SOM2LM: Self-Organized Multi-Modal Longitudinal Maps

Jiahong Ouyang[1], Qingyu Zhao[2], Ehsan Adeli[1], Greg Zaharchuk[1⋆], and Kilian M. Pohl[1⋆⋆]

[1] Stanford University, Stanford CA 94305, USA
[2] Cornell University, Ithaca NY 14850, USA

Abstract. Neuroimage modalities acquired by longitudinal studies often provide complementary information regarding disease progression. For example, amyloid PET visualizes the build-up of amyloid plaques that appear in earlier stages of Alzheimer's disease (AD), while structural MRIs depict brain atrophy appearing in the later stages of the disease. To accurately model multi-modal longitudinal data, we propose an interpretable self-supervised model called Self-Organized Multi-Modal Longitudinal Maps (SOM2LM). SOM2LM encodes each modality as a 2D self-organizing map (SOM) so that one dimension of each modality-specific SOMs corresponds to disease abnormality. The model also regularizes across modalities to depict their temporal order of capturing abnormality. When applied to longitudinal T1w MRIs and amyloid PET of the Alzheimer's Disease Neuroimaging Initiative (ADNI, N=741), SOM2LM generates interpretable latent spaces that characterize disease abnormality. When compared to state-of-art models, it achieves higher accuracy for the downstream tasks of cross-modality prediction of amyloid status from T1w-MRI and joint-modality prediction of individuals with mild cognitive impairment converting to AD using both MRI and amyloid PET. The code is available at https://github.com/ouyangjiahong/longitudinal-som-multi-modality.

1 Introduction

Multi-modal neuroimaging can play a crucial role in diagnosing diseases, such as structural MRI and amyloid PET for Alzheimer's disease (AD) [5]. Specifically, amyloid PET scans visualize the build-up of amyloid plaques that appear in earlier stages of the disease, while structural MRIs depict brain atrophy appearing at later stages of AD[5]. As a result, the longitudinal multi-modal monitoring of at-risk and diseased individuals allows for a more comprehensive understanding of the progression of AD [5].

However, accurately modeling multi-modal, longitudinal neuroimages remains under-explored as most analyses simply combine measurements extracted from different modalities into a single vector [2,9,14]. This simple fusion strategy

⋆ co-founder, equity Subtle Medical
⋆⋆ corresponding author

ignores cross-modal relationships (e.g., correspondence of disease abnormality measured across modalities), which are essential to modeling disease progression. To capture these cross-modal relationships, deep learning models often map scans into modality-specific latent spaces and then *align* these latent spaces [1,15,8]. While these approaches are generally confined to cross-sectional settings, the Longitudinal Correlation Analysis (LCA) [17] jointly disentangles one linear direction in each modality-specific latent space such that intra-subject changes along those directions are maximally correlated between modalities. However, the directions must be linear, which limits the accuracy of LCA. Furthermore, the model does not explicitly learn the time shift between modalities, i.e., the order among modalities of displaying disease abnormality. Here, we propose to incorporate temporal dependencies in multi-modal encoding by first learning clusters stratified by disease abnormality in each modality-specific latent space and then aligning clusters across modalities so that the time shift is properly encoded.

Specifically, we choose to generate modality-specific clusters based on the idea of Longitudinal Self-Organized Representation (LSOR) [10], which relies on self-supervision to organize the clusters into a 2D self-organizing map (SOM). To ease interpretation, the SOM is organized as a 2D grid so that each node is associated with a cluster and its corresponding SOM representation. Each scan can now be encoded by a SOM similarity map, which records the scan's distance to each SOM representation in the latent space. unlike LSOR [10], we derive modality-specific SOMs that can further account for cross-modality relationships. We do so by encouraging one direction of the modality-specific SOMs to correspond to disease abnormality so that the temporal dependencies of disease abnormality are captured across modalities. For example, the abnormality is depicted first in the amyloid PET-specific SOM and later in the structural MRI-specific one [5]. Named **S**elf-**O**rganized **M**ulti-**M**odal **L**ongitudinal **M**aps (SOM2LM), we apply our method to the longitudinal T1-weighted MRIs and amyloid PETs of 741 ADNI participants. We show that the resulting 4-by-8 modality-specific SOM grids are stratified by markers related to disease abnormality, such as percentage of dementia cases, cognitive measure, and amyloid status. Then, we demonstrate that the modality-specific similarity maps visually represent disease abnormality by measuring the similarity between the latent representation of a scan and SOM representations associated with different disease stages. Furthermore, we illustrate that trajectories of estimated disease abnormalities (computed from SOM similarity maps) align with clinical findings of AD progression. Lastly, we evaluate our representations for cross- and joint-modality predictions. The cross-modality task uses the structural MRI to predict amyloid status measured from amyloid PET while the joint-modality task uses both modalities to predict which individuals suffering from MCI will converter to AD. On these tasks, the accuracy of predictors based on SOM2LM is higher than those based on other state-of-the-art self-supervised representations of multi-modal longitudinal neuroimages.

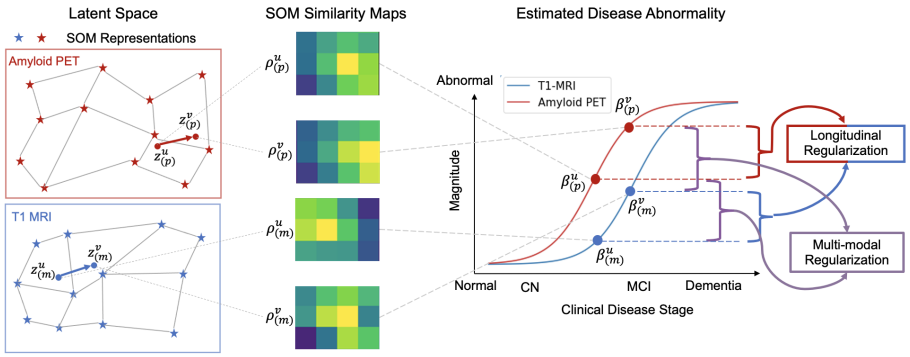

Fig. 1: Overview of SOM2LM. Left: Latent spaces of two modalities with SOM representations $\mathcal{G}_{(m)}$ (blue stars) and $\mathcal{G}_{(p)}$ (red stars) organized in 2D grids. Latent representations of a multi-modal longitudinal pair $(x^u_{(m)}, x^u_{(p)}, x^v_{(m)}, x^v_{(p)})$ are shown in red and blue arrows; Middle: SOM similarity maps of each representation with bright color suggesting high similarity; Right: Longitudinal regularization applied on the estimated disease abnormality $(\beta^u_{(\circ)}, \beta^v_{(\circ)})$ (red and blue arrows) and multi-modal regularization applied on $(\beta^\times_{(m)}, \beta^\times_{(p)})$ (purple arrows).

2 Constructing Self-Organized, Multi-Modal, Longitudinal Maps

For an individual, let $x^u_{(\circ)}$ be the brain measurements extracted from modality $\circ$ of assessment u and $\mathcal{U}_{(\circ)} = \{x^u_{(\circ)}\}$ be the set of assessments (of a modality) in the training set. As in [11,18,10], a modality-specific encoder $F_{(\circ)}$ maps input $x^u_{(\circ)}$ to the latent space, i.e., $z^u_{(\circ)} := F_{(\circ)}(x^u_{(\circ)})$, and a modality-specific decoder $H_{(\circ)}$ reconstructs the input $x^u_{(\circ)}$ from the latent representation $z^u_{(\circ)}$, i.e., $\tilde{x}^u_{(\circ)} := H_{(\circ)}(z^u_{(\circ)})$. In the remainder of this section, we first describe how to generate a modality-specific SOM and corresponding SOM similarity map. We then introduce a longitudinal regularization on those SOM similarity maps to encourage one direction of a modality-specific SOM to correspond to disease abnormality. In addition, we regularize across the modality-specific SOM similarity maps to preserve their temporal ordering of recording abnormality as defined by the clinical literature. All the resulting loss functions are then combined into a final objective function.

Creating Modality-Specific SOMs. We create a SOM for each modality deriving from [10]. Specifically, for each modality $\circ$, SOM representations (i.e., cluster centroids) are organized in a N_r by N_c grid (denoted as SOM grid) $\mathcal{G}_{(\circ)} = \{g_{(\circ),i,j}\}_{i=1,j=1}^{N_r,N_c}$, where $g_{(\circ),i,j}$ denotes the SOM representation corresponds to the node on the i-th *row* and j-th *column* in this grid. This easy-to-visualize grid preserves the high-dimensional relationships between the clusters. As in Fig. 1 (left), gray lines connect neighboring SOM representations in the

grid. In order to create the modality-specific SOMs, the objective of the models consists of three components.

Reconstruction Loss: given the latent representation $z^u_{(\circ)}$, its closest SOM representation is denoted as $g_{(\circ),\epsilon^u_{(\circ)}}$, where $\epsilon^u_{(\circ)} := argmin_{(i,j)} \parallel z^u_{(\circ)} - g_{(\circ),i,j} \parallel_2$ is its 2D grid index in $\mathcal{G}_{(\circ)}$ and $\| \cdot \|_2$ is the Euclidean norm. The reconstruction loss encourages both the latent representation and its closet SOM representation to be descriptive of the input, i.e.,

$$L_{recon,(\circ)} := \mathbb{E}_{x^u_{(\circ)} \sim \mathcal{U}_{(\circ)}} \left(\parallel x^u_{(\circ)} - \tilde{x}^u_{(\circ)} \parallel_2^2 \right),$$

where $\tilde{x}^u_{(\circ),g} = H_{(\circ)}(g_{(\circ),\epsilon^u_{(\circ)}})$ and $\mathbb{E}(\cdot)$ is the expected value.

Commitment Loss: the function explicitly promotes the closeness between the latent representation and its closet SOM representation, i.e.,

$$L_{commit,(\circ)} := \mathbb{E}_{x^u_{(\circ)} \sim \mathcal{U}_{(\circ)}} \left(\parallel z^u_{(\circ)} - g_{(\circ),\epsilon^u_{(\circ)}} \parallel_2^2 \right).$$

Proximity Loss: the models updates all SOM representations $g_{(\circ),i,j}$ by incorporating a soft weighting scheme. Specifically, a weight $w^u_{(\circ),i,j}$ defines how much $g_{(\circ),i,j}$ should be updated with respect to $z^u_{(\circ)}$ based on its proximity to the grid location $\epsilon^u_{(\circ)}$, i.e.,

$$L_{prox,(\circ)} := \mathbb{E}_{x^u_{(\circ)} \sim \mathcal{U}_{(\circ)}} \left(\sum_{g_{(\circ),i,j} \sim \mathcal{G}_{(\circ)}} \left(w^u_{(\circ),i,j} \cdot \parallel sg[z^u_{(\circ)}] - g_{(\circ),i,j} \parallel_2^2 \right) \right).$$

where $w^u_{(\circ),i,j} := \delta \left(e^{-\frac{\| \epsilon^u_{(\circ)} - (i,j) \|_1^2}{2\tau}} \right)$. $\delta(w) := \frac{w}{\sum_{i,j} w_{i,j}}$ ensures that the scale of weights is constant during training. τ is a scaling factor so that the weights gradually concentrate on SOM representations closer to $\epsilon^u_{(\circ)}$ as training proceeds. Specifically, $\tau(t) := N_r \cdot N_c \cdot \tau_{max} \left(\frac{\tau_{min}}{\tau_{max}} \right)^{t/T}$ with τ_{min} and τ_{max} being the minimum and maximum standard deviation in the Gaussian kernel. t and T represent the current and the maximum iteration. $sg[\cdot]$ is a stop-gradient operator [16], preventing the undesirable scenario where $z^u_{(\circ)}$ is pulled towards a naive solution [10].

SOM Similarity Map: once constructed, the interpretation of the SOM partly relies on computing the 2D SOM similarity map $\rho^u_{(\circ)}$ for the latent representation $z^u_{(\circ)}$ of each assessment. Specifically, we compute the similarity (i.e., closeness) between $z^u_{(\circ)}$ and the SOM representations $\mathcal{G}_{(\circ)}$, i.e., $\rho^u_{(\circ)} := softmax(- \parallel z^u_{(\circ)} - \mathcal{G}_{(\circ)} \parallel_2^2 / \gamma)$ with $\gamma := std(\parallel z^u_{(\circ)} - \mathcal{G}_{(\circ)} \parallel_2^2)$, where std denotes the standard deviation of the distance between the latent representation to all SOM representations. As visualized in Fig. 1 (middle), brighter colors represent higher similarity to the SOM representation than duller ones.

Stratifying Modality-Specific SOMs by Disease Abnormality via Longitudinal Regularization. Different from [10], we derive an abnormality-stratified

SOM grid with increasing disease abnormality along one direction (e.g., towards the right) of the grid. It is regularized by enforcing the latter assessment to have "high similarity" (bright color in Fig. 1 (middle)) along this direction in the SOM similarity map compared to the prior one. Specifically, for a similarity map $\rho^u_{(\circ)}$, we estimate disease abnormality $\beta^u_{(\circ)}$. To do so, we first compute the sum of the similarity in each column of the SOM similarity map $\rho^u_{(\circ)}$ because $\beta^u_{(\circ)}$ is invariant to the distribution of similarity within a column. Then we define the estimated disease abnormality $\beta^u_{(\circ)}$ as the weighted sum of each column's similarity: $\beta^u_{(\circ)} := \sum_{j=1}^{N_c} \left(j \cdot \sum_{i=1}^{N_r} \rho^u_{(\circ),i,j} \right)$. Now let be u and v be two assessments in a longitudinal scan (with u acquired before v), we then further enforce v to have more severe abnormality comparing to u by regularizing a hinge loss to encourage $\beta^v_{(\circ)}$ to be larger than $\beta^u_{(\circ)}$, i.e.,

$$L_{long,(\circ)} := \mathbb{E}_{(x^u_{(\circ)}, x^v_{(\circ)}) \sim \mathcal{S}_{(\circ)}} \left(max(0, \beta^u_{(\circ)} - \beta^v_{(\circ)} + \alpha_{(\circ)}) \right)$$

where $\mathcal{S}_{(\circ)} = \{(x^u_{(\circ)}, x^v_{(\circ)})\}$ is the set of modality-specific pairs of intra-subject assessments from all training samples. $\alpha_{(\circ)}$ is the threshold for the minimal increasing abnormality in the longitudinal regularization.

Enforcing Correspondence Across Modality-specific SOMs via Multi-modal Regularization. We propose to embed prior knowledge about disease progression into the model design. With respect to AD, amyloid PET reveals abnormality years before any are shown in structural MRI. We achieve this by incorporating a multi-modal regularization on modality-specific SOM similarity maps. Specifically, at a given assessment $\times$, amyloid PET should display a larger magnitude of abnormality than T1 MRI (red and blue dots in Fig. 1 (right)). Let $\circ = m$ denote T1w-MRI and $\circ = p$ stands for amyloid PET, then this relation is regularized by another hinge loss on the estimated disease abnormality $\beta^\times_{(m)}$ and $\beta^\times_{(p)}$, i.e.,

$$L_{multi} := \mathbb{E}_{(x^u_{(m)}, x^u_{(p)}, x^v_{(m)}, x^v_{(p)}) \sim \mathcal{S}_{(m,p)}} \sum_{\times \in \{u,v\}} \left(max(0, \beta^\times_{(m)} - \beta^\times_{(p)} + \alpha_{(m,p)}) \right)$$

Here, we define $\mathcal{S}_{(m,p)}$ as the set of all longitudinal pairs of the same subject with both MRI and PET at each assessment. $\alpha_{(m,p)}$ is the threshold for the minimal increasing abnormality between two modalities.

Objective Function. The complete objective function is the weighted combination of the prior losses with weighing parameters $\lambda_{commit,(\circ)}$, $\lambda_{prox,(\circ)}$, $\lambda_{long,(\circ)}$, and λ_{multi}:

$$\begin{aligned} L := \lambda_{multi} \cdot L_{multi} + \sum_{\circ \in \{m,p\}} & L_{recon,(\circ)} + \lambda_{commit,(\circ)} \cdot L_{commit,(\circ)} \\ & + \lambda_{prox,(\circ)} \cdot L_{prox,(\circ)} + \lambda_{long,(\circ)} \cdot L_{long,(\circ)} \end{aligned} \quad (1)$$

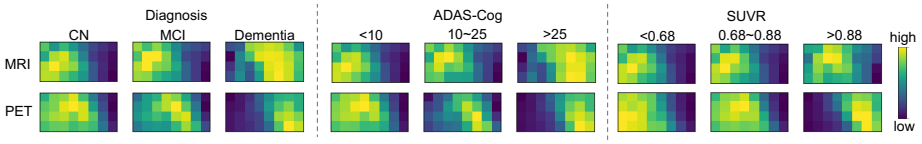

Fig. 2: The average SOM similarity maps ρ over all subjects with respect to diagnosis, ADAS-Cog scores, and amyloid summary SUVR. More severe groups and PET (compared to MRI) have higher similarity towards the right.

The objective function encourages an interpretable SOM grid for each modality with the horizontal direction being linked to disease progression and the temporal order of disease process across multi-modal SOM similarity maps.

3 Experiments

3.1 Experimental Setting

Dataset. We evaluate the proposed method on all subjects of ADNI-1, 2, 3, GO [12] that have at least two visits, and those assessments are at least half a year apart from each other. For T1-weighted MRI, this selection criteria resulted in 1194 subjects and a total of 5802 T1w MRIs. Each MRI was reduced to the z-score of 313 ROI measurements[3]. Specific to amyloid PET, our analysis is based on the 1977 PET from 676 subjects. 160 ROI features [7] were used as input $x_{(p)}$ with each representing the Standardized Uptake Value Ratio (SUVR) to the composite regions [7]. For multi-modal longitudinal data, we included all 953 visits from the 406 subjects with both modalities across multiple assessments. In downstream tasks, we also include those subjects that only have one assessment with both modalities so that the dataset increases to 1272 assessments from 741 subjects. 591 of those visits are labeled amyloid positive, i.e., they have a summary SUVR equal to or larger than 0.78 [7]. The age difference between amyloid positive (age: 74.7 $\pm$ 7.2) and negative (age: 72.4 $\pm$ 7.3) is significant ($p < 0.05$, two-sample t-test). Among those 741 subjects, 377 stayed Mild Cognitive Impaired (MCI) for 5 years (age: 72.6 $\pm$ 7.6) and 50 converted to dementia (age: 74.1 $\pm$ 7.4). The two groups had no significant age difference ($p = 0.12$, two-sample t-test).

Implementation Details. The encoders and decoders are both multilayer perceptrons (MLP) (details in Table. S1) with the dimension of latent representations being 64. SOM representations were randomly initialized. Next, a network specific to each modality was first trained based on $\mathcal{S}_{(m)}$ and $\mathcal{S}_{(p)}$ and then the networks were trained together via L_{multi} using $\mathcal{S}_{(m,p)}$. To accommodate the difference in the range of values of measurements extracted from MRI and PET, different weighing parameters λ were used to balance loss components in Eq. 1 as in [10]. Details are summarized in Table S2 and Table S3.

Evaluation. We performed the five-fold cross-validation (folds split based on subjects) using 10% of the training subjects for validation. Note that the same

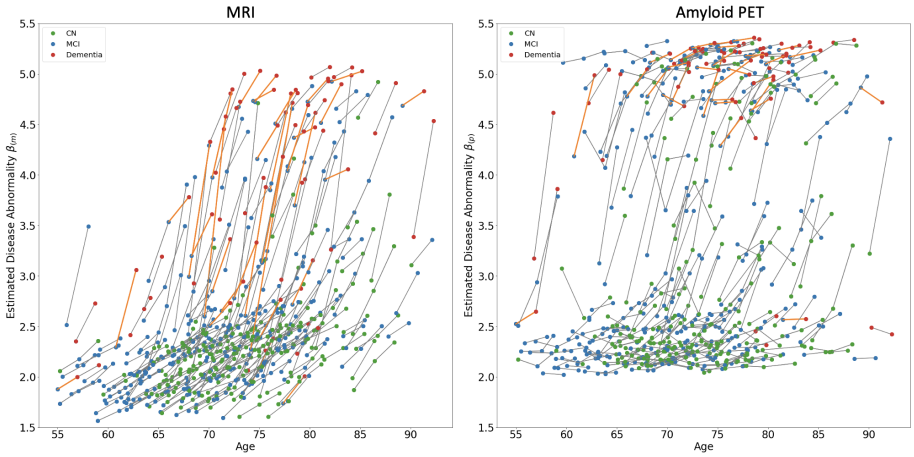

Fig. 3: Trajectories of estimated disease abnormalities (left: MRI $\beta_{(m)}$, right: amyloid PET $\beta_{(p)}$). Green dots represent cognitive normal subjects, blue are those that have MCI, and red are diagnosed with dementia. The conversion times (from MCI to Dementia) are highlighted by orange lines.

fold split was used for pre-training and downstream tasks. We illustrated the interpretability and capability in disease modeling of our method by assessing the correlation of the SOM similarity maps to markers related to disease abnormality (e.g., diagnosis, the Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog), and amyloid summary SUVR). Furthermore, we visualized the trajectories of estimated disease abnormality. Then, we quantitatively evaluated the quality of the representations by applying them to two downstream tasks. The first task focuses on cross-modality prediction, i.e., predicting amyloid status from T1-weighted MRI. This task is of clinical interest as (if successful) it enables the evaluation of amyloid status (one standard biomarker in diagnosing AD [6]) by MRI, resolving the problems of acquiring PET scans (e.g., high cost, limited accessibility, and radiation exposure). The age difference in amyloid positive and negative cohorts was resolved by regressing out age from the representation (details in Table S3);

The second focuses on joint-modality prediction, i.e., predict which individuals suffering from MCI will convert to AD using both MRI and PET. Detailed training setup is described in Table S3. We measured the classification accuracy via Balanced accuracy (BACC) and Area Under Curve (AUC). We compared the accuracy metrics to models using the same architecture with encoders pre-trained by other longitudinal self-supervised learning (SSL) methods (LVAE [13], LSOR [10]), multi-modal SSL (CLIP [4], ContIG [15]), and multi-modal longitudinal SSL (LCA [17]). All methods used the same experimental setup with the sole difference coming from the self-supervised regularization.

Type	Methods	Amyloid Status				MCI converter			
		Frozen		Fine-tuned		Frozen		Fine-tuned	
		BACC	AUC	BACC	AUC	BACC	AUC	BACC	AUC
N	No pretrain	-	-	0.67	0.73	-	-	0.65	0.73
L	LVAE [13]	0.59	0.67	0.69	0.72	0.63	0.70	0.65	0.73
	LSOR [10]	0.60	0.66	0.68	0.73	0.62	0.69	0.64	0.73
M	CLIP [4]	0.64	0.71	0.71	0.75	0.59	0.64	0.61	0.68
	ContIG [15]	0.63	0.69	0.69	0.74	0.60	0.66	0.62	0.71
ML	LCA [17]	0.64	0.70	0.70	0.77	0.63	0.70	0.62	0.73
	SOM2LM	**0.66**	**0.75**†	**0.74**	**0.80**†	**0.67**	**0.74**†	**0.67**	**0.75**

Table 1: Supervised downstream tasks. Types are no pre-training(N), and pre-training on longitudinal (L), multi-modal (M), and multi-modal longitudinal (ML) data. SOM2LM is more accurate than other state-of-the-art self-supervised methods († : $p < 0.05$, Delong's test).

3.2 Results

Interpretability of SOM similarity maps. The average similarity maps ρ shown in Fig. 2 reveal that higher similarity (yellow) gradually shifts towards the right for more severe groups. This observation is confirmed by the strong correlation between the SOM grid index with dementia diagnosis (MRI: 0.71, PET: 0.73), ADAS-Cog (MRI: 0.80, PET: 0.78), and SUVR (MRI: 0.82, PET: 0.91) shown in Fig. S1. Moreover, high similarity (yellow) in the SOM similarity maps of PET is on the right compared to those from MRIs. Specifically, while there is hardly any difference between Cognitive Normal (CN) and MCI captured from MRIs, the SOM similarity maps based on PET clearly distinguish the different stages.

Interpretability of estimated disease abnormality. Fig. 3 plots trajectories of the estimated disease abnormality (left: $\beta_{(m)}$, right: $\beta_{(p)}$) with respect to chronological age. As expected, most of the cognitively normal cases (green) stay at the bottom (small estimated abnormality), while dementia cases (red) appear on the top (large estimated abnormality), suggesting the effectiveness of the estimated disease abnormality. For both modalities, the model automatically learns the typical "sigmoid" shape associated with AD progression [5]. Compared to MRI, disease abnormality estimated from amyloid PET has a larger magnitude and conversion time (MCI to dementia, highlighted by orange lines) in PET are on the top (saturated regions of sigmoid). It suggests the temporal ordering of capturing abnormality: accumulation of amyloid plaque in the preclinical stage (cognitively normal) and gradually converges with progressing to the dementia stage; fast progressing brain atrophy happens with the worsening of the cognition [5].

Downstream Tasks. To evaluate the quality of the learned representations, we evaluate two downstream tasks with both frozen and fine-tuned encoders. For estimating amyloid status from the corresponding MRI (Table 1), the proposed method is significantly ($p < 0.05$, DeLong's test) more accurate for both frozen

(BACC: 0.66, AUC: 0.75) and fine-tuned encoder (BACC: 0.74, AUC: 0.80) than state-of-the-art multi-modal longitudinal self-supervised method such as LCA [17], which also jointly models the cross-modal relationship. With respect to predicting MCI conversion, the proposed method is again significantly ($p < 0.05$, DeLong's test) more accurate (BACC: 0.67, AUC: 0.74) with the frozen encoder and better accuracy compared with all other methods in the fine-tuned setting. The ablation study in Table S4 demonstrates that regularizing the cross-modal relationship via L_{multi} significantly contributes to the accuracy of both tasks.

4 Conclusion

In this work, we proposed SOM2LM, the first interpretable, multi-modal longitudinal self-supervised method that explicitly embeds domain knowledge about disease progression (i.e., temporal dependency of modalities showing disease abnormalities) in the modal design. Modality-specific SOMs yielded interpretable latent spaces and SOM similarity maps. By incorporating the longitudinal regularization, one direction of modality-specific SOM captures the longitudinal changes, which allows for the estimation of disease abnormality using SOM similarity maps. By regularizing a larger magnitude of abnormality in amyloid PET than in MRI, SOMs incorporated the clinical knowledge about the disease progression across different modalities. Note, ROI measurements were used as the input of the model, which can potentially improved by using images to obtain more informative and generalizable representations. As a result, the interpretability of the representations was confirmed by the correlation between the SOM grid and disease abnormality measures. When evaluated on downstream tasks of cross-modality prediction of amyloid status and joint-modality prediction of MCI converters, SOM2LM was more accurate than other state-of-the-art methods. In conclusion, SOM2LM can generate interpretable latent representations, encoding disease progression captured by longitudinal multi-modal neuroimaging, and yield valuable representations for downstream tasks.

Acknowledgement

This work was partly supported by funding from the National Institute of Health (AA010723 and AA028840), the DGIST R&D program of the Ministry of Science and ICT of KOREA (22-KUJoint-02), Stanford's Department of Psychiatry & Behavioral Sciences Faculty Development & Leadership Award, and by Stanford HAI Google Cloud Credit.

References

1. Duan, J., Chen, L., Tran, S., Yang, J., Xu, Y., Zeng, B., Chilimbi, T.: Multi-modal alignment using representation codebook. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15651–15660 (2022)

2. El-Sappagh, S., Abuhmed, T., Islam, S.R., Kwak, K.S.: Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. Neurocomputing **412**, 197–215 (2020)
3. Hartig, M., Truran-Sacrey, D., Raptentsetsang, S., Simonson, A., Mezher, A., Schuff, N., Weiner, M.: UCSF FreeSurfer method (2023), https://ida.loni.usc.edu/download/files/study/39ce6665-fbcf-4943-a2f2-ba40d163867b/file/adni/UCSF_FreeSurfer_Methods_and_QC_OFFICIAL_20140131.pdf
4. Huang, W.: Multimodal contrastive learning and tabular attention for automated alzheimer's disease prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2473–2482 (2023)
5. Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q.: Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology **9**(1), 119–128 (2010)
6. Jack Jr, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al.: NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia **14**(4), 535–562 (2018)
7. Lee, J., Murphy, A., Ward, T., Harrison, T., Landau, S., Jagust, W.: Amyloid PET processing methods (2023), https://ida.loni.usc.edu/download/files/study/fb22b321-5461-4555-b2a8-cef3f3c4709b/file/adni/ADNI_UCBerkeley_AmyloidPET_Methods_v2_2023-06-29.pdf
8. Liu, Y., Fan, L., Zhang, C., Zhou, T., Xiao, Z., Geng, L., Shen, D.: Incomplete multi-modal representation learning for Alzheimer's disease diagnosis. Medical Image Analysis **69**, 101953 (2021)
9. Lu, L., Elbeleidy, S., Baker, L.Z., Wang, H.: Learning multi-modal biomarker representations via globally aligned longitudinal enrichments. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 817–824 (2020)
10. Ouyang, J., Zhao, Q., Adeli, E., Peng, W., Zaharchuk, G., Pohl, K.M.: LSOR: Longitudinally-consistent self-organized representation learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science. pp. 279–289. Springer (2023)
11. Ouyang, J., et al.: Self-supervised longitudinal neighbourhood embedding. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science. vol. 12902, pp. 80–89. Springer (2021)
12. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., et al.: Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology **74**(3), 201–209 (2010)
13. Sauty, B., Durrleman, S.: Progression models for imaging data with longitudinal variational auto encoders. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science. pp. 3–13 (2022)
14. Tabarestani, S., Aghili, M., Eslami, M., Cabrerizo, M., Barreto, A., Rishe, N., Curiel, R.E., Loewenstein, D., Duara, R., Adjouadi, M.: A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. NeuroImage **206**, 116317 (2020)
15. Taleb, A., Kirchler, M., Monti, R., Lippert, C.: Contig: Self-supervised multimodal contrastive learning for medical imaging with genetics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20908–20921 (2022)

16. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Advances in neural information processing systems **30** (2017)
17. Zhao, Q., Adeli, E., Pohl, K.M.: Longitudinal correlation analysis for decoding multi-modal brain development. In: Medical Image Computing and Computer Assisted Intervention, Lecture Notes in Computer Science. pp. 400–409. Springer (2021)
18. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. Medical Image Analysis **71**, 102051 (2021)