Clinically-guided Prototype Learning and Its Use for Explanation in Alzheimer's Disease Identification

Ahmad Wisnu Mulyadi[2], **Wonsik Jung**[2], **Kwanseok Oh**[1], **Jee Seok Yoon**[2], **Heung-Il Suk**[1,2*]
[1]Department of Artificial Intelligence, Korea University
[2]Department of Brain and Cognitive Engineering, Korea University
{wisnumulyadi, ssikjeong1, ksohh, wltjr1007, hisuk}@korea.ac.kr

Abstract

Identifying Alzheimer's disease (AD) involves a deliberate diagnostic process owing to its innate traits of irreversibility with subtle and gradual progression, hampering the AD biomarker identification from structural brain imaging (*e.g.*, structural MRI) scans. We propose a novel deep-learning approach through eXplainable AD Likelihood Map Estimation (XADLiME) for AD progression modeling over 3D sMRIs using clinically-guided prototype learning. Specifically, we establish a set of topologically-aware prototypes onto the clusters of latent clinical features, uncovering an AD spectrum manifold. We then measure the similarities between latent clinical features and these prototypes to infer a "pseudo" AD likelihood map. Considering this pseudo map as an enriched reference, we employ an inferring network to estimate the AD likelihood map over a 3D sMRI scan. We further promote the explainability of such a likelihood map by revealing a comprehensible overview from two perspectives: clinical and morphological.

1 Introduction

Alzheimer's disease (AD) is widely known as a neurodegenerative disease distinguished by its traits of irreversibility, gradual yet subtle progression, and prevailing cause of dementia. Despite the potential benefits of exploiting structural magnetic resonance imaging (sMRI)[16], it is still challenging to identify the risk of AD at the earliest possible stage and time, owing to its innate traits with a diverse inter-subject progression rate [1] and a high possibility of entangled with normal aging [14].

A prototype-based network belongs to the example-based explanation approach under "explainable artificial intelligence" that learns a set of class-specific representative samples [10, 15]. Through prototype learning for AD progression modeling (ADPM), a subtle progression of AD can be imitated by maintaining the relations across prototype units in the AD spectrum [13]. It is also possible to dispense multiple samples per clinical stage to reconcile with diverse inter-subject variances and accommodate the multi-pathological pathways. Nevertheless, sMRI-driven prototype learning for ADPM is still relatively limited and under-explored.

We argue that a satisfactory ADPM must fulfill the following criteria: (i) it intuitively considers the natural traits of AD (*e.g.*, irreversibility and progressiveness); (ii) it carries out rigorous diagnostic performance; (iii) it provides explainability in the form of a set of representative features to reliably portray the progression facts of AD (clinically and/or morphologically). With these criteria in mind, we propose a framework for ADPM using a novel approach through eXplainable AD Likelihood Map Estimation (XADLiME)[1] over a whole-brain 3D sMRI scan via clinically-guided prototype learning.

*Corresponding author

[1]Available at https://github.com/ku-milab/XADLiME

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

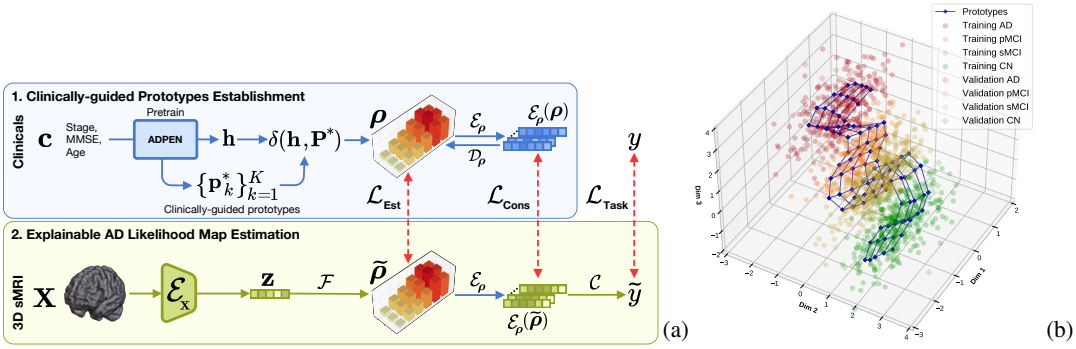


Figure 1: (a) Overview of XADLiME in estimating AD likelihood map $\widetilde{\boldsymbol{\rho}}$ over a 3D sMRI. (b) Visualization of latent clinical features and clinically-guided prototypes $\mathbf{P}^*$ with 5×20 arrangement.

2 Proposed Method

As depicted in Fig. 1a, our XADLiME is comprised of two subsequent streams. During the initial stream, we aim to establish K number of prototypes $\mathbf{P} = \{\mathbf{p}_k\}_{k=1}^K$ over the latent clinical features $\mathbf{h}$ given the composite, concatenated clinical information $\mathbf{c} = [\mathbf{y} \circ s \circ a]$ (*i.e.*, clinical stage $\mathbf{y}$, cognitive score s, and age a) with a predetermined topological arrangement (*e.g.*, 1D, 2D, or 3D) through our proposed AD-spectrum-aware prototypical embedding network (ADPEN). It is built upon a variational autoencoder (VAE) [7] along with two key components: (i) an AD-spectrum-aware ordering loss inspired by [17] to impose the features in comprehending the AD progression consistently and (ii) a self-organizing map (SOM) [8, 13] to establish a set of topological-aware prototypes adequately. Note that we train all modules within ADPEN jointly in an unsupervised manner, eliminating the necessity to predetermine the number of class-specific prototypes. Upon optimizing this network, we acquire well-established, clinically-guided prototypes $\mathbf{P}^*$ as an enriched clinical reference. As an auxiliary advantage, our ADPEN has a way to reconstruct each prototype $\{\mathbf{p}_k^*\}_{k=1}^K$ by employing the well-trained decoding network to observe their clinical states (*i.e.*, prototypical clinical states).

Subsequently, we utilize the pre-trained, clinically-guided prototypes $\mathbf{P}^*$ to infer a "pseudo" AD likelihood map $\boldsymbol{\rho} \in \mathbb{R}^K$ through $\rho_k = \frac{\exp(-\delta(\mathbf{h}, \mathbf{p}_k^*)/\gamma)}{\sum_j \exp(-\delta(\mathbf{h}, \mathbf{p}_j^*)/\gamma)}$ with $k \in \{1, \ldots, K\}$, δ denotes a distance function (e.g., squared Euclidean), and γ as the distilling temperature parameter [3]. Intuitively, we obtain such an AD likelihood map as a set of scores in highlighting which prototype most resembles the clinical information in question. Given a 3D sMRI scan $\mathbf{X}$, we infer the latent features $\mathbf{z} \in \mathbb{R}^\ell$ through a convolutional-based feature extractor $\mathcal{E}_\mathbf{X}$ as $\mathbf{z} = \mathcal{E}_\mathbf{X}(\mathbf{X})$. We then estimate the AD likelihood map $\widetilde{\boldsymbol{\rho}} \in \mathbb{R}^K$ via a feed-forward networks $\mathcal{F}$ as $\widetilde{\boldsymbol{\rho}} = \sigma(\mathcal{F}(\mathbf{z}))$, with σ denoting a sigmoid activation function. To ensure the inherent agreement between pseudo and the estimated AD likelihood maps, we employ a lightweight, pretrained encoder $\mathcal{E}_{\boldsymbol{\rho}}$ which is priorly optimized via $\sum_{n=1}^N \|\boldsymbol{\rho}_n - \mathcal{D}_{\boldsymbol{\rho}}(\mathcal{E}_{\boldsymbol{\rho}}(\boldsymbol{\rho}_n))\|_2^2$. We further utilize the likelihood map $\widetilde{\boldsymbol{\rho}}$ as subsidiary features over 3D sMRI scan for various downstream diagnostic tasks in estimating the predicted label through $\widetilde{\mathbf{y}} = \mathcal{C}(\mathcal{E}_{\boldsymbol{\rho}}(\widetilde{\boldsymbol{\rho}}))$. Finally, the entire networks are optimized through a composite loss that evaluates the element-wise estimation error $\mathcal{L}_{\text{Est}}$ and topological consistency of the inherent agreement $\mathcal{L}_{\text{Cons}}$ between two paired likelihood maps (both using MSE and MAE). In addition, it also incorporates task-dependent loss $\mathcal{L}_{\text{Task}}$ for the respective downstream task, *i.e.*, cross-entropy for the clinical stage classification and MSE for the cognitive score and age prediction.

3 Experimental Results

Dataset and Data Preprocessing. We conducted exhaustive experiments by utilizing first-visit (baseline) 3D brain sMRI scans on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [12] with a cohort comprised of 1,540 first-visit subjects in total. Each T1-weighted 3D brain sMRI sample in such a cohort was coupled with the corresponding clinical information, containing a clinical-stage, a mini-mental state examination (MMSE) score, and an age. We split the cohort using stratified five-fold cross-validation. Furthermore, we applied a series of pre-processing procedures upon 3D sMRI scans [5, 4], obtaining the pre-processed images, each with a dimension of $96 \times 114 \times 96$.

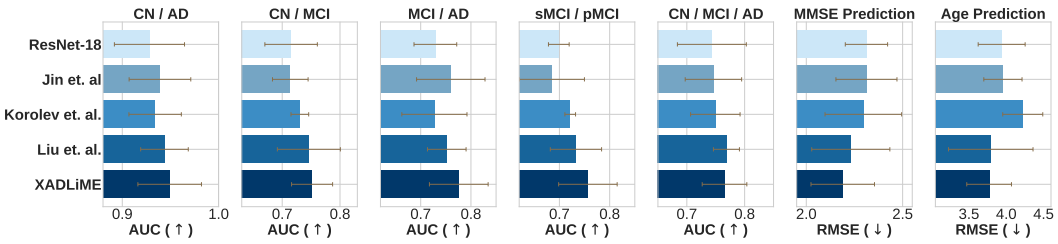
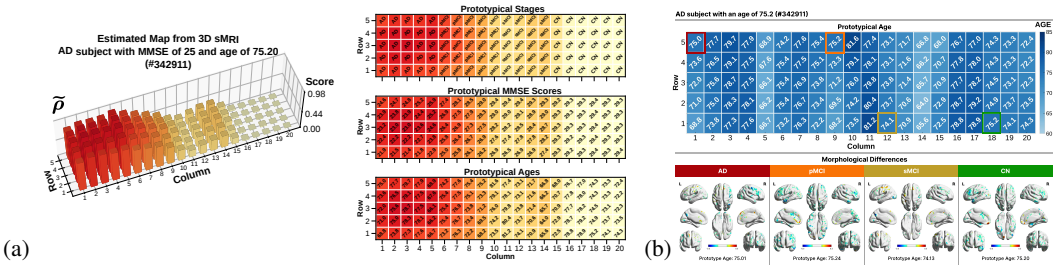


Figure 2: Experimental results on clinical stage classification, MMSE score, and age prediction. CN: cognitively normal, MCI: mild cognitive impairment, sMCI: stable MCI, pMCI: progressive MCI.



(a) (b)

Figure 3: AD likelihood map of an AD subject from (a) clinical and (b) morphological viewpoints.

Ablation Studies We conducted ablation studies on the establishment of clinically-guided prototypes with the number of prototypes $K = \{64, 100\}$ and their topological arrangements {1D, 2D, 3D}. Our XADLiME achieved the highest performance with 5×20 prototypes. We argue that a 2D grid-like topology with a sufficient number of prototypes exhibited a favorable mechanism in discovering a set of prototypes across the AD spectrum while accommodating the intra-stage variations as meaningful information to estimate the AD likelihood map and draw clinical outcomes.

Downstream Clinical Tasks. We evaluated the effectiveness of XADLiME compared to several baselines [6, 9, 11, 2] in carrying out a series of downstream clinical tasks, including clinical stage classification, MMSE score, and age prediction. Our proposed XADLiME considerably achieved greater classification and prediction performances in almost all scenarios (Fig. 2). Thereby, we verified that the estimated AD likelihood map accomplished its role as the concise substitute over the 3D sMRIs and provided discriminative features for drawing more accurate clinical verdicts.

Explainability Analysis. We ensured that our ADPEN effectively discovered the AD progression manifold by establishing the prototypes with a 5×20 2D topological arrangement (Fig 1b). Furthermore, we magnified the explainability from the clinical perspective of the estimated AD likelihood map (Fig. 3a left) by merging it with the prototypical clinical states (Fig. 3a right). This comprehensive (heat) map provided beneficial highlights in explaining the existing clinical conditions of a subject. In addition, as each prototype enclosed a tuple of clinical information, we were able to retrieve the nearest respective values over the training samples and eventually obtain the corresponding sMRI scans (*i.e.*, prototypical brains). We further analyzed the morphological changes map by subtracting the unseen sMRI sample from those sets of selected prototypical brains, treating the differences in cortical thickness as morphological changes (Fig. 3b).

Longitudinal Analysis. We performed an additional feasibility investigation of our XADLiME in real-world clinical applications in estimating the AD likelihood map from a single subject by utilizing a series of longitudinal 3D sMRI scans throughout his/her follow-up years of visits. Here, we treated such scans as the testing data without additional training over such longitudinal data. We illustrated a longitudinal study case of a healthy subject who progressed towards AD during their clinical visits in Fig. 4.

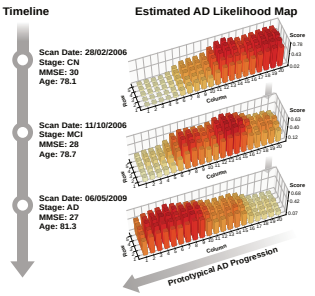


Figure 4: Estimated AD likelihood maps on longitudinal data.

4 Conclusion

We proposed XADLiME as a novel, explainable predictive framework for modeling AD progression assisted by clinically-guided prototypes. Such prototypes were established in an AD spectrum manifold through our proposed ADPEN. We further employed a deep estimation model to transfigure 3D sMRI into an estimated AD likelihood map. Performance comparison on the downstream clinical tasks over the ADNI cohort demonstrated the effectiveness of our XADLiME in contrast with existing diagnostics methods while simultaneously providing ease of interpretation as an extra benefit.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) No. 2022-0-00959 ((Part 2) Few-Shot Learning of Causal Inference in Vision and Language for Decision Making) and No. 2019-0-00079 (Artificial Intelligence Graduate School Program (Korea University)). This research was further supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2022R1A4A1033856).

Potential Negative Societal Impacts

We employed prototype-based networks to address the AD progression modeling through our proposed framework. As the currently utilized cohort of ADNI went through the de-identification procedures, thus, the usage of the proposed framework will not raise any issues. Despite that, we should emphasize that prototype learning picked a set of clinical-related data from representative patients, which might potentially violate the patient's privacy and sensitive information in unobserved scenarios in real-world clinical applications. Thus, this particular aspect regarding cohort preparation should undergo special attention concerning privacy-preserving for patients.

On the other hand, by coining the term prototypical brains (briefly mentioned in explainability analysis), we refer to those sets of sMRI scans that can be considered as the representative samples in an AD spectrum manifold, reflecting a group of subjects given a population. Through these prototypical brains, we could further analyze the morphological changes map, which complements the clinical AD likelihood maps offered by our proposed XADLiME framework. Implicitly, we could regard the subject with the closest distances in terms of their clinical and morphological similarities to the selected prototypical brains shall be treated by means of equivalent medical care. In the real clinical setting, however, a thorough individual medical check-up shall still be administered carefully.

References

[1] Matthew Davis, Thomas O Connell, Scott Johnson, Stephanie Cline, Elizabeth Merikle, Ferenc Martenyi, and Kit Simpson. Estimating Alzheimer's disease progression rates from normal cognition through mild cognitive impairment and stages of dementia. *Current Alzheimer research*, 15(8):777–788, 2018.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[3] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv:1503.02531*, 2015.

[4] Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, Heinz-Peter Schlemmer, Sabine Heiland, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein, and Philipp Kickingereder. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, 40(17):4952–4964, 2019.

[5] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782 – 790, 2012.

[6] D. Jin, J. Xu, K. Zhao, F. Hu, Z. Yang, B. Liu, T. Jiang, and Y. Liu. Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration. In *2019 IEEE 16th International Symposium on Biomedical Imaging*, pages 1047–1051, 2019.

[7] Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *Proceedings of International Conference on Learning Representation*, 2014.

[8] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[9] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova. Residual and plain convolutional neural networks for 3D brain MRI classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging*, pages 835–838, 2017.

[10] Oscar Li, Hao Liu, Chaofan Chen, and Cynthia Rudin. Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[11] Sheng Liu, Chhavi Yadav, Carlos Fernandez-Granda, and Narges Razavian. On the design of convolutional neural networks for automatic detection of Alzheimer's disease. In *Proceedings of the Machine Learning for Health NeurIPS Workshop*, volume 116, pages 184–201, 2020.

[12] Susanne G. Mueller, Michael W. Weiner, Leon J. Thal, Ronald C. Petersen, Clifford Jack, William Jagust, John Q. Trojanowski, Arthur W. Toga, and Laurel Beckett. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869 – 877, 2005.

[13] Ahmad Wisnu Mulyadi and Heung-Il Suk. ProtoBrainMaps: Prototypical brain maps for Alzheimer's disease progression modeling. *Medical Imaging with Deep Learning*, 2021.

[14] Raphaël Sivera, Hervé Delingette, Marco Lorenzi, Xavier Pennec, and Nicholas Ayache. A model of brain morphological changes related to aging and Alzheimer's disease from cross-sectional assessments. *NeuroImage*, 198:255 – 270, 2019.

[15] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[16] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, and The Alzheimer's Disease Neuroimaging Initiative. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function*, 220(2):841–859, 2015.

[17] Qingyu Zhao, Zixuan Liu, Ehsan Adeli, and Kilian M. Pohl. Longitudinal self-supervised learning. *Medical Image Analysis*, 71:102051, 2021.