

Uncertainty-Guided View-Strength-Aware Feature Utilization for Multi-View Classification

Li Lv^{1,2}, Qian Guo³, Li Zhang¹, Liang Du¹, Bingbing Jiang⁴, Lu Chen¹, Xinyan Liang^{1,2*}

¹Institute of Big Data Science and Industry, Shanxi University

²State Key Laboratory of AI Safety, Beijing, 100086

³Shanxi Key Laboratory of Big Data Analysis and Parallel Computing, Taiyuan University of Science and Technology

⁴School of Information Science and Technology, Hangzhou Normal University

{lvli924, czguoqian, liangxinyan48}@163.com, jiangbb@hznu.edu.cn, {duliang, chenlu}@sxu.edu.cn

Abstract

In multi-view classification tasks (MVC), each view provides an unique perspective on the data, offering complementary information that can improve classification performance when properly integrated. However, traditional methods typically adopt a uniform processing strategy for all views before fusion, overlooking the fact that different views may require different treatments due to variations in their quality and informativeness. To address this limitation, we propose a novel framework called Uncertainty-Guided View-Strength-Aware Feature Utilization (UVF) for multi-view classification. Our approach introduces a view uncertainty estimation module to quantify the discriminative strength of each view. Based on this estimation, a Differentiated Feature Selector (DFS) adaptively selects features, retaining informative dimensions in weak views while preserving original features in strong views. Furthermore, we employ an uncertainty-guided fusion strategy that assigns dynamic weights to each view's contribution based on its uncertainty score, enhancing the robustness and reliability of the final decision. Experimental results on benchmark datasets demonstrate that our method significantly outperforms conventional approaches, achieving better classification accuracy and interpretability through strength-aware feature processing and fusion.

Code — <https://github.com/lvli-rise/UVF>

Introduction

With rapid development of multimedia and representation learning methods, data are generally represented with multiple groups of features (Jiang et al. 2021; Liang et al. 2025d; Zhang et al. 2024a; Jin et al. 2025; Lu et al. 2025). Compared with single-view data, multi-view data is more informative and covers a wider range of information dimensions and diversity. By leveraging the complementarity and redundancy across views, multi-view learning methods can enhance the generalization ability of models and facilitate better decision-making in real-world scenarios. Hence, it is important to use multiple view data to perceive the world. With the development of deep learning techniques, many

multi-view fusion networks have been proposed (Liang et al. 2025b, 2024). They often employ different joint training strategies such as feature interaction (Zhu et al. 2022) (i.e., two modalities interact from the input level) and prediction ensemble (Zhang et al. 2024c) (i.e., two modalities interact from the prediction level), all while optimizing an unified learning objective.

While multi-view learning has demonstrated considerable success across a variety of domains, from image classification to biomedical analysis, many existing approaches operate under a critical limitation: they treat all views as equally informative and reliable throughout the learning process. As shown in Fig. 1, these methods typically adopt the same strategy to process all views before fusion. In real-world applications, the quality of each view is often far from uniform. Variations arise from multiple sources, including sensor fidelity, feature extraction pipelines, missing modalities, environmental noise, and domain shifts. As a result, the discriminative power and semantic relevance of each view can differ substantially.

Treating all views equally before fusion may lead to performance degradation due to the inclusion of redundant or misleading information. Simply aggregating features from all available views—regardless of their reliability—can propagate noise into the joint representation and hinder model interpretability and generalization. To address this issue, we propose a novel Uncertainty-Guided View-Strength-Aware Feature Utilization (UVF) framework for multi-view classification. The core idea is to estimate the discriminative strength of each view prior to fusion and adopt differentiated processing strategies accordingly. Meanwhile, this view strength information is also incorporated into the fusion stage to enhance the effectiveness of view integration and improve final classification performance.

Specifically, we first introduce an *Uncertainty Estimation Module*, which quantifies the reliability of each view—lower uncertainty implies stronger view quality. Based on these estimates, we design a *Differentiated Feature Selector (DFS)* that adaptively processes each view. Strong views (with low uncertainty) are retained in full to preserve informative patterns, while weak views (with high uncertainty) are gated to suppress noise and irrelevant features. Finally, an *Uncertainty-Guided Fusion Module* is employed to assign

*The corresponding author.

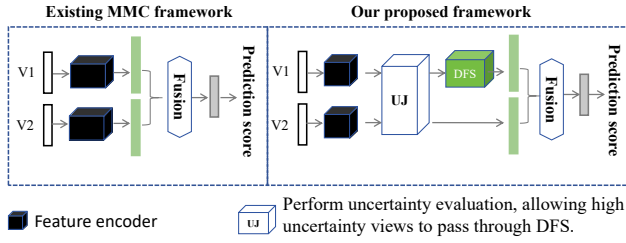


Figure 1: The differences between our proposed UVF and the existing MMC methods.

dynamic weights to each view based on its estimated uncertainty, leading to more reliable decision-making.

Our main contributions are summarized as follows:

- An uncertainty-guided view-strength-aware feature utilization (UVF) framework is proposed. The UVF incorporates a gating mechanism guided by the estimated uncertainty of each view. In particular, views with low uncertainty (strong views) retain most of their original features to preserve comprehensive representation capacity, while views with high uncertainty (weak views) undergo adaptive feature filtering to suppress noise and eliminate redundant or misleading information.
- We design a novel uncertainty-guided fusion strategy that leverages view-wise uncertainty scores to adaptively weigh the contributions of different views during decision-making.
- The extensive comparison experiments on public datasets show that UVF achieves competitive performance compared to the state-of-the-art MVC methods.

Related Work

Multi-view classification (MVC) is an effective paradigm that integrates multiple views of data to improve the robustness and accuracy of classification tasks. Most existing MVC methods are grounded in two fundamental principles: (i) the *consensus principle*, which encourages consistency among different views, and (ii) the *diversity principle*, which emphasizes the unique and complementary information provided by each view (Li and Tang 2024). For instance, Liang et al. (Liang et al. 2022) introduced association-based fusion by modeling cross-modality associations in a learnable and interpretable way. However, their method treats all modalities equally during the early processing stage, without considering the potential differences in informativeness across views. Similarly, Chen et al. (Chen et al. 2023) developed a joint deep learning framework to extract unified feature representations from heterogeneous views, but it also relies on shared representations that may suppress the unique contributions of stronger views.

Fusion strategies in MVC can generally be categorized into feature-level and decision-level methods (Guo et al. 2024; Zhang et al. 2024b). Feature-level fusion-based methods (FW) assign weights to features from different views during the representation learning stage. For example, AWDR (Yang, Deng, and Nie 2019) further refines fea-

ture contributions by incorporating view weights in a square-root form to emphasize more informative views. In contrast, decision-level methods (DW) assign weights after individual classifiers produce predictions. For example, TMC (Han et al. 2022) applies Dempster-Shafer theory to model the reliability of each view through an evidence-based confidence mechanism.

While these methods have shown promising results, they often ignore the inherent uncertainty and dynamic discriminative power of each view, which may vary across instances due to noise, redundancy, or sparsity.

The Proposed Method

Basic Setting

The remainder of this section uses the following notation. Let $\mathcal{X} = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \dots \times \mathbb{R}^{m_{|V|}}$ represent the instance space (or feature space) of representations from $|V|$ views, where m_i ($1 \leq i \leq |V|$) denotes the feature dimension of the i -th view, and let $\mathcal{Y} = \{l_1, l_2, \dots, l_q\}$ represent the label space with q class labels. Let \mathcal{D} denote an unknown distribution over $\mathcal{X} \times \mathcal{Y}$. A training set $\mathcal{D} = \{(\mathbf{x}_i^v, y_i) | 1 \leq v \leq |V|, 1 \leq i \leq n\} \in (\mathcal{X} \times \mathcal{Y})^n$ is drawn independently and identically from \mathcal{D} , where $\mathbf{x}_i^v = (x_{i1}^v, x_{i2}^v, \dots, x_{im_v}^v) \in \mathbb{R}^{m_v}$ is the feature vector of the v -th view with dimension m_v , and $y_i \in \mathcal{Y}$ is the known label associated with \mathbf{x}_i^v . The task of multi-view classification is to learn a prediction function $f : \mathcal{X} \rightarrow \mathcal{Y}$ from \mathcal{D} , which can assign an appropriate label $f(\mathbf{x}) \in \mathcal{Y}$ to an unseen instance \mathbf{x} .

Uncertainty Estimation Module

In multi-view data, each sample in the same dataset has multiple related feature representations which leads to differences in the dimensionality of the feature vectors of samples across different views. This requires unifying the dimensionality of feature vectors from different views to facilitate subsequent computations. Therefore, the first step is to map instances from different views in the original feature space to the same latent space, as shown in Fig. 2.

$$\mathbf{x}_z^v = W\mathbf{x}^v + b, \quad (1)$$

where $W \in \mathbb{R}^{d \times m_v}$ and $b \in \mathbb{R}^d$ are learnable parameters.

After projecting the features into a unified latent space, we aim to evaluate the uncertainty of each view with respect to the classification task. Intuitively, a view that consistently provides confident predictions (i.e., predictions with skewed class probability distributions) is considered strong, while a view that yields ambiguous or flat distributions is considered weak. To quantify this uncertainty, we calculate the deviation of the predicted probability distribution from the ideal uniform distribution. Specifically, we pass the latent feature \mathbf{x}_z^v through a classifier to obtain a class probability vector $\mathbf{p}^v \in \mathbb{R}^C$, where C is the number of categories. The uncertainty score for view v is then computed as:

$$s^v = \sum_{i=1}^C |p_i^v - \mu|, \quad \mu = \frac{1}{C}, \quad (2)$$

where μ denotes the probability value of a uniform distribution across all classes.

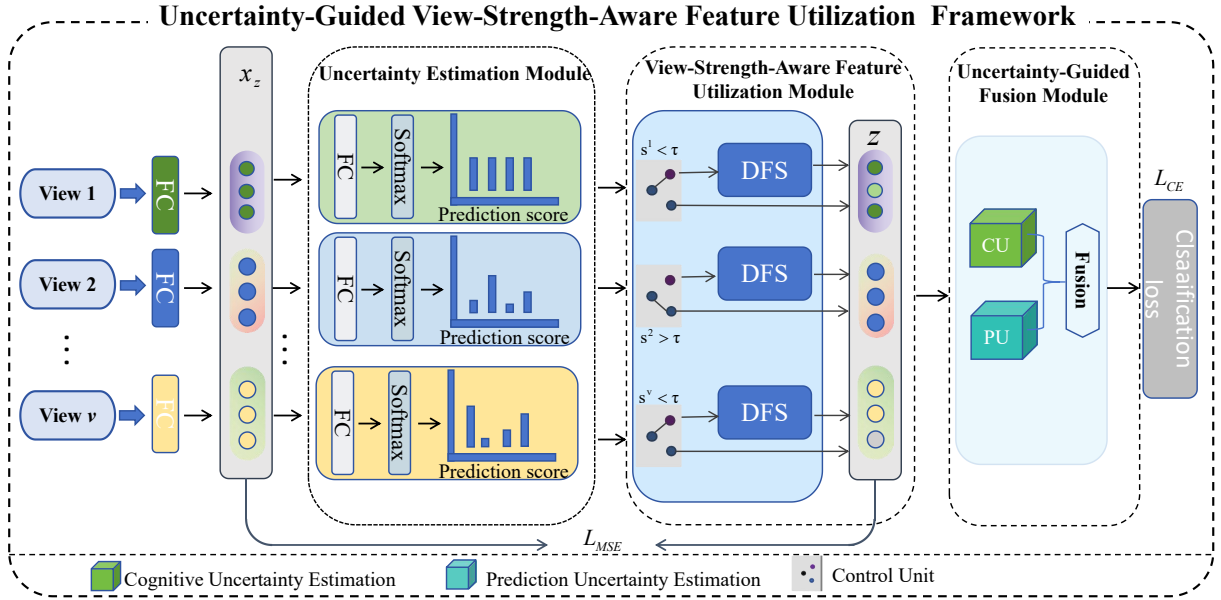


Figure 2: The whole framework of the proposed UVF.

View-Strength-Aware Feature Utilization Module

In traditional multi-view learning, all views are often processed equally, ignoring the potential quality differences between views. However, in real-world scenarios, different views may have varying discriminative abilities due to noise, redundancy, or data sparsity. To address this issue, we propose a gating mechanism with a dynamic selection strategy.

Differentiated Feature Selector (DFS). We design a learnable gate network to perform feature selection for each view. Given a feature vector $x_z^v \in \mathbb{R}^d$, a two-layer MLP with sigmoid activation is applied to compute a gating vector $\mathbf{g}^v \in \mathbb{R}^d$:

$$\mathbf{g}^v = \text{sigmoid}(W_2^v \cdot \text{ReLU}(W_1^v \cdot \mathbf{x}_z^v)). \quad (3)$$

Then, the gated feature vector \mathbf{z}^v is obtained by:

$$\mathbf{z}^v = \mathbf{g}^v \odot \mathbf{x}_z^v, \quad (4)$$

which selectively retains important feature dimensions while suppressing irrelevant or noisy ones. The gate is fully differentiable and allows end-to-end training.

Strength-Guided Gating Strategy. To prevent over-processing of high-quality views, we introduce a dynamic gating strategy guided by view uncertainty. After computing uncertainty scores $\{s^1, s^2, \dots, s^v\}$ for all V views, we sort them and define a threshold τ as the K -th smallest score. For each view v , we determine whether to apply the DFS as follows:

$$\mathbf{z}^v = \begin{cases} \text{DFS}(\mathbf{x}^v), & \text{if } s^v \leq \tau, \\ \mathbf{x}^v, & \text{if } s^v > \tau. \end{cases} \quad (5)$$

To encourage the model to select only the most discriminative features, we impose an ℓ_1 -based regularization on the

gate outputs:

$$\mathcal{L}_{L1} = \frac{1}{V} \sum_{v=1}^V \mathbb{E}_{x \sim \mathcal{D}} [\|\mathbf{g}^v\|_1]. \quad (6)$$

To prevent excessive distortion of the input features, we further introduce an MSE constraint that encourages the gated features to remain close to the original input:

$$\mathcal{L}_{MSE} = \frac{1}{B} \sum_{i=1}^B \|\mathbf{x}_z^v - \mathbf{z}^v\|_2^2. \quad (7)$$

Uncertainty-Guided Fusion Module

To effectively integrate information from multiple views, we propose an uncertainty fusion mechanism that estimates and utilizes both cognitive uncertainty and prediction uncertainty of each view to determine its contribution to the final representation. Unlike traditional fusion approaches that assign equal or fixed weights to all views, our method adaptively adjusts the fusion weights based on how confident the model is about the predictions made from each view.

Cognitive Uncertainty Estimation. Cognitive uncertainty, also known as epistemic uncertainty, reflects the uncertainty in the model parameters and can be reduced with more data. In this work, we approximate cognitive uncertainty using Monte Carlo Dropout (MC Dropout), which is a practical Bayesian approximation. Specifically, for each view v , we pass the gated feature vector $\mathbf{z}^{(v)}$ through a classifier $f^{(v)}$ with dropout enabled during T stochastic forward passes:

$$p_t^v = \text{Softmax}(f^{(v)}(\mathbf{z}^{(v)}); \text{Dropout}). \quad (8)$$

Accuracy							
Groups	Methods	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
Feature	EmbraceNet(IF19)	84.97 ± 0.23	72.43 ± 0.38	80.07 ± 0.21	83.58 ± 0.25	81.74 ± 0.34	80.90 ± 1.04
	AWDR(PR19)	90.46 ± 0.06	72.44 ± 0.66	79.69 ± 0.27	83.32 ± 0.32	91.08 ± 0.09	85.11 ± 0.15
	RAMC(INS22)	90.63 ± 0.13	72.51 ± 0.67	79.84 ± 0.25	83.48 ± 0.25	91.54 ± 0.11	85.21 ± 0.17
Decision	BV(TEVC21)	88.65 ± 0.43	68.69 ± 0.59	80.61 ± 0.25	83.98 ± 0.14	63.25 ± 0.14	82.01 ± 0.18
	SSV(TEVC21)	82.37 ± 1.26	63.70 ± 0.64	79.51 ± 0.41	84.71 ± 0.22	85.10 ± 0.23	84.43 ± 0.31
	MR(TEVC21)	87.10 ± 0.64	64.39 ± 0.85	78.24 ± 0.45	84.17 ± 0.19	79.92 ± 0.29	84.78 ± 0.21
	TMOA(AAAI22)	89.17 ± 0.31	72.60 ± 0.48	79.11 ± 0.43	84.19 ± 0.27	84.72 ± 0.21	84.35 ± 0.25
	TMC(ICLR22)	88.59 ± 0.25	72.73 ± 0.30	79.60 ± 0.56	84.23 ± 0.35	73.13 ± 0.15	71.18 ± 2.27
	ETMC(TPAMI23)	88.24 ± 0.17	73.05 ± 0.67	79.80 ± 0.41	84.24 ± 0.42	88.70 ± 0.15	79.63 ± 1.89
	ECML(AAAI24)	80.51 ± 0.41	72.53 ± 0.55	81.39 ± 0.18	85.88 ± 0.29	89.06 ± 0.21	81.95 ± 0.20
	RMVC(IF25)	81.49 ± 0.31	60.61 ± 0.54	78.89 ± 0.20	82.97 ± 0.19	88.34 ± 0.09	81.56 ± 0.28
	TUNED(AAAI25)	89.05 ± 0.45	74.08 ± 0.36	81.65 ± 0.32	86.02 ± 0.69	91.67 ± 0.30	84.79 ± 0.33
AsoDMVC(IJCAI2025)	<u>90.86 ± 0.19</u>	<u>74.62 ± 0.15</u>	<u>81.79 ± 0.20</u>	<u>86.04 ± 0.57</u>	<u>93.85 ± 0.05</u>	<u>86.21 ± 0.15</u>	
Ours	UVF	91.05 ± 0.14	75.04 ± 0.26	82.01 ± 0.10	86.09 ± 0.08	94.18 ± 0.24	86.38 ± 0.09
Precision							
Groups	Methods	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
Feature	EmbraceNet(IF19)	82.14 ± 0.57	71.73 ± 0.32	80.42 ± 0.25	83.77 ± 0.34	80.95 ± 0.46	83.71 ± 1.10
	AWDR(PR19)	89.32 ± 0.33	72.71 ± 0.61	79.87 ± 0.30	83.49 ± 0.34	91.83 ± 0.11	89.94 ± 0.32
	RAMC(INS22)	89.41 ± 0.38	72.82 ± 0.64	80.12 ± 0.27	83.70 ± 0.28	92.19 ± 0.06	<u>90.64 ± 0.08</u>
Decision	BV(TEVC21)	86.57 ± 0.46	70.98 ± 0.95	80.77 ± 0.19	84.13 ± 0.19	64.63 ± 0.63	84.34 ± 0.61
	SSV(TEVC21)	82.76 ± 1.10	67.23 ± 0.58	80.19 ± 0.49	85.16 ± 0.20	84.44 ± 0.18	94.13 ± 0.37
	MR(TEVC21)	85.44 ± 0.64	64.90 ± 0.81	78.21 ± 0.48	84.25 ± 0.13	78.85 ± 0.29	86.56 ± 0.58
	TMOA(AAAA22)	88.15 ± 0.62	72.73 ± 0.53	79.89 ± 0.72	84.40 ± 0.23	84.38 ± 0.30	87.59 ± 0.28
	TMC(ICLR22)	87.76 ± 0.40	72.71 ± 0.22	79.86 ± 0.46	84.43 ± 0.49	73.26 ± 0.34	82.53 ± 2.01
	ETMC(TPAMI23)	87.68 ± 0.63	72.39 ± 0.64	79.99 ± 0.33	84.38 ± 0.37	87.28 ± 0.15	83.40 ± 2.33
	ECML(AAAI24)	86.27 ± 1.22	73.05 ± 0.26	81.52 ± 0.18	85.81 ± 0.27	74.21 ± 0.46	84.34 ± 0.38
	RMVC(IF25)	81.35 ± 0.23	60.09 ± 0.41	80.32 ± 0.45	82.28 ± 0.49	89.26 ± 0.19	81.32 ± 0.23
	TUNED(AAAI25)	89.02 ± 0.12	74.12 ± 0.33	81.12 ± 0.22	85.79 ± 0.12	92.01 ± 0.23	85.35 ± 0.15
AsoDMVC(IJCAI2025)	<u>89.79 ± 0.23</u>	<u>75.00 ± 0.13</u>	<u>82.03 ± 0.20</u>	<u>85.89 ± 0.23</u>	<u>93.95 ± 0.13</u>	<u>86.52 ± 0.40</u>	
Ours	UVF	91.03 ± 0.14	75.13 ± 0.41	82.26 ± 0.16	86.20 ± 0.07	94.37 ± 0.22	86.47 ± 0.08

Table 1: Comparison results with SOTA methods on the accuracy and precision. The best and the second best results are highlighted by **boldface** and underlined, respectively.

Here, $p_t^{(v)} \in \mathbb{R}^C$ represents the predicted class probability distribution at the t -th pass. We then compute the average prediction and the variance across these T passes:

$$\bar{p}^v = \frac{1}{T} \sum_{t=1}^T p_t^v, \quad (9)$$

$$\text{Var}^v = \frac{1}{T} \sum_{t=1}^T (p_t^v - \bar{p}^v)^2. \quad (10)$$

To obtain the overall cognitive uncertainty score CU^v for the v -th view, we aggregate the variance across all C classes and average over the batch:

$$cus^v = \mathbb{E}_{\text{batch}} \left[\sum_{c=1}^C \text{Var}_c^v \right]. \quad (11)$$

$$CU^v = \frac{\exp(-cus^v)}{\sum_{i=1}^V \exp(-cus^i)}. \quad (12)$$

This score captures how unstable the model’s predictions are when given input from view v , reflecting the model’s confidence.

Prediction Uncertainty Estimation. In the multi-view classification, different views may contain varying levels of noise, redundancy, or informativeness. To account for this, we introduce an uncertainty estimation mechanism that evaluates the reliability of each view’s prediction. We first compute an uncertainty score for each view to quantify its prediction ambiguity:

$$us^v = \sum_{i=1}^C |\text{Softmax}(f^v(z^v))_i - \mu|, \quad \mu = \frac{1}{C}, \quad (13)$$

where f^v denotes the classifier associated with the v -th view, implemented as a fully connected layer followed by

Recall							
Groups	Methods	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
Feature	EmbraceNet(IF19)	80.04 ± 0.59	72.04 ± 0.34	79.85 ± 0.26	83.46 ± 0.21	78.36 ± 0.34	80.65 ± 1.13
	AWDR(PR19)	86.86 ± 0.20	71.87 ± 0.62	79.59 ± 0.23	83.30 ± 0.29	87.26 ± 0.13	83.57 ± 0.30
	RAMC(INS22)	87.08 ± 0.42	71.92 ± 0.65	79.73 ± 0.23	83.45 ± 0.23	87.95 ± 0.11	83.35 ± 0.27
Decision	BV(TEVC21)	85.72 ± 0.57	67.67 ± 0.57	80.52 ± 0.29	83.91 ± 0.11	57.79 ± 0.14	81.05 ± 0.35
	SSV(TEVC21)	77.28 ± 1.45	60.52 ± 0.63	79.08 ± 0.40	84.48 ± 0.25	81.07 ± 0.26	80.80 ± 0.53
	MR(TEVC21)	83.55 ± 0.77	63.10 ± 0.91	78.11 ± 0.45	84.12 ± 0.26	75.36 ± 0.32	83.87 ± 0.31
	TMOA(AAAI22)	83.62 ± 0.91	71.81 ± 0.49	78.85 ± 0.30	84.25 ± 0.30	81.54 ± 0.26	82.63 ± 0.39
	TMC(ICLR22)	84.47 ± 0.54	71.70 ± 0.43	79.60 ± 0.56	84.19 ± 0.29	64.06 ± 0.12	68.50 ± 2.77
	ETMC(TPAMI23)	83.52 ± 0.51	72.35 ± 0.79	79.74 ± 0.52	83.51 ± 0.51	85.85 ± 0.22	81.51 ± 0.26
	ECML(AAAI24)	84.34 ± 0.88	71.55 ± 0.68	81.52 ± 0.16	85.81 ± 0.27	74.53 ± 0.46	80.59 ± 0.18
	RMVC(IF25)	80.01 ± 0.49	59.80 ± 1.46	78.09 ± 0.85	84.24 ± 0.56	87.14 ± 0.09	82.36 ± 0.13
	TUNED(AAAI25)	87.14 ± 0.73	73.25 ± 0.55	81.34 ± 0.41	84.78 ± 0.44	91.46 ± 0.32	84.78 ± 0.47
AsoDMVC(IJCAI2025)	<u>87.98 ± 0.21</u>	<u>74.01 ± 0.15</u>	<u>81.57 ± 0.15</u>	<u>86.01 ± 0.18</u>	<u>91.55 ± 0.11</u>	86.93 ± 0.23	
Ours	UVF	90.98 ± 0.16	74.89 ± 0.29	81.99 ± 0.11	86.07 ± 0.07	94.11 ± 0.2	<u>86.27 ± 0.07</u>
F1							
Groups	Methods	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
Feature	EmbraceNet(IF19)	80.60 ± 0.62	71.78 ± 0.36	80.07 ± 0.22	83.59 ± 0.25	78.64 ± 0.41	81.61 ± 0.99
	AWDR(PR19)	87.72 ± 0.21	72.16 ± 0.62	79.71 ± 0.27	83.37 ± 0.30	88.57 ± 0.13	85.51 ± 0.12
	RAMC(INS22)	87.92 ± 0.30	72.21 ± 0.65	79.90 ± 0.25	83.54 ± 0.24	89.20 ± 0.10	86.69 ± 0.17
Decision	BV(TEVC21)	85.94 ± 0.50	68.64 ± 0.63	80.61 ± 0.24	83.90 ± 0.11	58.34 ± 0.23	82.49 ± 0.25
	SSV(TEVC21)	78.82 ± 1.45	62.13 ± 0.69	79.49 ± 0.42	84.75 ± 0.21	81.75 ± 0.23	86.55 ± 0.27
	MR(TEVC21)	84.14 ± 0.73	62.96 ± 0.93	78.11 ± 0.46	84.16 ± 0.19	75.88 ± 0.30	85.03 ± 0.29
	TMOA(AAAI22)	83.65 ± 0.86	72.02 ± 0.49	79.14 ± 0.49	84.24 ± 0.24	82.02 ± 0.33	84.85 ± 0.25
	TMC(ICLR22)	85.28 ± 0.54	71.84 ± 0.31	79.52 ± 0.57	84.22 ± 0.38	65.22 ± 0.09	71.92 ± 2.06
	ETMC(TPAMI23)	84.60 ± 0.49	72.19 ± 0.68	79.72 ± 0.40	84.24 ± 0.42	86.03 ± 0.20	80.97 ± 1.48
	ECML(AAAI24)	84.82 ± 1.05	72.01 ± 0.52	81.35 ± 0.16	85.89 ± 0.28	75.87 ± 0.35	82.30 ± 0.15
	RMVC(IF25)	80.62 ± 0.59	59.97 ± 1.46	79.21 ± 0.85	83.24 ± 0.70	88.14 ± 0.09	81.66 ± 0.15
	TUNED(AAAI25)	88.11 ± 0.73	73.65 ± 0.55	81.46 ± 0.45	85.24 ± 0.44	91.84 ± 0.42	85.06 ± 0.57
AsoDMVC(IJCAI2025)	<u>88.38 ± 0.30</u>	<u>74.50 ± 0.15</u>	<u>81.77 ± 0.15</u>	<u>85.93 ± 0.20</u>	<u>92.73 ± 0.11</u>	86.72 ± 0.40	
Ours	UVF	90.94 ± 0.13	75.03 ± 0.34	82.08 ± 0.07	86.12 ± 0.06	94.13 ± 0.24	<u>86.34 ± 0.07</u>

Table 2: Comparison results with SOTA methods on the recall and F1. The best and the second best results are highlighted by **boldface** and underlined, respectively.

Batch Normalization and a ReLU activation function. Subsequently, we normalize the uncertainty scores across all views to obtain fusion weights:

$$PU^v = \frac{us^v}{\sum_{i=1}^V us^i}. \quad (14)$$

To obtain an overall uncertainty score for each view, we combine the prediction uncertainty (PU) and cognitive uncertainty (CU) using a weighted linear formulation:

$$w^v = \lambda_1 \cdot PU^v + \lambda_2 \cdot CU^v, \quad (15)$$

where λ_1 and λ_2 are hyperparameters that control the contributions of prediction and cognitive uncertainties, respectively. Finally, the fused representation is computed as a weighted sum of the selected features from all views:

$$\hat{y} = \sum_{i=1}^V w^i \cdot z^i. \quad (16)$$

Overall Objective Function

The proposed model is optimized in an end-to-end manner using a composite loss function that integrates classification accuracy, feature sparsity, and information preservation. Specifically, the overall objective is defined as:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{L1} + \beta \mathcal{L}_{MSE}, \quad (17)$$

where \mathcal{L}_{L1} is an ℓ_1 regularization term imposed on the gating masks to encourage feature sparsity, and \mathcal{L}_{MSE} is a mean squared error term that preserves the similarity between original and gated features. The hyperparameters λ and β are used to balance the contributions of the sparsity and reconstruction losses and \mathcal{L}_{ce} denotes the cross entropy loss, formulated as:

$$\mathcal{L}_{ce} = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}). \quad (18)$$

Experiments

Datasets

Our experiments are conducted on six challenging multi-view classification datasets which include image, text, audio, depth and video datasets. (1) Animals with Attributes (AWA)(Lampert, Nickisch, and Harmeling 2013) dataset, which includes 30,475 images from 50 categories with seven view features. (2) NUS-WIDE-128 (NUS)(Tang et al. 2016) dataset, which includes 43,800 samples from 128 categories with seven view features. (3) Reuters (Liang et al. 2025a) dataset, which includes 111,740 samples from six categories with five multilingual view features. (4) MVoxCeleb (Liang et al. 2025a), which includes 153,516 samples from 1,251 categories with five audio view features. (5) YoutubeFace dataset, which includes 3,425 videos from 1,595 different people with five view features. According to (Wang et al. 2022), we use a subset of 31 categories from this dataset, with a total of 101,499 frames.

Experimental Results with Other Methods

To validate the effectiveness of the our method, comprehensive comparison experiments are conducted with eight related weighting-based multi-view classification methods. The compared methods can be classified into the following two groups according to the level of weighting:

1. The first category is the *feature* level including EmbraceNet, AWDR and RAMC (Jiang et al. 2022). EmbraceNet assigns 1 to the weight value of one view while 0 to others for each example according to a multinomial distribution. AWDR is an adaptive-weighting discriminative regression approach. Following (Yang, Deng, and Nie 2019), the parameter λ is chosen from the set $\{10^{-3}, 10^{-2}, \dots, 10^3\}$, while k varies within the range $\{1, 3, \dots, 9\}$. RAMC employs an $L_{2,1}$ -norm loss function to acquire a joint weighted projection space across all views. This method preserves the correlation and diversity among views through a self-supervised weighting strategy. Similarly, the parameter λ is chosen from the set $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ and k ranges from the range $\{1, 3, \dots, 9\}$.
2. The second category is the *decision* level including BV, SSV, MR (Liang et al. 2021), TMC (Han et al. 2022), TMOA (Liu et al. 2022), ETMC (Han et al. 2023), ECML (Xu et al. 2024), RMVC (Yue et al. 2025), TUNED (Huang et al. 2025) and AssoDMVC (Liang et al. 2025c). BV assigns 1 to the weight value of the view with the best performance while 0 to others according to whole classification performance of each view. MR assigns 1 to the weight value of the view with the best performance while 0 to others for each example according to the classification performance of each view of each example. SSV assigns the same values to all views. TMC, TMOA, ETMC, ECML and RMVC are trusted fusion methods.

The results in Tables 1 and 2, are presented through the mean metric value and the standard deviation obtained from 5-fold cross-validation. From Tables 1 and 2, the following observations can be made:

1. In summary, the UVF method almost achieve the best performance on all datasets. For instance, on the Vox-Celeb and NUS datasets, UVF achieves approximately 0.33% and 0.42% accuracy improvements compared to the second-best model. These improvements highlight the significant advantage of the UVF in ensuring high performance in decision making. Additionally, UVF consistently maintained top or near-top performance across most datasets, including AWA, Reuters5, Reuters3, and YoutubeFace, achieving 91.05%, 82.01%, 86.09%, and 86.38%, respectively.
2. In addition to accuracy, we also observe consistent improvements in other metrics such as precision, recall, and F1-score across all datasets. These comprehensive gains indicate that UVF not only improves the model’s overall correctness but also enhances its robustness in detecting minority classes and handling class imbalance. For example, on the Reuters5 dataset, UVF achieves an F1-score of 82.08%, surpassing other methods by a considerable margin, which suggests its strong discriminative capability in complex real-world text scenarios.

Overall, the experimental results clearly demonstrate that UVF provides a robust and effective solution for multi-view classification, especially in scenarios where the views differ in quality and relevance.

Further Analysis

Ablation Experiments. To comprehensively evaluate the contribution of each component in the proposed UVF framework, we perform ablation experiments by selectively enabling the Differentiated Feature Selector (DFS), the Strength-Guided Gating Strategy (SG), and the Uncertainty-Guided Fusion Module (UG). The results on six datasets are summarized in Table 3.

When DFS is added alone, the model performance improves significantly. Specifically, the model performs better with the DFS, achieving 89.69% on the AWA dataset, which is significantly higher than the 82.27% without it. Moreover, the integration of SG and DFS results in a 2.06% accuracy gain on the VoxCeleb dataset. When DFS is used in conjunction with UG, the model performance is further improved. Specifically, the accuracy on the AWA dataset increases from 89.69% to 90.98% when the UG is added.

These results verify the effectiveness and synergy of each proposed module. Among them, DFS contributes the most significant improvement, while SG and UG provide further complementary gains when integrated together.

DFS	SG	UG	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
×	×	×	82.27	72.13	77.48	83.23	90.06	84.58
×	×	✓	82.46	72.99	77.84	83.30	90.61	85.05
✓	×	×	89.69	74.48	80.59	84.65	91.17	85.59
✓	✓	×	89.69	74.66	81.07	85.19	93.24	85.59
✓	×	✓	90.98	74.59	81.43	85.51	92.35	85.32
✓	✓	✓	91.05	75.04	82.01	86.09	94.18	86.38

Table 3: Ablation results for different components.

Uncertainty-View Strength Consistency Validation. To assess whether the uncertainty estimation accurately reflects the relative strength of different views, we conduct a consistency validation experiment. Specifically, for each view, we compute two quantities:

- **Single-view accuracy.** The classification accuracy when the model uses only the features from that view.
- **Uncertainty fusion weight.** The weight assigned to the view during uncertainty-guided fusion, computed from its uncertainty score.

Metric	View 1	View 2	View 3	View 4	View 5
Accuracy (%)	41.29	38.86	57.37	67.07	81.11
Uncertainty Weight	0.13	0.12	0.17	0.19	0.39
Rank	4	5	3	2	1

Table 4: Comparison between single-view classification accuracy and uncertainty-derived weights on Reuters5 dataset.

Intuitively, a stronger view (i.e., more informative and reliable) should achieve higher single-view accuracy and be assigned a higher fusion weight (i.e., lower uncertainty). Therefore, we compare the ranking of the views based on their single-view accuracy with the ranking of their uncertainty weights. As presented in Table 4, the results demonstrate that views achieving higher single-view accuracy tend to receive higher fusion weights through uncertainty-based estimation. This suggests that the proposed uncertainty-guided fusion mechanism is capable of reliably reflecting the relative informativeness of each view. As shown in Table 5, we also analyze how two uncertainty-based weighting methods affect the model.

CU	PU	AWA	NUS	Reuters5	Reuters3	VoxCeleb	YoutubeFace
×	×	89.69	74.66	81.07	85.19	93.24	85.59
✓	×	90.17	74.83	81.30	85.09	93.82	85.69
×	✓	91.04	74.72	81.88	85.68	94.12	85.71
✓	✓	91.05	75.04	82.01	86.09	94.18	86.38

Table 5: Effect of CU and PU in terms of accuracy.

Effectiveness of DFS on Single-View Classification. To further validate the effectiveness of the Differentiated Feature Selector (DFS), we conduct a comparison experiment under a single-view setting on the Reuters5. Specifically, we evaluate the classification performance of each individual view both with and without DFS applied. As shown in Fig. 3, introducing DFS consistently improves the accuracy of each view. Notably, the performance gains range from 1.22% to 3.54%, demonstrating that DFS effectively enhances view-specific discriminative representation even when no cross-view information is utilized.

DFS Output Visualization Analysis. To investigate how the gating network selects features, we visualize its output to understand the importance distribution learned by the model. During inference, we collect the DFS output values

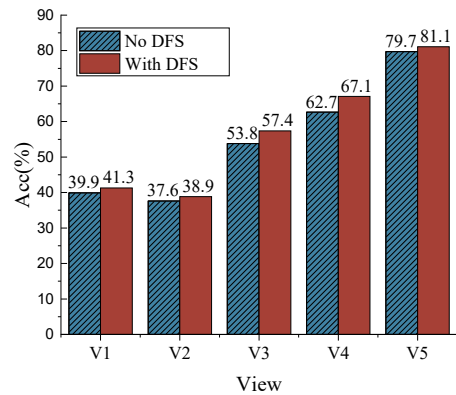


Figure 3: Effective analysis of DFS.

(after the sigmoid activation). We visualize these values using a heatmap, where each row represents a sample, and each column corresponds to a feature. As shown in Fig. 4, the visualization results demonstrate that DFS prioritizes retaining the majority of features from strong views by assigning them relatively high weights, whereas for weak views, it conducts adaptive feature selection to emphasize the most informative feature.

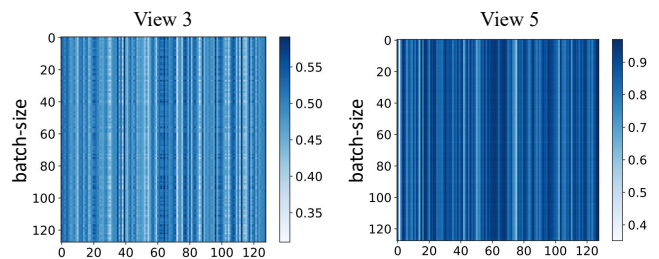


Figure 4: Visualization of DFS on YoutubeFace.

Conclusion

In this paper, we propose a novel framework, Uncertainty-Guided View-Strength-Aware Feature Utilization for multi-view classification tasks. The proposed method addresses the challenge of unequal view quality by introducing three key components: a Differentiated Feature Selector (DFS) to capture view-specific discriminative features, a Strength-Guided Gating Strategy (SG) to model and leverage the contribution strength of each view, and an Uncertainty-Guided Fusion Module (UG) to adaptively weight view information based on cognitive and predictive uncertainty. By integrating uncertainty estimation with strength-adaptive feature utilization, our method offers a principled framework for robust multi-view fusion and paves the way for further exploration into uncertainty-aware fusion strategies.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 62406218, 62306171, T2495250, T2495251, 62373233, 62376146), the Science and Technology Major Project of Shanxi (No. 202201020101006), the Self-initiated Research Project of the Foundation of Shanxi Key Laboratory of Big Data Analysis and Parallel Computing (No. BDPC-ZZ-23-001), and the Open Project of State Key Laboratory of AI Safety (No. 2025-12).

References

- Chen, Z.; Fu, L.; Yao, J.; Guo, W.; Plant, C.; and Wang, S. 2023. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95: 109–119.
- Guo, Q.; Liang, X.; Qian, Y.; Cui, Z.; and Wen, J. 2024. A progressive skip reasoning fusion method for multi-modal classification. In *Proceedings of the 32nd ACM international conference on multimedia*, 429–437.
- Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted multi-view classification. In *International Conference on Learning Representations*, 1–11.
- Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2023. Trusted Multi-View Classification With Dynamic Evidential Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2551–2566.
- Huang, H.; Qin, C.; Liu, Z.; Ma, K.; Chen, J.; Fang, H.; Ban, C.; Sun, H.; and He, Z. 2025. Trusted unified feature-neighborhood dynamics for multi-view classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Jiang, B.; Xiang, J.; Wu, X.; He, W.; Hong, L.; and Sheng, W. 2021. Robust Adaptive-weighting Multi-view Classification. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3117–3121.
- Jiang, B.; Xiang, J.; Wu, X.; Wang, Y.; Chen, H.; Cao, W.; and Sheng, W. 2022. Robust multi-view learning via adaptive regression. *Information Sciences*, 610: 916–937.
- Jin, Z.; Qian, Y.; Liang, X.; and Geng, H. 2025. A Multi-view Fusion Approach for Enhancing Speech Signals via Short-time Fractional Fourier Transform. In Kwok, J., ed., *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 5508–5516. International Joint Conferences on Artificial Intelligence Organization. Main Track.
- Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2013. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3): 453–465.
- Li, S.; and Tang, H. 2024. Multimodal Alignment and Fusion: A Survey. *arXiv preprint arXiv:2411.17040*.
- Liang, X.; Fu, P.; Guo, Q.; Zheng, K.; and Qian, Y. 2024. DC-NAS: Divide-and-Conquer Neural Architecture Search for Multi-Modal Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13754–13762.
- Liang, X.; Fu, P.; Qian, Y.; Guo, Q.; and Liu, G. 2025a. Trusted multi-view classification via evolutionary multi-view fusion. In *Proceedings of the 13th International Conference on Learning Representations*, 1–14.
- Liang, X.; Guo, Q.; Qian, Y.; Ding, W.; and Zhang, Q. 2021. Evolutionary deep fusion method and its application in chemical structure recognition. *IEEE Transactions on Evolutionary Computation*, 25(5): 883–893.
- Liang, X.; Li, S.; Guo, Q.; Qian, Y.; Jiang, B.; Luo, T.; and Du, L. 2025b. Evolutionary Multi-View Classification via Eliminating Individual Fitness Bias. In *Proceedings of the Thirty-ninth Annual Conference on Neural Information Processing Systems*.
- Liang, X.; Lv, L.; Guo, Q.; Jiang, B.; Li, F.; Du, L.; and Chen, L. 2025c. View-Association-Guided Dynamic Multi-View Classification. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 5680–5688.
- Liang, X.; Qian, Y.; Guo, Q.; Cheng, H.; and Liang, J. 2022. AF: An Association-Based Fusion Method for Multi-Modal Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9236–9254.
- Liang, X.; Wang, S.; Qian, Y.; Guo, Q.; Du, L.; Jiang, B.; Luo, T.; and Li, F. 2025d. Trusted Multi-View Classification with Expert Knowledge Constraints. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, 37409–37426.
- Liu, W.; Yue, X.; Chen, Y.; and Denooux, T. 2022. Trusted multi-view deep learning with opinion aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7585–7593.
- Lu, J.; Buntine, W.; Qi, Y.; Dipnall, J.; Gabbe, B.; and Du, L. 2025. Navigating Conflicting Views: Harnessing Trust for Learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, 40411–40435.
- Tang, J.; Shu, X.; Qi, G.-J.; Li, Z.; Wang, M.; Yan, S.; and Jain, R. 2016. Tri-clustered tensor completion for social-aware image tag refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8): 1662–1674.
- Wang, S.; Liu, X.; Liu, L.; Tu, W.; Zhu, X.; Liu, J.; Zhou, S.; and Zhu, E. 2022. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9776–9785.
- 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, 16129–16137.
- Yang, M.; Deng, C.; and Nie, F. 2019. Adaptive-weighting discriminative regression for multi-view classification. *Pattern Recognition*, 88: 236–245.
- Yue, X.; Dong, Z.; Chen, Y.; and Xie, S. 2025. Evidential dissonance measure in robust multi-view classification to resist adversarial attack. *Information Fusion*, 113: 102605.
- Zhang, C.; Fang, Y.; Liang, X.; Zhang, H.; Zhou, P.; Wu, X.; Yang, J.; Jiang, B.; and Sheng, W. 2024a. Efficient

Multi-view Unsupervised Feature Selection with Adaptive Structure Learning and Inference. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, 5443–5452.

Zhang, C.; Liang, X.; Zhou, P.; Ling, Z.; Zhang, Y.; Wu, X.; Sheng, W.; and Jiang, B. 2024b. Scalable Multi-view Unsupervised Feature Selection with Structure Learning and Fusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 5479–5488.

Zhang, C.; Zhu, X.; Wang, Z.; Zhong, Y.; Sheng, W.; Ding, W.; and Jiang, B. 2024c. Discriminative Multi-View Fusion via Adaptive Regression. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Zhu, T.; Li, L.; Yang, J.; Zhao, S.; Liu, H.; and Qian, J. 2022. Multimodal sentiment analysis with image-text interaction network. *IEEE Transactions on Multimedia*, 25: 3375–3385.