Biased Incomplete Multi-View Learning

Haishun Chen, Cai Xu, Ziyu Guan[*]**, Wei Zhao, Jinlong Liu**

School of Computer Science and Technology, Xidian University, China
{chenhaishun@stu., cxu@, zyguan@, ywzhao@mail., 20069100042@stu.}xidian.edu.cn

Abstract

Considering the ubiquitous phenomenon of missing views in multi-view data, incomplete multi-view learning is a crucial task in many applications. Existing methods usually follow an impute-then-predict strategy for handling this problem. However, they often assume that the view-missing patterns are uniformly random in multi-view data, which does not agree with real-world scenarios. In practice, view-missing patterns often vary across different classes. For example, in the medical field, patients with rare diseases would take more examinations than those with common diseases; in the financial field, high-risk customers tend to receive evaluations from more views than ordinary ones. Hence, we often observe that data-rich classes suffer limited views while data-poor classes suffer limited samples. Previous methods would typically fail due to such biased view-missing patterns. This motivates us to delve into a new biased incomplete multi-view learning problem. To this end, we develop a Reliable Incomplete Multi-view Learning (RIML) method. RIML is a simple yet effective learning-free imputation framework that goes beyond the conventional approaches by considering information from all classes, rather than just relying on individual views or within-class samples. Specifically, we utilize an inter-class association matrix that allows data-poor classes to refer the knowledge from data-rich classes. This enables the construction of more reliable view-specific distributions, from which we perform multiple samplings to recover missing views. Additionally, to obtain a reliable multi-view representation for downstream tasks, we develop an enhanced focal loss with a category-aware marginal term to learn a more distinguishable feature space. Experiments on five multi-view datasets demonstrate that RIML significantly outperforms existing methods in both accuracy and robustness.

Introduction

Multi-view learning has gained increasing importance in real-world applications (Chen et al. 2021; Fan et al. 2022; Zhang et al. 2024; Xu et al. 2024). For example, in the medical domain, physicians often depend on multiple perspectives of data to achieve a comprehensive understanding of a patient's condition. Combining consistent and complementary information from multiple views enables a more
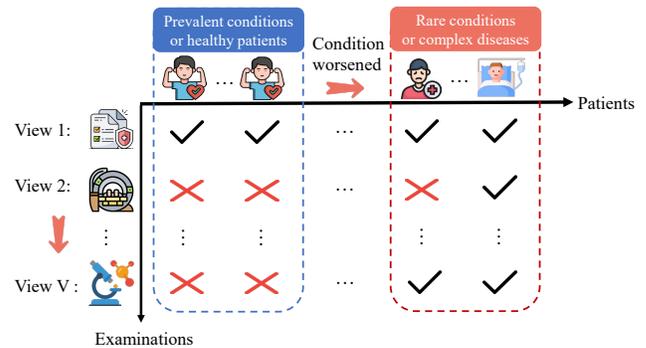
---

[*]Corresponding author.

Figure 1: Visualization of the non-uniform view-missing patterns in multi-view data: "✓" indicates that patients had this view examination, while "✗" indicates that patients did not have this view examination.

complete representation of data instances, thereby enhancing various tasks such as clustering (Ling et al. 2023; Huang et al. 2023; Cui et al. 2024). However, in real-life cases, partial views can be missing due to factors like sensor failures, which leads most multi-view learning methods to suffer notable performance decrease.

In addressing the challenges of incomplete multi-view learning (IMVL), existing methods primarily rely on two key principles: Cross-View Consistency (CVC) and Intra-Class Similarity (ICS). For instance, certain methods (Liu et al. 2022; Xia et al. 2022) exploit the connotative consistency among available views within certain instances to identify potential cross-view correlation patterns. These patterns are then used to predict or reconstruct the missing views in incomplete instances. Alternatively, other approaches (Tang and Liu 2022; Xie et al. 2023) analyze the common characteristics shared by instances within the same category to predict missing views based on the premise that instances from the same class exhibit similar view patterns.

Although these methods achieve promising performance, it is worth noting that most prior studies (Huang et al. 2021; Xu et al. 2023; Tang et al. 2024) assume that view-missing patterns are entirely arbitrary. However, in practice, the view-missing patterns are often closely associated with specific categories, making it systematically biased rather than
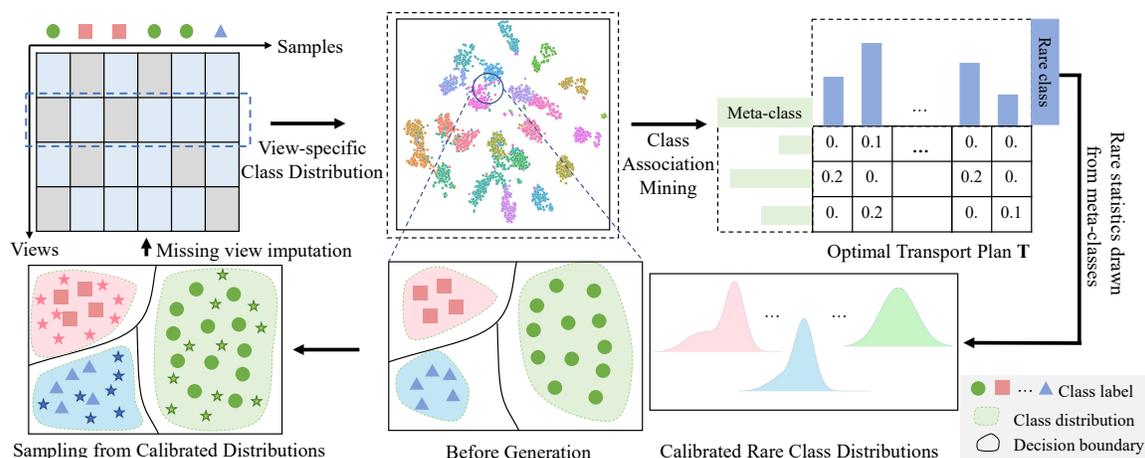
Figure 2: Illustration of the proposed method. We first construct view-specific feature distributions, which are estimated using available data. Then we calibrate distributions with a learned class-association matrix, which provides the relations among different classes. Finally, we can obtain complete instances by sampling from calibrated distribution of missing views.

random. For instance, as illustrated in Figure 1, the number of views required for diagnosis frequently varies based on the severity and type of disease. Prevalent conditions or healthy patients typically need only one or two views for accurate diagnosis, whereas rare or complex diseases often require all available diagnostic views to ensure an informed decision. This observation indicates that the pattern of missing views is intrinsically tied to the specific disease type, suggesting that missing patterns are class-dependent rather than uniformly random. Specifically, data-rich classes tend to have higher missing rates, while data-poor classes with limited samples often have more complete views available. This discovered phenomenon would bring tremendous challenges for existing methods: (1) Methods relying on CVC struggle in data-rich classes because the limited number of available views prevents the identification of consistent cross-view patterns. In contrast, methods relying on ICS fail in data-poor classes due to the insufficient number of instances to capture reliable intra-class similarities, resulting in poor generalization across classes. (2) The class imbalance skews model training, causing the model to focus excessively on data-rich classes, which leads to class ambiguities near boundaries and results in unreliable predictions.

In this paper, we propose a Reliable Incomplete Multi-view Learning (RIML) method to address the Biased Incomplete Multi-view Learning (BIML) problem. Distinct from previous methods that often focus on CVC or ICS principles, RIML introduces a novel learning-free imputation framework that utilizes inter-class associations to guide the imputation of missing views. Specifically, as illustrated in Figure 2, We first characterize the view-specific feature distribution using the available data. Due to the disparity in sample sizes, these distributions may be unreliable. To address this, we learn a class-association matrix to allow data-poor classes to reference statistical information from data-rich classes, improving the reliability of the estimated distributions. From these calibrated distributions, we perform mul-

tiple samplings to recover the missing views and generate completed instances. To further alleviate the deficiency of class imbalance, we introduce an enhanced focal loss with a category-aware marginal term that adaptively adjusts the decision boundary, making the model learn a more distinct feature representation, especially for data-poor classes.

The main contributions of this work are summarized as follows: (1) we discover an neglected but widespread phenomenon in incomplete multi-view data, where the view-missing pattern in data is usually biased rather than arbitrary. Thus, we propose a biased incomplete multi-view learning problem. (2) we propose the RIML method to tackle this problem, it leverages underlying inter-class associations to guide the imputing of missing views. (3) Extensive experiments on five real-world multi-view datasets demonstrate the superiority of the RIML method over existing methods, showing significant improvements in accuracy, reliability and robustness.

Related Work

Incomplete Multi-view Learning. Incomplete Multi-view Learning has attracted significant attention due to the prevalence of missing data in real-world. Most existing methods works mainly rely on two principles: cross-view consistency and intra-class similarity. For instance, (Zhang et al. 2019) utilize a cross-partial network architecture to effectively integrate information from available views, ensuring data consistency across incomplete multi-view datasets. (Zhang et al. 2022) employs low-rank tensor regularization to reconstruct missing views by capturing the underlying low-dimensional structure across multiple views. (Xu et al. 2021; Zhang et al. 2020) employ Generative Adversarial Networks (GANs) to generate missing views, ensuring data consistency across different views through a consistency constraint mechanism. (Jin et al. 2023) achieves missing view recovery by aligning partial samples and class prototypes across views. In addition to these learning-based im-

putations, learning-free imputation methods have also been proposed (Zhou, Wang, and Yang 2019; Xie et al. 2023), which builds a distribution by analyzing the statistical characteristics of existing views and then performs imputation by sampling from this distribution, offering more imputation through samples with a simpler way. However, these methods often require sufficient views and samples to perform effectively, while in real-world settings, multi-view data usually suffer from biased missing patterns and limited intra-class samples, their performance may be heavily degraded.

Imbalance Multi-view Learning. In recent years, there has been a growing interest in addressing the challenges posed by imbalanced multi-view data, with existing approaches generally categorized into resampling-based methods and cost-sensitive learning methods. Resampling-based methods aim to achieve class balance by either augmenting samples (Tan and Zhao 2022; Wang et al. 2020b) for minority classes or removing samples (Zhang and Dong 2020) from majority classes. Cost-sensitive learning methods (Kim and Sohn 2020; Miao, Yao, and Zhao 2022; Tang et al. 2023) tackle the imbalance by assigning different weight to misclassifications of different classes, thereby encouraging the model to focus more on the minority classes. The methods achieve promising performance main rely on complete views to accurately balance class distributions and enforce penalties (Wang and Zhou 2021; Hu et al. 2024). However, the missing of key views may introduce noise in augmenting and impact the accurate assignment of misclassification costs, affecting their effectiveness in dealing with incomplete multi-view data. These limitations motivate us to delve into the problem of both view incompleteness and class imbalance in multi-view learning.

Method

In this section, we first define the biased incomplete multi-view learning problem, then present our proposed method in detail, together with its implementation.

Notations and Problem Statement

In the context of incomplete multi-view learning, each instance is characterized by multiple views, while some of them may be missing. To clarity, suppose we are given a dataset containing K multi-view instances with C distinct classes. Every observation $\{\mathbf{x}_k^v\}_{v=1}^V$ is composed with possible views. In our settings, the missing pattern should be associated with specific class. Therefore, the missing rate(MR) varies in different classes, which can be calculated as:

$$\mathbf{MR}_c = \frac{\sum_{n=1}^{N_c} \sum_{v=1}^{V} \mathbb{I}\left(\mathbf{x}_n^v = \emptyset\right)}{K \times V}$$

where $\mathbb{I}(\cdot)$ is the indicator function that equals 1 if the v-th view of the k-th instance is missing and 0 otherwise. N_c is the number of instances that belong to c-th class, satisfying $\sum_{c=1}^{C} N_c = K$. As shown in Figure 3, to present the view status (present/absent) for corresponding instances, a class-specific indicator matrix $\mathbf{M_c}$ is also defined.

$$\mathbf{M}_c^{vn} = \begin{cases} 1 \text{ if the } n\text{-th instance has the } v\text{-th view} \\ 0 \text{ otherwise} \end{cases}$$
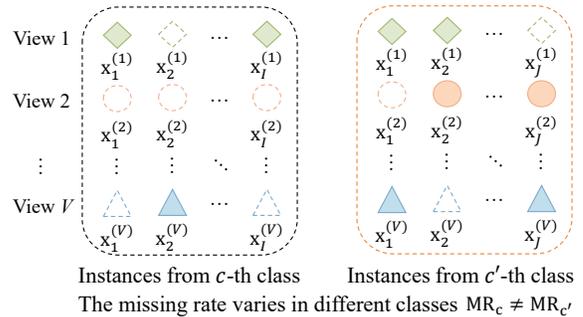

Figure 3: Notations for biased incomplete multi-view data. The left and right parts show instances from two different classes, and they suffer different missing patterns.

Our goal is to develop a robust and reliable classifier that can classify all the K instances into C distinct classes.

Biased Incomplete Multi-view Learning

Considering that a comprehensive multi-view representation should contain complete view information, and each observation $\{\mathbf{x}_n^v\}_{v=1}^V$ contains only partial view information. Distilling insights from few-shot learning (Guo et al. 2022), which augments the training set by sampleing the estimated distribution of observed data, we propose an efficient incomplete multi-view learning framework, named Reliable Incomplete Multi-View Learning. As shown in Figure 2, the overall framework consists of two components: class association mining and missing-views imputation. In the first component, we characterize the view-specific feature distribution, which can be estimated using the available data in each category. However, this approach may be hindered by the limited number of samples, leading to potentially unreliable class distributions. To address this, we learn the class-association matrix by formulating an optimal transport (OT) problem, which measures the similarity or closeness between classes. In the second component, we calibrate the distributions using the learned class-association matrix, which explicitly captures the relationships between different classes. This calibration helps reduce confusion caused by unreliable distributions. We then perform multiple samplings from the calibrated distributions to impute the missing views. Further details are provided below.

Class Association Mining. Most existing methods construct complete multi-view observations based on imputation techniques, which typically rely on exploiting cross-view consistency and intra-class similarity inherent in multi-view data. However, these methods face limitations in accurately generating missing views. This is primarily because they require a sufficient number of views and instances to perform effectively, leading to imprecise imputations.

To address this issue, we explore the possibility of applying similarities between pairs of classes, enabling the imputation of missing views to maximize the effective utilization of all available data. In this context, we begin by defining an optimal transport (OT) problem, where the ob-

jective is to find the most efficient way to align different classes. This alignment captures the underlying associations between classes, which then guides the imputation process, ensuring that the generated missing views are more reliable.

Specifically, for the optimal transport problem, we introduce the concept of meta-classes, which refer to a group of classes with a larger number of instances, determined by a predefined threshold. In contrast, classes with fewer instances are referred to as rare classes. Building on these definitions, we define $\mathbf{P^v}$ as a discrete distribution for the v-th view over N meta-classes:

$$\mathbf{P^v} = [p_1^v, p_2^v, \cdots, p_n^v, \cdots] \tag{1}$$

where p_n^v represents the probability of the n-th meta-class from the v-th view, and the sum of all probabilities in $\mathbf{P^v}$ is 1, ensuring that it forms a valid probability distribution. Assuming that each rare class has one sample for the v-th view, we represent $\mathbf{Q^v}$ as a uniform discrete distribution over M rare classes:

$$\mathbf{Q^v} = [q_1^v, q_2^v, \cdots, q_m^v, \cdots] \tag{2}$$

where q_m^v typically being $1/m$ in a uniform distribution.

To capture the associations between meta-classes and rare classes, we difine the regularized OT distance between $\mathbf{P^v}$ and $\mathbf{Q^v}$, where the entropic regularization term promotes the transport matrix to approach a uniform distribution:

$$\mathrm{OT}(\mathbf{P^v}, \mathbf{Q^v}) \overset{\text{def.}}{=} \min_{\mathbf{T} \in \Pi(\mathbf{P^v}, \mathbf{Q^v})} \langle \mathbf{T}, \mathbf{C} \rangle + \sum_{i=1, j=1}^{n,m} T_{ij} \ln T_{ij}$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius dot product, and $\mathbf{C}^v \in \mathbb{R}^{n \times m}$ is the cost matrix of the transport, indicating the cost between the n-th meta-class and the m-th rare class, and $\mathbf{T}^v \in \mathbb{R}^{n \times m}$ denotes the transport plan to be learned, which should satisfy:

$$\Pi(\mathbf{P^v}, \mathbf{Q^v}) := \left\{ \mathbf{T} \mid \sum_{i=1}^{n} T_{ij} = 1/m, \sum_{j=1}^{m} T_{ij} = p_n^v \right\} \tag{3}$$

To compute the view-specific OT problem, we define the cost function. Intuitively, $\mathbf{C}_{ij}^v$ will be small if two classes are similar. We use (1-cosine similarity) as the distance metric, which ensures non-negativity and preserves monotonicity by assigning smaller distances to more similar classes. Moreover, since each view represents different features of the same class in multi-view learning, we leverage the consistency principle across these views to ensure coherence in the transport matrix for all views.

$$\mathcal{L}_{\mathrm{con}} = \frac{1}{V} \sum_{v=1}^{V} \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \left| T_{ij}^v - \hat{T}_{ij} \right| \right) \tag{4}$$

where $\hat{T}_{nm} = \left(\sum_{v=1}^{V} T_{nm}^v \right) / V$, Therefore, the final optimization objective for our multi-view OT problem is:

$$\mathrm{OT}(\mathbf{P}, \mathbf{Q}) = \sum_{v=1}^{V} \mathrm{OT}(\mathbf{P^v}, \mathbf{Q^v}) + \mathcal{L}_{\mathrm{con}} \tag{5}$$

By minimizing Eq.(5) using the Sinkhorn algorithm (Sinkhorn 1967), we obtain the view-specific optimal transport matrix $\mathbf{T}^v$, which acts as a weight matrix to achieve optimal calibration by measuring relevance among classes.

Missing-views Imputation. Instead of using generative models to recover missing views as deterministic values, we propose to impute missing views as distributions, which allows us to capture the inherent variability in the data and thereby imporve generalization capabilities of our model. Moreover, sampling from distributions provides a non-parametric, non-learning-based strategy for imputation, significantly reducing computational demands.

Given that meta-classes have a larger number of instances, we can extrapolate the distribution of a rare class by taking a weighted average of the distributions of all meta-classes. This can be framed as transfer learning, where knowledge from the meta-classes is transferred to the rare classes. Specifically, for the n-th meta-class, we assume that its samples are generated from a Gaussian distribution, with mean and covariance matrix calculated as:

$$\boldsymbol{\mu}_n^v = \frac{\sum_{i=1}^{B} \widetilde{\mathbf{x}}_i^v}{B} \tag{6a}$$

$$\boldsymbol{\Sigma}_n^v = \frac{1}{B-1} \sum_{i=1}^{B} \left(\widetilde{\mathbf{x}}_i^v - \boldsymbol{\mu}_n^v \right) \left(\widetilde{\mathbf{x}}_i^v - \boldsymbol{\mu}_n^v \right)^T \tag{6b}$$

where $\widetilde{\mathbf{x}}_i^v$ is the feature vector from the n-th meta-class, and B represents the number of samples in the n-th meta-class.

For the incomplete instances from the meta-class, we can obtain complete instances $\mathbf{X_k}$ by sampling from the specific-view distributions of the missing view. For the j-th incomplete rare instance with label y, we use a heuristic approach to calibrate the Gaussian parameters as follows:

$$\widetilde{\mu}_j^v = \frac{N \sum_{n=1}^{N} T_{nj}^v \mu_n^v + \widetilde{\mathbf{x}}_j^v}{N+1}, \tag{7a}$$

$$\widetilde{\Sigma}_j^v = \frac{N \sum_{n=1}^{N} T_{nj}^v \boldsymbol{\Sigma}_n^v}{N} + \delta \tag{7b}$$

where T_{nm}^v indicates the importance of the n-th meta-class for the rare instance $\widetilde{\mathbf{x}}_j^v$, and δ is a hyperparameter that controls the degree of dispersion of the sampled features. Next, for class y, which contains n rare samples, we derive a set of specific-view calibrated distributions:

$$\mathbb{S}_y^v = \left\{ \left(\widetilde{\mu}_1^v, \widetilde{\Sigma}_1^v \right), \ldots, \left(\widetilde{\mu}_z^v, \widetilde{\Sigma}_z^v \right) \right\} \tag{8}$$

Therefore, a rare incomplete instance with label y can be imputed through multiple samplings from the calibrated distributions. More specifically, we can obtain N_s complete training instances $\{\mathbf{X}_s\}_{s=1}^{N_s}$ via N_s samplings.

$$\mathbb{D}_y = \left\{ \left(\{\mathbf{x}^v\}_{v=1}^{V}, y \right) \mid \mathbf{x}^v \sim \mathcal{N}(\mu^v, \Sigma^v), \mathcal{N} \in \mathbb{S}_y^v \right\}. \tag{9}$$

Multiple samplings from calibrated distributions effectively augment the available training data, enabling the model to train on a larger dataset and enhancing its generalization capability. During testing, where labels are unavailable, missing views are imputed by identifying and utilizing the k-nearest neighbors (Zhang 2016) in the dataset.

Loss function. In this subsection, we introduce the training of a DNN to obtain the multi-view representation. For conventional neural network-based classifiers, the cross-entropy loss is commonly used. However, in biased incomplete multi-view learning, data often exhibits varying levels of quality and informativeness due to incomplete views, which can obscure the underlying class characteristics and lead to incorrect predictions.

For each completed training instance $(\{\mathbf{x}_i^v\}_{v=1}^V, y_i)$, $\mathbf{e}_i^v = f(\mathbf{x}_i^v)$ represents the specific-view feature vector predicted by the network. To address the issue of uncertain quality from completed samples and ensure the model learns effectively from all samples, especially those that are harder to learn, we integrate a focal mechanism (Lin et al. 2017) into our loss function, as described below:

$$\mathcal{L}_{\text{focal}}(\mathbf{e}_i^v) = -\sum_{y=1}^C (1 - p_y)^\gamma \log(p_y) \qquad (10)$$

where p_y is the predicted probability of instance assigned to c-th class, γ is the focusing parameter that can be tuned. A higher value of the γ lowers the loss of the easy samples, which enables the model to turn its attention toward more harder samples.

Considering the inherent class imbalance in biased incomplete multi-view data, this issue is exacerbated after imputing missing views, as data-rich classes often suffer from a higher missing rate. This skews the decision boundaries towards data-rich classes, making the classifier's outputs less deterministic (Chen, Han, and Debattista 2024). Therefore, we introduce an additional regularization term in the loss function, which encourages the model to balance per-class margins, called Category-Aware Marginal Adjustment:

$$\mathcal{L}_{\text{margin}}(\mathbf{e}_i^v) = \log(1 + \sum_{y \neq y_i} e^{\delta_{y_n}} \cdot e^{p_y - p_{y_i}}) \qquad (11)$$

where $\delta_{y_i} \propto \mathbb{P}(y_i)^{-1/4}$, a larger δ_{y_n} means the model encounters greater resistance when predicting these classes, manifested as larger logit differences. It would effectively push the decision boundary further away from the rare and harder-to-classify categories, ensuring better class separation and robustness in classification. Hence the overall loss for our model, $\mathcal{L}_\mathcal{A}$, is the sum of the above two terms:

$$\mathcal{L}_\mathcal{A}(\mathbf{e}_i^v) = \mathcal{L}_{\text{focal}}(\mathbf{e}_i^v) + \beta \mathcal{L}_{\text{margin}}(\mathbf{e}_i^v) \qquad (12)$$

where β are considered hyperparameters which can be adjusted. We will investigate the impact of this parameter on model performance in future experiments. Additionally, to ensure that view-speicfic and aggregated feature receive the same supervised guidance. We apply a multitasking strategy, considering both single and multiple views:

$$\mathcal{L} = \mathcal{L}_\mathcal{A}(\mathbf{e}_i) + \lambda \sum_{v=1}^V \mathcal{L}_\mathcal{A}(\mathbf{e}_i^v) \qquad (13)$$

where $\mathbf{e}_i$ represents the aggregated features from multiple views, λ is a changeable parameter.

Experiment

In this section, we conduct extensive experiments on five multi-view datasets. General implementation details are provided in supplementary material.

Experimental Setup

Dataset. **HandWritten** (Perkins and Theiler 2003) comprises 2,000 instances of handwritten numerals ranging from '0' to '9', which is a six-view dataset. **Scene15** (Fei-Fei and Perona 2005) consists of 4,485 images associated from 15 different categories with three views. **Animal** (Lampert, Nickisch, and Harmeling 2013) is a animal dataset containing 10,158 instances from 50 categories, we utilize two types of features using DECAF and VGG19. **YalB** (Georghiades, Belhumeur, and Kriegman 2001) is a three views dataset contains 10 categories, with a total of 650 facial images. **BRCA** (Wang et al. 2020a) is a three views dataset for classifying five Breast Invasive Carcinoma subtypes.

Compared Methods. We compare the proposed method with the following methods: (1) **UIMC** (Xie et al. 2023) explores and exploits uncertainty to enhance incomplete multi-view classification by modeling the uncertainty in missing views and integrating it into the learning process. (2) **MI-WAE** (Mattei and Frellsen 2019) employ deep generative modeling to impute missing data by training an importance-weighted autoencoder. (3) **CPSPAN** (Jin et al. 2023) performs deep incomplete multi-view clustering by aligning cross-view partial samples and prototypes to handle missing data effectively. (4) **DeepIMV** (Lee and Van der Schaar 2021) integrates multi-views data by utilizing a variational information bottleneck approach to effectively impute missing data and learn a shared latent representation with the available data. (5) **CPM-Nets** (Zhang et al. 2019) directly leverage partial views to learn a common latent space for arbitrary missing data by integrating information across views.

Implementation Details. To create the biased incomplete multi-view dataset, we apply varying missing rates across different categories in the training set, with the highest missing rates assigned to categories with the largest number of samples. For the test set, a uniform missing rate of 0.1 is applied. For our proposed model, we use a fully connected network to extract view-specific features. The network is trained using the Adam optimizer with L2-norm regularization. The learning rate is initialized at 0.001 and decays by a factor of 0.1 if the validation loss does not improve over five epochs. To prevent overfitting, we incorporate a dropout rate of 0.2 and employ early stopping during training. We use the same hyperparameter values across all datasets: $\delta = 0.21$ in Eq. (7b), $N_s = 10$ in Eq. (9), $\gamma = 2$ in Eq. (10), and $k = 10$ for k-nearest neighbors.

We create class-imbalance version of the above datasets by sampling a subset from the original dataset following the Pareto distribution. Each dataset consists of an imbalanced training set and a balanced test set. We determine the meta-classes based on a predefined threshold for the number of samples in each dataset. Details of each dataset are available in the supplementary material.

Datasets	Methods	Maximum missing rates					
		$\eta = 0$	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.3$	$\eta = 0.4$	$\eta = 0.5$
Handwritten	UIMC	0.4020 ± 0.01	0.3680 ± 0.02	0.359 ± 0.02	0.3450 ± 0.02	0.3170 ± 0.02	0.3130 ± 0.01
	MIWAE	0.5540 ± 0.03	0.5190 ± 0.05	0.5100 ± 0.01	0.4930 ± 0.03	0.4850 ± 0.05	0.4770 ± 0.04
	CPSPAN	0.6850 ± 0.02	0.6596 ± 0.04	0.6596 ± 0.04	0.6581 ± 0.04	0.6554 ± 0.03	0.6249 ± 0.04
	DeepIMV	0.6270 ± 0.02	0.6210 ± 0.01	0.6170 ± 0.02	0.6030 ± 0.02	0.6040 ± 0.02	0.6030 ± 0.02
	CMP-Nets	0.8705 ± 0.01	0.8615 ± 0.01	0.8615 ± 0.02	0.8615 ± 0.01	0.8530 ± 0.02	0.8455 ± 0.01
	Ours	**0.9650 ± 0.00**	**0.9500 ± 0.01**	**0.9440 ± 0.01**	**0.9430 ± 0.02**	**0.9380 ± 0.01**	**0.9290 ± 0.02**
Scene15	UIMC	0.4539 ± 0.02	0.3555 ± 0.01	0.3201 ± 0.01	0.2825 ± 0.01	0.2250 ± 0.01	0.2190 ± 0.01
	MIWAE	0.2349 ± 0.02	0.2229 ± 0.03	0.2190 ± 0.02	0.2171 ± 0.02	0.1949 ± 0.01	0.1911 ± 0.02
	CPSPAN	0.3217 ± 0.03	0.3183 ± 0.02	0.3133 ± 0.01	0.3096 ± 0.03	0.3060 ± 0.03	0.3143 ± 0.02
	DeepIMV	0.4876 ± 0.02	0.4819 ± 0.02	0.4717 ± 0.02	0.4609 ± 0.02	0.4590 ± 0.01	0.4392 ± 0.02
	CMP-Nets	0.4260 ± 0.04	0.4127 ± 0.05	0.4073 ± 0.05	0.4022 ± 0.01	0.3968 ± 0.04	0.3886 ± 0.01
	Ours	**0.7034 ± 0.00**	**0.6819 ± 0.00**	**0.6565 ± 0.01**	**0.6425 ± 0.03**	**0.6241 ± 0.03**	**0.5954 ± 0.02**
Animal	UIMC	0.3970 ± 0.02	0.3480 ± 0.02	0.3280 ± 0.01	0.3180 ± 0.01	0.3100 ± 0.01	0.2980 ± 0.01
	MIWAE	0.4180 ± 0.02	0.3890 ± 0.01	0.3690 ± 0.02	0.3640 ± 0.02	0.3560 ± 0.02	0.3400 ± 0.02
	CPSPAN	0.4329 ± 0.01	0.4240 ± 0.02	0.4264 ± 0.02	0.4126 ± 0.02	0.4161 ± 0.03	0.3995 ± 0.03
	DeepIMV	0.6100 ± 0.01	0.6050 ± 0.01	0.5800 ± 0.03	0.5710 ± 0.01	0.5650 ± 0.01	0.5650 ± 0.02
	CMP-Nets	0.6905 ± 0.07	0.6795 ± 0.05	0.6720 ± 0.05	0.6630 ± 0.03	0.6385 ± 0.04	0.5870 ± 0.05
	Ours	**0.7870 ± 0.00**	**0.7750 ± 0.01**	**0.7620 ± 0.02**	**0.7670 ± 0.02**	**0.7440 ± 0.01**	**0.7400 ± 0.02**
YaleB	UIMC	0.4314 ± 0.05	0.3271 ± 0.03	0.3272 ± 0.04	0.3257 ± 0.03	0.2571 ± 0.03	0.2700 ± 0.03
	MIWAE	0.4840 ± 0.03	0.4680 ± 0.03	0.4320 ± 0.04	0.4480 ± 0.03	0.4460 ± 0.04	0.4180 ± 0.03
	CPSPAN	0.3004 ± 0.02	0.2926 ± 0.02	0.2826 ± 0.02	0.2740 ± 0.03	0.2715 ± 0.01	0.2681 ± 0.02
	DeepIMV	0.6600 ± 0.02	0.6514 ± 0.02	0.6486 ± 0.03	0.6371 ± 0.03	0.6229 ± 0.06	0.5571 ± 0.03
	CMP-Nets	0.7572 ± 0.02	0.7457 ± 0.04	0.7343 ± 0.02	0.7172 ± 0.04	0.7029 ± 0.04	0.6929 ± 0.05
	Ours	**0.9285 ± 0.01**	**0.9142 ± 0.00**	**0.8857 ± 0.01**	**0.8571 ± 0.01**	**0.8285 ± 0.02**	**0.8143 ± 0.01**
BRCA	UIMC	0.6240 ± 0.01	0.6160 ± 0.01	0.6000 ± 0.01	0.5800 ± 0.01	0.5520 ± 0.03	0.5500 ± 0.03
	MIWAE	0.6440 ± 0.02	0.6300 ± 0.03	0.5960 ± 0.04	0.5960 ± 0.04	0.5900 ± 0.03	0.5840 ± 0.02
	CPSPAN	0.6288 ± 0.04	0.6259 ± 0.04	0.5973 ± 0.06	0.5690 ± 0.05	0.5968 ± 0.05	0.5931 ± 0.05
	DeepIMV	0.7400 ± 0.03	0.7200 ± 0.03	0.7040 ± 0.05	0.6760 ± 0.05	0.6720 ± 0.01	0.6720 ± 0.03
	CMP-Nets	0.7900 ± 0.03	0.8000 ± 0.03	0.7740 ± 0.02	0.7760 ± 0.04	0.7720 ± 0.04	0.7600 ± 0.03
	Ours	**0.8720 ± 0.01**	**0.8600 ± 0.01**	**0.8480 ± 0.01**	**0.8440 ± 0.01**	**0.8420 ± 0.02**	**0.8360 ± 0.02**

Table 1: Accuracy of ours and other compared methods on five datasets with different maximum missing rates, where $\eta = 0$ denotes all views are compelte. The best results are highlighted by boldface.

Experiment Results

Perfomance Comparison. Table 1 shows the performance of our proposed method and other methods across six maximum missing rates η, ranging from 0 to 0.5. We can observe the following points: (1) Even when facing the complete data, the performance of RIML surpasses that of other methods. For example, on the Animal dataset, RIML achieves an approximately 9.7% improvement compared to the second-best model. This superior performance is largely attributed to the integration of an enhanced focal loss, which enhances the model to make more reliable predictions under imbanlance class distributions. (2) In general, the performance of all methods decreases clearly as the missing rate increase. However, our model is rather robust to view-missing data, since our method consistently performs well even with a high missing rate. (3) We find the methods, which obtain the recovered missing views from latent representations perform better than other methods. Specifically,

DeepIMV and CMP-Nets are consistently ranked in the top-3 across all datasets, which demonstrates their effectiveness in handling biased incomplete views. (4) We test our method on datasets such as Scene15, with less clear class associations, and Animal, with a larger number of classes. Despite the performance drop compared to datasets with clearer associations, our method still outperforms other methods, demonstrating its robustness in more challenging scenarios.

Imputed complete data visualization. In order to further explore the quality of completion view, we visualized the imputed complete data through t-SNE. We conduct experiments on the HandWritten dataset with maximum missing rate of 0.5. As shwon in Figure 4, our proposed method achieves superior class separability, where data points from the same class naturally cluster together, and the separation between different groups is clear. It demonstrates our imputation strategy is rather robust to non-uniform view-missing patterns. Moreover, the clarity of clusters suggests that our

(a). UIMC   (b). MIWAE   (c). CPSPAN

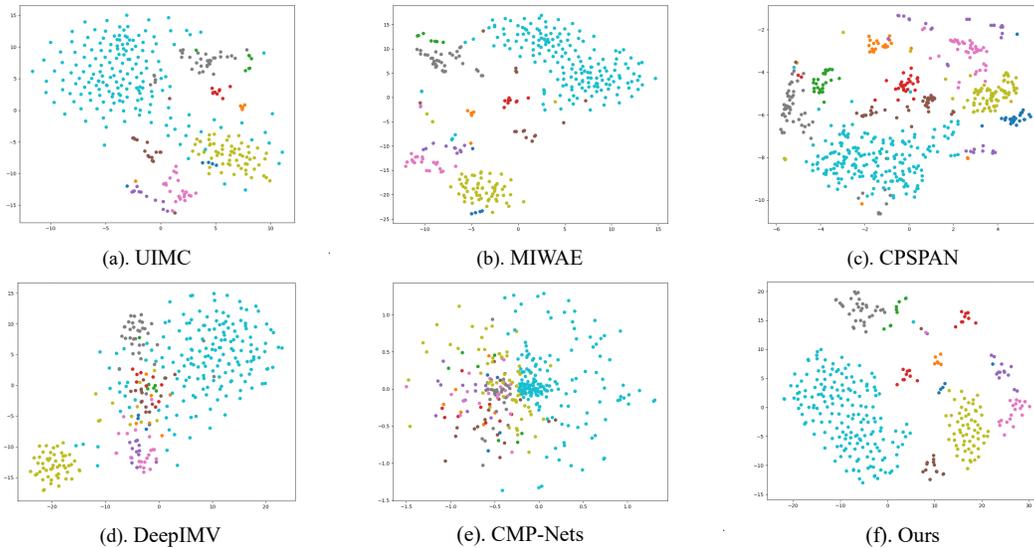(d). DeepIMV   (e). CMP-Nets   (f). Ours

Figure 4: Visualization of imputed complete data (Handwritten) under $\eta = 0.5$, where multiple views are concatenated.
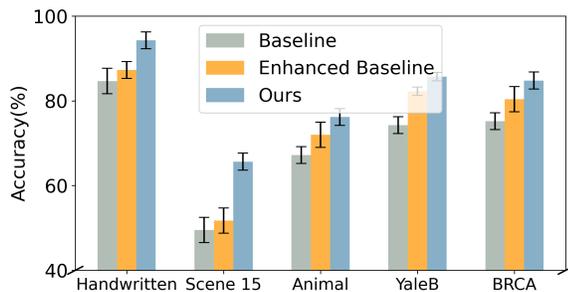

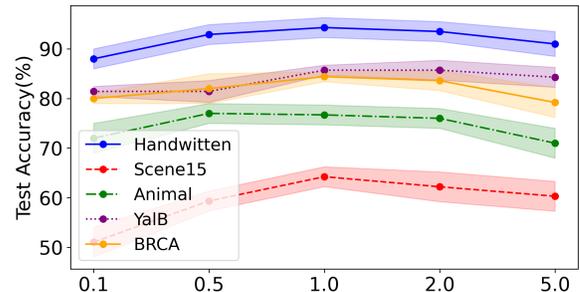Figure 5: Ablation study results on datasets with $\eta = 0.3$.

Figure 6: Accuracy(%) when adjusting β on datasets.

method not only recovers missing data effectively but also maintains the integrity of the underlying data structure.

Ablation study. To explore where our improvement is located, we design ablation experiments across all datasets. Specifically, we compare ours with a baseline imputation method, which fills the missing views by samplings from the view-specific distribution under the same class without calibration, and we train classifiers using the standard focal loss. In addition, we also expand our baseline model by introduce a category-aware focal loss to determine the deficiency of the original focal loss. We refer to this model as the enhanced baseline. From the results in Figure 5, we have the following observations: (1) The results shed light on the impact of imputation quality in incomplete multi-view learning, the perfromance of downstream tasks are primarily dependent on the reliability of imputation. (2) In the presence of class imbalance, adjusting decision boundaries is crucial, as it effectively addresses the issue of poorly distinguishable features between classes. (3) Our method yields more stable results, as it introduces less noise during the view completion phase compared to other methods.

Parameter Analysis. We analyze the sensitivity of thehyperparamter β on all five datasets with the maximum missing rate of 0.3. we set the β ranging from 0.1 to 5.0 in experiments. The results are shown in Figure 6. We can see that RIML achieves consistently good performance when β is around 1.0. Therefore, we draw a conclusion that assigning equal weight to both loss components yield better results.

Conclusion

In this paper, we propose the RIML method to address the challenges posed by the biased incomplete multi-view learning problem. RIML is built on a learning-free imputation framework that imputes missing views using calibrated view-specific distributions. These distributions are constructed by leveraging inter-class associations through optimal transport, facilitating the transfer of knowledge from data-rich to data-poor classes. Furthermore, we integrated a category-sensitive focal loss to learn a more distinguishable feature space, thus enhancing the model's robustness against inherent class imbalances often encountered in real-world datasets. The experimental results confirm the effectiveness of RIML compared to other IMVL methods.

Acknowledgments

References

Chen, C.; Han, J.; and Debattista, K. 2024. Virtual Category Learning: A Semi-Supervised Learning Method for Dense Prediction With Extremely Limited Labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5595–5611.

Chen, H.; Wang, Y.; Yang, X.; and Li, J. 2021. Captioning transformer with scene graph guiding. In *2021 IEEE international conference on image processing (ICIP)*, 2538–2542. IEEE.

Cui, C.; Ren, Y.; Pu, J.; Li, J.; Pu, X.; Wu, T.; Shi, Y.; and He, L. 2024. A novel approach for effective multi-view clustering with information-theoretic perspective. *Advances in Neural Information Processing Systems*, 36.

Fan, G.; Zhang, C.; Wang, K.; and Chen, J. 2022. MV-HAN: A hybrid attentive networks based multi-view learning model for large-scale contents recommendation. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 1–5.

Fei-Fei, L.; and Perona, P. 2005. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, 524–531. IEEE.

Georghiades, A. S.; Belhumeur, P. N.; and Kriegman, D. J. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6): 643–660.

Guo, D.; Tian, L.; Zhao, H.; Zhou, M.; and Zha, H. 2022. Adaptive distribution calibration for few-shot learning with hierarchical optimal transport. *Advances in neural information processing systems*, 35: 6996–7010.

Hu, Y.; Wang, J.; Zhu, H.; Li, J.; and Shi, J. 2024. Cost-Sensitive Weighted Contrastive Learning Based on Graph Convolutional Networks for Imbalanced Alzheimer's Disease Staging. *IEEE Transactions on Medical Imaging*.

Huang, Z.; Ren, Y.; Pu, X.; and He, L. 2021. Non-linear fusion for self-paced multi-view clustering. In *Proceedings of the 29th ACM International Conference on Multimedia*, 3211–3219.

Huang, Z.; Ren, Y.; Pu, X.; Huang, S.; Xu, Z.; and He, L. 2023. Self-supervised graph attention networks for deep weighted multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 7936–7943.

Jin, J.; Wang, S.; Dong, Z.; Liu, X.; and Zhu, E. 2023. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11600–11609.

Kim, K. H.; and Sohn, S. Y. 2020. Hybrid neural network with cost-sensitive support vector machine for class-imbalanced multimodal data. *Neural Networks*, 130: 176–184.

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.

Lee, C.; and Van der Schaar, M. 2021. A variational information bottleneck approach to multi-omics data integration. In *International Conference on Artificial Intelligence and Statistics*, 1513–1521. PMLR.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Ling, Y.; Chen, J.; Ren, Y.; Pu, X.; Xu, J.; Zhu, X.; and He, L. 2023. Dual label-guided graph refinement for multi-view graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8791–8798.

Liu, S.; Liu, X.; Wang, S.; Niu, X.; and Zhu, E. 2022. Fast incomplete multi-view clustering with view-independent anchors. *IEEE Transactions on Neural Networks and Learning Systems*.

Mattei, P.-A.; and Frellsen, J. 2019. MIWAE: Deep generative modelling and imputation of incomplete data sets. In *International conference on machine learning*, 4413–4423. PMLR.

Miao, F.; Yao, L.; and Zhao, X. 2022. Adaptive margin aware complement-cross entropy loss for improving class imbalance in multi-view sleep staging based on eeg signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30: 2927–2938.

Perkins, S.; and Theiler, J. 2003. Online feature selection using grafting. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 592–599.

Sinkhorn, R. 1967. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4): 402–405.

Tan, Y.; and Zhao, G. 2022. Multi-view representation learning with Kolmogorov-Smirnov to predict default based on imbalanced and complex dataset. *Information Sciences*, 596: 380–394.

Tang, H.; and Liu, Y. 2022. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, 21090–21110. PMLR.

Tang, J.; Hou, Z.; Yu, X.; Fu, S.; and Tian, Y. 2023. Multi-view cost-sensitive kernel learning for imbalanced classification problem. *Neurocomputing*, 552: 126562.

Tang, J.; Yi, Q.; Fu, S.; and Tian, Y. 2024. Incomplete multi-view learning: Review, analysis, and prospects. *Applied Soft Computing*, 111278.

Wang, H.; and Zhou, Z. 2021. Multi-view learning based on maximum margin of twin spheres support vector machine. *Journal of Intelligent & Fuzzy Systems*, 40(6): 11273–11286.

Wang, T.; Shao, W.; Huang, Z.; Tang, H.; Zhang, J.; Ding, Z.; and Huang, K. 2020a. Moronet: multi-omics integration via graph convolutional networks for biomedical data classification. *BioRxiv*, 2020–07.

Wang, Z.; Chen, L.; Zhang, J.; Yin, Y.; and Li, D. 2020b. Multi-view ensemble learning with empirical kernel for heart failure mortality prediction. *International journal for numerical methods in biomedical engineering*, 36(1): e3273.

Xia, W.; Gao, Q.; Wang, Q.; and Gao, X. 2022. Tensor completion-based incomplete multiview clustering. *IEEE Transactions on Cybernetics*, 52(12): 13635–13644.

Xie, M.; Han, Z.; Zhang, C.; Bai, Y.; and Hu, Q. 2023. Exploring and exploiting uncertainty for incomplete multi-view classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19873–19882.

Xu, C.; Liu, H.; Guan, Z.; Wu, X.; Tan, J.; and Ling, B. 2021. Adversarial incomplete multiview subspace clustering networks. *IEEE Transactions on Cybernetics*, 52(10): 10490–10503.

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*, 16129–16137.

Xu, J.; Ren, Y.; Shi, X.; Shen, H. T.; and Zhu, X. 2023. UN-TIE: Clustering analysis with disentanglement in multi-view information fusion. *Information Fusion*, 100: 101937.

Zhang, C.; Cui, Y.; Han, Z.; Zhou, J. T.; Fu, H.; and Hu, Q. 2020. Deep partial multi-view learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(5): 2402–2415.

Zhang, C.; Han, Z.; Fu, H.; Zhou, J. T.; Hu, Q.; et al. 2019. CPM-Nets: Cross partial multi-view networks. *Advances in Neural Information Processing Systems*, 32.

Zhang, C.; Li, H.; Chen, C.; Jia, X.; and Chen, C. 2022. Low-rank tensor regularized views recovery for incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhang, H.; and Dong, J. 2020. Application of sample balance-based multi-perspective feature ensemble learning for prediction of user purchasing behaviors on mobile wireless network platforms. *EURASIP Journal on Wireless Communications and Networking*, 2020: 1–26.

Zhang, Q.; Wei, Y.; Han, Z.; Fu, H.; Peng, X.; Deng, C.; Hu, Q.; Xu, C.; Wen, J.; Hu, D.; and Zhang, C. 2024. Multimodal Fusion on Low-quality Data: A Comprehensive Survey. arXiv:2404.18947.

Zhang, Z. 2016. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11).

Zhou, W.; Wang, H.; and Yang, Y. 2019. Consensus graph learning for incomplete multi-view clustering. In *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part I 23*, 529–540. Springer.