

Exploring and Exploiting Uncertainty for Incomplete Multi-View Classification

Mengyao Xie[‡], Zongbo Han[‡], Changqing Zhang^{*}, Yichen Bai, Qinghua Hu
 College of Intelligence and Computing, Tianjin University

zhangchangqing@tju.edu.cn

Abstract

Classifying incomplete multi-view data is inevitable since arbitrary view missing widely exists in real-world applications. Although great progress has been achieved, existing incomplete multi-view methods are still difficult to obtain a trustworthy prediction due to the relatively high uncertainty nature of missing views. First, the missing view is of high uncertainty, and thus it is not reasonable to provide a single deterministic imputation. Second, the quality of the imputed data itself is of high uncertainty. To explore and exploit the uncertainty, we propose an Uncertainty-induced Incomplete Multi-View Data Classification (UIMC) model to classify the incomplete multi-view data under a stable and reliable framework. We construct a distribution and sample multiple times to characterize the uncertainty of missing views, and adaptively utilize them according to the sampling quality. Accordingly, the proposed method realizes more perceivable imputation and controllable fusion. Specifically, we model each missing data with a distribution conditioning on the available views and thus introducing uncertainty. Then an evidence-based fusion strategy is employed to guarantee the trustworthy integration of the imputed views. Extensive experiments are conducted on multiple benchmark data sets and our method establishes a state-of-the-art performance in terms of both performance and trustworthiness.

1. Introduction

Learning from multiple complementary views has the potential to yield more generalizable models. Benefiting from the power of deep learning, multi-view learning has further exhibited remarkable benefits against the single-view paradigm in clustering [1–3], classification [4, 5] and representation learning [6, 7]. However, real-world data are usually incomplete. For instance, in the medical field, patients with same condition may choose different medical examinations producing incomplete/unaligned multi-view

data; similarly, sensors in cars at times may be out of order and thus only part of information can be collected. Flexible and trustworthy utilization of incomplete multi-view data is still very challenging due to the high uncertainty of missing issues.

For the task of incomplete multi-view classification (IMVC), there have been plenty of studies which could be roughly categorized into two main lines. The methods [8, 9] only use available views without imputation to conduct classification. While the other line [10–13] reconstructs missing data based on deep learning methods such as autoencoder [14, 15] or generative adversarial network (GAN) [16], and then utilizes the imputed complete data for classification. Although significant progress has been achieved by existing IMVC methods, there are still limitations: (1) Methods that simply neglect missing views are usually ineffective especially under high missing rate due to the limitation in exploring the correlation among views; (2) Methods that impute missing data based on deep learning methods are short in interpretability and the deterministic imputation way fails to characterize the uncertainty of missing resulting in unstable classification; (3) Few IMVC methods can handle multi-view data with complex missing patterns especially for the data with more than two views, which makes these methods inflexible.

In view of above limitations, we propose a simple yet effective, stable and flexible incomplete multi-view classification model. First, the proposed model characterizes the uncertainty of each missing view by imputing a distribution instead of a deterministic value. The necessity of characterizing the uncertainty for missing data on single view has been well recognized in recent works [17, 18]. Second, we conduct sampling multiple times from the above distribution and each one is combined with the observed views to form multiple completed multi-view samples. The quality of the sampled data is of high uncertainty, and thus we adaptively integrate them according to their quality on the single view and multi-view fusion. For single view, the uncertainty of the low-quality sampled data tends to be large and should not affect the learning of other views. Therefore, we construct an evidence-based classifier for each view to obtain the

[‡]Equal contribution.

^{*}Corresponding author.

opinion including subjective probabilities and uncertainty masses. While for multi-view fusion, the opinions from multiple views are integrated based on DS rule, which ensures the trustworthy utilization of arbitrary-quality views. The main contributions of this work are summarized as follows:

- We propose an exploration-and-exploitation strategy for classifying incomplete multi-view data by characterizing the uncertainty of missing data, which promotes both effectiveness and trustworthiness in utilizing imputed data. To the best of our knowledge, the proposed UIMC is the first work asserting uncertainty in incomplete multi-view classification.
- To fully exploit the high-quality imputed data and reduce the affect of the low-quality imputed views, we propose to weight the imputed data from two aspects, avoiding the negative effect on single view and multi-view fusion. The uncertainty-aware training and fusion significantly ensure the effectiveness and reliability of integrating uncertain imputed data.
- We conduct experiments on classification for data with multiple types of features or modalities, and evaluate the results with diverse metrics, which validates that the proposed UIMC outperforms existing methods in the above tasks and is trustworthy with reliable uncertainty.

2. Related Work

2.1. Incomplete Multi-View Learning

Incomplete multi-view learning can be generally divided into two main lines in terms of how to handle missing views. Specifically, existing works mainly focus on neglecting or complementing the missing views based on deep-learning methods. **Methods without imputation.** The methods only use present views and directly learn the common latent subspace or representation for all views in clustering [19–21] and classification [8, 9]. **Generative Methods.** The methods impute missing views with the present views and then utilize the reconstructed complete data to conduct downstream tasks [10, 11, 22–28]. Specifically, one of the most popular ways is applying the structure of variational auto-encoder on partial multi-view data to reconstruct missing views [10, 11, 22]. Generative adversarial network is also used to generate missing views [26–28]. Besides, there are some methods to obtain imputations based on kernel CCA [23], spectral graph [24], and information theory [25]. Compared with the above algorithms, our method obtains multiple imputations instead of single imputation and then dynamically evaluates the imputation quality. Thus the more reliable downstream classification tasks can be performed.

2.2. Uncertainty Estimation

One of the key points of our method is to explore and exploit the uncertainty in missing data. To achieve high-

quality uncertainty estimation, many approaches have been proposed [29–31]. The uncertainty in deep learning can be generally divided into aleatoric uncertainty and epistemic uncertainty [32–34]. Aleatoric uncertainty refers to the uncertainty caused by data and it measures the inherent noise of data. Aleatoric uncertainty can be further divided into homoscedastic uncertainty and heteroscedastic uncertainty, while the first one which varies with different tasks is usually used to estimate the uncertainty in multi-task learning [35, 36], and the latter one which varies with input is useful when the input space includes variable noise [37, 38]. On the other hand, epistemic uncertainty refers to uncertainty caused by insufficient model training and can be eliminated in theory. It can be estimated by predicting an uncertain observation using models with different parameters, the instability of predicting results just reflects the epistemic uncertainty [32, 39]. In this work, we estimate the aleatoric uncertainty of imputations by adopting subjective logic [40] and Dempster-Shafer theory [41] to construct a trustworthy and reliable multi-view classification network.

3. Method

The key goal of the proposed method is to explore and exploit the uncertainty of incomplete multi-view data, promoting both effectiveness and trustworthiness of the model. We first introduce the background of incomplete multi-view classification in Sec. 3.1 and then present how to characterize the uncertainty of imputed view in Sec. 3.2, exploit the imputation uncertainty and integrate the uncertain decisions in Sec. 3.3, and finally demonstrate how to obtain classification predictions with multiple imputations in Sec. 3.4.

3.1. Background

Given N training inputs $\{\mathbb{X}_n\}_{n=1}^N$ with V views, i.e., $\mathbb{X} = \{\mathbf{x}^v\}_{v=1}^V$, and the corresponding class labels $\{\mathbf{y}_n\}_{n=1}^N$, multi-view classification aims to construct a mapping between input and label by exploiting the complementary multi-view data. In this paper, we focus on the incomplete multi-view classification task defined in Def. 3.1.

Definition 3.1 (*Incomplete Multi-View Classification*) *Formally, a complete multi-view sample is composed of V views $\mathbb{X} = \{\mathbf{x}^v\}_{v=1}^V$ and the corresponding class label \mathbf{y} . An incomplete multi-view observation $\bar{\mathbb{X}}$ is a subset of the complete multi-view observation (i.e., $\bar{\mathbb{X}} \subseteq \mathbb{X}$) with arbitrary possible \bar{V} views, where $1 \leq \bar{V} \leq V$. Given an incomplete multi-view training dataset $\{\bar{\mathbb{X}}_n, \mathbf{y}_n\}_{n=1}^N$ with N samples, incomplete multi-view classification aims to learn a mapping between the incomplete multi-view observation $\bar{\mathbb{X}}$ and the corresponding class label \mathbf{y} .*

There are two main lines to solve the incomplete multi-view classification problem, including imputing and neglecting the missing views. Specifically, the imputation-based

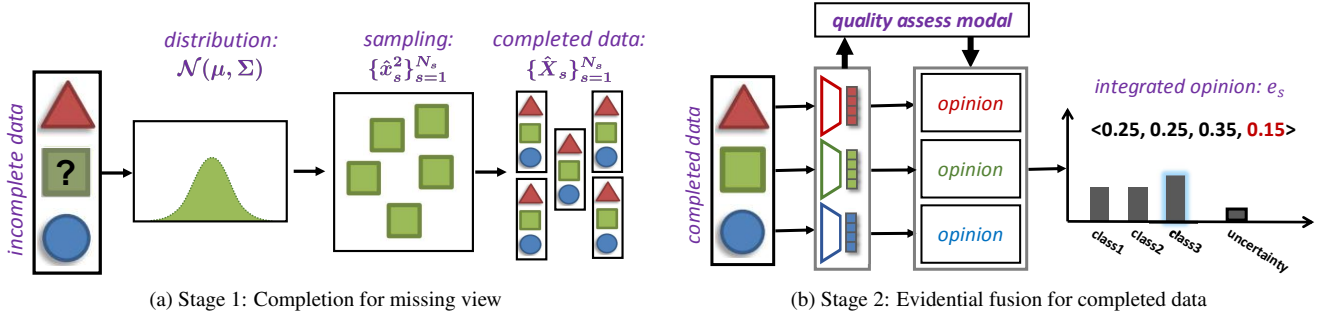


Figure 1. Illustration of the proposed UIMC. For clarity, we use 3-view data with one missing view as example in the Fig. 1a. The overall process of UIMC consists of two stages. In the stage 1, we characterize the underlying uncertainty of missing view with a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ based on the non-parameterized nearest-neighbor strategy. Then we can obtain N_s complete instances $\{\hat{\mathbb{X}}_s\}_{s=1}^{N_s}$ by N_s samplings from the distribution of the missing view. In the stage 2, for instance $\hat{\mathbb{X}}_s$, to alleviate the negative impact from low-quality imputations, we exploit the uncertainty of each view with evidential classification. Then the Dempster’s combination rule introduced in Def. 3.2 is adopted to integrate the subjective opinions from multiple views, which enables the model to trustworthily exploit the imputation data.

methods [10–13] complete the missing views based on observed data, but they basically ignore that the untrustworthy imputation will harm downstream tasks. In contrast, the methods neglecting the missing views [8, 9] train the classification model only using the available views, which usually lack of the exploration of the correlation among different views. The proposed method aims to learn a more trustworthy model by exploring and exploiting the uncertainty of the missing views.

3.2. Imputing Missing Views

Given an incomplete multi-view data-instance $\bar{\mathbb{X}}$, we consider reasonably imputing its missing views based on the observed views, which could potentially promote the downstream classification model.

Characterizing uncertainty for missing views. Most of existing incomplete multi-view classification methods [10–13] impute the missing views with single imputation. Formally, they usually construct a deterministic mapping from the incomplete multi-view data-instance $\bar{\mathbb{X}}$ to complete data, i.e., $f: \bar{\mathbb{X}} \rightarrow \mathbb{X}$. The single imputation ignores the high uncertainty nature of the imputed data, which might negatively impact downstream classification tasks due to the untrustworthy imputation. This phenomenon has also been well recognized in single-view classification tasks with missing attributes [18, 42]. Instead of imputing missing views with deterministic mapping, we characterize the underlying uncertainty of missing views with a distribution based on the nearest-neighbor-based non-parameterized strategy.

Constructing the nearest neighbor set. Without loss of generality, we consider that there is an incomplete multi-view training instance $\bar{\mathbb{X}}$ and its corresponding label \mathbf{y} with m -th view missing, we aim to impute the missing view \mathbf{x}^m based on the information of other training samples whose

m -th view is available. Inspired by the classical classification algorithm k -nearest neighbors [43], we employ a non-parameterized method to construct the distribution of missing view \mathbf{x}^m by exploring the neighbors of $\bar{\mathbb{X}}$. Specifically, for the v -th view ($v \neq m$), given the available observation $\mathbf{x}^v \in \bar{\mathbb{X}}$, we construct a neighbor set by finding its k -nearest neighbors in other samples with the same label in the following way. Firstly, we construct the distance set \mathbb{D}^v by computing the distance between \mathbf{x}^v and \mathbf{x}_n^v , where \mathbf{x}_n^v is the data of v -th view of the samples available for m -th view and $\mathbf{y} = \mathbf{y}_n$,

$$\mathbb{D}^v = \left\{ -\|\mathbf{x}^v - \mathbf{x}_n^v\|^2 \mid \mathbf{x}_n^v, \mathbf{x}_n^m \text{ are available and } \mathbf{y} = \mathbf{y}_n \right\}. \quad (1)$$

Then we can obtain the nearest neighbor indicator set \mathbb{I}^v with

$$\mathbb{I}^v = \left\{ i \mid -\|\mathbf{x}^v - \mathbf{x}_i^v\|^2 \in \text{topk}(\mathbb{D}^v) \right\}, \quad (2)$$

where $\text{topk}(\cdot)$ is an operator to select the k -nearest neighbors from the training set according to the distance set \mathbb{D}^v . Note that, during the test time, the missing view \mathbf{x}^m is imputed by selecting the nearest neighbors from the training set being independent with labels.

Imputation with statistical information. Given the nearest neighbor indicator set $\mathbb{I} = \{\mathbb{I}^v\}_{v \neq m}$, we characterize the distribution of the missing view \mathbf{x}^m with its neighbor set $\{\mathbf{x}_i^m\}_{i \in \mathbb{I}}$ statistically. Formally, we assume that the missing view \mathbf{x}^m follows a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix calculated from neighbor set with

$$\boldsymbol{\mu} = \frac{\sum_{i \in \mathbb{I}} \mathbf{x}_i^m}{|\mathbb{I}|}, \boldsymbol{\Sigma} = \frac{1}{|\mathbb{I}| - 1} \sum_{i \in \mathbb{I}} (\mathbf{x}_i^m - \boldsymbol{\mu})(\mathbf{x}_i^m - \boldsymbol{\mu})^T. \quad (3)$$

Then \mathbf{x}^m can be imputed by taking multiple samplings from the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. More specifically, given an incomplete training instance $\bar{\mathbf{X}}$, we can obtain N_s complete training instances $\{\hat{\mathbf{X}}_s\}_{s=1}^{N_s}$ via N_s samplings. The imputation framework can be referred to Fig. 1a.

3.3. Classifying Imputed Multi-View Samples

After obtaining imputed multi-view training dataset $\{\{\hat{\mathbf{X}}_{n,s}\}_{n=1}^N\}_{s=1}^{N_s}$, we consider alleviating the negative impact from low-quality imputations by exploiting the uncertainty of imputed samples. Due to inherent noise in imputation, imputed multi-view data are usually of high uncertainty. Therefore we first estimate the uncertainty of different views based on an evidential multi-view learning framework and then conduct uncertainty-based decision fusion. The overall framework is shown in Fig. 1b.

Different from traditional classification algorithm, evidential classification [44, 45] defines a theoretical framework to obtain the subjective opinions $\mathcal{S} = \{\{b_k\}_{k=1}^K, u\}$ including subjective probabilities (belief masses) $\{b_k\}_{k=1}^K \geq 0$ and overall uncertainty mass $u \geq 0$, where K denotes the number of classes and $\sum_{k=1}^K b_k + u = 1$. The subjective opinion is associated with a Dirichlet distribution with parameters $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]$. Specifically, a belief mass b_k can be derived easily from the parameters of the corresponding Dirichlet distribution with $b_k = (\alpha_k - 1)/\alpha_0$, where $\alpha_0 = \sum_{k=1}^K \alpha_k$ is the Dirichlet strength.

To obtain the Dirichlet distribution parameters for each view $\hat{\mathbf{x}}^v \in \hat{\mathbf{X}}$, we construct the evidence neural network by replacing *softmax* layer of traditional neural network classifier with activation function such as *softplus* so that the positive output can be regarded as evidence vector $\mathbf{e}^v = [e_1^v, \dots, e_K^v]$. Then the parameters of Dirichlet distribution $Dir(\mathbf{p}^v | \boldsymbol{\alpha}^v)$ for view $\hat{\mathbf{x}}^v$ are obtained with $\alpha_k^v = e_k^v + 1$. Under the Variational framework, the evidential classification loss function \mathcal{L}^v is a combination of classification loss \mathcal{L}_c^v and a regularization term \mathcal{L}_r^v [46]. More specifically, the classification objective function can be regarded as an integral of the traditional cross-entropy loss on the simplex determined by $Dir(\mathbf{p}^v | \boldsymbol{\alpha}^v)$. Specifically, it is induced as:

$$\mathcal{L}_c^v(\boldsymbol{\alpha}^v | \mathbf{x}^v) = \sum_{k=1}^K y_k (\psi(\alpha_0^v) - \psi(\alpha_k^v)), \quad (4)$$

where y_k is the k -th element of \mathbf{y} represented by a one-hot vector and $\psi(\cdot)$ is the digamma function. To obtain a reasonable Dirichlet distribution, we employ the following equation to add a prior for the Dirichlet distribution as the regularization term:

$$\mathcal{L}_r^v(\boldsymbol{\alpha}^v | \mathbf{x}^v) = D_{KL}[Dir(\mathbf{p}^v | \tilde{\boldsymbol{\alpha}}^v) \| Dir(\mathbf{p}^v | [1, \dots, 1])], \quad (5)$$

where $\tilde{\boldsymbol{\alpha}}^v = \mathbf{y} + (1 - \mathbf{y}) \odot \boldsymbol{\alpha}^v$ is the Dirichlet distribution after replacing the α_k corresponding to the label with 1. $Dir(\mathbf{p}^v | [1, \dots, 1])$ is the uniform Dirichlet distribution. Then for each view the overall loss function can be written as

$$\mathcal{L}^v(\boldsymbol{\alpha}^v | \mathbf{x}^v) = \mathcal{L}_c^v(\boldsymbol{\alpha}^v | \mathbf{x}^v) + \lambda \mathcal{L}_r^v(\boldsymbol{\alpha}^v | \mathbf{x}^v), \quad (6)$$

where λ is the annealing coefficient gradually changing from 0 to 1 during training to control constraint strength. We could obtain the Dirichlet distribution $Dir(\mathbf{p}^v | \boldsymbol{\alpha}^v)$ by minimizing \mathcal{L}^v . Then the corresponding subjective opinions could be derived with $b_k^v = (\alpha_k^v - 1)/\alpha_0^v$ and $u^v = 1 - \sum_{k=1}^K b_k^v$.

Now we consider integrating the subjective opinions from different views according to the uncertainty. Dempster-Shafer theory allows to integrate subjective opinions from different sources to produce a more comprehensive opinion [47]. Specifically, the Dempster's combination rule for two different views is defined in Def. 3.2. Accordingly, the integrated multi-view subjective opinion can be derived with $\mathcal{S}^m = \mathcal{S}^1 \oplus \mathcal{S}^2 \oplus \dots \oplus \mathcal{S}^V = \{\{b_k^m\}_{k=1}^K, u^m\}$. The corresponding parameters of the integrated multi-view Dirichlet distribution $Dir(\mathbf{p}^m | \boldsymbol{\alpha}^m)$ are obtained with $\alpha_k^m = b_k^m \times \alpha_0^m + 1$.

Definition 3.2 (*Dempster's Combination Rule.*) *The subjective opinion $\mathcal{S} = \{\{b_k\}_{k=1}^K, u\}$ after integration can be obtained from two subjective opinion $\mathcal{S}^1 = \{\{b_k^1\}_{k=1}^K, u^1\}$ and $\mathcal{S}^2 = \{\{b_k^2\}_{k=1}^K, u^2\}$ with $\mathcal{S} = \mathcal{S}^1 \oplus \mathcal{S}^2$. Specifically, it can be written as:*

$$b_k = \frac{1}{1 - C} (b_k^1 b_k^2 + b_k^1 u^2 + b_k^2 u^1), u = \frac{1}{1 - C} u^1 u^2, \quad (7)$$

where $C = \sum_{i \neq j} b_i^i b_j^j$.

To obtain reliable both single-view and integrated Dirichlet distribution, we employ a multi-task strategy. Specifically, the final loss function in classification period composes of which of single and multiple views:

$$\mathcal{L} = \sum_{n=1}^N \sum_{s=1}^{N_s} \left\{ \mathcal{L}^m(\boldsymbol{\alpha}^m) + \sum_{v=1}^V \mathcal{L}^v(\boldsymbol{\alpha}^v | \mathbf{x}_{n,s}^v) \right\}, \quad (8)$$

where

$$\begin{aligned} \mathcal{L}^m(\boldsymbol{\alpha}^m) &= \sum_{k=1}^K y_k (\psi(\alpha_0^m) - \psi(\alpha_k^m)) \\ &+ \lambda D_{KL}[Dir(\mathbf{p}^m | \tilde{\boldsymbol{\alpha}}^m) \| Dir(\mathbf{p}^m | [1, \dots, 1])]. \end{aligned} \quad (9)$$

The pseudo-code of UIMC is provided in Alg. 1.

Algorithm 1: Training Process of UIMC

Input: Incomplete multi-view training set $\{\tilde{\mathbf{X}}_n, \mathbf{y}_n\}_{n=1}^N$.
Initialize: Initialize the parameters of UIMC.
for $n = 1 : N$ **do**
 Impute missing view x_n^m by taking N_s samplings from distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with Eq. 3.
end
while *not converged* **do**
 for $v = 1 : V$ **do**
 Obtain evidence e_n^v with DNN;
 Obtain $\alpha_n^v = e_n^v + 1$ and $\mathcal{S}_n^v = \{\{b_{n,k}^v\}_{k=1}^K, u_n^v\}$.
 end
 Obtain $\alpha_n = [\alpha_{n,1}, \dots, \alpha_{n,K}]$ and $\mathcal{S}_n = \{\{b_{n,k}\}_{k=1}^K, u_n\}$ through Def. 3.2.
 Calculate overall loss \mathcal{L} in Eq. 8;
 Use an optimized algorithm to update the parameters.
end
Output: Parameters of model.

3.4. Predicting with Multiple Imputations

During the test stage, for each incomplete sample, we could obtain N_s imputed multi-view samples $\{\tilde{\mathbf{X}}_s\}_{s=1}^{N_s}$. Then we employ a voting strategy to acquire the classification prediction with N_s imputations for each sample. Specifically, given the N_s imputed test instances $\{\tilde{\mathbf{X}}_s\}_{s=1}^{N_s}$, we can obtain N_s label predictions $\{\hat{y}_s\}_{s=1}^{N_s}$. The final predict label for multi-view test instance $\tilde{\mathbf{X}}$ would be the most frequent element in $\{\hat{y}_s\}_{s=1}^{N_s}$.

4. Experiments

In this section, we conduct extensive experiments on multiple multi-view datasets with various missing rates $\eta = \frac{\sum_{v=1}^V M^v}{V \times N}$ to answer the following questions, where M^v indicates the number of observations without the v -th view. Q1 Effectiveness (I). Is the proposed UIMC superior to other methods? Q2 Effectiveness (II). How about the quality of the imputed complete multi-view data of different methods? Q3 Ablation study (I). Is multiple-imputation strategy really better than single-imputation strategy? Q4 Ablation study (II). Do the uncertainty-guided classification and fusion strategy help with performance? Q5 Stability. How about the stability of the proposed method under the sampling operation?

4.1. Datasets and Comparison Methods

To validate the effectiveness of UIMC, we conduct experiments on five datasets. **YaleB** [48] is a 3-view dataset

contains 10 categories, and there are 65 facial images in each category. **ROSMAP** [49] is a 3-view dataset contains two categories: Alzheimer’s disease (AD) patients with 182 samples and normal control (NC) with 169 samples. **Hand-written** [50] is a 6-view dataset contains 10 categories from digit “0” to “9”. The number of samples in each category is 200. **BRCA** [49] is a 3-view dataset for Breast Invasive Carcinoma (BRCA) subtype classification, and it contains 5 categories. The number of samples in each category is between 46 and 436. **Scene15** [51] is a 3-view dataset contains 15 categories for scene classification, and the number of samples in each category is between 210 and 410.

We compare the proposed UIMC with the following methods: (1) **Mean-Imputation** simply imputes missing view \mathbf{x}^m with the mean of all available observations on the m -th view. (2) **GCCA** [52] extends Canonical Correlation Analysis (CCA) [53] to handle data with more than two views. (3) **TCCA** [54] obtains a common subspace shared by all views by maximizing the canonical correlation of multiple views. (4) **MVAE** [10] extends variational autoencoder to multi-view data and adopts product-of-experts strategy to obtain a common latent subspace. (5) **MIWAE** [11] extends importance-weighted autoencoder to multi-view data to impute missing data. (6) **CPM-Nets** [9] directly learns the joint latent representations for all views with available data, and maps the latent representation to classification predictions. (7) **DeepIMV** [8] applies the information bottleneck (IB) framework to obtain marginal and joint representations with the available data, and constructs the view-specific and multi-view predictors to obtain the classification predictions.

4.2. Experiment Results

We now provide detailed empirical results to investigate the above key questions, which can validate the effectiveness and trustworthiness of our model.

Q1 Effectiveness (I). We evaluate our algorithm by comparing it with the state-of-the-art incomplete multi-view classification methods with different missing rates $\eta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$. We employ classification accuracy as the evaluation metric following prior works [55]. For the Mean-Imputation method, we impute the missing views with corresponding means and then train classifiers based on the completed multi-view data. For multi-view learning methods including GCCA and TCCA, we first impute the missing views with the corresponding means, and then train GCCA/TCCA to obtain the joint concatenated representation for all views. Finally, we train a classifier based on the multi-view representation. For incomplete multi-view generative methods including MVAE and MIWAE, we take an impute-then-classify strategy. Specifically, we train MVAE/MIWAE to obtain the imputed multi-view data, and then train classifiers both on single-view and multi-view data based on the imputation. For the classification methods that

Table 1. Comparison in terms of classification accuracy (mean±std) with $\eta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$ on five datasets.

Datasets	Methods	Missing rates					
		$\eta = 0$	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.3$	$\eta = 0.4$	$\eta = 0.5$
YaleB	Mean-Imputation	1.0000 ± 0.00	0.9923 ± 0.00	0.9769 ± 0.01	0.9769 ± 0.01	0.9692 ± 0.01	0.9615 ± 0.01
	GCCA	0.9692 ± 0.00	0.9385 ± 0.01	0.9077 ± 0.02	0.8615 ± 0.03	0.8385 ± 0.02	0.8231 ± 0.02
	TCCA	0.9846 ± 0.00	0.9625 ± 0.00	0.9492 ± 0.01	0.90.77 ± 0.01	0.8846 ± 0.02	0.8615 ± 0.01
	MVAE	1.0000 ± 0.00	0.9969 ± 0.00	0.9861 ± 0.00	0.9831 ± 0.01	0.9692 ± 0.02	0.9599 ± 0.01
	MIWAE	1.0000 ± 0.00	0.9923 ± 0.00	0.9923 ± 0.01	0.9903 ± 0.01	0.9846 ± 0.01	0.9692 ± 0.03
	CPM-Nets	0.9915 ± 0.02	0.9862 ± 0.01	0.9800 ± 0.01	0.9700 ± 0.02	0.9469 ± 0.01	0.9100 ± 0.02
	DeepIMV	1.0000 ± 0.00	0.9846 ± 0.03	0.9231 ± 0.02	0.9154 ± 0.08	0.8923 ± 0.02	0.8718 ± 0.06
	Ours	1.0000 ± 0.00	1.0000 ± 0.00	0.9981 ± 0.00	0.9962 ± 0.01	0.9847 ± 0.01	0.9769 ± 0.01
ROSMAP	Mean-Imputation	0.7429 ± 0.03	0.6809 ± 0.01	0.6714 ± 0.02	0.6571 ± 0.07	0.6429 ± 0.02	0.6072 ± 0.05
	GCCA	0.6953 ± 0.03	0.6571 ± 0.02	0.6429 ± 0.04	0.6143 ± 0.03	0.5714 ± 0.06	0.5429 ± 0.06
	TCCA	0.7143 ± 0.03	0.7072 ± 0.01	0.6857 ± 0.05	0.6500 ± 0.05	0.6286 ± 0.03	0.6036 ± 0.06
	MVAE	0.7429 ± 0.02	0.7286 ± 0.05	0.7143 ± 0.05	0.6786 ± 0.03	0.6786 ± 0.06	0.6524 ± 0.06
	MIWAE	0.7429 ± 0.03	0.7286 ± 0.02	0.6714 ± 0.05	0.6571 ± 0.03	0.6571 ± 0.05	0.6357 ± 0.08
	CPM-Nets	0.7840 ± 0.05	0.7517 ± 0.04	0.7394 ± 0.06	0.7183 ± 0.04	0.6901 ± 0.08	0.6409 ± 0.08
	DeepIMV	0.7607 ± 0.03	0.7429 ± 0.01	0.7143 ± 0.05	0.6643 ± 0.05	0.6524 ± 0.06	0.6250 ± 0.06
	Ours	0.8714 ± 0.00	0.8429 ± 0.03	0.7714 ± 0.05	0.7464 ± 0.03	0.7214 ± 0.03	0.7143 ± 0.02
Handwritten	Mean-Imputation	0.9800 ± 0.00	0.9750 ± 0.00	0.9700 ± 0.01	0.9700 ± 0.01	0.9500 ± 0.01	0.9100 ± 0.01
	GCCA	0.9500 ± 0.01	0.9350 ± 0.02	0.9100 ± 0.01	0.8875 ± 0.02	0.8425 ± 0.02	0.8200 ± 0.03
	TCCA	0.9725 ± 0.00	0.9650 ± 0.00	0.9575 ± 0.02	0.9350 ± 0.01	0.9200 ± 0.01	0.9100 ± 0.02
	MVAE	0.9800 ± 0.00	0.9750 ± 0.01	0.9700 ± 0.00	0.9650 ± 0.01	0.9575 ± 0.01	0.9500 ± 0.01
	MIWAE	0.9800 ± 0.00	0.9800 ± 0.00	0.9725 ± 0.00	0.9650 ± 0.00	0.9475 ± 0.01	0.9375 ± 0.02
	CPM-Nets	0.9550 ± 0.01	0.9475 ± 0.01	0.9375 ± 0.01	0.9300 ± 0.02	0.9225 ± 0.01	0.9125 ± 0.01
	DeepIMV	0.9908 ± 0.04	0.9883 ± 0.02	0.9850 ± 0.04	0.9750 ± 0.02	0.9625 ± 0.04	0.9450 ± 0.06
	Ours	0.9825 ± 0.00	0.9800 ± 0.00	0.9800 ± 0.00	0.9775 ± 0.00	0.9700 ± 0.01	0.9600 ± 0.01
BRCA	Mean-Imputation	0.7885 ± 0.02	0.7143 ± 0.03	0.7000 ± 0.04	0.6571 ± 0.02	0.6429 ± 0.03	0.6286 ± 0.02
	GCCA	0.7371 ± 0.03	0.7143 ± 0.03	0.6971 ± 0.04	0.6762 ± 0.02	0.6514 ± 0.03	0.6381 ± 0.04
	TCCA	0.7543 ± 0.02	0.7314 ± 0.03	0.7238 ± 0.04	0.7129 ± 0.03	0.6857 ± 0.04	0.6743 ± 0.03
	MVAE	0.7885 ± 0.03	0.7691 ± 0.02	0.7347 ± 0.01	0.6968 ± 0.03	0.6633 ± 0.05	0.6388 ± 0.03
	MIWAE	0.7885 ± 0.02	0.7352 ± 0.03	0.7314 ± 0.03	0.7105 ± 0.02	0.7029 ± 0.02	0.6857 ± 0.04
	CPM-Nets	0.7388 ± 0.02	0.7317 ± 0.04	0.7107 ± 0.08	0.7233 ± 0.04	0.6980 ± 0.05	0.6788 ± 0.03
	DeepIMV	0.7686 ± 0.03	0.7614 ± 0.02	0.7457 ± 0.02	0.7414 ± 0.02	0.7400 ± 0.02	0.6714 ± 0.04
	Ours	0.8286 ± 0.01	0.7943 ± 0.01	0.7771 ± 0.01	0.7657 ± 0.02	0.7543 ± 0.02	0.7429 ± 0.02
Scene15	Mean-Imputation	0.7681 ± 0.02	0.6912 ± 0.01	0.6477 ± 0.01	0.6098 ± 0.01	0.5864 ± 0.02	0.5106 ± 0.02
	GCCA	0.6611 ± 0.02	0.6511 ± 0.01	0.6176 ± 0.01	0.5708 ± 0.01	0.5385 ± 0.02	0.5006 ± 0.02
	TCCA	0.6878 ± 0.02	0.6644 ± 0.01	0.6566 ± 0.01	0.6187 ± 0.01	0.5741 ± 0.01	0.5563 ± 0.02
	MVAE	0.7681 ± 0.00	0.7346 ± 0.01	0.7157 ± 0.01	0.6689 ± 0.01	0.6444 ± 0.01	0.6098 ± 0.01
	MIWAE	0.7681 ± 0.03	0.7179 ± 0.01	0.6990 ± 0.01	0.6566 ± 0.01	0.6265 ± 0.02	0.5875 ± 0.02
	CPM-Nets	0.6990 ± 0.02	0.6566 ± 0.02	0.6388 ± 0.00	0.6265 ± 0.01	0.5903 ± 0.01	0.5708 ± 0.01
	DeepIMV	0.7124 ± 0.00	0.6934 ± 0.02	0.6656 ± 0.01	0.6410 ± 0.00	0.5853 ± 0.02	0.5719 ± 0.01
	Ours	0.7770 ± 0.00	0.7581 ± 0.01	0.7347 ± 0.00	0.6990 ± 0.01	0.6689 ± 0.01	0.6254 ± 0.02

can directly work on incomplete multi-view data including CPM-Nets and DeepIMV, we train the classifiers on the incomplete data without imputing missing views. We adopt the same network architecture for all methods, and the comparison results are shown in Table 1. From the empirical results, it is observed: (1) UIMC achieves competitive performance on complete multi-view data ($\eta = 0$). (2) UIMC outperforms all comparison methods on most datasets when $\eta \neq 0$ especially when the task is difficult. For example,

the performance of the proposed method on ROSMAP from $\eta = 0.1$ to $\eta = 0.5$ is significantly superior to other methods. (3) UIMC achieves the highest classification accuracy with relatively high missing rates which validates that UIMC is very robust to incomplete multi-view data. For example, the accuracy declines only 2.28% on BRCA from $\eta = 0.3$ to $\eta = 0.5$ compared with 7% of other methods.

Q2 Effectiveness (II). We conduct experiments for visualizing the imputed complete multi-view data on Handwritten

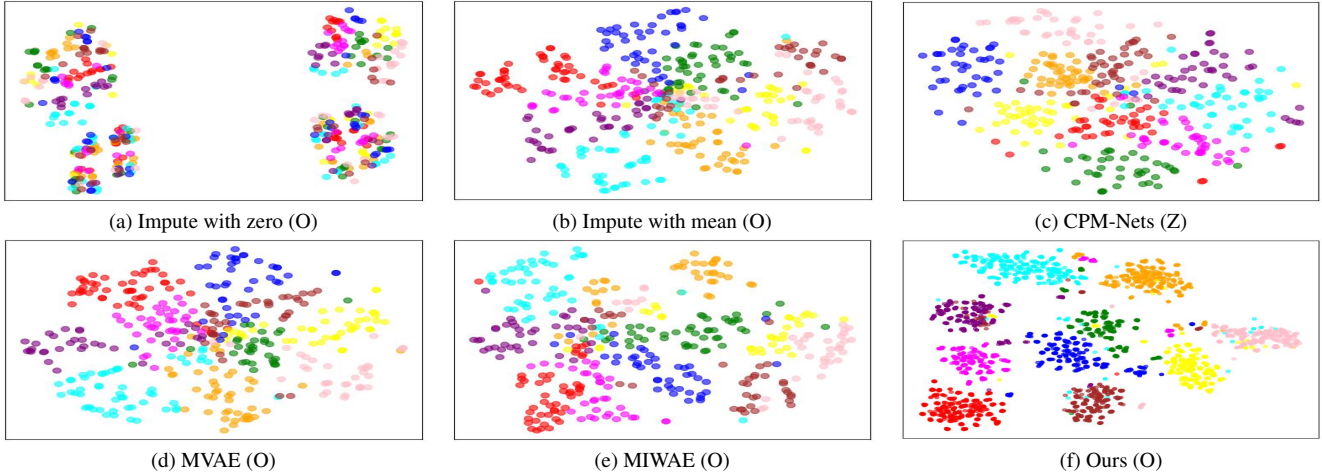


Figure 2. Visualization of imputed complete data (Handwritten) under $\eta = 0.5$, where multiple views are concatenated. “O” and “Z” indicate concatenation with original data (and imputed views) and learned common latent representation from multiple available views, respectively.

with $\eta = 0.5$. Specifically, we take $N_s = 30$ samplings from the distribution according to the available views to obtain the imputed complete data. As shown in Fig. 2, the margins between different classes are much clearer since UIMC explores the correlation among different views. Although our imputation strategy may introduce noise, our model still achieves the best performance.

Q3 Ablation study (I). To demonstrate the necessity of conducting multiple samplings to impute missing views instead of deterministic imputation, we conduct ablation experiment to compare the multiple samplings with naive deterministic imputation. As shown in Fig. 3, the method termed single-imputation is designed as imputing the missing views with the mean of neighbors. The superior performance of UIMC verifies that our multiple-imputation strategy can achieve more excellent performance than single-imputation empirically.

Q4 Ablation study (II). We conduct ablation experiment to investigate the effect of exploiting uncertainty in guiding multi-view fusion and classification. Specifically, we compare ours with a naive version termed Naive-IMVC which ignores the uncertainty introduced with imputation. Concretely, the key difference between Naive-IMVC and UIMC is that Naive-IMVC trains classifiers on each single-view and the integrated multi-view data optimized with cross-entropy loss $\mathcal{L}_{ce} = -\sum_{k=1}^K y_k \log(p_k)$, where p_k is the predicted probability for class k , while UIMC adopts an uncertainty-guided multi-view fusion and classification strategy. As shown in Fig. 3, we compare UIMC with Naive-IMVC on Handwritten, YaleB, and Scene15 with $\eta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$. The classification accuracy of UIMC is superior to Naive-IMVC, which indicates that our method can exploit the uncertainty of imputations and make predictions more efficiently and credibly.

Q5 Stability. We impute missing views by taking multiple samplings from the estimated multivariate Gaussian distributions. Since the randomness is involved in sampling process, each sampling operation yields different complete data and then the corresponding classification predictions might be different. To obtain stable and robust prediction, our method adopts a voting strategy to obtain the final classification, which can mitigate the negative impact of the uncertainty from sampling operation. To investigate the prediction variance from the sampling process, we fix the missing patterns of incomplete multi-view data (i.e., the missing views of each sample is fixed), and then conduct $N_i = 10$ samplings to obtain 10 sets of different complete multi-view data. Then we test these 10 imputed multi-view data sets on a pre-trained classification model to obtain 10 prediction results. We conduct this experiment on ROSMAP, BRCA, Handwritten, and Scene15 whose number of classes are 2, 5, 10 and 15, respectively. As shown in Fig. 4, the relatively small proportion of “Inconsistent” indicates that the prediction of our model is quite stable, where “Consistent” and “Inconsistent” indicate the conditions of 10 tests with the same and different predictions, respectively, .

5. Conclusion

In this work, we propose a novel framework for incomplete multi-view classification termed Uncertainty-induced Incomplete Multi-View Data Classification (UIMC), which can elegantly explore and exploit the uncertainty arising from imputation, producing effectiveness and trustworthiness. Specifically, UIMC focuses on imputing missing views in consideration of the imputation quality. Since UIMC conducts sampling multiple times from the estimated distribution of missing view, it inevitably introduces uncertainty.

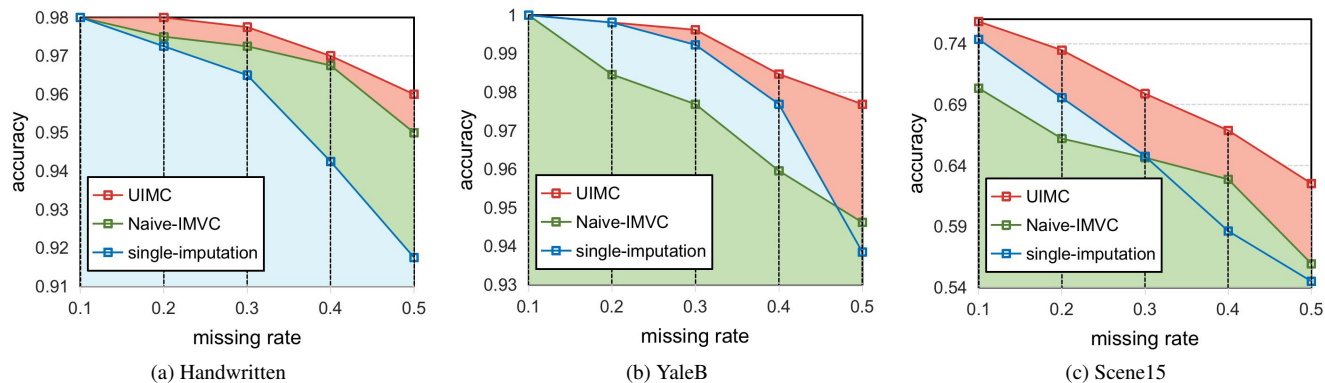


Figure 3. Classification performance of UIMC, single-imputation, and Naive-IMVC on three datasets with $\eta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$.

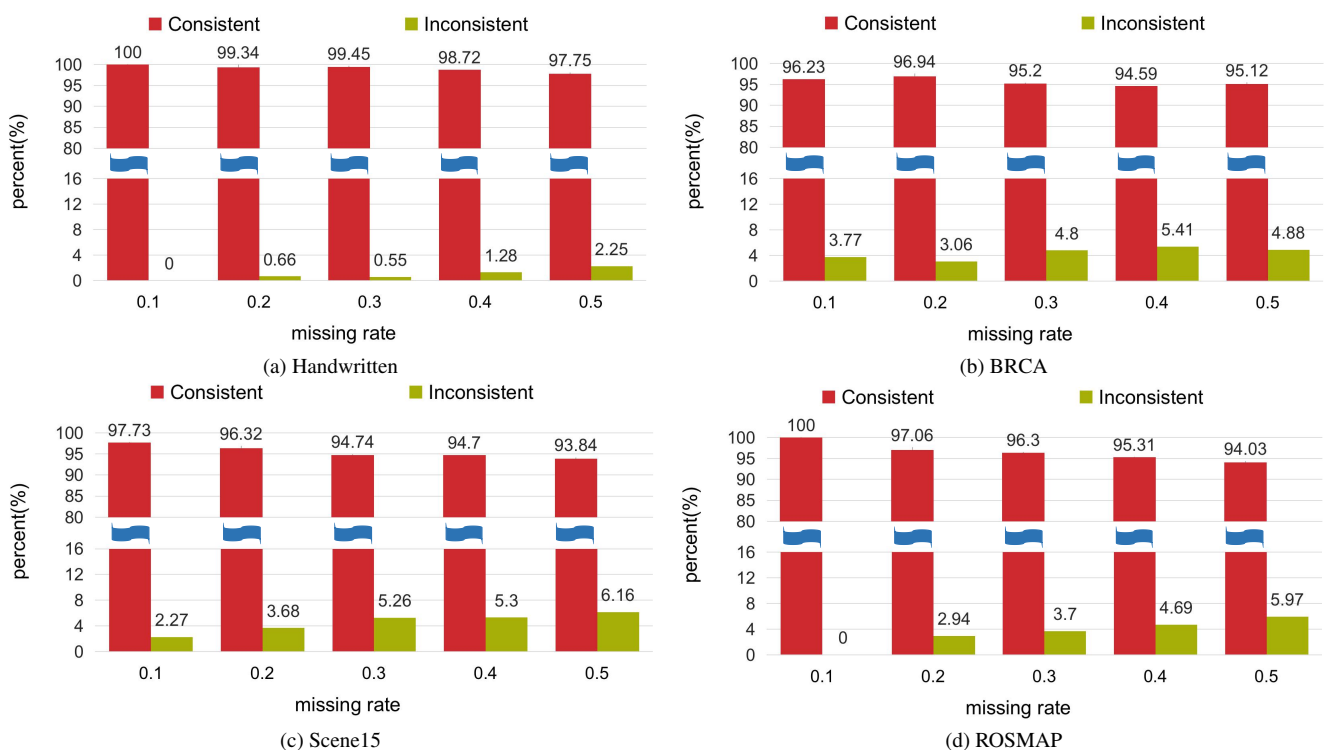


Figure 4. Evaluating the stability of UIMC on four datasets with $\eta = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$.

We employ evidential classifier to characterize the view-specific uncertainty and further utilize Dempster’s combination rule to fuse the uncertain opinions of multiple (imputed) views. We conduct extensive experiments and the empirical results solidly validate the effectiveness and stability of the proposed UIMC. The limitation of this work is that the imputation distributions have been determined before multi-view fusion and classification. Thus, our future work will focus on jointly learning the parameterized distributions of the missing views and the multi-view classifier, and in this way the two components could promote mutually. Another

interesting line is theoretical analysis to show the necessity of introducing uncertainty for missing views.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (61976151). The authors gratefully acknowledge the support of MindSpore, CANN (Compute Architecture for Neural Networks), and Ascend AI Processor used for this research. We also want to use UIMC on MindSpore, which is a new deep learning computing framework. These problems are left for future work.

References

- [1] Zhiqiang Tao, Hongfu Liu, Sheng Li, Zhengming Ding, and Yun Fu. From ensemble clustering to multi-view clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2017. 1
- [2] Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *Proceedings of the AAAI conference on artificial intelligence*, 2017. 1
- [3] Hong Tao, Chenping Hou, Xinwang Liu, Tongliang Liu, Dongyun Yi, and Jubo Zhu. Reliable multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. 1
- [4] Chunjie Zhang, Jian Cheng, and Qi Tian. Multi-view image classification with visual, semantic and view consistency. *IEEE Transactions on Image Processing*, 29:617–627, 2019. 1
- [5] Jinglin Xu, Wenbin Li, Xinwang Liu, Dingwen Zhang, Ji Liu, and Junwei Han. Deep embedded complementary and interactive information for multi-view classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 6494–6501, 2020. 1
- [6] Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering*, 31(10):1863–1883, 2018. 1
- [7] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. Flexible multi-view representation learning for subspace clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2916–2922, 2019. 1
- [8] Changhee Lee and Mihaela van der Schaar. A variational information bottleneck approach to multi-omics data integration. In *The 24th International Conference on Artificial Intelligence and Statistics*, 2021. 1, 2, 3, 5
- [9] Changqing Zhang, Zongbo Han, Yajie Cui, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. Cpm-nets: Cross partial multi-view networks. In *Advances in Neural Information Processing Systems*, pages 557–567, 2019. 1, 2, 3, 5
- [10] Mike Wu and Noah D. Goodman. Multimodal generative models for scalable weakly-supervised learning. In *Advances in Neural Information Processing Systems*, pages 5580–5590, 2018. 1, 2, 3, 5
- [11] Pierre-Alexandre Mattei and Jes Frellsen. Miwae: Deep generative modelling and imputation of incomplete data sets. In *International conference on machine learning*, pages 4413–4423, 2019. 1, 2, 3, 5
- [12] Kim-Han Thung, Pew-Thian Yap, and Dinggang Shen. Joint robust imputation and classification for early dementia detection using incomplete multi-modality data. In *International Workshop on Predictive Intelligence In Medicine*, pages 51–59, 2018. 1, 3
- [13] Lei Cai, Zhengyang Wang, Hongyang Gao, Dinggang Shen, and Shuiwang Ji. Deep adversarial learning for multi-modality missing data completion. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1158–1166, 2018. 1, 3
- [14] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on Machine Learning*, 2011. 1
- [15] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. 1
- [16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2014. 1
- [17] Niels Ipsen, Pierre-Alexandre Mattei, and Jes Frellsen. How to deal with missing data in supervised deep learning? In *The Tenth International Conference on Learning Representations*, 2020. 1
- [18] Alma B Pedersen, Ellen M Mikkelsen, Deirdre Cronin-Fenton, Nickolaj R Kristensen, Tra My Pham, Lars Pedersen, and Irene Petersen. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, 9:157, 2017. 1, 3
- [19] Linlin Zong, Xianchao Zhang, Xinyue Liu, and Hong Yu. Multi-view clustering on data with partial instances and clusters. *Neural Networks*, 129:19–30, 2020. 2
- [20] Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014. 2
- [21] Handong Zhao, Hongfu Liu, and Yun Fu. Incomplete multi-modal visual data grouping. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2392–2398, 2016. 2
- [22] HyeonJoo Hwang, Geon-Hyeong Kim, Seunghoon Hong, and Kee-Eung Kim. In *Advances in Neural Information Processing Systems*, pages 12194–12207. 2
- [23] Piyush Rai, Anusua Trivedi, Hal Daumé III, and Scott L DuVall. Multiview clustering with incomplete views. In *Proceedings of the NIPS Workshop on Machine Learning for Social Computing*, 2010. 2
- [24] Hang Gao, Yuxing Peng, and Songlei Jian. Incomplete multi-view clustering. In *International Conference on Intelligent Information Processing*, pages 245–255, 2016. 2
- [25] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11174–11183, 2021. 2
- [26] Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE Transactions on pattern analysis and machine intelligence*, 44(5):2402–2415, 2022. 2
- [27] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial incomplete multi-view clustering. pages 3933–3939. *Proceedings of the International Joint Conference on Artificial Intelligence*, 2019. 2

- [28] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Partial multi-view clustering via consistent gan. In *IEEE International Conference on Data Mining*, pages 1290–1295, 2018. 2
- [29] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. volume 30, 2017. 2
- [30] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. volume 31, 2018. 2
- [31] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *International conference on machine learning*, pages 9690–9700, 2020. 2
- [32] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? volume 30, 2017. 2
- [33] Durga Rao Karanki, S Rahman, Vinh N Dang, and O Zerkak. Epistemic and aleatory uncertainties in integrated deterministic and probabilistic safety assessment: Tradeoff between accuracy and accident simulations. *Reliability Engineering & System Safety*, 162:91–102, 2017. 2
- [34] Jon C Helton, Jay D Johnson, William L Oberkampf, and Cedric J Sallaberry. Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems*, 39(6):605–646, 2010. 2
- [35] Rafael Peres da Silva, Chayaporn Suphavilai, and Niranjan Nagarajan. Task uncertainty loss reduce negative transfer in asymmetric multi-task feature learning (student abstract). In *Proceedings of the AAI Conference on Artificial Intelligence*, pages 15867–15868, 2021. 2
- [36] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Conference on Computer Vision and Pattern Recognition*, pages 7482–7491, 2018. 2
- [37] Bindya Venkatesh and Jayaraman J Thiagarajan. Heteroscedastic calibration of uncertainty estimators in deep learning. *The Computing Research Repository*, 2019. 2
- [38] Murat Seckin Ayhan and Philipp Berens. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In *International conference on Medical Imaging with Deep Learning*, 2018. 2
- [39] Jakob Gawlikowski, Cedrique Rovile Njiteucheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *The Computing Research Repository*, 2021. 2
- [40] Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001. 2
- [41] Arthur P Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968. 2
- [42] Marine Le Morvan, Julie Josse, Erwan Scornet, and Gael Varoquaux. What’s a good imputation to predict with missing values? In *Advances in Neural Information Processing Systems*, volume 34, pages 11530–11540, 2021. 3
- [43] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967. 3
- [44] Murat Sensoy, Lance M. Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31:3183–3193, 2018. 4
- [45] Audun Jøsang. *Subjective Logic - A Formalism for Reasoning Under Uncertainty*. Artificial Intelligence: Foundations, Theory, and Algorithms. 2016. 4
- [46] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 4
- [47] Arthur P Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968. 4
- [48] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. 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, 2001. 5
- [49] Tongxin Wang, Wei Shao, Zhi Huang, Haixu Tang, Jie Zhang, Zhengming Ding, and Kun Huang. Moronet: multi-omics integration via graph convolutional networks for biomedical data classification. *bioRxiv*, 2020. 5
- [50] Simon Perkins and James Theiler. Online feature selection using grafting. In *Proceedings of the 20th International Conference on Machine Learning*, pages 592–599, 2003. 5
- [51] Pietro Perona FeiFei Li. A bayesian hierarchical model for learning natural scene categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 5
- [52] Jon R Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971. 5
- [53] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1935. 5
- [54] Yong Luo, Dacheng Tao, Kotagiri Ramamohanarao, Chao Xu, and Yonggang Wen. Tensor canonical correlation analysis for multi-view dimension reduction. *IEEE transactions on Knowledge and Data Engineering*, 27(11):3111–3124, 2015. 5
- [55] Junhyun Nam, Hyuntak Cha, Sungsoo Ahn, Jaeho Lee, and Jinwoo Shin. In *Advances in Neural Information Processing Systems*, pages 20673–20684. 5