Patch-based Intuitive Multimodal Prototypes Network (PIMPNet) for Alzheimer's Disease classification

Lisa Anita De Santi [1,2], Jörg Schlötter [3,4], Meike Nauta [5], Vincenzo Positano [2], Christin Seifert [3]

[1] Department of Information Engineering, University of Pisa, Pisa, Italy
[2] Fondazione Toscana G Monasterio - Bioengineering Unit, Pisa, Italy
[3] University of Marburg, Marburg, Germany
[4] University of Mannheim, Mannheim, Germany
[5] Datacation, Eindhoven, Netherlands

Introduction

Early detection of **Alzheimer's Disease** (AD) is crucial to mitigate the cognitive decline of affected patients, but **diagnosis** is still **challenging**. This has raised interest in supporting AD diagnosis with Deep Learning (DL) models [1]. Diagnostic guidelines often integrate clinical evaluation with **structural Magnetic Resonance Imaging** (sMRI), such AD subjects typically report **pathological brain patterns** like grey matter atrophy. However, information collected from sMRI should be interpreted together with the patient's **age**, as there are anatomical **brain changes due** to the physiological **ageing process**[2].

DL might facilitate the analysis of sMRI, identify unconventional AD subtypes, and extract yet unknown biomarkers[3], but their black-box nature poses controversy in high-stakes scenarios[4]. **Prototypical-Part (PP) networks** combine the advantages of DL models in an interpretable architecture and have collected interesting performances in medical imaging[5].

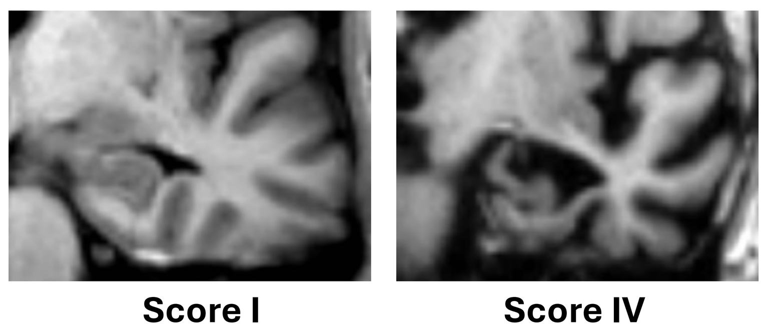

Fig. I – Medial Temporal Atrophy in sMRI assessed with Scheltens scale. From https://alzimaging.com/

However, existing PP models cannot be directly applied to interpret sMRI with patients' demographics to discern age-related image alteration from pathological ones, and effectively combining non-image prototypes (**ps**) to the standard PP architecture is a non-trivial task, such that there are no unique strategies available[6].

We present **PIMPNet**, the first **multi-modal prototype classifier** which learns prototypical **3D sMRI** image **patches** and **age values** to predict **AD** cognitive decline.

Method

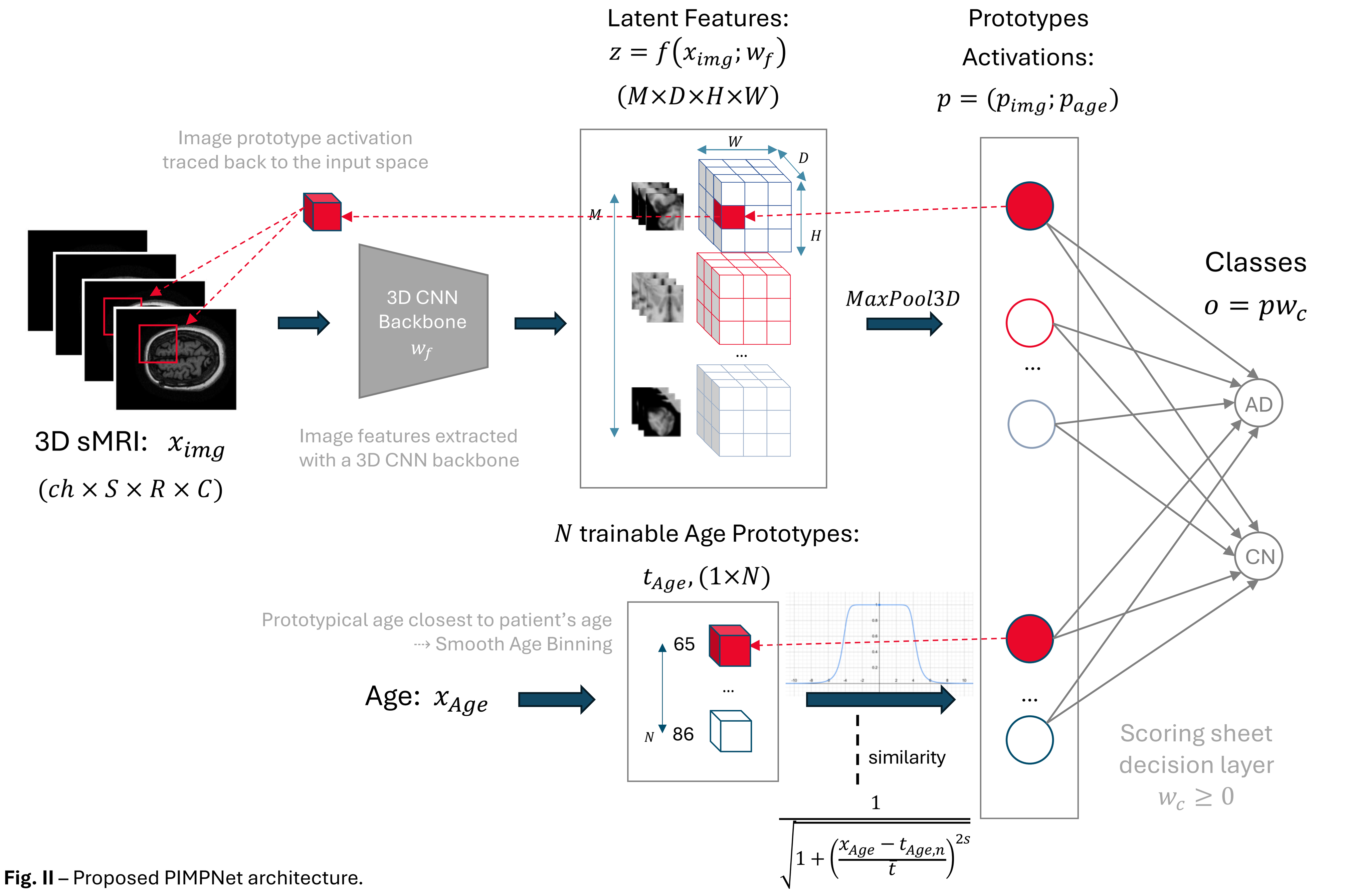

Fig. II – Proposed PIMPNet architecture.

Age-prototypes Layer: Computes similarity between input age and every age prototype and selects the most similar age prototype.

Why trainable age values?

- Relevant ages might not be equally distributed and/or known in prior
- To not assign different age bins to patients of similar ages close to the bins' boundary (smooth age binning)

Training Process

I) Self-supervised pre-pretraining of image ps

Learn a semantically meaningful image representation independently from the downstream classification task.

$$\min_{(w_f)} \lambda_A \mathcal{L}_A + \lambda_T \mathcal{L}_T$$

II) PIMPNet Training

Learn meaningful age values and image representation for the downstream classification task.
Optimise classification performances.

$$\min_{(w_f, t_{age}, w_c)} \lambda_A \mathcal{L}_A + \lambda_T \mathcal{L}_T + \lambda_C \mathcal{L}_C$$

Where:
- $\mathcal{L}_A$: Alignment Loss, $\mathcal{L}_A(w_f)$
- $\mathcal{L}_T$: Tanh-Loss, $\mathcal{L}_T(w_f)$
- $\mathcal{L}_C$: Log-likelihood, Loss $\mathcal{L}_C(w_f, t_{age}, w_c)$

Evaluation

We collected the *"ADNI1 Standardized Screening Data Collection for 1.5T"* sMRI and the corresponding ages from the Alzheimer's Disease Neuroimaging Initiative (ADNI) obtaining 307 Cognitively Normal (CN) and 243 AD subjects.

- The **ResNet-18 3D** backbone **performs better** than ConvNeXt-tiny 3D.
- Learned **age ps** do **not improve** classification **results** in both cases.

Table I – **Performance** comparison (Average ± Std Dev over 5 folds) of PIPNet and PIMPNet.

Model	Acc	Bal Acc	SENS	SPEC	F1
PIPNet (3D sMRI only)					
ResNet-18 3D	83 ± 04	83 ± 04	86 ± 06	79 ± 07	81 ± 05
ConvNeXt-Tiny 3D	65 ± 12	66 ± 09	56 ± 32	76 ± 15	66 ± 05
PIMPNet (3D sMRI + Age)					
ResNet-18 3D	84 ± 04	83 ± 04	89 ± 03	77 ± 08	81 ± 05
ConvNeXt-Tiny 3D	72 ± 04	70 ± 04	86 ± 10	55 ± 14	63 ± 09

Accuracy (**Acc**)
Balanced Accuracy (**Bal Acc**)
Sensitivity (**SENS**): CN Acc
Specificity (**SPEC**): AD Acc
F1 score (**F1**)

Table II – Learned **age ps** t_{age} in five different folds (denoted as Mx, x = current fold).

Fold	$t_{Age,1}$	$t_{Age,2}$	$t_{Age,3}$	$t_{Age,4}$	$t_{Age,5}$	$t_{Age,1}$	$t_{Age,2}$	$t_{Age,3}$	$t_{Age,4}$	$t_{Age,5}$
	ResNet-18 3D					*ConvNeXt-Tiny 3D*				
M1	65.77	65.81	66.14	76.81	80.99	56.81	65.00	64.96	74.13	85.80
M2	68.46	69.40	70.38	77.04	82.38	55.75	58.39	64.96	74.32	85.59
M3	66.37	67.27	67.91	75.87	81.96	54.86	56.63	65.21	74.40	85.11
M4	66.72	66.72	67.07	77.07	79.75	58.22	58.59	66.50	75.88	89.09
M5	66.51	66.52	67.23	77.37	80.00	57.79	66.94	65.44	72.55	84.58

- Image ps are generally **consistently located** in the same anatomical brain regions (low LCp). The CNet is **more compact** (lower GS, LS, higher Sp) and its ps are **purer** (lower Hp) than RNet. This higher purity is related to a **higher** percentage of **background voxels** included in the image ps (clinically-irrelevant regions).

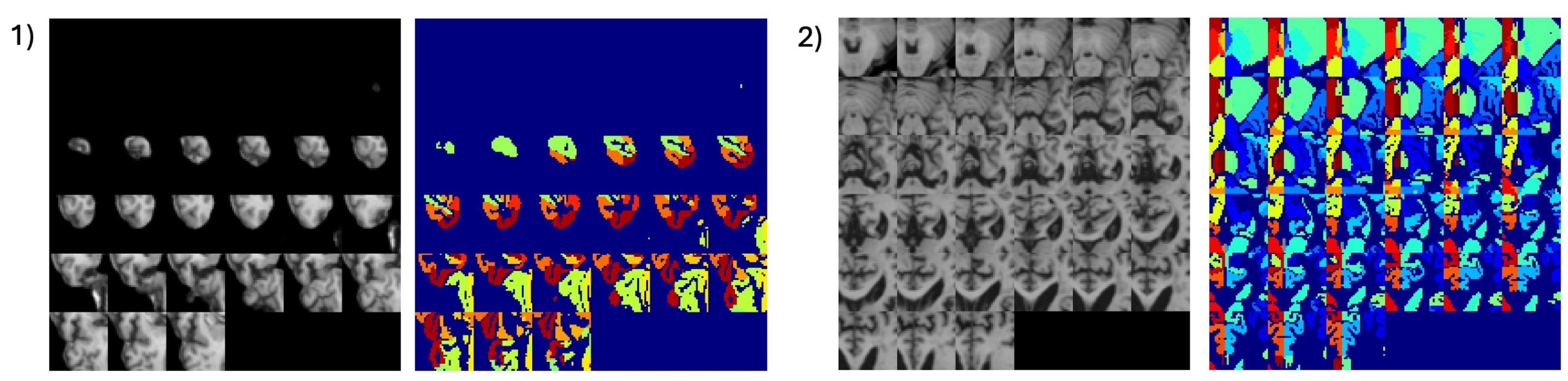

Fig. III – Examples of brain ps and corresponding CerebrA regions with different % of background included (higher (1), lower (2)).

Table III - Functionally grounded **metrics** of **explainability**. ↑ and ↓: tendency for better value.

Model	GS ↓	LS ↓	Sp ↑	LC_p ↓	H_p ↓
ResNet-18 3D					
PIPNet	149 ± 18	73 ± 10	0.855 ± 0.018	0.008 ± 0.006	2.474 ± 0.249
PIMPNet	143 ± 35	74 ± 20	0.861 ± 0.033	0.006 ± 0.006	2.424 ± 0.162
ConvNeXt-Tiny 3D					
PIPNet	4 ± 2	2 ± 1	0.997 ± 0.001	0.000 ± 0.000	1.803 ± 0.999
PIMPNet	10 ± 9	4 ± 4	0.993 ± 0.002	0.000 ± 0.000	1.543 ± 0.626

Global size (**GS**): #ps
Local size (**LS**): #ps in prediction
Sparsity (**Sp**) %zero-weights in decision layer
Ps Localization Consistency (**LC$_p$**): differences in the image ps center
Ps Brain Entropy (**H$_p$**): ps purity in terms of regions included w.r.t. the *CerebrA* brain atlas.

Conclusion

PIMPNet is the first PP network which performs an interpretable classification by detecting ps learned from different data modalities (3D images and ages).
In binary AD classification from 3D sMRI and age, age ps does not improve performance (3D sMRI only). This defines our future work directions:

- ***Model training*** *→ Include an* ***age ps pre-training*** *step w.r.t.* $\mathcal{L}_C$
- ***Model design*** *→ Do not simply concatenate image and age ps, but combine them using a different (still interpretable) classifier.*

[1] Ebrahimighahnavieh et al., Comp. Met. And Prog. In Biomed., 187 (2020)
[2] Zhao et al., Cerebral Cortex, 29 (2019)
[3] Böhle et al., Front. Aging Neurosci., 10 (2019)
[4] Longo et al., Information Fusion, 106 (2024)
[5] Nauta et al., Artificial Intelligence, ECAI 2023 Int. Workshop.
[6] Ma et al., Knowledge-based Systems, 250 (2022)

Contact: lisa.desanti@phd.unipi.it
desanti@monasterio.it