Adversarial and Implicit Modality Imputation with Applications to Depression Early Detection

Yuzhou Nie[1⋆], Chengyue Huang[1⋆], Hailun Liang[2], and Hongteng Xu[3†]

[1] School of Statistics, Renmin University of China
[2] School of Public Administration and Policy, Renmin University of China
[3] Gaoling School of Artificial Intelligence, Renmin University of China

Abstract. Depression early detection is a significant healthcare task that heavily relies on high-quality multi-modal medical data. In practice, however, learning a robust detection model is challenging because real-world data often suffers from serious modality-level missing issues caused by imperfect data collection and strict data sharing policies. In this study, we propose an Adversarial and Implicit Modality Imputation (AIMI) method to resolve this challenge. In particular, when training multi-modal predictive models, we learn an implicit mechanism to impute the missing modalities of training data at the same time. These two learning objectives are achieved jointly in an adversarial learning framework. Based on the UK Biobank dataset, we demonstrate the effectiveness and superiority of our method on the early detection of depression. Codes are available at `https://github.com/rucnyz/AIMI`.

Keywords: Multi-modal learning · Adversarial learning· Implicit data imputation· Depression early detection.

1 Introduction

Depression is a leading cause of disability and is a major contributor to the overall global burden of disease. According to WHO, approximately 5.0% of the adult population suffer from depression [23]. Due to the high burden of brain health disorders, depression early detection and screening are valuable, which could slow the progression of the disease through earlier access to treatment, advice and support [17,11].

Many indicators of presence or severity of depression are observable. While one single modality of indicators rarely provides complete information, the diversity of modalities brings added value. Recently, artificial intelligence-assisted techniques have been developed for the early detection of depression [9,14,15,5]. Essentially, these methods leverage a multi-modal paradigm to represent and fuse the healthcare information collected by different ways (e.g., physical exams, standard questionnaire, lab tests, the medical images like CT and MRI, genetic

⋆ Equal contribution.
† Corresponding author: H. Xu <hongtengxu313@gmail.com>.

tests, etc.) and predict the risk of depression based on the fused information. However, these methods often ignore the fact that real-world data often miss some modalities because of various reasons. For example, many patients merely do a part of lab tests or medical images because of the lack of medical insurance coverages. Even for the patients having complete modalities, their multi-modal data may be stored in different hospitals and cannot be fully-accessible because of the privacy and security issues. The above phenomena are common in the early stage of depression — the patients may just take some tests and thus only some modalities are available for disease prediction. Facing to the above modality-missing issue, most existing multi-modal learning methods often simply discard the data with missing modalities, which results in the loss of information and sub-optimal models [4,28].

Related Work. Several methods are proposed to resolve the modality-missing issue. One direct way is to impute the missing modalities, and then adopt the on-shelf multi-modal learning algorithms. The early work in [7] simply fills the missing views with a single value (e.g. mean value based on available samples within the same class). The matrix completion methods, such as SVD imputation [20] and Bayesian principal component analysis [16], infer the missing values based on the low-rank assumption of the latent data structure. However, the missing values would simultaneously appear in one modality instead of randomly distributed, which makes the above conventional methods no longer applicable. Focusing on the imputation of missing modalities, the CRA (Cascaded Residual Autoencoder) in [19] predicts the missing modalities from the observed one. Following the same strategy, more sophisticated modality prediction mechanisms are considered, e.g. the grouping strategy in [26] and the merging strategy in [3]. More recently, deep learning-based methods [25,24] impute missing data by adapting the well-known Generative Adversarial Network (GAN) [10], treating observed data as real data and generating missing data as fake ones. Synthetic EHR methods like medGAN [6] and its variants [2] combine GAN with an autoencoder(AE) [12] to learn the data distribution of real EHRs. However, without any feedback mechanisms, imputed modalities achieved by the above methods are unchanged and may lead to catastrophic error propagation.

To solve the above issues, we propose a novel **A**dversarial and **I**mplicit **M**odality **I**mutation method, called AIMI for short, which exploits multi-modal data with arbitrary modality-missing patterns both effectively and flexibly. As illustrated in Fig. 1, our AIMI consists of an AE [12] and a GAN [10]. For the autoencoder, its encoder projects samples with arbitrary modality-missing patterns into a common latent space and connects to a classifier, and the decoder reconstructs complete multi-view data from the latent codes. The decoder works as the generator in the framework of GAN, which imputes missing modalities via generated ones. Accordingly, besides training the encoder and the decoder to minimizes the reconstruction error of the observed modalities, we further consider two more objectives: 1) improving the classification accuracy of the associated classifier; and 2) cheating a discriminator that checks whether a modality is observed or generated.

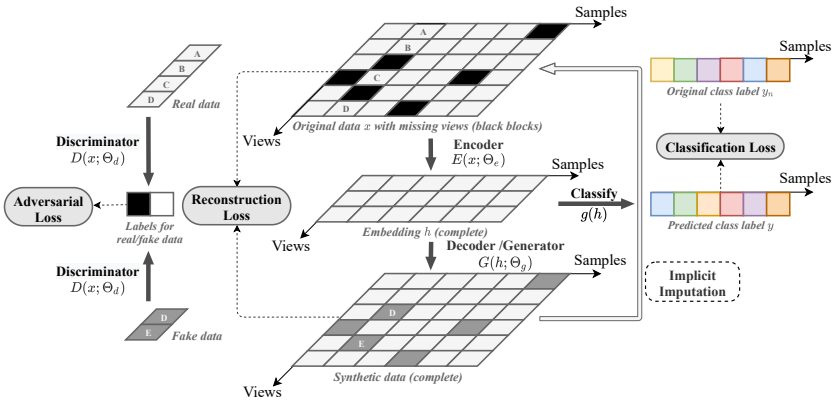

Fig. 1: An illustration of our AIMI method.

Differing from most existing data imputation methods, our method contains an implicit feedback mechanism, adjusting imputed modalities during training. In particular, after imputing the missing modalities, AIMI takes them as the input of the encoder and thus train the model under the guidance of the imputed data. This feedback mechanism achieves a new kind of data augmentation. As a result, the encoder is trained to be robust to the uncertainty of the imputation, which tends to preserve the classification accuracy of the imputed data, while the decoder also takes the imputed data into account when reconstructing modalities and cheating the discriminator.

We demonstrate the effectiveness of AIMI in the depression early detection task. based on the UK Biobank (UKB) database [18], which contains approximately 480,000 EHRs with 8 modalities, we train a multi-modal predictive model for depression early detection. Experimental results show that our AIMI is robust to the modality-missing issue and mitigates the associated problems like error propagation and over-fitting. In particular, in the experiments with missing modalities, our AIMI and its variants outperform state-of-the-art methods in both supervised and unsupervised settings.

2 Proposed Method

2.1 Learning multi-modal representations via auto-encoding

Facing the data with missing modalities, we design a model with auto-encoding architecture. For the multi-modal data with V modalities, the representation model consists of V encoder-decoder pairs. For the v-th modality, its encoder is denoted as $E_v(\cdot)$, which is often parameterized by a neural network. Given the data of the v-th modality $\mathbf{x}^{(v)}$, the encoder projects it to a latent code in the d-dimensional latent space, i.e., $\mathbf{h}^{(v)} = E_v(\mathbf{x}^{(v)}) \in \mathbb{R}^d$.

We obtain a fused latent code via pooling all the modalities' latent codes:

$$\mathbf{h} = \text{Pooling}(\{\mathbf{h}^{(v)}\}_{v=1}^{V}). \tag{1}$$

Here, various pooling operations are applicable, e.g., the commonly-used mean-pooling or the attention-pooling in [8], e.g.,

$$\mathbf{h} = \sum\nolimits_{v=1}^{V} \sum\nolimits_{i=1}^{V} \frac{\exp(\mathbf{q}_v^T \mathbf{k}_i)}{\sqrt{d} \sum_{j=1}^{V} \exp(\mathbf{q}_v^T \mathbf{k}_j)} \mathbf{v}_i, \tag{2}$$
$$\text{where } \mathbf{q}_v = f_Q(\mathbf{h}^{(v)}), \mathbf{k}_v = f_K(\mathbf{h}^{(v)}), \mathbf{v}_v = f_V(\mathbf{h}^{(v)}), \quad \forall v = 1, ..., V.$$

Note that, for the data with the missed modalities, we derive the latent codes of the observed modalities and fuse them by the pooling operation in the beginning of the training process. During the training process, the missed modalities are imputed by our AIMI method and the pooling operation is applied to both observed and imputed modalities, which will be introduced in details in the following content. Accordingly, the decoder (generator) of each modality, denoted as $G_v(\cdot)$ for $v = 1, ..., V$, reconstructs the data of the v-th view from the complete latent code $\mathbf{h}$. For the missed modalities, the decoder estimates them from the complete latent code, which achieves the imputation of incomplete data.

Similar to classic autoencoders [12], given N samples of V modalities, i.e., $\{\mathbf{x}_n^{(v)}\}_{n,v=1}^{N,V}$, we can learn the model via minimizing the reconstruction error of each modality given a metric (e.g., the Euclidean distance used below):

$$\mathcal{L}_r = \sum\nolimits_{n=1}^{N} \sum\nolimits_{v=1}^{V} \underbrace{m_n^{(v)} \|G_v(\mathbf{h}_n) - \mathbf{x}_n^{(v)}\|^2}_{\ell_r^{(v)}(S_n; \{G_v, E_v\}_{v=1}^{V})}, \tag{3}$$

where $m_n^{(v)} \in \{0, 1\}$ is an indicator of the availability for the v-th modality of the n-th sample. Therefore, the learning problem above just considers the reconstruction of the observed modalities. It should be noted here that in the CPM-Net [27], the randomly initialized latent embedding $\mathbf{h}_n$ is directly used to represent the incomplete data (available modalities). For us, at the beginning of the training process, only available modalities are involved in feed-forward networks by considering $m_n^{(v)}$. After the first imputation, all modalities (including the synthetic parts) are trained to obtain $\mathbf{h}_n$ no matter what $m_n^{(v)}$ is, though $\mathcal{L}_r$ is still calculated by available modalities.

Besides the above reconstruction loss, we further consider the classification loss used in the CPM-Net [27]. Specifically, denote the label of the n-th multi-modal sample as y_n. The classification loss is defined as

$$\mathcal{L}_c = \sum\nolimits_{n=1}^{N} \underbrace{\Delta(y_n, \arg\max\nolimits_{y \in \mathcal{Y}} g(\mathbf{h}_n, y)) + \max\nolimits_{y \in \mathcal{Y}} g(\mathbf{h}_n, y) - g(\mathbf{h}_n, y_n)}_{\ell_c(y_n; \{E_v\}_{v=1}^{V})}, \tag{4}$$

where $g(\mathbf{h}_n, y) = \mathbb{E}_{\mathbf{h} \sim \mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_n)$, $\mathcal{T}(y)$ is the set of latent codes from class y, $F(\mathbf{h}, \mathbf{h}_n) = \mathbf{h}^T \mathbf{h}_n$, and $\Delta(y, y_n) = 0$ if $y = y_n$ and 1 otherwise. This loss leads to a classification scheme enhancing the clustering structure of the latent codes.

Taking equations (1)$\sim$(4) into account, we obtain the proposed multi-modal predictive model. Specifically, by jointly considering the complementary information of multi-modalities and the distribution of classes, our model makes

them mutually improve each other to obtain the representation reflecting the underlying patterns, thus promoting the prediction performance.

2.2 Adversarial and implicit modality imputation (AIMI)

Adversarial learning Besides the above two learning objectives, we further consider a generative adversarial loss to improve the quality of the imputed modalities. In particular, for each multi-modal sample with missing modalities, we impute the unavailable modalities by the outputs of the corresponding decoders. Given the observed real modalities and the estimated modalities, denoted as $\{\mathbf{x}_n^{(v)}\}_{v:m_n^{(v)}=1}$ and $\{G_v(\mathbf{h}_n)\}_{v:m_n^{(v)}=0}$, we would like to train a discriminator to check whether a given modality is real or fake, whose objective is

$$\mathcal{L}_a = \sum\nolimits_{n=1}^{N} \sum\nolimits_{v=1}^{V} \underbrace{m_n^{(v)} \log D_v(\mathbf{x}_n^{(v)}) + (1 - m_n^{(v)}) \log(1 - D_v(G_v(\mathbf{h}_n)))}_{\ell_a^{(v)}(S_n; \{G_v, D_v\}_{v=1}^V)}, \quad (5)$$

where $D_v(\cdot)$ is the discriminator for the v-th modality.

At the same time, we train the decoder to improve the quality of the imputed modalities, which aims at cheating the discriminator. Such a generative adversarial learning strategy leads to a "min-max" game. As a result, the optimization problem of our learning task becomes

$$\mathcal{L} = \min_{\{G_v, E_v\}_{v=1}^V} \max_{\{D_v\}_{v=1}^V} \mathcal{L}_r + \lambda \mathcal{L}_c + \mu \mathcal{L}_a \quad (6)$$

where $\lambda, \mu > 0$ control the significance of different objectives. When attention-pooling is applied, the parameters of the attention layer, i.e., $\{f_Q, f_K, f_V\}$ are considered as a part of the encoder module.

Implicit imputation with a feedback loop For each missed modality $\mathbf{x}_n^{(v)}$, we fill it with the output of the corresponding decoder, i.e., $\mathbf{x}_n^{(v)} = G_v(\mathbf{h}_n; \mathbf{\Theta}_g^{(v)})$. Here, given a batch of data, both the latent code and the decoder are updated during the training progress, while their imputed modalities are fixed for the batch, which leads to sub-optimal updating of the model. Therefore, we consider an implicit imputation method, which applies several inner iterations to adjust the imputed modalities — for the t-th inner iteration, we will take the imputed modalities obtained in the previous iteration as the input of the encoders and update them by the auto-encoding architecture. Accordingly, we obtain an iterative adjustment of the imputed modalities: for the missed modality $x_n^{(v)}$, we set $\mathbf{x}_n^{(v),0} = \mathbf{0}$ and update it via

$$\mathbf{h}_n^{t-1} = \text{Pooling}(\{E_v(\mathbf{x}_n^{(v),t-1})\}_{v=1}^V), \quad \mathbf{x}_n^{(v),t} = G_v(\mathbf{h}_n^{t-1}) \quad \text{for } t = 1, 2, ... \quad (7)$$

Such an iterative adjustment leverages the feedback of current model, which improve the quality of the imputations.

The rationality of this implicit imputation mechanism can be verified in both feed-forward computation and backpropagation. Firstly, this mechanism leads to

Algorithm 1: Algorithm for AIMI

Input: Multi-modal data $\{S_n, y_n\}_{n=1}^N$. Hyperparameters s and T
Initialize $\{E_v, G_v, D_v\}_{v=1}^V$ randomly.
while *not converged* **do**
  Given a batch of samples $\{S_n, y_n\}_{n\in\mathcal{B}}$.
  For $v = 1 : V$, update the discriminator D_v with gradient descent.
  For $v = 1 : V$, update the decoder G_v with gradient descent.
  For $v = 1 : V$, update the encoder E_v with gradient descent.
  If #epochs $> s$, impute missing modalities via (7) with T iterations.
end
Output: $\{E_v, G_v, D_v\}_{v=1}^V$.

iterative auto-encoding, which can be treated as an iterative function system. As a result, with the increase of the inner iterations, the imputed modalities may converge to a fixed point, which suppress the uncertainty of the imputed modalities. Secondly, when updating our model in each outer iteration, we unroll the inner iterations to compute the gradients. According to the chain rule, the gradients accumulates the gradients corresponding to the intermediate imputations, which are more robust than merely considering the gradients given one-step imputations. From this viewpoint, our implicit imputation mechanism can be treated as an augmentation method for the gradients. Taking the above implicit imputation mechanism into account, we illustrate the scheme of the proposed AIMI method in Fig. 1 and show its implementation details in Algorithm 1.

Compared with existing partial multi-modal learning methods, e.g., the Cross Partial Multi-View Network (CPM-Net) [27], our AIMI method owns several improvements and advantages. In particular, AIMI leverages an encoder to derive latent codes of different modalities and fuse them together, whose training can be achieved by stochastic gradient descent (SGD). We do not enforce the latent codes of different modalities to have the same latent distribution, which enhances the flexibility of model. Moreover, AIMI leverages an implicit imputation mechanism to achieve robust modality adjustment with a feedback loop.

3 Experiments

3.1 Experimental setup

We demonstrate the effectiveness of our AIMI method and apply it to depression early detection. The data in this study is derived from the UK Biobank (UKB) database[1]. We leverage the database consisting of the diagnoses and the multi-modal clinical records of depression, and aim to train a binary classifier

[1] The UK Biobank is an open access resource. Data are available on application to the UK Biobank (`www.ukbiobank.ac.uk/`). This research has been conducted using the UK Biobank resource under application number 44430.

Table 1: Eight modalities associated with depression diagnosis.

Modalities (# of features)	**Demography** (4)	**Sociology** (6)	**Lifestyle** (7)	**Blood** (19)
Features	Age	Low income	Healthy	Red blood cell count
	Gender	Work status	Healthy diet	White blood cell count
	Screening	...	...	...
	Family history	Housing tenure	Healthy score	Lymphocytes percentage
Modalities (# of features)	**Metabolism** (7)	**Urine** (4)	**Gene** (38)	**Others** (2)
Features	Glucose	Creatinine	rs159963_ac	Non-cancerous diseases
	Total cholesterol	Microalbumin	rs1432639_ac	Medication
	...	Potassium	...	
	Apolipoprotein A	Sodium	rs5758265_ag	

to achieve the early detection of depression. In particular, we manually grouped the clinical records into 8 modalities, namely demography, sociology, lifestyle, blood, metabolism, urine, gene and others. The diagnoses are used as binary labels indicating whether the corresponding subjects are depressed or healthy. Table 1 shows the number of features in each modality.

We processed the UKB database ($\text{UKB}_{Original}$) into three parts: UKB_{All}, $\text{UKB}_{Balanced}$ and $\text{UKB}_{Complete}$. UKB_{All} is a real-world modality missing dataset (missing rate=76%), and we use it for our final evaluation. Note that we randomly select the negative samples from $\text{UKB}_{Original}$ to make the number of them equal to positive ones. We obtain $\text{UKB}_{Complete}$ dataset by removing all the "NA" values from $\text{UKB}_{Original}$ for evaluation, and $\text{UKB}_{Balanced}$ has equivalent positive and negative samples selected from $\text{UKB}_{Complete}$. The number of samples in $\text{UKB}_{Original}$, UKB_{All}, $\text{UKB}_{Complete}$ and $\text{UKB}_{Balanced}$ are 461,289, 27,616, 34,240 and 2,802 respectively. For each dataset, we choose 80% samples for training and the remaining 20% samples for testing and evaluation. Additionally, to imitate the real-world scenarios that have missed modalities, we simulate the missed modalities of each samples by setting a missing rate α and removing some modalities randomly based on the rate. The missing rate (α) is defined as $\alpha = \frac{\sum_v M_v}{V \times N}$, where M_v indicates the number of samples without the v^{th} view. $\alpha \in \{0, 0.1, 0.3, 0.5\}$. When $\alpha = 0$, the dataset is complete without missed modalities.

For each dataset, we train a multi-modal predictive model based on our AIMI method, in which we set 400 epochs with the batch size of 800. The key hyperparameters of our method include *i*) the latent space dimension (d), *ii*) the starting epoch (s) for imputation, and *iii*) the number of imputation times (T). Here, we apply grid search to find the best configuration for our AIMI methods, where $d \in \{64, 128, 256\}$, $s \in \{-1, 10, 20, 30, 40, 50, 60, 100\}$, and $T \in \{0, 1, 2, 3, 4, 5\}$. s=-1 means we do not impute during the entire training. Additionally, we implement our AIMI method using mean-pooling and attention-pooling, denoted as $\textbf{AIMI}_{mean}$ and $\textbf{AIMI}_{att}$, respectively.

Table 2: Comparisons for various methods under different missing rates (α).

Dataset	$\text{UKB}_{Complete}$				$\text{UKB}_{Balanced}$				UKB_{All}
α ($\times 100\%$)	0	10	30	50	0	10	30	50	76
FeatCon	87.95	81.44	77.85	74.69	62.21	61.67	60.07	58.28	58.85
Fusion_{mean}	92.83	88.54	87.80	87.15	61.67	62.21	61.32	57.04	59.86
Fusion_{att}	90.30	89.77	88.78	82.64	62.21	61.85	60.96	55.79	59.63
DCCA	92.22	89.94	88.73	79.03	66.67	65.78	62.57	60.07	60.05
LMNN	75.70	73.08	73.16	69.75	62.92	64.17	62.38	59.89	58.12
CPM-Net	**95.83**	93.03	89.12	84.24	58.47	60.07	57.93	51.87	61.27
AIMI_{mean}	**95.83**	95.63	93.79	93.49	66.13	65.24	62.92	61.14	61.69
AIMI_{att}	95.49	**95.83**	**95.60**	**94.78**	**68.27**	**66.84**	**63.96**	**61.50**	**61.73**

3.2 Comparisons on robust multi-modal depression diagnosis

Superiority to baselines. First, the superiority of our AIMI methods are sufficiently investigated. We compare it with the following typical multi-modal representation learning methods using $\text{UKB}_{Complete}$ and $\text{UKB}_{Balanced}$ datasets under various missing rates α: (1) **FeatCon** simply concatenates multiple types of features from different modalities. (2) **DCCA** [1] (Deep Canonical Correlation Analysis) learns low-dimensional features with neural networks and concatenates them. (3) **LMNN** [22] (Large Margin Nearest Neighbors) searches a Mahalanobis distance metric to optimize the k-nearest neighbours classifier. (4) **Average-based Fusion** simply average all the modalities. (5) **Attention-based Fusion** fuse all the modalities by the attention mechanism. For the baselines, the missed modalities are filled statically with average values according to available samples within the same modality.

We report the AUC in Table 2 to verify the usefulness of our AIMI method. Specifically, we can find that

- No matter whether the dataset has missed modalities or not, our method AIMI and its variant achieve competitive performances on $\text{UKB}_{Complete}$ and outperform other baselines on $\text{UKB}_{Balanced}$.
- On $\text{UKB}_{Complete}$ dataset, our methods are relatively robust to modality-missing dataset. The degradation of the baselines' performance is much larger than ours, and the performance gaps between our methods and the baselines are widened with the increase of missing rate.
- We evaluate our algorithm on UKB_{All} dataset, and the result demonstrates our model's ability to better represent the partial modality data and stability when facing severe missing problem.

Robustness to hyperparameters. We further check the robustness of our method to its hyperparameters (i.e., s and T) on the $\text{UKB}_{Balanced}$ dataset. In Fig. 2, we show the prediction accuracy achieved by our methods under different settings. We can find that

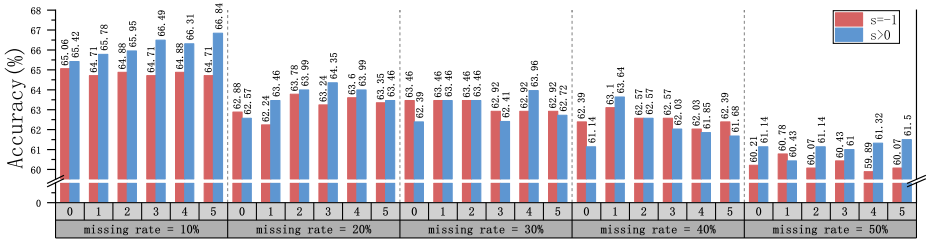

Fig. 2: Robustness analysis on the $\text{UKB}_{Balanced}$ dataset.

- With the increase of the missing rate, the performance of our method degrades monotonously, which reflects the negative influence of the missed modalities on the training of the multi-modal predictive model.
- Imputing missed modalities in the early time (i.e., $s = 1$) may bring in unstable noises because the model is not well-trained at the early stage. Therefore, we prefer imputing missed modalities after several epochs (i.e., $s > 0$) in most situations, in which the model trained in the previous epochs and with missed modalities achieves a warm-up of our AIMI method.
- Our implicit imputation method works. As Fig. 2 shows, under different missing rates α, the best performance is achieved by applying our imputation method more than once. In most cases, the prediction accuracy of our model ascends as the imputation times (T) increases.

3.3 Advantages on modality imputation and representation

Apart from supervised learning scenarios, we also design experiments to validate our advantages on modality imputation and representation in unsupervised setting. For the missed modalities, besides our AIMI, we further consider the following imputation methods as baselines: (1) **Denoising Autoencoder (DAE)** [21] corrupts the data on purpose by randomly removing some of the input values. (2) **LRMC** [13] is a low-rank matrix completion method by iterative soft thresholding of singular value decomposition.

We conduct unsupervised experiments on the $\text{UKB}_{Balanced}$ dataset with different missing rates α to investigate the performance of our proposed method, especially the quality of the latent codes we learned. In particular, for the latent codes of different modalities, we evaluate their rationality and semantics by the following two measurements.

- **Imputation fidelity:** Given the latent codes, we check whether it is possible to reconstruct the corresponding missed data from the decoder. The imputation performance is evaluated by estimating the missing part of the origin data and calculating the RMSE (Root Mean Square Error [13]).
- **Clustering accuracy:** Given the latent codes, we check whether the multi-modal latent codes can reflect the clustering structure of the samples. The

Table 3: Comparisons on clustering accuracy and RMSE of imputed data

	Clustering Accuracy (ACC)					RMSE
α ($\times 100\%$)	10	20	30	40	50	50
DAE	57.70	57.11	56.32	54.58	54.26	0.5944
LRMC	56.68	58.29	57.58	55.97	55.79	0.3961
AIMI$_{mean}$	**60.61**	53.48	57.04	56.86	53.30	0.1027
AIMI$_{att}$	59.44	**58.82**	**58.69**	**58.46**	**56.68**	**0.0947**

clustering accuracy is evaluated by clustering the resulting representations and compute the accuracy(ACC).

Table 3 shows the RMSE with $\alpha = 0.5$ and ACC with respect to different α for the unsupervised baselines and our method. We can find that the proposed AIMI and its variant outperform other methods consistently in all missing rate cases, for both ACC and RMSE. The variant using attention-pooling performs better when the missing rate is high (e.g. $\alpha = 0.4, 0.5$). We suppose the attention mechanism captures the interdependence between different modalities, thus contributing to our model's robustness and superiority.

4 Conclusion

In this work, we propose a novel multi-modal data imputation method AIMI and apply it to real-world depression early detection task. Our method imputes modalities via an adversarial network with an implicit imputation mechanism. The analytic experiments demonstrate the rationality of our method and the comparison experiments show its superiority to the baselines.

AIMI also opens the door to many interesting future directions. Firstly, justification is needed for whether adversarial learning can guarantee sufficient predictive performance for our datasets with high missing rates, since adversarial learning requires a large number of training samples and time complexity in the training phase. On the other hand, we may improve the implicit feedback mechanism to avoid the risk of repeatedly performing unnecessary (non-informative) label prediction: While the best performance is achieved by applying our imputation method more than once, the prediction accuracy does not ascend monotonously as the imputation times increases.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (No. 62106271, 71804183); the Research Seed Funds of School of Interdisciplinary Studies, Renmin University of China; and the Research Funds of Renmin University of China (the Fundamental Research Funds for the Central Universities). Dr. Hongteng Xu also would like to thank the supports from the Beijing Key Laboratory of Big Data Management and Analysis Methods and the Public Policy and Decision-making Research Lab of Renmin University of China.

References

1. Andrew, G., Arora, R., Bilmes, J., Livescu, K.: Deep canonical correlation analysis. In: International conference on machine learning. pp. 1247–1255. PMLR (2013)
2. Baowaly, M.K., Lin, C.C., Liu, C.L., Chen, K.T.: Synthesizing electronic health records using improved generative adversarial networks. Journal of the American Medical Informatics Association **26**(3), 228–241 (12 2018). https://doi.org/10.1093/jamia/ocy142, https://doi.org/10.1093/jamia/ocy142
3. Boulahia, S.Y., Amamra, A., Madi, M.R., Daikh, S.: Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. Machine Vision and Applications **32**(6), 1–18 (2021)
4. Bzdok, D., Meyer-Lindenberg, A.: Machine learning for precision psychiatry: opportunities and challenges. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **3**(3), 223–230 (2018)
5. Cheng, B., Liu, M., Zhang, D., Shen, D.: Robust multi-label transfer feature learning for early diagnosis of alzheimer's disease. Brain imaging and behavior **13**(1), 138–153 (2019)
6. Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W.F., Sun, J.: Generating multi-label discrete patient records using generative adversarial networks. In: Machine learning for healthcare conference. pp. 286–305. PMLR (2017)
7. Donders, A.R.T., van der Heijden, G.J., Stijnen, T., Moons, K.G.: Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology **59**(10), 1087–1091 (2006). https://doi.org/https://doi.org/10.1016/j.jclinepi.2006.01.014, https://www.sciencedirect.com/science/article/pii/S0895435606001971
8. Er, M.J., Zhang, Y., Wang, N., Pratama, M.: Attention pooling-based convolutional neural network for sentence modelling. Information Sciences **373**, 388–403 (2016). https://doi.org/https://doi.org/10.1016/j.ins.2016.08.084, https://www.sciencedirect.com/science/article/pii/S0020025516306673
9. Fritsch, J., Wankerl, S., Nöth, E.: Automatic diagnosis of alzheimer's disease using neural network language models. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5841–5845. IEEE (2019)
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks (2014)
11. Guest, F.L.: Early detection and treatment of patients with alzheimer's disease: Future perspectives. Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders pp. 295–317 (2019)
12. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. science **313**(5786), 504–507 (2006)
13. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936), http://www.jstor.org/stable/2333955
14. Kar, S., Majumder, D.D.: A novel approach of diffusion tensor visualization based neuro fuzzy classification system for early detection of alzheimer's disease. Journal of Alzheimer's Disease Reports **3**(1), 1–18 (2019)
15. Martino-IST, I., Navarra, E.: Machine learning based analysis of fdg-pet image data for the diagnosis of neurodegenerative diseases. In: Applications of Intelligent Systems: Proceedings of the 1st International APPIS Conference 2018. vol. 310, p. 280. IOS Press (2018)

16. Oba, S., Sato, M.a., Takemasa, I., Monden, M., Matsubara, K.i., Ishii, S.: A bayesian missing value estimation method for gene expression profile data. Bioinformatics **19**(16), 2088–2096 (2003)
17. Rentz, D.M., Parra Rodriguez, M.A., Amariglio, R., Stern, Y., Sperling, R., Ferris, S.: Promising developments in neuropsychological approaches for the detection of preclinical alzheimer's disease: a selective review. Alzheimer's research & therapy **5**(6), 1–10 (2013)
18. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al.: Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine **12**(3), e1001779 (2015)
19. Tran, L., Liu, X., Zhou, J., Jin, R.: Missing modalities imputation via cascaded residual autoencoder. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4971–4980 (2017). https://doi.org/10.1109/CVPR.2017.528
20. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B.: Missing value estimation methods for dna microarrays. Bioinformatics **17**(6), 520–525 (2001)
21. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning. pp. 1096–1103 (2008)
22. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of machine learning research **10**(2) (2009)
23. World Health Organisation: Depression (2021), `https://www.who.int/news-room/fact-sheets/detail/depression`, Accessed 12 December 2021
24. Xu, Y., Zhang, Z., You, L., Liu, J., Fan, Z., Zhou, X.: scIGANs: single-cell RNA-seq imputation using generative adversarial networks. Nucleic Acids Research **48**(15), e85–e85 (06 2020). https://doi.org/10.1093/nar/gkaa506, `https://doi.org/10.1093/nar/gkaa506`
25. Yoon, J., Jordon, J., van der Schaar, M.: GAIN: Missing data imputation using generative adversarial nets. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 5689–5698. PMLR (10–15 Jul 2018), `https://proceedings.mlr.press/v80/yoon18a.html`
26. Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J.: Multi-source learning for joint analysis of incomplete multi-modality neuroimaging data. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1149–1157 (2012)
27. Zhang, C., Han, Z., cui, y., Fu, H., Zhou, J.T., Hu, Q.: Cpm-nets: Cross partial multi-view networks. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019), `https://proceedings.neurips.cc/paper/2019/file/11b9842e0a271ff252c1903e7132cd68-Paper.pdf`
28. Zhou, T., Thung, K.H., Liu, M., Shi, F., Zhang, C., Shen, D.: Multi-modal latent space inducing ensemble svm classifier for early dementia diagnosis with neuroimaging data. Medical image analysis **60**, 101630 (2020)