Multimodal Learning with Uncertainty Quantification based on Discounted Belief Fusion

Grigor Bezirganyan
Aix-Marseille Univ, LIS

Sana Sellami
Aix-Marseille Univ, LIS

Laure Berti-Équille
IRD, ESPACE-DEV

Sébastien Fournier
Aix-Marseille Univ, LIS

Abstract

Multimodal AI models are increasingly used in fields like healthcare, finance, and autonomous driving, where information is drawn from multiple sources or modalities such as images, texts, audios, videos. However, effectively managing uncertainty—arising from noise, insufficient evidence, or conflicts between modalities—is crucial for reliable decision-making. Current uncertainty-aware machine learning methods leveraging, for example, evidence averaging, or evidence accumulation underestimate uncertainties in high-conflict scenarios. Moreover, the state-of-the-art evidence averaging strategy is not order invariant and fails to scale to multiple modalities. To address these challenges, we propose a novel multimodal learning method with order-invariant evidence fusion and introduce a conflict-based discounting mechanism that reallocates uncertain mass when unreliable modalities are detected. We provide both theoretical analysis and experimental validation, demonstrating that unlike the previous work, the proposed approach effectively distinguishes between conflicting and non-conflicting samples based on the provided uncertainty estimates, and outperforms the previous models in uncertainty-based conflict detection.

1 INTRODUCTION

Recently, the use of Artificial Intelligence (AI) has surged dramatically for both autonomous and assisted decision-making. In fields such as healthcare, finance, and autonomous driving, information is often distributed across multiple modalities or views[1]. To effectively leverage this diverse information, multimodal AI models have been developed (Li et al., 2024; Xiao et al., 2020; Lee and Yoo, 2020; Chergui et al., 2024; Krones et al., 2025). However, deep neural networks are prone to making wrong decisions with very high confidence (Guo et al., 2017; Hein et al., 2019). In safety-critical areas, such as healthcare or autonomous driving, this kind of overconfidence can lead to catastrophic results. Hence, when deploying these models in safety-critical areas, it is essential to understand the true uncertainty of the predictions.

Uncertainty in deep learning can appear from multiple sources. Firstly, data can have an inherent amount of randomness (aleatoric uncertainty) that can affect the quality of the model's decisions. Secondly, there can be not enough evidence in the data to make a confident prediction (epistemic uncertainty) (Kiureghian and Ditlevsen, 2009). Another source of uncertainty that is unique to multimodal networks is the conflict between modalities. We define conflict as confident disagreement between several opinions. Conflicts arise from having unreliable sources of information (Martin, 2019). In case of multimodal data, the uncertainties may also come from misaligned modalities or malicious injection of incorrect data in the modalities.

While some multimodal approaches (Marcos Alvarez et al., 2013; Hou et al., 2020) opt to remove conflicting samples from the dataset, in many scenarios, it is more beneficial to make decisions based on existing evidences while providing relatively higher uncertainty estimates. Xu et al. (2024) has proposed an approach for multi-view conflicting information fusion by evidence averaging (RCML). The authors demonstrate the advantage of their fusion method by showcasing superior classification accuracy through experimental validation under both conflicting and non-conflicting settings. However, **the evidence averaging operation is defined only for two views and is not**

---

[1]The terms modality and view will be used interchangeably in this paper

associative (Jøsang, 2016), which limits its application to more than 2 views or modalities, making it non-order-invariant. Another shortcoming of evidence averaging is that **under high levels of conflict, the uncertainty does not increase**, which means that a decision based on two conflictive sources is as reliable as the one based on two non-conflictive sources. This is counter-intuitive, as for example, a decision based on conflicting opinions, shall be less reliable than the decision based on total agreement.

To address these issues, we propose a novel uncertainty aware multimodal fusion approach: Discounting Belief Fusion (DBF), which utilizes conflictive discounting of unreliable modalities, is order-invariant and can effectively scale to multiple modalities. Additionally, the proposed approach assigns higher uncertainties to conflictive decisions, which improves the reliability assessment of the final decision. We provide an experimental validation of our proposed approach, and open-source the source code in a public repository[2].

This paper is structured as follows. Section 2 presents the related work in unimodal and multimodal uncertainty quantification, section 3 presents our proposed approach of discounted conflictive fusion, section 4 presents our experiments and their results, and we conclude the paper with section 5.

2 RELATED WORK

2.1 Uncertainty Quantification in Deep Learning

Gawlikowski et al. (2023) classified uncertainty quantification (UQ) techniques in deep learning into (1) Bayesian Methods (Gal and Ghahramani, 2016; Wilson and Izmailov, 2020), (2) Ensemble Methods (Lakshminarayanan et al., 2017), (3) Single Network Deterministic methods (Malinin and Gales, 2018; Sensoy et al., 2018), and (4) Test-time augmentation methods (Lyzhov et al., 2020). Bayesian methods model uncertainty by placing distributions over weights, which provides better-calibrated predictions and robustness to overfitting, especially in low-data situations. Nevertheless, they often increase the training and inference complexity, require approximations and can be sensitive to the choice of priors (Ovadia et al., 2019). Deep ensembles provide reliable and well-calibrated uncertainty estimates by combining predictions from multiple models, capturing both epistemic and aleatoric uncertainties effectively. However, training and storing multiple independent models in a deep ensemble can be significantly more computationally expensive in terms of memory, training and inference times (Lak-

shminarayanan et al., 2017). Test-time Augmentation naturally introduces variability through multiple input transformations, which can provide a straightforward estimate of prediction uncertainty by measuring the variability in outputs (Ayhan and Berens, 2018). However they significantly increase inference time, do not estimate the epistemic uncertainty and are highly dependent on the augmentation strategy. Single Network Deterministic methods are computationally very efficient and require minimal changes to the network architecture. However, they can struggle to capture the full range of epistemic uncertainty (Ulmer et al., 2023).

Evidential Deep Learning (EDL) (Sensoy et al., 2018), a single-network deterministic method, has recently gained popularity due to its easy adaptation to existing networks and its low training and testing times (Gao et al., 2024). EDL replaces the categorical distribution of the last layers of neural networks with a Dirichlet distribution and uses subjective logic (Jøsang, 2016) to optimize the network parameters. Thanks to its ease of adaptation to existing architectures, and computational efficiency, it was also adapted to multi-modal models, which will be described in the next subsection.

2.2 Uncertainty Quantification in Multimodal and Multi-view Learning

While there are many different approaches for unimodal uncertainty quantification, the multimodal UQ, where we can also have misalignment and conflict between modalities, is relatively less studied. One of the pioneering works, Trusted Multi-view Classification (TMC) (Han et al., 2021) employs Evidential Deep Learning and uses Dempster's combination (or belief constraint fusion in subjective logic) to combine decisions from multiple modalities. However, as shown by Zadeh (1979), Dempster's combination rule can yield counter-intuitive results under a high degree of conflict between information sources. Liu et al. (2022) address this issue by using evidence accumulation (or cumulative belief fusion in subjective logic) instead of Dempster's combination rule. This type of fusion is suitable for fusing independent information sources, where each source contributes new evidence and thus can reduce uncertainty (Jøsang, 2016).

In these approaches, incorporating a new modality decreases uncertainty regardless of the conflict between modalities, while a decision based on conflicting view can be less reliable in practice. To alleviate this issue, Xu et al. (2024) propose Evidential Conflictive Multi-view Learning (ECML), which introduces an average pooling operation (averaging belief fusion in subjective logic). This method reduces the combined un-

---

certainty if the new view has lower uncertainty and increases it if the new view has higher uncertainty. For two views, the combined uncertainty is the harmonic mean of the individual view uncertainties. However, it is possible to have two conflicting views with low uncertainty, resulting in low combined uncertainty when higher uncertainty would be expected. Additionally, the proposed fusion operation is not associative (Jøsang, 2016), making the result of fusing more than two modalities highly dependent on the order of fusion. Huang et al. (2025) proposed a learnable discounting factor for modality reliability, but it uses a fixed value per modality and class, making it less adaptable to discrepancies between training and deployment. In contrast, our method computes reliability per sample, dynamically adjusting to misalignments and noise that may arise in real-world settings. Similar to these works, we also base our approach on evidential deep learning methods due to their computational efficiency and ease of adaptation. Nevertheless, we address shortcomings of the current approaches by presenting an order-invariant and scalable approach that effectively distinguishes between conflictive and non-conflictive samples based on uncertainty estimates.

Jung et al. (2022) and Jung et al. (2024) propose alternative multimodal UQ approaches based on Gaussian processes and neural processes, respectively. While these methods perform well, the Multimodal Gaussian Process is computationally expensive due to its non-parametric nature. The Multimodal Neural Process is relatively faster; however, its results are highly dependent on the chosen context set, and there are currently no theoretically guaranteed methods to obtain an optimal context set.

3 METHODOLOGY

In this section we will first provide a background on subjective logic, motivate and present the generalized belief averaging fusion operator (Jøsang et al., 2017), and finally present our method called Discounted Belief Fusion.

3.1 Subjective Logic and Uncertainty

Evidential Deep Learning uses Subjective Logic to model classification uncertainty by learning evidence scores for each class. Subjective Logic (Jøsang, 2016) is the extension of Dempster–Shafer's theory of belief functions (Dempster, 1968; Shafer, 1976) to allow a bijective mapping between a *subjective opinion* and the parameters of Dirichlet distribution. Specifically, for a classification problem of K classes, the subjective opinion $\boldsymbol{\omega}$ is defined as a triplet $\boldsymbol{\omega} = (\boldsymbol{b}, u, \boldsymbol{a})$,

where the $\boldsymbol{b} = (b_1, \ldots, b_K)^T$ represents the *belief mass* for each class, u represents the *uncertainty mass* and $\boldsymbol{a} = (a_1, \ldots, a_K)^T$ represents the *base rate* or prior probability distribution over the classes. In scenarios where no prior knowledge about the classes is available (as is typical in most deep learning applications), a uniform base rate can be assumed. All these terms are strictly non-negative and are less than or equal to one. Unlike probability theory, the belief masses do not sum up to one, however the additivity property dictates that $\sum_{i=1}^{K} b_i + u = 1$. The base rates can usually be set to $1/K$ in case of absence of a prior belief on the classes. The *projected probability distribution* for class k can be obtained with $P_k = b_k + a_k u, \forall k \in [1, \ldots, K]$.

As mentioned, subjective logic provides a bijective mapping between the multinomial opinions and the parameters of Dirichlet distribution. The Dirichlet distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ is defined by

$$D(\mathbf{p} \mid \boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1}, & \text{for } \mathbf{p} \in \mathcal{S}_K \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $\mathbf{p}$ is the probability of each class, B is the multivariate beta function and $\mathcal{S}_K$ is the K dimensional unit simplex. The Dirichlet distribution is a conjugate prior for categorical distribution, meaning that a prior belief encoded by a Dirichlet distribution can be continuously updated by using data with categorical likelihood to obtain updated Dirichlet posterior distribution. The Dirichlet distribution is a distribution over categorical distributions on $K - 1$ dimensional simplex, and can be used to understand the uncertainties present in the data and the model.

The bijective mapping between the multinomial opinion and Dirichlet distribution can be given with:

$$b_k = \frac{e_k}{S} = \frac{\alpha_k - 1}{S}, u = \frac{K}{S}, \quad (2)$$

where $S = \sum_{k=1}^{K} (e_k + 1) = \sum_{k=1}^{K} \alpha_k$ is called a Dirichlet strength and $\boldsymbol{e} = (e_1, \ldots, e_K)^T = \boldsymbol{\alpha} - 1$ is the accumulated evidence that shows the support for each class. Having these parameters, the probability of each class can be computed by the expectation of the Dirichlet distribution with $p_k = \alpha_k / S$. In the next subsections we will adapt and use different fusion strategies from subjective logic to fuse the evidences collected from various modalities.

3.2 Generalized Belief Averaging

In this subsection, we will theoretically show the shortcoming of the ECML approach (Xu et al., 2024) of using belief averaging fusion and motivate the use of the generalized belief averaging.

Let the dataset be defined as $(x_n^v, y_n)_{n=1}^N$, $v \in [1, \ldots, V]$, where $x_n^v \in X$ represents the n-th data sample from the v-th modality, $y_n \in [0, 1]^K$ is the corresponding one-hot encoded label vector, V is the number of modalities, N is the total number of samples, and K is the number of classes. Having V neural networks $f^v(x^v)$ for each view, we obtain the view-specific multinomial opinions $\boldsymbol{\omega}^v = f^v(x^v)$, and the goal is to fuse these opinions into one final opinion: $\boldsymbol{\omega}^{\diamond V} = \boldsymbol{\omega}^1 \diamond \cdots \diamond \boldsymbol{\omega}^V$.

The Evidential Conflictive Multi-view Learning (ECML) (Xu et al., 2024) uses belief averaging fusion (Jøsang, 2016) for fusing information from multiple views. The averaging operation for two views is defined as:

$$\boldsymbol{\omega}^{1\diamond2} = \boldsymbol{\omega}^1 \diamond \boldsymbol{\omega}^2 = (\boldsymbol{b}^{1\diamond2}, u^{1\diamond2}, \boldsymbol{a}^{1\diamond2})$$
$$b_k^{1\diamond2} = \frac{b_k^1 u^2 + b_k^2 u^1}{u^1 + u^2}, \quad u^{1\diamond2} = \frac{2u^1 u^2}{u^1 + u^2}, \quad (3)$$
$$\boldsymbol{a}_k^{A\diamond B} = \frac{\boldsymbol{a}_k^A + \boldsymbol{a}_k^B}{2}$$

In terms of evidences, the fusion can be defined as

$$\boldsymbol{e}^{1\diamond2} = \frac{1}{2}\left(\boldsymbol{e}^1 + \boldsymbol{e}^2\right). \quad (4)$$

For multiple views, the fusion can be performed with $\boldsymbol{\omega}^{\diamond V} = \boldsymbol{\omega}^1 \diamond \boldsymbol{\omega}^2 \diamond \ldots \diamond \boldsymbol{\omega}^V$. This averaging belief fusion operator, however, is not associative (Jøsang, 2016).

Proposition 1. *When using an averaging belief fusion operator, the evidence associated with previously fused terms is reduced by a factor of two each time a new term is incorporated, relative to the evidence of the newly fused term.*

Proof. The evidence after sequentially fusing V terms using the averaging belief fusion operator, denoted as $\boldsymbol{e}^{\diamond V}$, can be computed as follows:

$$\boldsymbol{e}^{\diamond V} = \left(\ldots\left(\left(\boldsymbol{e}^1 + \boldsymbol{e}^2\right)\frac{1}{2} + \boldsymbol{e}^3\right)\frac{1}{2} + \cdots + \boldsymbol{e}^V\right)\frac{1}{2}$$
$$= \frac{1}{2^{V-1}}\boldsymbol{e}^1 + \frac{1}{2^{V-1}}\boldsymbol{e}^2 + \frac{1}{2^{V-2}}\boldsymbol{e}^3 + \cdots + \frac{1}{2}\boldsymbol{e}^V. \quad (5)$$

The recursive fusion process is initiated by fusing the first two terms, $\boldsymbol{e}^1$ and $\boldsymbol{e}^2$, with equal weights $\frac{1}{2}$. Each subsequent term $\boldsymbol{e}^v$ ($v > 2$) is then added and averaged iteratively. This results in the evidence weight of each previously fused term being halved with every new fusion step, leading to the final expression where the evidence of each term $\boldsymbol{e}^v, v \neq 1$ is scaled by $\frac{1}{2^{V-v+1}}$, and $\boldsymbol{e}^1$ is scaled by $\frac{1}{2^{V-1}}$. □

As an example of Proposition 1, consider a scenario with three modalities. For a specific class k, the first

modality provides an evidence value of 3, the second modality an evidence of 5, and the third modality an evidence of 10. When fusing these evidences using BAF, we compute:

$$e_k^{1\diamond2\diamond3} = \frac{\left(\frac{3+5}{2}\right) + 10}{2} = \frac{3+5}{4} + \frac{10}{2} = 7$$

Here, the evidence values from the first two modalities (3 and 5) are divided by 4, while the evidence value from the third modality (10) is divided by 2. Consequently, the contributions from the first two modalities are weighted only half as much as the third modality. This imbalance leads to a lack of commutativity in the fusion process, meaning that: $e_k^{1\diamond2\diamond3} \neq e_k^{3\diamond2\diamond1}$. To illustrate, if we change the fusion order:

$$e_k^{3\diamond2\diamond1} = \frac{\left(\frac{10+5}{2}\right) + 3}{2} = 5.25$$

This discrepancy is counter-intuitive and significantly hinders the scalability of the algorithm.

To make the fusion process scalable and efficient for many sources, we suggest to use the generalized version of the belief averaging fusion for multiple sources (Jøsang et al., 2017), with:

$$\boldsymbol{b}^{\diamond V} = \frac{\sum_{v=1}^V \left(\boldsymbol{b}^v \prod_{i \neq v} u^i\right)}{\sum_{v=1}^V \left(\prod_{i \neq v} u^i\right)},$$
$$u^{\diamond V} = \frac{V \prod_{v=1}^V u^v}{\sum_{v=1}^V \left(\prod_{i \neq v} u^i\right)}, \quad \boldsymbol{a}^{\diamond V} = \frac{\sum_{v=1}^V \boldsymbol{a}^v}{V}. \quad (6)$$

The generalized belief averaging fusion in terms of evidences will then be:

$$\boldsymbol{e}^{\diamond V} = \frac{\sum_{v=1}^V \boldsymbol{e}^v}{V}. \quad (7)$$

As we can see, unlike the standard pairwise averaging approach, the generalized belief averaging fusion for multiple sources simultaneously incorporates all V sources, ensuring a uniform weighting scheme.

3.3 Discounted Belief Fusion

In this subsection, we demonstrate that the fusion approach described earlier does not adequately address the increased uncertainty in conflicting samples and propose the Discounted Belief Fusion (DBF) solution to mitigate this issue. The general pipeline of the DBF approach is shown in Figure 1.

Proposition 2. *The uncertainty estimate $u^{\diamond V}$ in equation 6 is the harmonic mean of the individual uncertainties $u^1, u^2, \ldots, u^V$.*
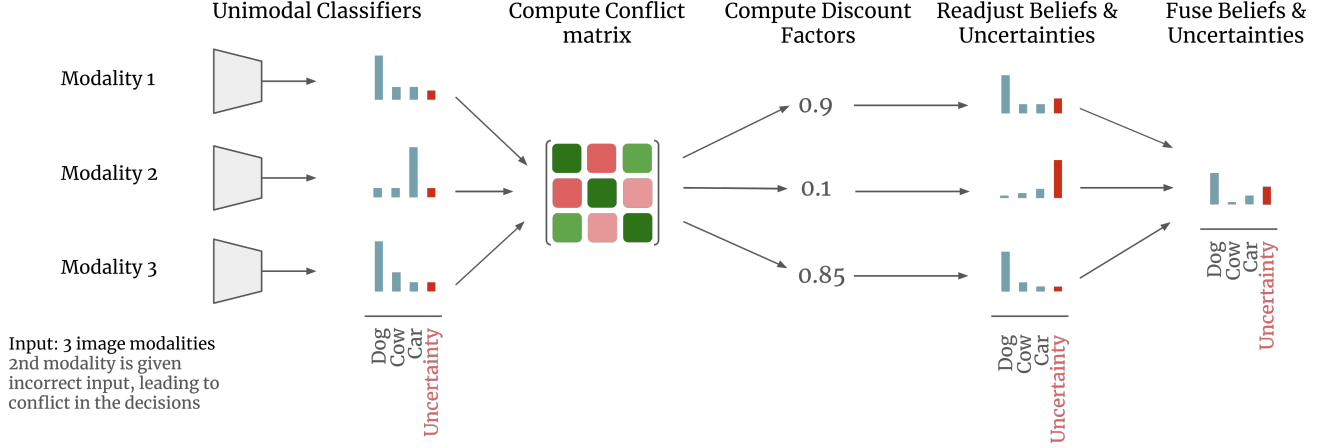
Figure 1: The general pipeline of the Discounted Belief Fusion: First, subjective opinions are being formed for each modality using unimodal classifiers. Then conflict matrix is computed, which is then used to compute discounting factors for each modality. Finally, the beliefs of each modality are being adjusted using the discounting factors and are fused together using generalized belief averaging.

Proof. From equation 6, the uncertainty after fusion is given by:

$$u^{\diamond V} = \frac{V \prod_{v=1}^{V} u^v}{\sum_{v=1}^{V} \left(\prod_{i \neq v} u^i \right)}. \qquad (8)$$

Next, observe that the denominator can be rewritten as:

$$\sum_{v=1}^{V} \left(\prod_{i \neq v} u^i \right) = \sum_{v=1}^{V} \frac{\prod_{i=1}^{V} u^i}{u^v}. \qquad (9)$$

Substituting into the expression for $u^{\diamond V}$, we obtain:

$$u^{\diamond V} = \frac{V \prod_{v=1}^{V} u^v}{\sum_{v=1}^{V} \frac{\prod_{i=1}^{V} u^i}{u^v}} = \frac{V}{\sum_{v=1}^{V} \frac{1}{u^v}}. \qquad (10)$$

The expression $\frac{V}{\sum_{v=1}^{V} \frac{1}{u^v}}$ is the definition of the harmonic mean of the individual uncertainties $u^1, \dots, u^V$. $\qquad \square$

Proposition 2 suggests that the fused uncertainty, as the harmonic mean of individual uncertainties, remains low even in cases of conflict between modalities with low uncertainty. However, in an ideal scenario, conflicting views should result in increased fused uncertainty, reflecting the difficulty and reduced reliability of making a decision in such cases.

Shafer (1976) introduced a discounting operation for combining beliefs in belief theory. The discounting operation is meant to reduce the belief masses of unreliable sources by a constant, and assign the discounted masses to the uncertainty. Since there exists a bijective

mapping between belief theory and subjective logic under uniform base rates, we can translate the discounting operation to subjective logic, which will become:

$$\boldsymbol{b}' = \eta \boldsymbol{b}, \qquad u' = 1 - \eta + \eta u. \qquad (11)$$

The η represents the degree of the reliability of the evidence. When $\eta = 0$, the updated uncertainty u' becomes 1, indicating complete uncertainty. Conversely, when $\eta = 1$, the uncertainty remains unchanged from its original value, u. For intermediate values of η, the uncertainty u' varies linearly between these two extremes. In our case, we assume that the reliability of the evidences are connected to the degree of conflict between them. In case of 2 views with high conflict we will regard both evidences as unreliable, since based on existing evidence it is not possible to correctly assess which source is truly the reliable one. If we have multiple views, we want to count how many other views each view conflicts with. If a view is in conflict with many views, it has to be regarded as more unreliable, and has to be discounted more, than views with less conflict. To assess the amount of conflict between the evidences, we will use the Degree of Conflict measure from subjective logic (Jøsang, 2016).

Definition 1 (Degree of Conflict). *The degree of conflict* DC *between two subjective opinions* $\boldsymbol{\omega}^1$ *and* $\boldsymbol{\omega}^2$ *is defined as:*

$$\begin{aligned} \mathrm{PD}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right) &= \frac{\sum_{k=1}^{K} \left| p_k^1 - p_k^2 \right|}{2}, \\ \mathrm{CC}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right) &= \left(1 - u^1\right)\left(1 - u^2\right), \\ \mathrm{DC}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right) &= \mathrm{PD}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right) \cdot \mathrm{CC}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right), \end{aligned} \qquad (12)$$

where p_k^v *denotes the projected probability for class* k *under opinion* v, *and* u^v *represents the uncertainty*
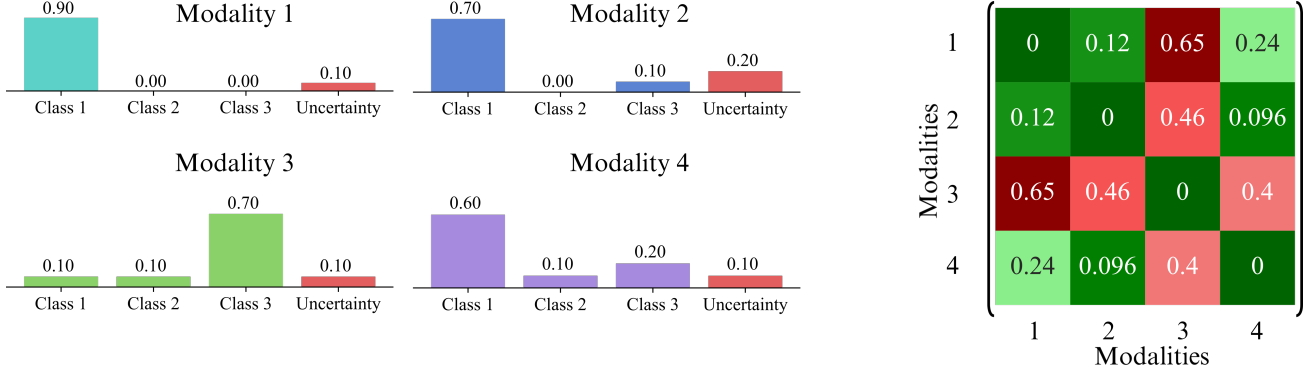
Figure 2: The Conflict Matrix $\boldsymbol{C}$ for 4 opinions.
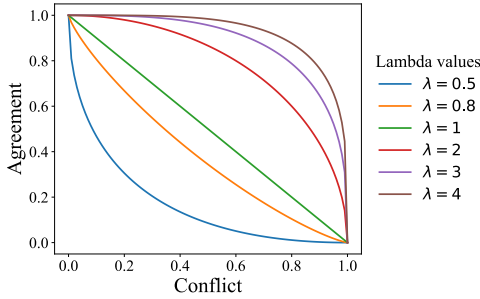
Figure 3: Conflict versus Agreement according to equation 14. Higher values of λ reduce discounting for lower levels of conflict.

of opinion v. The function $\text{PD}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right)$ *is the projected distance between two opinions, which measures the dissimilarity between the projected probabilities of the opinions and equals zero only when both sets of probabilities are identical. The function* $\text{CC}\left(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2\right)$ *is the conjunctive certainty between the two opinions, which takes a value of one when both opinions are fully confident (i.e., $u^1 = u^2 = 0$) and approaches zero as the uncertainty of one or both opinions increases.*

Analyzing the Degree of Conflict, we observe that it is significantly high only when two conditions are met: (1) both opinions exhibit low uncertainty, and (2) the projected probabilities of these opinions differ substantially, indicating a high distance between them. **Consequently, uncertain modalities cannot contribute to high conflict, as conflict arises predominantly between opinions that are both confident and significantly divergent.**

Equation 12 defines the degree of conflict for two opinions. We will use this metric to understand how conflicting an opinion is with the other opinions. For that, we compute the pairwise degree of conflicts between each pair of opinions, and store them in a matrix $\boldsymbol{C}$ (see Figure 2 as an example). Then, we can compute

an agreement matrix with $\boldsymbol{A} = 1 - \boldsymbol{C}$, and we can compute the discounting factor η^v as the product of the row elements of the agreement matrix with:

$$C_{ij} = \text{DC}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j), \quad \boldsymbol{A} = 1 - \boldsymbol{C},$$
$$\eta^v = \prod_{i=1}^{V} A_{vi}, \quad \forall v \in [1, \dots V]. \tag{13}$$

Looking back to the example in Figure 2, we can see that by using equation 13 to compute the discounting factor, the opinions that are in conflict with many other views (opinion 3 in the example), will be discounted much more than the other opinions. We can compute the updated belief masses and uncertainties with equation 13, and fuse the results with generalized belief averaging (equation 6).

However, due to the product in the equation 13, having multiple views can cause the discounting factor get too small, and hence it can increase the uncertainty substantially, even if there is very little conflict between the modalities. To overcome this issue, following the work of Martin (2019) we modify the agreement matrix calculation equation 13 to be:

$$\boldsymbol{A} = (1 - \boldsymbol{C}^\lambda)^{1/\lambda}, \tag{14}$$

where λ is a hyperparameter controlling how strong the discounting shall be relative to conflict. In Figure 3 the relation between conflict and agreement under different values of λ are shown.

To illustrate the differences between the fusion functions, we will use the famous example of Zadeh (1979), adapted to a multimodal setting. Here, two conflicting modalities allocate belief masses as follows: the first source assigns 0.99 to class 1 and 0.01 to class 3, while the second assigns 0.99 to class 2 and 0.01 to class 3. As we can see in Table 1, despite class 3 being unlikely for both, Dempster's rule (Belief Constrained Fusion) combines them into class 3 with full confidence, ignoring the conflict. The Cumulative Belief Fusion and

Table 1: Beliefs and Uncertainty values for fusing two modalities' opinions with Belief Constraint Fusion (BCF), Cumulative Belief Fusion (CMF), Belief Averaging Fusion (BAF), and our proposed Discounting Belief Fusion (DBF).

Fusion Method	b_1	b_2	b_3	u
Modality 1	0.99	0.00	0.01	0
Modality 2	0.00	0.99	0.01	0
BCF	0.0000	0.0000	1.0000	0
CBF	0.4950	0.4950	0.0100	0
BAF	0.4950	0.4950	0.0100	0
DBF $\lambda = 1$	0.0050	0.0050	0.0001	0.9900
DBF $\lambda = 3$	0.1533	0.1533	0.0031	0.6903

the Averaging Belief Fusion provide better results in this case, by splitting most of the belief mass between the classes 1 and 2, however, the uncertainty is still 0, which means they are still confident in the decision. In contrast, discounting fusion shifts most belief mass to uncertainty, indicating the unreliability of the evidences acquired, while the parameter λ in equation 14 controls the strength of discounting based on conflict.

3.4 Training Multimodal Evidential Neural Networks

To fuse information from different modalities as described in the previous subsections, we first need to extract subjective opinions from each modality. For that, we construct evidential neural networks for each modality, which output non-negative evidence values to be later combined using a chosen fusion function. Simultaneous availability of the modalities during training is preferred but not mandatory. If simultaneous availability is not possible, one can effectively integrate new modalities without complete retraining, by training only a classifier for the new modality, and if necessary, fine-tuning the other modalities.

Transforming a traditional deep neural network into an evidential one is straightforward. It involves replacing the softmax activation function in the output layer with a non-negative, monotonically increasing activation function. Common choices include ReLU, softplus, and exponential functions. While Xu et al. (2024) utilized a softplus activation layer, our experiments revealed that the slow growth of the softplus function resulted in small evidence values, leading to consistently high uncertainties across modalities. Recently, Pandey and Yu (2023) demonstrated the superiority of the exponential activation function over other alternatives for evidential networks. Hence, we employ the exponential activation function to obtain evidence

values. To mitigate numerical instability, we cap the activation values at 10^{13} with $f(x) = \frac{10^{13}}{1+10^{13}e^{-x}}$.

For training the neural network, we follow Xu et al. (2024) to use 3 part loss, since it also aims to minimize the conflict between modalities. The first part, is the adapted cross-entropy loss per modality and for fused result:

$$L_{ace}(\boldsymbol{\alpha}_n) = \int \left[\sum_{j=1}^{K} -y_{nj} \log p_{nj} \right] \frac{\prod_{j=1}^{K} p_{nj}^{\alpha_{nj}-1}}{B(\boldsymbol{\alpha}_n)} d\mathbf{p}_n$$

$$= \sum_{j=1}^{K} y_{nj} \left(\psi(S_n) - \psi(\alpha_{nj}) \right), \tag{15}$$

where $\psi(\cdot)$ is the digamma function, and $\boldsymbol{\alpha}_n$ are the Dirichlet parameters of the n-th sample. To force the evidences of incorrect labels to be lower, a regularization term is employed to minimize the Kullback-Leiber divergence of modified Dirichlet distribution parameters and the uniform Dirichlet distribution:

$$L_{KL}(\boldsymbol{\alpha}_n) = KL\left[D(\boldsymbol{p}_n \mid \tilde{\boldsymbol{\alpha}}_n) \| D(\boldsymbol{p}_n \mid \mathbf{1}) \right]$$

$$= \log \left(\frac{\Gamma\left(\sum_{k=1}^{K} \tilde{\alpha}_{nk} \right)}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\tilde{\alpha}_{nk})} \right)$$

$$+ \sum_{k=1}^{K} (\tilde{\alpha}_{nk} - 1) \left[\psi(\tilde{\alpha}_{nk}) - \psi\left(\sum_{j=1}^{K} \tilde{\alpha}_{nj} \right) \right], \tag{16}$$

where $\tilde{\alpha}_n = \mathbf{y}_n + (\mathbf{1} - \mathbf{y}_n) \odot \boldsymbol{\alpha}_n$ are the Dirichlet parameters after removing the non-misleading evidence, and $\Gamma(\cdot)$ is the gamma function. We can gradually increase the effect of the KL Divergence loss with an annealing coefficient, to allow the network to explore the parameter, and avoid early convergence of the misclassified samples to the uniform distribution (Sensoy et al., 2018). The annealing coefficient σ_t can be chosen as $\sigma_t = \min(1.0, t/T) \in [0, 1]$, where t is the current epoch and T is the annealing step.

$$L_{acc}(\boldsymbol{\alpha}_n) = L_{ace}(\boldsymbol{\alpha}_n) + \sigma_t L_{KL}(\boldsymbol{\alpha}_n) \tag{17}$$

Finally, a consistency loss is employed to minimize the degree of conflict between the modalities:

$$L_{con} = \frac{1}{V-1} \sum_{p=1}^{V} \left(\sum_{q \neq p}^{V} DC(\boldsymbol{\omega}_n^p, \boldsymbol{\omega}_n^q) \right) \tag{18}$$

The total loss is calculated with:

$$L = L_{acc}(\boldsymbol{\alpha}_n) + \beta \sum_{v=1}^{V} L_{acc}(\boldsymbol{\alpha}_n^v) + \gamma L_{con}. \tag{19}$$

4 EXPERIMENTS

4.1 Experimental setup

Five multimodal / multi-view datasets will be used for comparison: HandWritten (Duin, 1998), CUB (Wah et al., 2011), Scene15 (Ali and Zafar, 2018), Caltech101 (Li et al., 2022) and PIE (Gross et al., 2008). Please check the supplementary material for the descriptions and details about the datasets. As baselines we use three evidential multi-view methods with different opinion aggregation strategies. The Trusted Multiview Classification (Han et al., 2021) uses Dempster's combination rule or Belief Constraint Fusion (BCF), Trusted Multiview Deep Learning with Opinion Aggregation (Liu et al., 2022) uses Cumulative Belief Fusion (CBF), Evidential Conflictive Multiview Learning (Xu et al., 2024) uses the Belief Averaging Fusion (BAF) and the Generalized Belief Averaging Fusion (GBAF, described in subsection 3.2).

We conduct all training and evaluation on a local cluster equipped with a combination of Nvidia GeForce RTX 2080 and RTX 4090 GPUs. Each experiment is repeated 10 times using different random seeds to ensure robustness. Hyperparameters are selected through 10-fold cross-validation prior to the main experiments. Further details on training procedures and hyperparameter configurations can be found in the supplementary material.

4.2 Experimental Results

We evaluate the proposed method by training baseline models on five datasets and testing them on clean and conflicting test sets. Conflicts are introduced by randomly replacing a modality in each sample with one from a different class, simulating misalignment. This setup helps assess whether models produce higher uncertainties on conflicting set compared to clean data.

Figure 4 and Table 2 show that on most of the benchmark dataset the proposed Discounted Belief Fusion achieves better separation between uncertainties of conflictive and non-conflictive sets compared to the baseline methods. However, for the CUB and Scene15 datasets, we observe less effective separation. To understand this, we analyze the uni-modal uncertainties before fusion for each dataset in Figure 5. For the CalTech dataset, the uni-modal uncertainties are very low, which leads to a high degree of conflict when misalignment is introduced, resulting in clear separation between conflictive and non-conflictive sets. In contrast, the Scene15 dataset shows high initial uncertainties across modalities, causing low degree of conflict even after misalignment. Thus, although there is misalignment, the conflict remains minimal, and fused uncer-
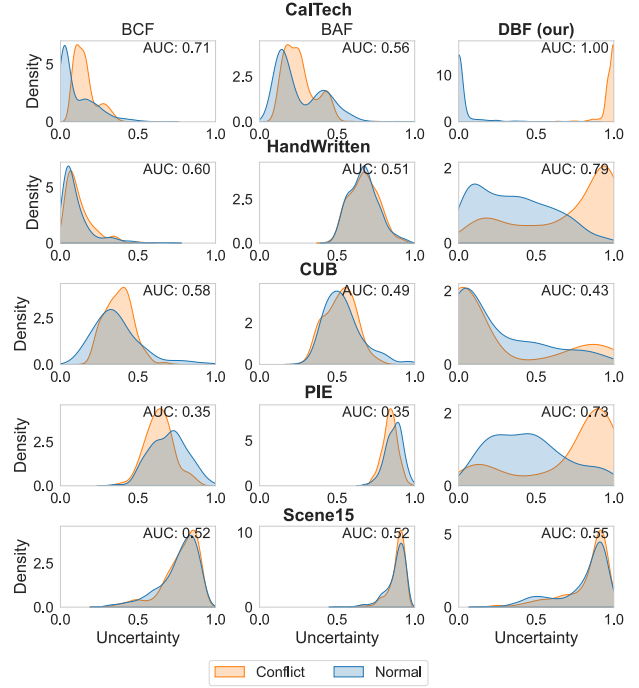

Figure 4: Uncertainty distributions on normal and conflictive test sets using Belief Constraint Fusion (BCF), Belief Averaging Fusion (BAF), and Discounted Belief Fusion (DBF). $\lambda = 1$ is used for all datasets.

tainties are high for both sets. For CUB dataset, we observe that one modality has low uncertainty while the other is highly uncertain, hence according to Equation 12, the degree of conflict is low since the uncertain modality contributes minimally to the final decision. These results indicate that the proposed approach provides higher uncertainties for conflictive sets when the degree of conflict between modalities is high.

Figure 6 demonstrates how the parameter λ controls the strictness of discounting. Lower λ values lead to more discounting and higher uncertainties. The choice of λ can be made using a validation set or based on problem-specific requirements, depending on the desired strictness against conflict.

In the ablation studies, we tested all models using both softplus (for consistency with prior work) and exponential activation functions (see Appendix E for details). Results show that while the exponential activation improves model performance, it alone does not effectively separate uncertainties between conflictive and non-conflictive samples.

Our experiments showed that the proposed approach does not impact classification accuracy (see Appendix D). In most cases, the fusion function had minimal effect on final accuracy, allowing the proposed method to

Table 2: AUC scores of conflictive sample detection based on the uncertainty estimates, using Belief Constraint Fusion (BCF), Cumulative Belief Fusion (CBF), Belief Averaging Fusion (BAF), Generalized Belief Averaging Fusion (GBAF), Discounted Belief Fusion (DBF). $\lambda = 1$ is used for all datasets.

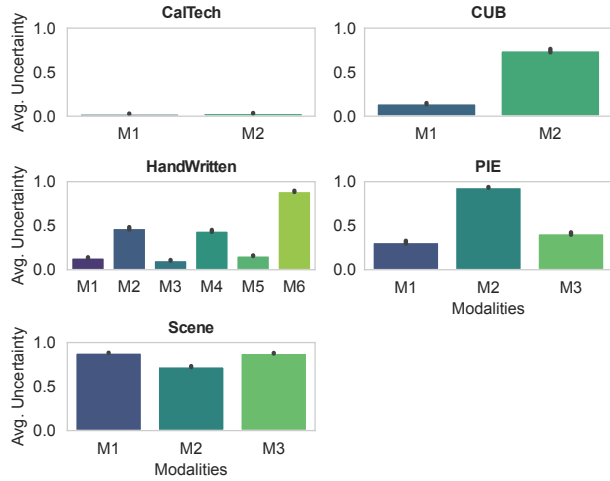|	CUB	CalTech	HandWritten	PIE	Scene15
BCF	0.53 ± 0.04	0.72 ± 0.02	0.61 ± 0.01	0.36 ± 0.02	0.54 ± 0.01
CBF	0.48 ± 0.05	0.56 ± 0.02	0.49 ± 0.02	0.35 ± 0.03	0.53 ± 0.02
BAF	0.50 ± 0.04	0.54 ± 0.02	0.51 ± 0.01	0.35 ± 0.03	0.53 ± 0.02
GBAF	0.48 ± 0.03	0.55 ± 0.02	0.51 ± 0.02	0.37 ± 0.03	0.53 ± 0.01
DBF (our)	$\mathbf{0.57 \pm 0.06}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.71 \pm 0.04}$	0.53 ± 0.01

Figure 5: Average uncertainty values across different modalities (denoted as M1, M2, etc.) for CalTech, CUB, HandWritten, Scene, and PIE datasets. Error bars indicate the standard deviation, highlighting the low variability in uncertainty measurements.

deliver improved uncertainty estimates without compromising the predictive performance of the models.

Table 3: Training times of different algorithms for each dataset. The means and standard deviations of 10 runs are reported in seconds.

Dataset	BAF	BCF	DBF (Our)
Scene15	757 ± 22	854 ± 25	798 ± 11
CalTech	413 ± 2	419 ± 5	455 ± 3
PIE	133 ± 2	146 ± 2	140 ± 2
HandWritten	607 ± 4	752 ± 6	602 ± 12
CUB	94 ± 0.5	99 ± 1	105 ± 1

As discussed in the Appendix A, the conflict matrix computation has a quadratic time complexity relative to the number of modalities. However, since the number of modalities is typically not large, the practical time overhead of using DBF remains low. To validate this, we present the training times of various baseline
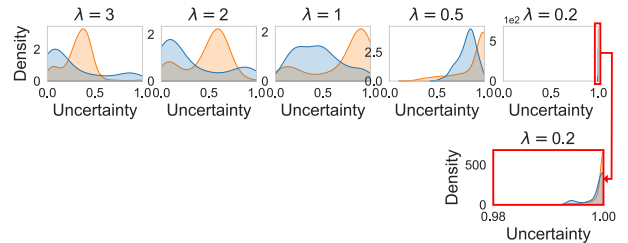

Figure 6: The uncertainty distributions for conflictive and non-conflictive sets with different values of λ on PIE dataset.

algorithms alongside the DBF method in Table 3. As observed, DBF is not consistently slower than other algorithms, and even when it is, the difference is not much.

5 CONCLUSION

We introduced Discounted Belief Fusion (DBF), which discounts beliefs from conflicting sources towards uncertainty, achieving better separation between conflictive and non-conflictive samples compared to existing baselines. Extensive experiments on benchmark datasets show that DBF yields higher uncertainty for conflictive samples, especially under high conflict. Our analysis reveals that DBF's performance adapts to initial modality uncertainty levels, making it suitable for varying alignment quality. Moreover, DBF maintains high classification accuracy, providing robust uncertainty estimates without compromising predictive performance, thereby enhancing interpretability in multimodal learning.

In future work, we plan to further validate our approach using larger benchmark datasets designed for uncertainty-aware multimodal learning, such as the LUMA dataset Bezirganyan et al. (2024). Additionally, we plan to explore modality interactions and dependency structures, which could further enhance our uncertainty quantification strategy.

References

Ali, N. and Zafar, B. (2018). 15-Scene Image Dataset.

Ayhan, M. S. and Berens, P. (2018). Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In *Medical Imaging with Deep Learning*.

Bezirganyan, G., Sellami, S., Berti-Équille, L., and Fournier, S. (2024). Luma: A benchmark dataset for learning from uncertain and multimodal data. *arXiv preprint arXiv:2406.09864*.

Chergui, A., Bezirganyan, G., Sellami, S., Berti-Équille, L., and Fournier, S. (2024). Mixmas: A framework for sampling-based mixer architecture search for multimodal fusion and learning. In *2024 IEEE International Conference on Big Data (BigData)*, pages 3254–3257. IEEE.

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232.

Duin, R. (1998). Multiple Features. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5HC70.

Gal, Y. and Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR.

Gao, J., Chen, M., Xiang, L., and Xu, C. (2024). A comprehensive survey on evidential deep learning and its applications. *arXiv preprint arXiv:2409.04720*.

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al. (2023). A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, pages 1–77.

Gross, R., Matthews, I., Cohn, J., Kanade, T., and Baker, S. (2008). Multi-pie. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1321–1330. PMLR.

Han, Z., Zhang, C., Fu, H., and Zhou, J. T. (2021). Trusted multi-view classification. *arXiv preprint arXiv:2102.02051*.

Hein, M., Andriushchenko, M., and Bitterwolf, J. (2019). Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hou, D., Cong, Y., Sun, G., Dong, J., Li, J., and Li, K. (2020). Fast multi-view outlier detection via deep encoder. *IEEE Transactions on Big Data*, 8(4):1047–1058.

Huang, L., Ruan, S., Decazes, P., and Denœux, T. (2025). Deep evidential fusion with uncertainty quantification and reliability learning for multimodal medical image segmentation. *Information Fusion*, 113:102648.

Jøsang, A. (2016). *Subjective logic*, volume 3. Springer.

Jøsang, A., Wang, D., and Zhang, J. (2017). Multi-source fusion in subjective logic. In *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–8.

Jung, M. C., Zhao, H., Dipnall, J., and Du, L. (2024). Beyond unimodal: Generalising neural processes for multimodal uncertainty estimation. *Advances in Neural Information Processing Systems*, 36.

Jung, M. C., Zhao, H., Dipnall, J., Gabbe, B., and Du, L. (2022). Uncertainty estimation for multi-view data: The power of seeing the whole picture. *Advances in Neural Information Processing Systems*, 35:6517–6530.

Kiureghian, A. D. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112.

Krones, F., Marikkar, U., Parsons, G., Szmul, A., and Mahdi, A. (2025). Review of multimodal machine learning approaches in healthcare. *Information Fusion*, 114:102690.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30.

Lee, S. I. and Yoo, S. J. (2020). Multimodal deep learning for finance: integrating and forecasting international stock markets. *The Journal of Supercomputing*, 76:8294–8312.

Li, F.-F., Andreeto, M., Ranzato, M., and Perona, P. (2022). Caltech 101.

Li, Y., Daho, M. E. H., Conze, P.-H., Zeghlache, R., Le Boité, H., Tadayoni, R., Cochener, B.,

Lamard, M., and Quellec, G. (2024). A review of deep learning-based information fusion techniques for multimodal medical image classification. *Computers in Biology and Medicine*, page 108635.

Liu, W., Yue, X., Chen, Y., and Denoeux, T. (2022). Trusted Multi-View Deep Learning with Opinion Aggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7):7585–7593.

Lyzhov, A., Molchanova, Y., Ashukha, A., Molchanov, D., and Vetrov, D. (2020). Greedy policy search: A simple baseline for learnable test-time augmentation. In *Conference on uncertainty in artificial intelligence*, pages 1308–1317. PMLR.

Malinin, A. and Gales, M. (2018). Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31.

Marcos Alvarez, A., Yamada, M., Kimura, A., and Iwata, T. (2013). Clustering-based anomaly detection in multi-view data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1545–1548.

Martin, A. (2019). Conflict management in information fusion with belief functions. *Information quality in information fusion and decision making*, pages 79–97.

Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., and Snoek, J. (2019). Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems*, 32.

Pandey, D. S. and Yu, Q. (2023). Learn to accumulate evidence from all training samples: theory and practice. In *International Conference on Machine Learning*, pages 26963–26989. PMLR.

Sensoy, M., Kaplan, L., and Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton university press.

Ulmer, D. T., Hardmeier, C., and Frellsen, J. (2023). Prior and posterior networks: A survey on evidential deep learning methods for uncertainty estimation. *Transactions on Machine Learning Research*.

Wah, C., Branson, S., Welinder, P., Perona, P., and Belongie, S. (2011). Caltech-ucsd birds-200-2011 (cub-200-2011). Technical Report CNS-TR-2011-001, California Institute of Technology.

Wilson, A. G. and Izmailov, P. (2020). Bayesian deep learning and a probabilistic perspective of general-ization. *Advances in neural information processing systems*, 33:4697–4708.

Xiao, Y., Codevilla, F., Gurram, A., Urfalioglu, O., and López, A. M. (2020). Multimodal end-to-end autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):537–547.

Xu, C., Si, J., Guan, Z., Zhao, W., Wu, Y., and Gao, X. (2024). Reliable conflictive multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16129–16137.

Zadeh, L. (1979). On the validity of dempster's rule of combination of evidence. Technical Report UCB/ERL M79/24.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes. Please refer to the Appendix A.

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Yes

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. Yes

   (b) Complete proofs of all theoretical results. Yes

   (c) Clear explanations of any assumptions. Yes

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). Yes, code included in an anonymized repository

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Yes. Please refer to Appednix B.

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Yes

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. Yes

   (b) The license information of the assets, if applicable. Yes

   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable

   (d) Information about consent from data providers/curators. Not Applicable

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. Not Applicable

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

Grigor Bezirganyan, Sana Sellami, Laure Berti-Équille, Sébastien Fournier

Supplementary Material

The supplementary material is organized as follows: In Section A, we present the algorithm for the proposed Discounted Belief Fusion (DBF) and its time complexity analysis. Section B provides details on the training process and the hyperparameters used. In Section C, we offer additional information about the benchmark datasets employed in the paper. Section D includes supplementary results that could not be included in the main paper, and Section E presents ablation studies.

A DISCOUNTED BELIEF FUSION ALGORITHM

Algorithm 1 Discounted Belief Fusion (DBF) Algorithm

1: **Input:** Set of V opinions $\leftarrow \{\boldsymbol{\omega}^1, \boldsymbol{\omega}^2, \ldots, \boldsymbol{\omega}^V\}$ with K classes, hyperparameter λ
2: **for** $i = 1$ to V **do**
3:    **for** $j = 1$ to V **do**
4:       Compute projected distance: $\text{PD}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j) \leftarrow \frac{1}{2} \sum_{k=1}^{K} |p_k^i - p_k^j|$
5:       Compute conjunctive certainty: $\text{CC}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j) \leftarrow (1 - u^i)(1 - u^j)$
6:       Compute degree of conflict: $\text{DC}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j) \leftarrow \text{PD}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j) \times \text{CC}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j)$
7:       Compute Conflict Matrix: $C_{ij} \leftarrow \text{DC}(\boldsymbol{\omega}^i, \boldsymbol{\omega}^j)$
8:       Compute Agreement Matrix: $A_{ij} \leftarrow (1 - (C_{ij})^\lambda)^{1/\lambda}$
9:    **end for**
10: **end for**
11: **for** $v = 1$ to V **do**
12:    Compute discounting factor: $\eta^v \leftarrow \prod_{i=1}^{V} A_{vi}$
13:    Apply discounting: $\boldsymbol{b}^{v'} \leftarrow \eta^v \boldsymbol{b}^v, \ u^{v'} \leftarrow 1 - \eta^v + \eta^v u^v$
14: **end for**
15: Compute fused belief masses: $\mathbf{b}^{\diamond V} \leftarrow \frac{\sum_{v=1}^{V} \left(\boldsymbol{b}^v \prod_{i \neq v} u^i \right)}{\sum_{v=1}^{V} \left(\prod_{i \neq v} u^i \right)}$
16: Compute fused uncertainty: $u^{\diamond V} \leftarrow \frac{V \prod_{v=1}^{V} u^v}{\sum_{v=1}^{V} \left(\prod_{i \neq v} u^i \right)}$
17: **return** Fused opinion $\boldsymbol{\omega}^\diamond$ with $\boldsymbol{b}^\diamond = [b_1, \ldots, b_K]$ and uncertainty u

---

The time complexity of the proposed Discounting Belief Fusion (DBF) described in Algorithm 1 is as follows. Line 4 has a complexity of $\mathcal{O}(K)$, while lines 5-8 each have $\mathcal{O}(1)$ complexity, resulting in a complexity of $\mathcal{O}(V^2 K)$ for lines 2-10. Lines 12-13 have a complexity of $\mathcal{O}(V + K)$, leading to a complexity of $\mathcal{O}(V^2 + VK)$ for lines 11-14. Lastly, lines 15-16 contribute $\mathcal{O}(V^2 K)$. Therefore, the overall time complexity of the algorithm is $\mathcal{O}(V^2 K)$.

B TRAINING DETAILS AND HYPERPARAMETERS

In this section, we provide additional details about the training process for both the baseline models and the proposed approach, with the aim of enhancing the reproducibility of the presented results.

We utilize 2-layer fully connected neural networks with ReLU activations for retrieving the evidences for each modality. In the final layer, we use either a softplus activation function (for baseline models, as proposed in their respective papers) or an exponential activation, capped at 10^{13} for numerical stability. To maintain differentiability, the activation function is modified to $f(x) = \frac{10^{13}}{1 + 10^{13} e^{-x}}$, which behaves exponentially for small inputs but gradually flattens as x increases. To determine the optimal hyperparameters, we conducted a hyper-

Table 4: Hyperparameters for Different Datasets

Dataset	Learning Rate	Annealing Step	Gamma	Weight Decay
CUB	0.003	10	1	0.0001
HandWritten	0.003	50	0.7	0.00001
PIE	0.003	50	0.5	0.00001
CalTech	0.0003	30	0.5	0.0001
Scene	0.01	30	0.5	0.00001

parameter search using 10-fold cross-validation. To minimize time and resource consumption, this search was only performed on the ECML model Xu et al. (2024). This ensures that any slight performance differences due to hyperparameters would favor the baseline rather than our proposed approach. The selected hyperparameters are presented in Table 4. All experiments have been performed 10 times with different random seeds, and the average and the standard deviation values have been reported in the results.

C   DETAILS ABOUT THE DATASETS

HandWritten (Duin, 1998) is a small dataset comprised of 2,000 instances of handwritten digits. Each digit is represented using 6 different features, and each feature will be used as a modality. The dataset is licensed under CC BY 4.0 License.

CUB (Wah et al., 2011) is a textual dataset comprised of descriptions of 11,788 bird belonging to 200 different categories. To be consistent with the previous research, we will use only the first 10 categories and extract the HOG and MBH features as two views of the dataset.

Scene15 (Ali and Zafar, 2018) consists of 4485 images of 15 indoor and outdoor scene categories. GIST, PHOG and LBP features will be used as different views of the dataset. The dataset is licensed under CC BY 4.0 License.

Caltech101 (Li et al., 2022) consists of 8677 images from 101 different classes. To be consistent with the previous research, we will use only the first 10 classes and use features extracted from DECAF and VGG19 models as different views of the dataset. The dataset is licensed under CC BY 4.0 License.

PIE (Gross et al., 2008) contains 680 instances from 68 different classes. The intensity, LBP and Gabor features are used as 3 different views of the dataset. The license of the dataset allows it to be used only for noncommercial research purposes.

D   ADDITIONAL RESULTS

In Tables 5 and 6 please find the classification accuracy scores on normal and conflictive test sets respectively. As we can see, the proposed belief constraint fusion does not affect the classification accuracy. In fact, the results indicate, that none of the fusion functions affects the classification accuracy in a statistically significant manner. The Discounted Belief Fusion, however, provides better separation for conflicting and non-conflicting uncertainties as seen in the main paper.

Table 5: Mean Accuracy (%) and Standard Deviation for Different Aggregations on Normal test set

	CUB	CalTech	HandWritten	PIE	Scene
BCF	91.08 ± 3.23	99.45 ± 0.20	97.97 ± 0.55	94.78 ± 1.75	72.03 ± 1.22
CBF	91.58 ± 3.50	99.52 ± 0.17	97.88 ± 0.52	92.57 ± 2.66	72.93 ± 1.69
BAF	91.92 ± 2.79	99.36 ± 0.23	97.53 ± 0.59	93.82 ± 1.68	71.52 ± 1.76
GBAF	90.25 ± 2.27	99.05 ± 0.49	97.85 ± 0.45	95.07 ± 1.09	72.26 ± 1.42
DBF (our)	89.33 ± 2.60	99.34 ± 0.34	97.60 ± 0.71	93.90 ± 1.68	73.14 ± 1.67

Table 6: Mean Accuracy (%) and Standard Deviation for Different Aggregations on Conflicting test set

	CUB	CalTech	HandWritten	PIE	Scene
BCF	52.42 ± 3.22	61.68 ± 1.47	96.97 ± 0.88	91.40 ± 2.13	57.17 ± 1.18
CBF	53.50 ± 4.94	61.00 ± 1.32	95.03 ± 0.99	90.81 ± 2.24	55.79 ± 1.04
BAF	50.83 ± 4.50	60.48 ± 1.93	90.17 ± 1.71	84.41 ± 2.65	56.38 ± 1.84
GBAF	49.25 ± 3.32	61.32 ± 1.68	95.67 ± 0.51	91.69 ± 1.68	56.15 ± 0.95
DBF (our)	52.33 ± 4.91	61.51 ± 2.10	97.58 ± 0.74	90.29 ± 1.70	57.76 ± 1.29

In Figure 7, the uncertainty distributions for normal and conflicting sets are shown across different fusion methods (plots for CBF and GBAF were omitted from the main paper due to space limitation). As we can see, the proposed DBF method continues to exhibit superior separation between conflicting and non-conflicting sets.
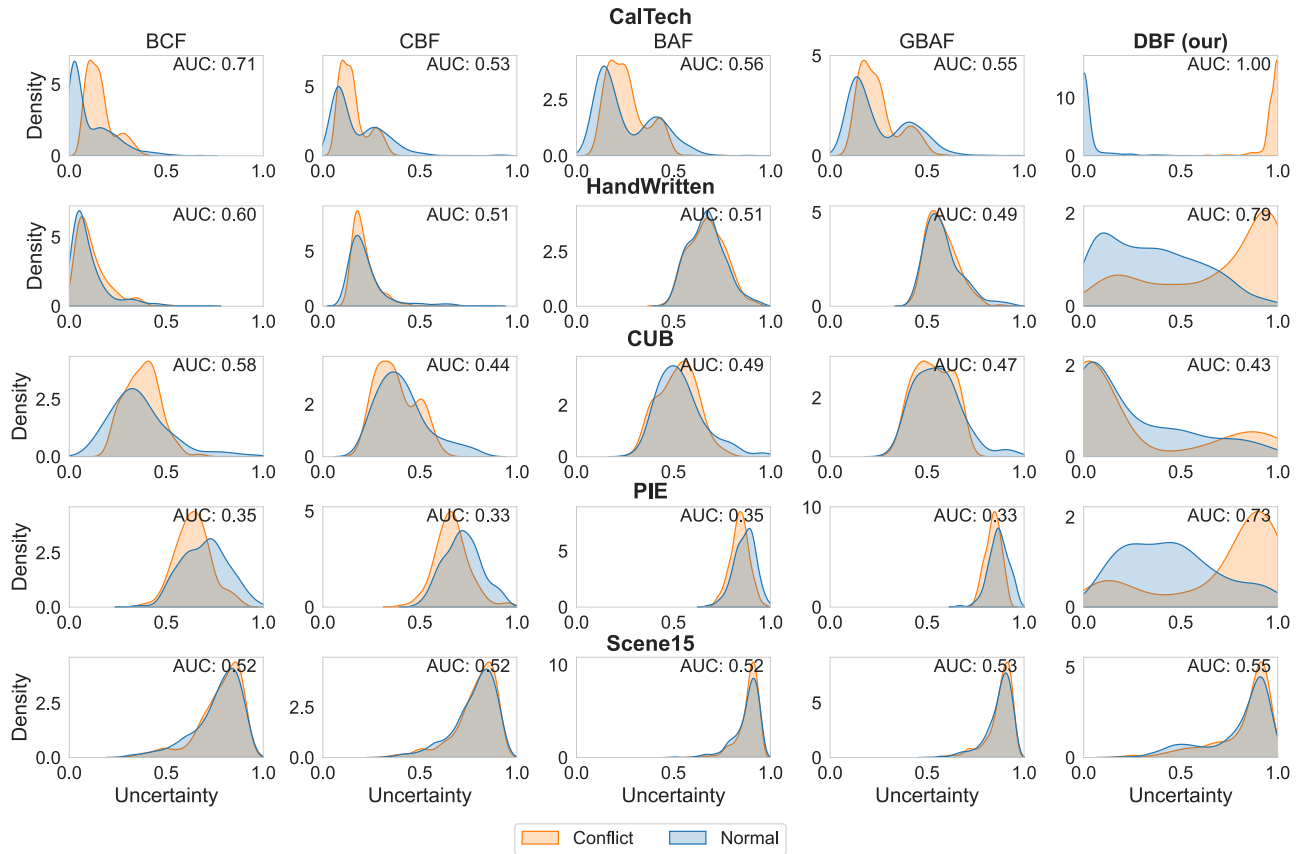

Figure 7: Uncertainty distributions on normal and conflictive test sets using Belief Constraint Fusion (BCF), Cumulative Belief Fusion (CBF), Generalized Belief Averaging Fusion (GBAF), Belief Averaging Fusion (BAF), and Discounted Belief Fusion (DBF). $\lambda = 1$ is used for all datasets.

E ABLATION STUDIES

In this section, we present ablation studies to analyze the effects of both the activation function and the fusion method. As shown in Table 7, the exponential activation function improves performance in Discounted Belief Fusion (DBF) in contrast with softplus activation function. However, it does not yield better results in Belief Averaging Fusion (BAF). This suggests that while the activation function plays a role, it is not solely responsible for better outcomes. The results emphasize the importance of the interaction between both the fusion method and the activation function in the proposed approach.

Table 7: Ablation study examining the impact of activation functions on the performance of different fusion methods.

Dataset	Discounted Belief Fusion		Belief Averaging Fusion	
	Exponential	Softplus	Exponential	Softplus
CUB	0.57 ± 0.06	0.58 ± 0.03	0.44 ± 0.03	0.50 ± 0.04
CalTech	1.00 ± 0.00	0.95 ± 0.01	0.41 ± 0.01	0.54 ± 0.02
HandWritten	0.80 ± 0.02	0.68 ± 0.02	0.47 ± 0.02	0.51 ± 0.01
PIE	0.71 ± 0.04	0.34 ± 0.03	0.31 ± 0.02	0.35 ± 0.03
Scene	0.53 ± 0.01	0.53 ± 0.01	0.52 ± 0.01	0.53 ± 0.02