Debaque, M. C. Florea, N. D.-Hindié, and A.-C. B.-Brisset, "Evidential reasoning for ship classification: Fusion of deep learning classifiers," in *Proc. 22th Int. Conf. Inf. Fusion*, 2019, pp. 1–8.

[92] Q. M. Zhou, L. Zhe, R. J. Brooke, M. M. Hudson, and Y. Yuan, "A relationship between the incremental values of area under the ROC curve and of area under the precision-recall curve," *Diagn. Prognostic Res.*, vol. 5, no. 13, pp. 1–15, 2021.

[93] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

[94] R. B. -Cerezo, F. Bianconi, A. Fernández, E. González, and F. D. Maria, "Experimental comparison of color spaces for material classification," *J. Electron. Imag.*, vol. 25, no. 6, Art. no. 061406.

[95] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE J. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[96] X. Cui, Q. Liu, S. Zhang, F. Yang, and D. N. Metaxas, "Temporal spectral residual for fast salient motion detection," *Neurocomputing*, vol. 86, pp. 24–32, 2012.

[97] Y. Fu, W. Li, S. Fan, Y. Jiang, and H. Bai, "CAL-Net: Conditional attention lightweight network for in-orbit landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4408515.

[98] J. Gawlikowski et al., "A survey of uncertainty in deep neural networks," *Artif. Intell. Rev.*, vol. 56, pp. 1513–1589, 2023.

[99] L. Golshani, E. Pasha, and G. Yari, "Some properties of Rényi entropy and Rényi entropy rate," *Inf. Sci.*, vol. 179, no. 14, pp. 2426–2433, 2009.

[100] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE. Trans. Inf. Theory*, vol. 60, no. 11, pp. 6801–6810, Nov. 2014.

Qi Cui received the B.S. degree in information and computational science in 2023 with the College of Science, Northeast Forestry University, Harbin, China. He is working toward the master's degree with the School College, Northeastern University, Shenyang, China.

His current research interests include deep learning and remote sensing scene image classification.

Xiaojian Ma received the B.S. degree in mathematics and applied mathematics from Northeast Forestry University, Harbin, China, in 2000, and the M.S. degree in applied mathematics from Harbin University of Science and Technology, Harbin, in 2003.

She is currently an Associate Professor with Northeast Forestry University. Her research interests include uncertainty theory, image processing, and applied statistics.

Jiaxu Zhang received the B.S. degree in information and computational science in 2023 with the College of Science, Northeast Forestry University, Harbin, China. He is currently working toward the master's degree in mathematics with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, China.

He is working on the research project of "multi-model sensors based" with Sun Yat-sen University. His current research interests include evidence theory, information fusion, and remote sensing.