OCL: Ordinal Contrastive Learning for Imputating Features with Progressive Labels

Seunghun Baek[1*], Jaeyoon Sim[1*], Guorong Wu[2], and Won Hwa Kim[1]

[1] Pohang University of Science and Technology, Pohang, South Korea
{habaek4, simjy98, wonhwa}@postech.ac.kr
[2] University of North Carolina at Chapel Hill, Chapel Hill, USA

Abstract. Accurately discriminating progressive stages of Alzheimer's Disease (AD) is crucial for early diagnosis and prevention. It often involves multiple imaging modalities to understand the complex pathology of AD, however, acquiring a complete set of images is challenging due to high cost and burden for subjects. In the end, missing data become inevitable which lead to limited sample-size and decrease in precision in downstream analyses. To tackle this challenge, we introduce a holistic imaging feature imputation method that enables to leverage diverse imaging features while retaining all subjects. The proposed method comprises two networks: 1) An encoder to extract modality-independent embeddings and 2) A decoder to reconstruct the original measures conditioned on their imaging modalities. The encoder includes a novel *ordinal contrastive loss*, which aligns samples in the embedding space according to the progression of AD. We also maximize modality-wise coherence of embeddings within each subject, in conjunction with domain adversarial training algorithms, to further enhance alignment between different imaging modalities. The proposed method promotes our holistic imaging feature imputation across various modalities in the shared embedding space. In the experiments, we show that our networks deliver favorable results for statistical analysis and classification against imputation baselines with Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

1 Introduction

Understanding the progression of neurodegenerative disease such as Alzheimer's Disease (AD) often requires a complicated integration of multiple neuroimaging modalities. Advanced imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) provide insights into the structural, functional and molecular changes in the brain associated with AD. However, in many clinical studies, it is difficult to collect a complete image set of various imaging modalities due to cost and participant burden. In the end, subjects with missing scans become inevitable that cannot be used for downstream analyses, e.g., training a disease prediction model, and they eventually become

---

[*] Seunghun Baek and Jaeyoon Sim contributed equally to this paper.

discarded leading to limited sample-size. To learn robust and effective representation to discriminate underlying patterns from progressive disease stages, it becomes imperative to accurately impute the measures of unobserved modalities.

Recent studies [13] translate imaging scans at different levels (e.g., pixel, voxel) by leveraging their correlations [17,7,18]. However, focusing on brain regions of interest (ROIs) offers more relevant and interpretable information as disease-specific symptoms manifest in arbitrary shapes. Region-based models enable the extraction of localized features indicative of AD pathology, such as hippocampal atrophy or cortical thinning [23,2,1,19]. Prioritizing these regions enhances the sensitivity of prediction models by effectively capturing subtle disease-specific changes. Therefore, leveraging region-based models for estimating missing scans holds promise for improving the robustness in AD classification.

When training a diagnostic model, employing Cross-entropy or Supervised Contrastive Learning (SCL) [9] often simply groups the samples by their labels in an embedding space. However, as each AD stage represents the ordinal disease progression such as Control (CN), Early Mild Cognitive Impairment (EMCI), Late MCI (LMCI) and AD, aligning the embedding along the progression continuum offers two key advantages. First, it comprehensively captures the underlying pathology from preclinical stages to full-blown AD, enriching the data representation rather than treating them entirely separate entities. Second, it enhances the model robustness to inter-individual variability in Alzheimer's research by focusing on consistent pathological changes rather than individual differences in symptom manifestation. To achieve this, we propose *ordinal contrastive learning* (OCL), which dynamically adjusts the degree of repulsion and attraction among embeddings based on their severity discrepancy informed by disease labels. Simultaneously, different modalities from the same individual are pulled together to ensure personalized embeddings in the modality-agnostic embedding space.

With the OCL, we design an architecture that projects each sample to a modality-agnostic but progress-aligned embedding space and then generates a target measure from the embedding conditioned on a specific imaging modality. Our goal is to holistically generate realistic data for all missing scans personalized to each subject, accurately reflecting modality and disease characteristics without exhaustively training one-to-one mapping between modalities. **Key Contributions: 1)** Our method accurately estimates unobserved imaging measures for individual subjects using their existing data to solidify downstream analyses. **2)** We introduce *ordinal contrastive learning*, which aligns samples in the embedding space based on their disease severity. **3)** The experiments on Alzheimer's Disease Neuroimaging Initiative (ADNI) data show that our method accurately translates data, capturing realistic information for subsequent analyses.

2 Method

In this section, we describe an unified imaging measure-to-measure translation framework for samples with missing image features. Fig. 1 illustrates the overview of our framework, which consists of two training phases and one infer-
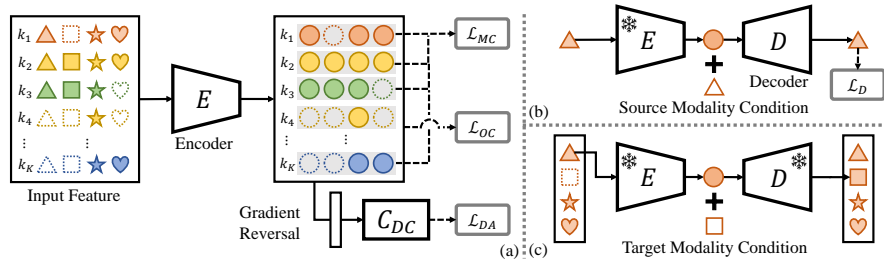
Fig. 1. Illustration of our framework. (a) An encoder E is trained to extract disease progression information across various input modalities through $\mathcal{L}_{DA}$ and $\mathcal{L}_{OC}$. Additionally, E is guided to maximize the similarity of embeddings from the same subject using $\mathcal{L}_{MC}$. (b) A decoder D is trained to reconstruct the embedding of a fixed E to its original input under its original modality condition, utilizing $\mathcal{L}_D$. (c) The trained E and D facilitate the translation of an input to the target modality when the corresponding condition is provided, while preserving disease progression information.

ence phase: (a) An encoder E is trained to map data into a modality-agnostic and disease-progression aligned embedding. (b) A decoder D, conditioned on the source (input) modality, is trained to reconstruct the original measure from the given embedding from (a). (c) Missing features are imputed from any existing feature of the subject by E and D, conditioned on the target (missing) modality.

### 2.1   Modality-agnostic Progressive Embedding

Let us consider data X consisting of K subjects, each with Q-dimensional measurements (e.g., Q ROIs) from S imaging features. The k-th subject $x_k \in \mathbb{R}^{S \times Q}$ has a diagnostic label $y_k \in \{1, \cdots, V\}$ ordered by severity (e.g., CN, EMCI, LMCI and AD), and each modality measurement is denoted as $x_{k,s} \in \mathbb{R}^Q$. For every k, the encoder E extracts disease progression information informed by y_k from $x_{k,s}$, regardless of the modality type s. This is accomplished by integrating three different guidance strategies as outlined below.

1) Domain Adversarial Training ($\mathcal{L}_{DA}$). We employ a domain adversarial training strategy, which derives an embedding $z_{k,s} = E(x_{k,s})$ with E to eliminate modality-specific information associated with s. This is achieved by a modality classifier C_{DC} and a gradient reversal layer [5] that reverses the sign of the gradient between E and C_{DC}. The modality adversarial loss $\mathcal{L}_{DA}$ is defined as

$$\mathcal{L}_{DA} = \mathcal{J}(s, C_{DC}(E(x_{k,s}))), \tag{1}$$

where $\mathcal{J}$ represents a suitable loss function (e.g., Cross-entropy). Through the gradient reversal layer, E is trained to maximize $\mathcal{L}_{DA}$, which leads to the discarding of modality-specific information, while C_{DC} learns to minimize $\mathcal{L}_{DA}$.

2) Ordinal Contrastive Learning ($\mathcal{L}_{OC}$). To accurately characterize measures of missing scans for downstream analyses, the encoder E should effectively
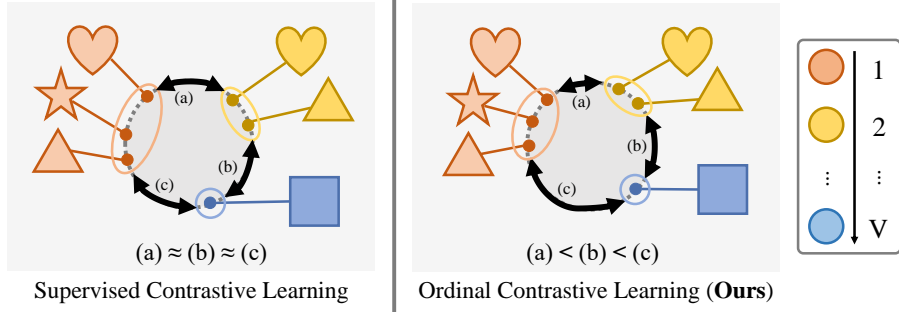
Fig. 2. Comparison of supervised (left) and ordinal (right) contrastive learning: Both approaches contrast the set of all samples from the same class as positives against the negatives from the rest of the batch. While supervised contrastive learning repels each negative without differentiation on labels denoted as $(a) \approx (b) \approx (c)$, ordinal contrastive learning assigns the penalizing strength based on the label distance.

extract disease progression information. E is designed to arrange each sample in the embedding space by the orders to accurately characterize disease progression.

The OCL is initiated from supervised contrastive learning (SCL) [9] which is briefly described below. For brevity, we omit the modality type $s \in \{1, \cdots, S\}$ (i.e., $x_{k,\cdot}$ denotes the k-th subject) as $\mathcal{L}_{DA}$ removes s-specific information in the embedding. In the context of SCL, within a batch I, let $i \in I \equiv \{1, ..., |I|\}$ represent the index of samples in the batch. Then, the embedding $z_{i,\cdot}$ of sample $x_{i,\cdot}$ is supervised by its label y_i through SCL. When $P(i) \equiv \{p \in P(i) : \hat{y}_p = \hat{y}_i\}$ and $N(i) \equiv \{n \in N(i) : \hat{y}_n \neq \hat{y}_i\}$ are the sets of indices for all positives and negatives in the batch distinct from i each, the loss of SCL $\mathcal{L}_{SC}$ is expressed as

$$\mathcal{L}_{SC} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum\limits_{p \in P(i)} \exp(z_i \cdot z_p / \tau) + \sum\limits_{n \in N(i)} \exp(z_i \cdot z_n / \tau)} \qquad (2)$$

where $\tau \in \mathbb{R}^+$ is a scalar temperature parameter, and $|P(i)|$ is its cardinality.

Notice that a single τ to control the strength of separation acts similar to classification loss, ignoring the degree of differences between each label. Considering that values of diagnostic label $y \in \{1, \cdots, V\}$ are aligned according to their severity (e.g., i-th subject is more severe than n-th subject if $y_i > y_n$), we define a function $d(y_i, y_n)$ measuring the distance between two labels as $|y_i - y_n|$. Higher $d(y_i, y_n)$ indicates greater diagnostic differences between z_i and z_n. Therefore, we make $\tau_{i,n}$ dependent on $y_{i,\cdot}$ and $y_{n,\cdot}$ as $\tau/d(i, n)$ to penalize greater label distance. By setting adaptive $\tau_{i,n}$ for each $z_{n,\cdot}$ and unique $\tau_{i,P}$ for every $z_{p,\cdot}$, we formulate our ordinal contrastive loss $\mathcal{L}_{OC}$ as

$$\mathcal{L}_{OC} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau_{i,P})}{\sum\limits_{q \in P(i)} \exp(z_i \cdot z_q / \tau_{i,P}) + \sum\limits_{n \in N(i)} \exp(z_i \cdot z_n / \tau_{i,n})}. \qquad (3)$$

To prevent the collapse or dispersion of the embedding space, the magnitude of gradient w.r.t positives and negatives should be the same [27]. By the gradient analysis detailed in the supplementary material, $\tau_{i,P}$ between z_i and z_p is set as

$$\tau_{i,P} = \frac{\sum\limits_{n \in N(i)} \exp(z_{i,\cdot} \cdot z_{n,\cdot}/\tau_{i,n})}{\sum\limits_{n \in N(i)} \exp(z_{i,\cdot} \cdot z_{n,\cdot}/\tau_{i,n})/\tau_{i,n}}. \tag{4}$$

3) $\mathcal{L}_{MC}$ (Maximize modality-wise coherence within a subject). While $\mathcal{L}_{DA}$ and $\mathcal{L}_{OC}$ require only unpaired data, identifying pairs originating from the same subject helps alignment by mapping those pairs to similar embeddings. To maximize the coherence between modalities of the same subject, we design the loss $\mathcal{L}_{MC}$ using a similarity function $sim(\cdot, \cdot)$ (e.g., cosine similarity) as

$$\mathcal{L}_{MC} = \frac{\sum_{k=1}^{K} \sum_{\substack{i,j \in \{1,\cdots,S\} \\ i \neq j}} -\delta_k(i,j) \cdot sim(x_{k,i}, x_{k,j})}{\sum_{k=1}^{K} \sum_{\substack{i,j \in \{1,\cdots,S\} \\ i \neq j}} \delta_k(i,j)} \tag{5}$$

where $\delta_k(i,j)$ is an indicator function defined as $\delta_k(i,j) = 1$ if both $x_{k,i}$ and $x_{k,j}$ exist for subject k, and $\delta_k(i,j) = 0$ otherwise. Our final loss to train E is the combination of the three introduced losses defined as

$$\mathcal{L}_E = \mathcal{L}_{DA} + \mathcal{L}_{OC} + \mathcal{L}_{MC} \tag{6}$$

where each loss term is equally contributing.

2.2   Modality-Conditioned Reconstruction from Embeddings

Similar to conditional generation methods which utilize a single generator to sample different distributions by a conditional vector [14,21], we treat the target modality type t as a one-hot condition vector $c_t \in \mathbb{R}^S$. Thus, the decoder D is tasked with estimating values of the target modality from a given embedding under a condition. Although modality-wise paired data from the same subject are limited, due to $\mathcal{L}_{DA}$ and $\mathcal{L}_{MC}$, the loss for the decoder $\mathcal{L}_D$ can be approximated to the self-reconstruction loss of unpaired data $x_{k,t}$ from the translation loss as

$$\mathcal{L}_D(x_{k,s}, x_{k,t}) = ||x_{k,t} - D([E(x_{k,s}), c_t])||^2 \approx \mathcal{L}_D(x_{k,t}) = ||x_{k,t} - D([E(x_{k,t}), c_t])||^2 \tag{7}$$

where $s, t \in \{1, \cdots, S\}$ correspond to distinct modality types of source and target each, and $[\cdot, \cdot]$ denotes to the concatenation of given elements.

2.3   Imputation Procedure

Suppose a subject k lacks a feature $x_{k,t}$. Our framework is capable of generating $\hat{x}_{k,t}$ based on existing $x_{k,s}$ where $s \neq t$, leveraging the trained encoder E and decoder D conditioned on the target modality t. This ensures that all K subjects possess feasible features across S modalities to maximize data utilization.

Table 1. Sample-size per modality of ADNI dataset.

| Label	CT	TAU	FDG	AMY	Common
CN	844	237	861	735	123
EMCI	490	186	597	833	102
LMCI	250	105	1138	447	40
AD	240	85	755	422	10
Total	1824	613	3351	2437	275

3   Experiment

Dataset. We validate our framework on four imaging measures from ADNI study [8]. Based on Destrieux atlas [3], each image was partitioned into 148 cortical and 12 sub-cortical regions. For each parcellation, region-specific imaging features including Standard Uptake Value Ratio (SUVR) [24] of β-amyloid protein (AMY), metabolism (FDG), and Tau protein (TAU) from PETs, along with cortical thickness (CT) from MRI scans, were measured. The diagnostic labels for each subject include CN, EMCI, LMCI and AD. $N{=}275$ subjects have complete image set denoted as 'Common' in Table 1, which serve as the test data in Group Comparison Analysis and Downstream Classification. To prevent double dipping, we exclusively utilize the remaining data to train our framework.

Baselines. Our baseline starts from *No Imputation* that utilizes only subjects with all imaging features. We adopt five established imputation baselines; Imputation with diagnostic group means per imaging feature (*Class-wise Mean*) [4], Linear model-based imputation with chained equations (*MICE*) [26], Imputation via non-parametric random forests (*MissForest*) [22], Imputation using generative adversarial networks [6] (*GAIN*) [29], and Imputation with trained models utilizing Sinkhorn divergences between random batches for similarity quantification (*Sinkhorn*) [15]. Additionaly, we employ *Pair-wise MLPs*, training one-to-one Multi-Layer Perceptrons (MLPs) for all pairs of modalities (i.e., $_4P_2$ pairs). For the purpose of ablation, we introduce two additional baselines from our framework (*Ours* ($\mathcal{L}_{OC} + \mathcal{L}_{MC}$)); (1) A model that replaces $\mathcal{L}_{OC}$ with $\mathcal{L}_{SC}$ and omits $\mathcal{L}_{MC}$ (*SCL*), and (2) a model that solely omit $\mathcal{L}_{MC}$ (*Ours* ($\mathcal{L}_{OC}$)).

Training. For each baselines and our method, we stack 2 layers with 128 hidden units for E and D each. AdamW [10] with learning rate 10^{-3} is used, decayed at 0.05 for every layer. We train each model 3000 epochs with a batch size of 4096.
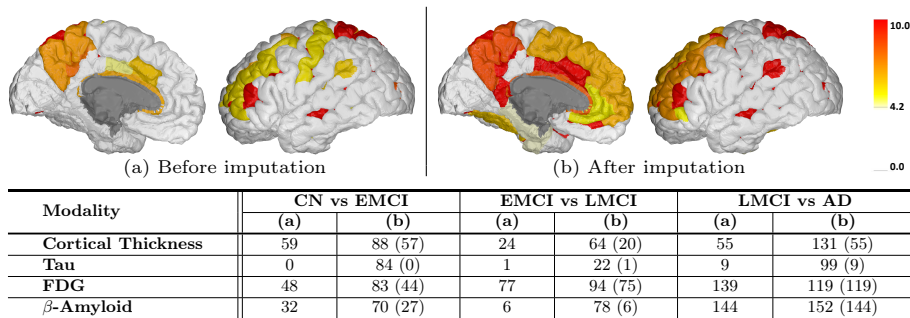
4   Result and Analysis

Embedding Space Analysis. To verify $\mathcal{L}_{OC}$ for aligning samples in the embedding space, we visualize the embeddings using t-SNE [11] compared to those from Cross-entropy $\mathcal{L}_{CE}$ and supervised contrastive learning $\mathcal{L}_{SC}$. As there are only 10 AD subjects in the test set (i.e., 'Common' in the Table 1), we randomly split the whole data in 8:2 ratio for train/test sets only in this visualization experiment. For every result in Fig. 3, $\mathcal{L}_{DA}$ was adopted together with each loss to utilize single E for every modality. As $\mathcal{L}_{CE}$ and $\mathcal{L}_{SC}$ simply separate the
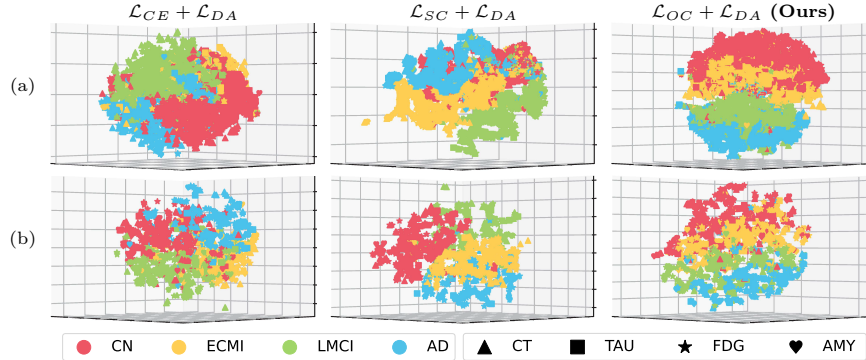
Fig. 3. Visualizations of embeddings under each loss by t-SNE [11]. Each individual encoder is trained with three distinct losses including Cross-Entropy $\mathcal{L}_{CE}$ (left), Supervised Contrastive Loss $\mathcal{L}_{SC}$ (center) and our Ordinal Contrastive Loss $\mathcal{L}_{OC}$ (right) along with domain adversarial loss $\mathcal{L}_{DA}$. (a) and (b) correspond to training and testing data respectively. (Color: AD-stage labels, Shape: imaging scan types.)

samples without differentiation under diagnostic label in the embedding space, resulting embeddings do not show any visible order. However, our $\mathcal{L}_{OC}$ aligns the embedding under the disease severity as we argued in section 2.1.

Statistical Analysis. To validate the robustness of our estimated data, we conduct statistical group comparisons at each ROI-level between consecutive groups (i.e., CN-EMCI, EMCI-LMCI, LMCI-AD). Our goal is to detect more statistically significant ROIs by imputing the missing features with our method. The increase indicates higher statistical sensitivity than using only real samples, as randomly generated data will not change the effect size between the groups.

Modality	CN vs EMCI		EMCI vs LMCI		LMCI vs AD	
	(a)	(b)	(a)	(b)	(a)	(b)
Cortical Thickness	59	88 (57)	24	64 (20)	55	131 (55)
Tau	0	84 (0)	1	22 (1)	9	99 (9)
FDG	48	83 (44)	77	94 (75)	139	119 (119)
β-Amyloid	32	70 (27)	6	78 (6)	144	152 (144)

Fig. 4. p-values from group comparisons with Bonferroni correction at $\alpha = 0.01$: (a) before imputation, (b) after imputation from our model. Top: Resultant p-value maps on a brain surface (left hemisphere) [12] in a $-log_{10}$ from CN and EMCI comparison with cortical thickness, and (b) shows higher sensitivity. Bottom: Number of significant ROIs. Number of common ROIs before-and-after imputation are in ().

Table 2. Classification performance on ADNI data with all imaging features.

Classifier	MLP (2 layers)			MLP (4 layers)		
Method	Accuracy	Precision	Recall	Accuracy	Precision	Recall
No Imputation	0.673±0.030	0.659±0.025	0.673±0.030	0.698±0.048	0.707±0.047	0.698±0.048
Class-wise Mean [4]	0.753±0.050	0.778±0.041	0.753±0.050	0.775±0.036	0.771±0.032	0.775±0.036
MICE [26]	0.739±0.043	0.761±0.046	0.739±0.043	0.814±0.043	0.761±0.046	0.814±0.043
MissForest [22]	0.721±0.061	0.753±0.080	0.721±0.061	0.832±0.025	0.844±0.024	0.832±0.025
Sinkhorn [15]	0.776±0.044	0.799±0.041	0.776±0.044	0.829±0.033	0.847±0.041	0.829±0.033
GAIN [29]	0.752±0.029	0.766±0.022	0.752±0.029	0.795±0.054	0.805±0.050	0.795±0.054
Pair-wise MLPs	0.756±0.030	0.782±0.036	0.756±0.030	0.782±0.062	0.799±0.060	0.782±0.062
SCL [9]	0.813±0.042	0.812±0.051	0.813±0.042	0.845±0.020	0.851±0.038	0.845±0.020
Ours ($\mathcal{L}_{OC}$)	0.826±0.029	0.829±0.021	0.826±0.029	0.851±0.046	0.862±0.051	0.851±0.046
Ours ($\mathcal{L}_{OC} + \mathcal{L}_{MC}$)	**0.829±0.042**	**0.839±0.041**	**0.829±0.042**	**0.854±0.025**	**0.862±0.024**	**0.854±0.025**

In Fig. 4, we summarize the count of significant ROIs, whose p-values survive multiple comparison correction i.e., Bonferroni correction at 0.01 [16]. With the imputation, the number of significant ROIs increases in most experiments depicted in Fig. 4 (bottom). The Fig. 4 (top) demonstrates the results of CN vs. EMCI analysis with CT on a cortical surface, which highlights the enhanced sensitivity (i.e., lower p-values) from the imputation in several frontal and temporal regions [20]. Since the resulting ROIs from our method subsume the majority of the ROIs from the real data analysis, as summarized in () of Fig. 4, it indicates that our approach preserves the characteristics of the real data as well. Comparing Fig. 4 (a) and (b) at the top, the ROIs newly identified includes *superior frontal gyrus*, *superior temporal gyrus* and *anterior cingulate cortex* which are major ROIs in preclinical AD analysis [25,28,30].

Downstream Classification Performance. We employ MLPs with 2 and 4 layers as downstream classifiers to demonstrate effectiveness of our imputation. These models take whole imaging measures as input after concatenation (e.g., $x_k \in \mathbb{R}^{S \cdot Q}$ for k-th subject). For unbiased results, we perform classification using cross-validation (CV). Beginning with the *No Imputation* scenario, we partition samples from common subjects into 5 folds, with one fold served as test data while the rest folds are utilized to train the classification model. For other methods, subjects with missing features are integrated as supplementary training data for the classification models after applying each distinctive imputation method.

In Table 2, we report the classification performance derived from the 5-fold CV. Imputation improved performance over *No Imputation* highlighting the advantage of utilizing incomplete subjects. Most of all, our framework outperformed all imputation baselines across both classifiers, showing superior results for all evaluation metrics. When comparing the use of $\mathcal{L}_{SC}$ (for SCL) and $\mathcal{L}_{OC}$, utilizing $\mathcal{L}_{OC}$ yields better performance by 1.3%p in accuracy, indicating that $\mathcal{L}_{OC}$ helps imputing data with ordinal labels. In addition, $\mathcal{L}_{MC}$ also improves the result by personalizing the embeddings from different modalities of the same subject, achieving almost 83% accuracy in the 4-way classification, which would remain as 67.3% without the our imputation.

5   Conclusion

In this work, we propose a promising framework that imputes unobserved imaging measures of subjects by translating their existing measures. To enable holistic imputation accurately reflecting individual disease conditions, our framework devises modality-invariant and disease-progress aligned latent space guided by **1)** domain adversarial training, **2)** maximizing modality-wise coherence, and **3)** ordinal contrastive learning. Experimental results on the ADNI study show that our model offers reliable estimations of unobserved modalities for individual subjects, facilitating the downstream AD analyses. Our work has potential to be adopted by other neuroimaging studies suffering from missing measures.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baek, S., Choi, I., et al.: Learning covariance-based multi-scale representation of neuroimaging measures for Alzheimer classification. In: IEEE International Symposium on Biomedical Imaging. pp. 1–5. IEEE (2023)
2. Cui, W., Yan, et al.: Bmnet: A new region-based metric learning method for early Alzheimer's disease identification with FDG-PET images. Frontiers in Neuroscience **16**, 831533 (2022)
3. Destrieux, C., et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage **53**(1), 1–15 (2010)
4. Donders, A.R.T., Heijden, V.D., et al.: A gentle introduction to imputation of missing values. Journal of clinical epidemiology **59**(10), 1087–1091 (2006)
5. Ganin, Y., Ustinova, et al.: Domain-adversarial training of neural networks. Journal of Machine Learning Research **17**(59), 1–35 (2016)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. Advances in Neural Information Processing Systems **27** (2014)
7. Harrison, T.M., Du, et al.: Distinct effects of beta-amyloid and tau on cortical thickness in cognitively healthy older adults. Alzheimer's & dementia **17**(7), 1085–1096 (2021)
8. Jack Jr, C.R., Bernstein, et al.: The Alzheimer's disease neuroimaging initiative (adni): MRI methods. Journal of Magnetic Resonance Imaging **27**(4), 685–691 (2008)
9. Khosla, P., Teterwak, et al.: Supervised contrastive learning. Advances in Neural Information Processing Systems **33**, 18661–18673 (2020)
10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. International Conference on Learning Representations (2017)
11. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research **9**(11) (2008)

12. Marinescu, R.V., et al.: BrainPainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes. In: Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop. pp. 112–120. Springer (2019)
13. McNaughton, J., Fernandez, J., Holdsworth, S., otheres: Machine learning for medical image translation: A systematic review. Bioengineering **10**(9), 1078 (2023)
14. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
15. Muzellec, B., Josse, J., Boyer, C., Cuturi, M.: Missing data imputation using optimal transport. In: International Conference on Machine Learning. pp. 7130–7140. PMLR (2020)
16. Napierala, M.A.: What is the bonferroni correction? Aaos Now pp. 40–41 (2012)
17. Ossenkoppele, R., Smith, et al.: Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. Neurology **92**(6), e601–e612 (2019)
18. Rubinski, A., Franzmeier, et al.: FDG-PET hypermetabolism is associated with higher tau-PET in mild cognitive impairment at low amyloid-PET levels. Alzheimer's Research & Therapy **12**, 1–12 (2020)
19. Sim, J., Jeon, S., Choi, I., Wu, G., Kim, W.H.: Learning to approximate adaptive kernel convolution on graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4882–4890 (2024)
20. Sintini, I., Graff-Radford, J., Senjem, et al.: Longitudinal neuroimaging biomarkers differ across Alzheimer's disease phenotypes. Brain **143**(7), 2281–2294 (2020)
21. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Advances in Neural Information Processing Systems **28** (2015)
22. Stekhoven, D.J., Bühlmann, P.: Missforest—non-parametric missing value imputation for mixed-type data. Bioinformatics **28**(1), 112–118 (2012)
23. Teipel, S.J., Born, C., et al.: Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage **38**(1), 13–24 (2007)
24. Thie, J.A.: Understanding the standardized uptake value, its methods, and implications for usage. Journal of Nuclear Medicine **45**(9), 1431–1434 (2004)
25. Valdés Hernández, M.d.C., et al.: Do 2-year changes in superior frontal gyrus and global brain atrophy affect cognition? Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **10**(1), 706–716 (2018)
26. Van Buuren, S., Groothuis-Oudshoorn, K.: mice: Multivariate imputation by chained equations in r. Journal of Statistical Software **45**, 1–67 (2011)
27. Wang, F., Liu, H.: Understanding the behaviour of contrastive loss. In: Computer Vision and Pattern Recognition. pp. 2495–2504 (2021)
28. Yang, H., Xu, H., Li, et al.: Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. gen psychiatr. 32 (2): e100005 (2019)
29. Yoon, J., Jordon, J., Schaar, M.: Gain: Missing data imputation using generative adversarial nets. In: International Conference on Machine Learning. pp. 5689–5698. PMLR (2018)
30. Yuan, Q., Liang, et al.: Altered anterior cingulate cortex subregional connectivity associated with cognitions for distinguishing the spectrum of pre-clinical Alzheimer's disease. Frontiers in Aging Neuroscience **14**, 1035746 (2022)