Rel-SA: Alzheimer's Disease Detection using Relevance-augmented Self Attention by Inducing Domain Priors in Vision Transformers

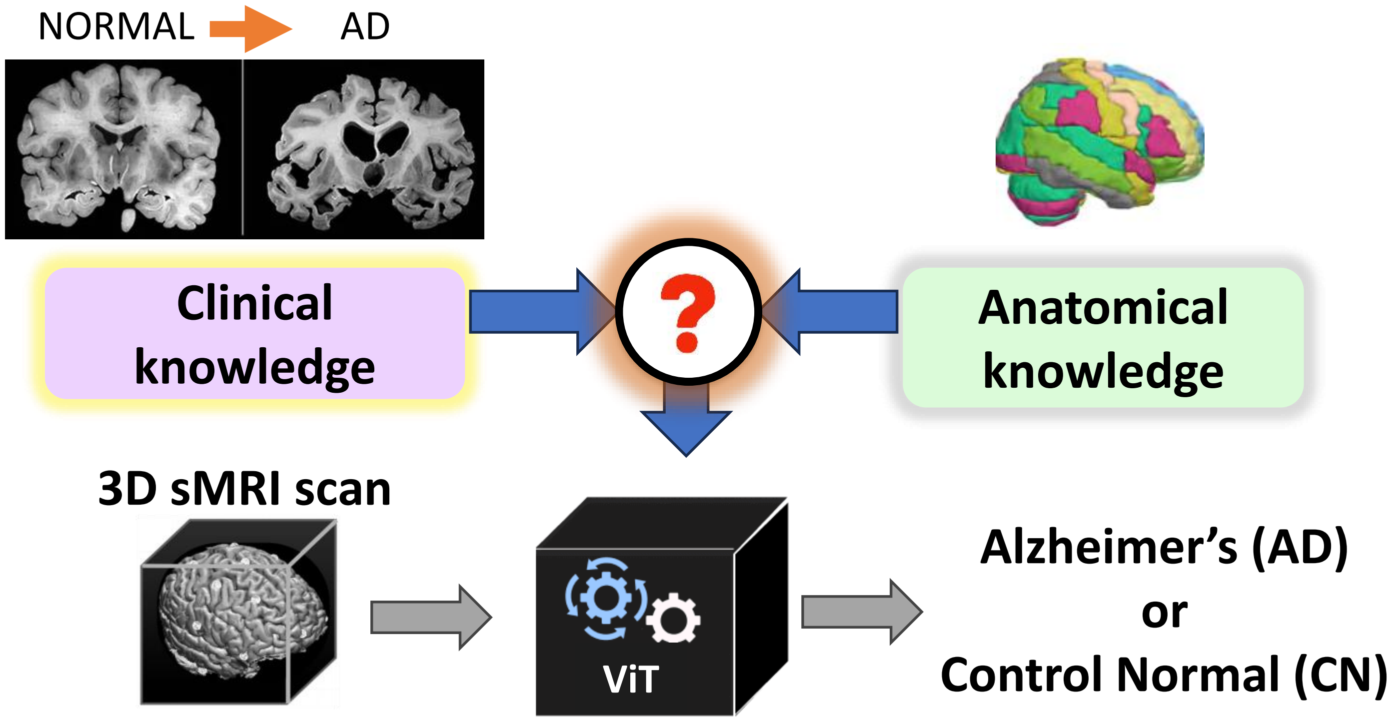
Madhumitha V*[,1] · Sunayna Padhye*[,1] · Shanawaj Sahebpatel Madarkar*[,1,2] ·
Susmit Agrawal[1] · Konda Reddy Mopuri[1]

[1] Indian Institute of Technology Hyderabad, India & [2] Indian Navy

* Equal Contributors

Motivation

Can we incorporate crucial domain-specific knowledge, both clinical and anatomical, as an inductive bias for ViT-based models to learn effectively?



Contributions

- **Relevance Augmented Self Attention (Rel-SA) module**, encoding clinical priors using Relevance Bias
- Unifying two clinically validated brain atlases: 1) AAL3v1 and 2) JHU White Matter atlas
- Qualitative Evaluation through**:**
  - **Leave-One-Out Analysis,** and
  - **Reverse Analysis**

Model Architecture Design

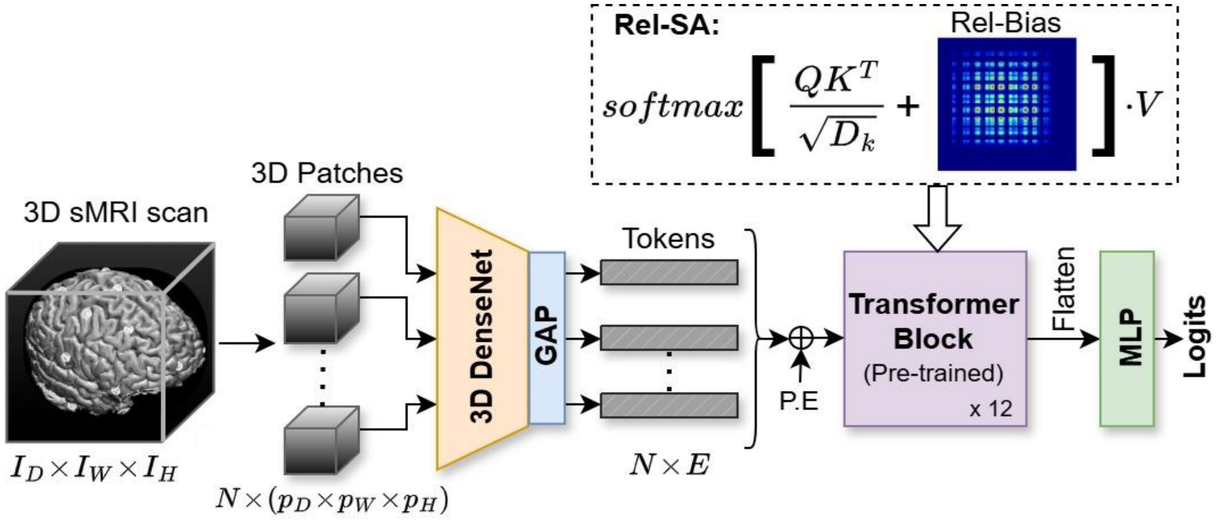


Fig 1: 3D Densenet+ViT Architecture with Rel-Bias in the Rel-SA module.

Rel-SA:

$$\boxed{Rel\text{-}SA(Q,K,V) = \sigma\left(\frac{QK^T}{\sqrt{d}} + \phi(M; w_{high}, \alpha)\right)V}$$

where, $Q = XW_Q, \mathrm{K} = XW_K, V = XW_V,$

$X \in \mathbb{R}^{N\times E},\ \ \phi(\cdot) \in \mathbb{R}^{N\times N}\colon Rel\text{-}Bias, \sigma(\cdot)\colon softmax$

Model Architecture Design

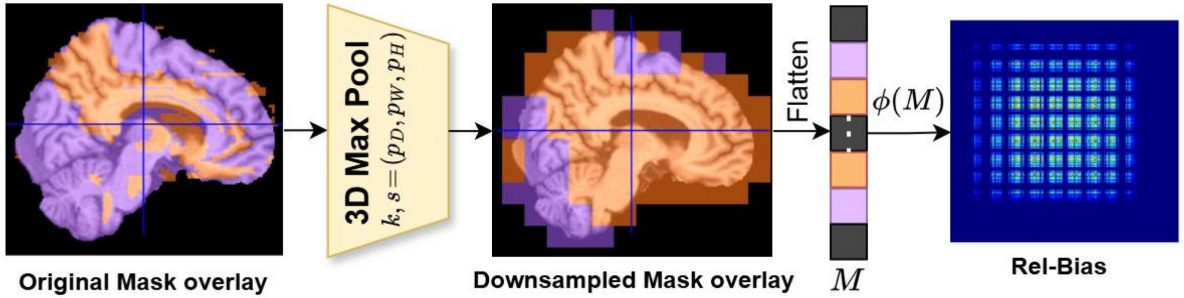


Fig 2: Rel-Bias calculation using relevance regions identified using Atlas.

Score vector:

$$S(M; w_{high}, \alpha) = w_{high} \cdot M_{high} + w_{low} \cdot M_{low},$$

$$w_{low} = \alpha \cdot w_{high}$$

where, $S \in \mathbb{R}^{N\times 1}, M \in \mathbb{R}^{N\times 1}, \alpha \in (0,1)$

Rel-Bias:

$$\boxed{\phi(M; w_{high}, \alpha) = \beta \cdot ||S(M)S(M)^T||_F}$$

where, β is a hyperparameter

Experiments

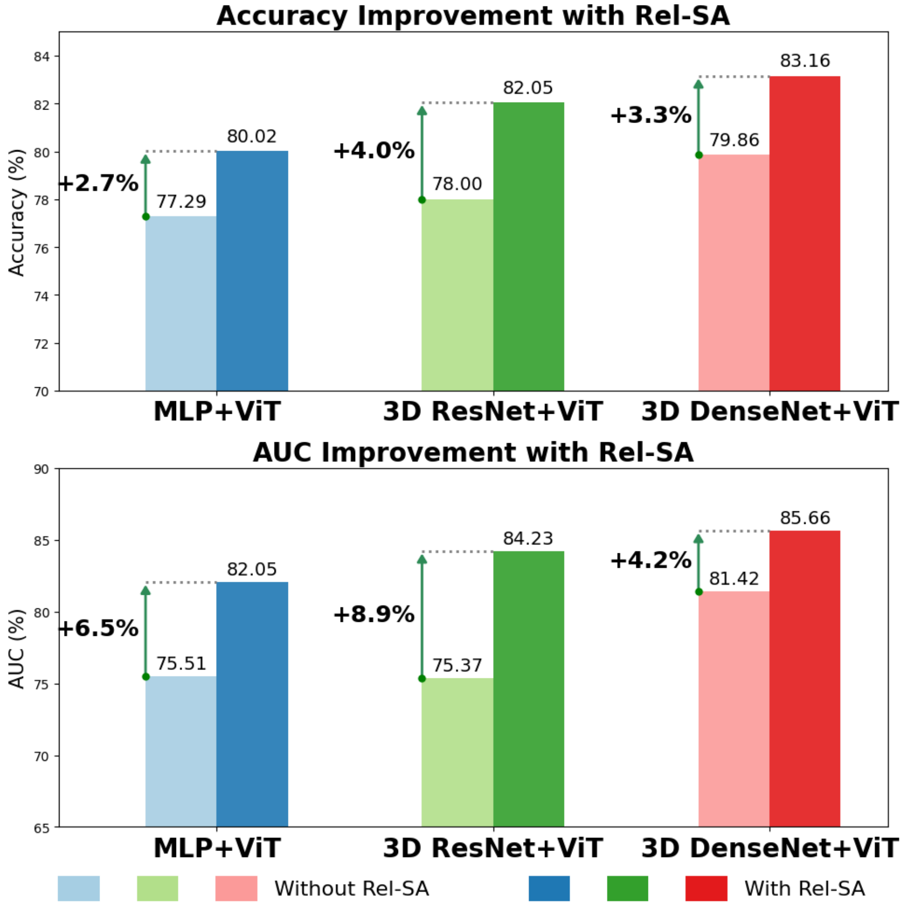


Fig 4: Performance analysis of ViTs w/ and w/o Rel-SA on ADNI dataset for AD-CN classification

Table 1: AD-CN classification analysis using Rel-SA on ADNI/AIBL

Dataset	Metric (%)	w/ Rel-SA	w/o Rel-SA
ADNI	Acc.	**83.16**$_{\pm 0.54}$	79.86$_{\pm 0.94}$
	AUC	**85.66**$_{\pm 0.87}$	81.42$_{\pm 1.83}$
AIBL	Acc.	**87.40**$_{\pm 2.58}$	82.81$_{\pm 1.55}$
	AUC	**83.57**$_{\pm 2.25}$	81.42$_{\pm 2.29}$

- Rel-SA's max boost to ViT-base: **Accuracy: ~4% & AUC: ~9%**
- Parametric overhead: just **24 additional parameters**!

Attention Rollout Visualizations

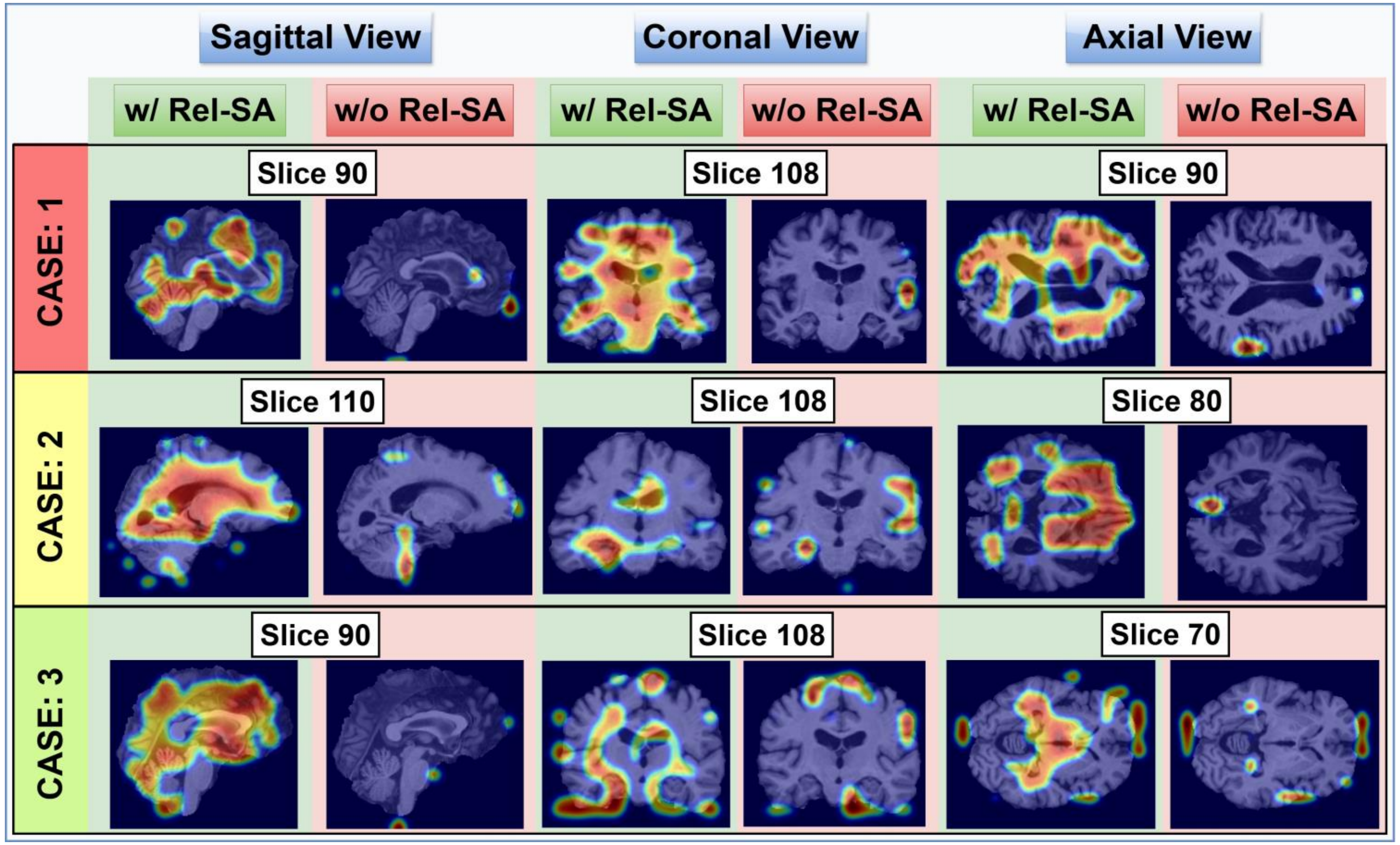


Fig 3: Case-wise Comparative Analysis of Attention Rollout visualizations. Case 1: GT = AD, both models predict AD; Case 2: GT = AD, Rel-SA predicts AD while vanilla ViT predicts CN; and Case 3: GT = CN, both models predict CN.

Post-hoc Analysis

Leave-One-Out Analysis

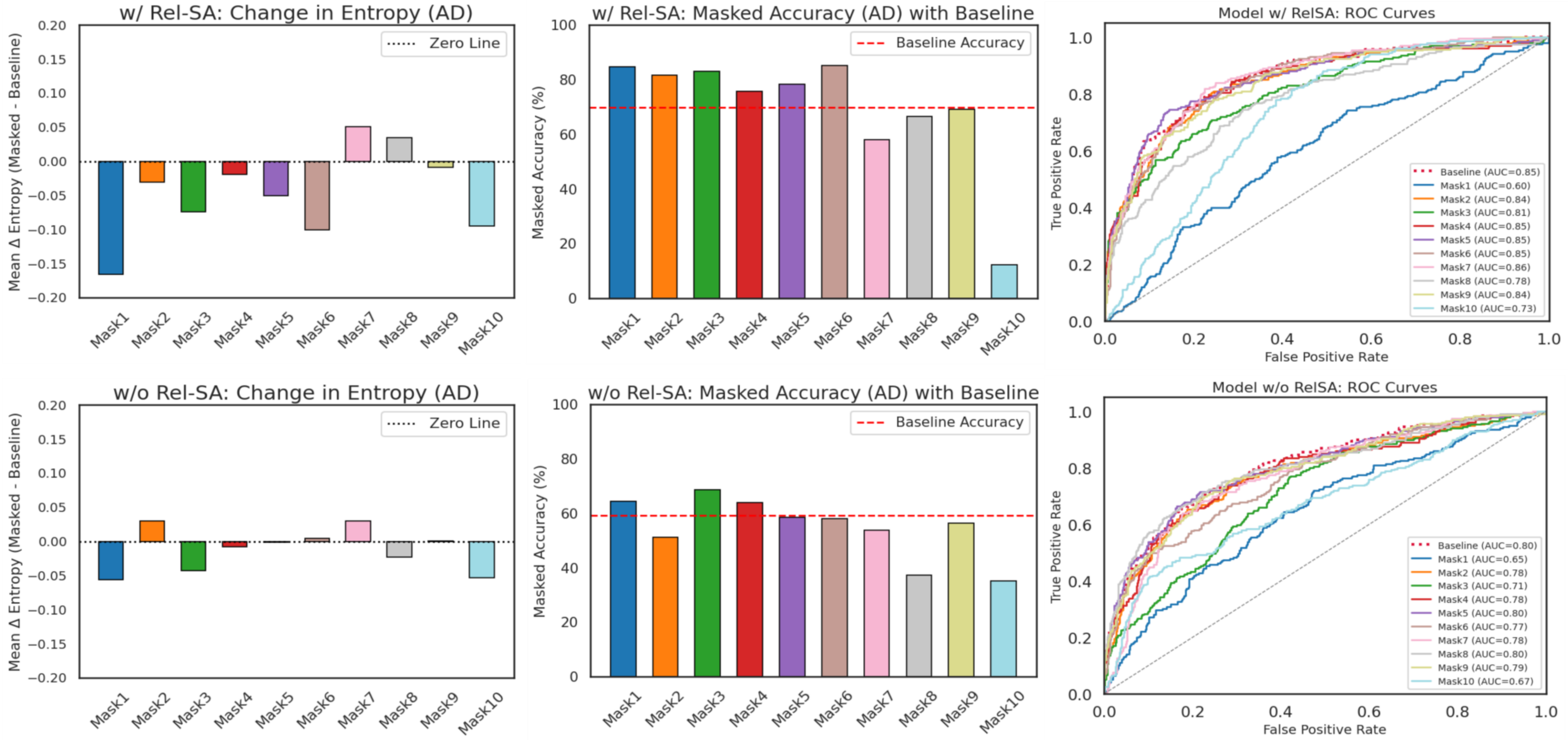


Fig 5: Leave-one-out Analysis of 3D Densenet+ViT model’s performance w/ and w/o Rel-SA

Post-hoc Analysis

Reverse Analysis

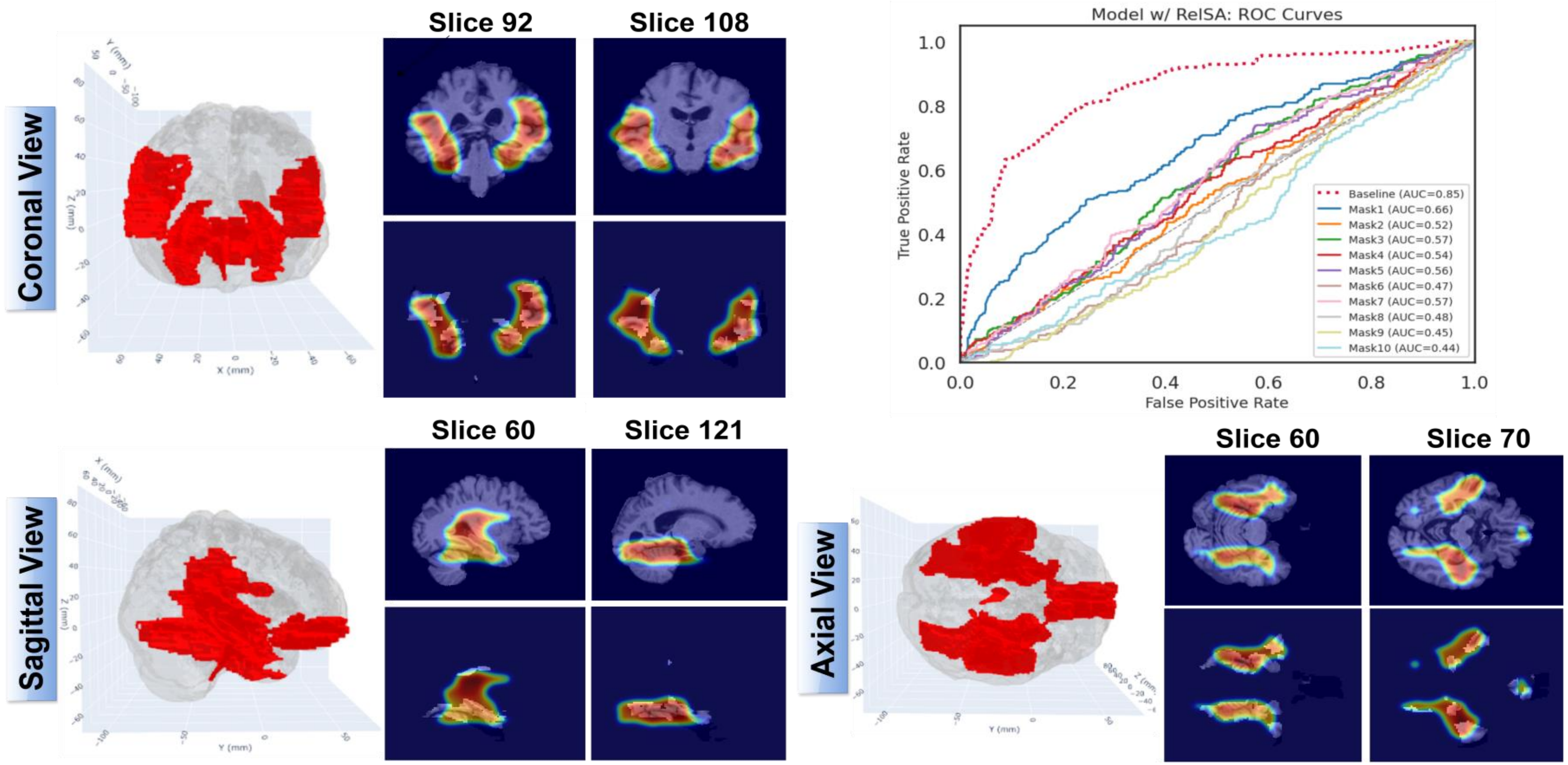


Fig 6: Reverse Analysis of 3D Densenet+ViT model's performance w/ and w/o Rel-SA using Mask1

Conclusion

- Our results depict that integrating clinical knowledge through Rel-SA can have a considerable impact both in terms of performance and faithful interpretations of the sMRI data consistent with Neuroscience literature.
- The key lies in leveraging domain knowledge as the right inductive bias.

Thank You

CONTACT:

Madhumitha V
ai23resch11004@iith.ac.in
https://madhuv29.github.io/

Dr. Konda Reddy Mopuri
krmopuri@ai.iith.ac.in
https://krmopuri.github.io/

PAPER

CODE