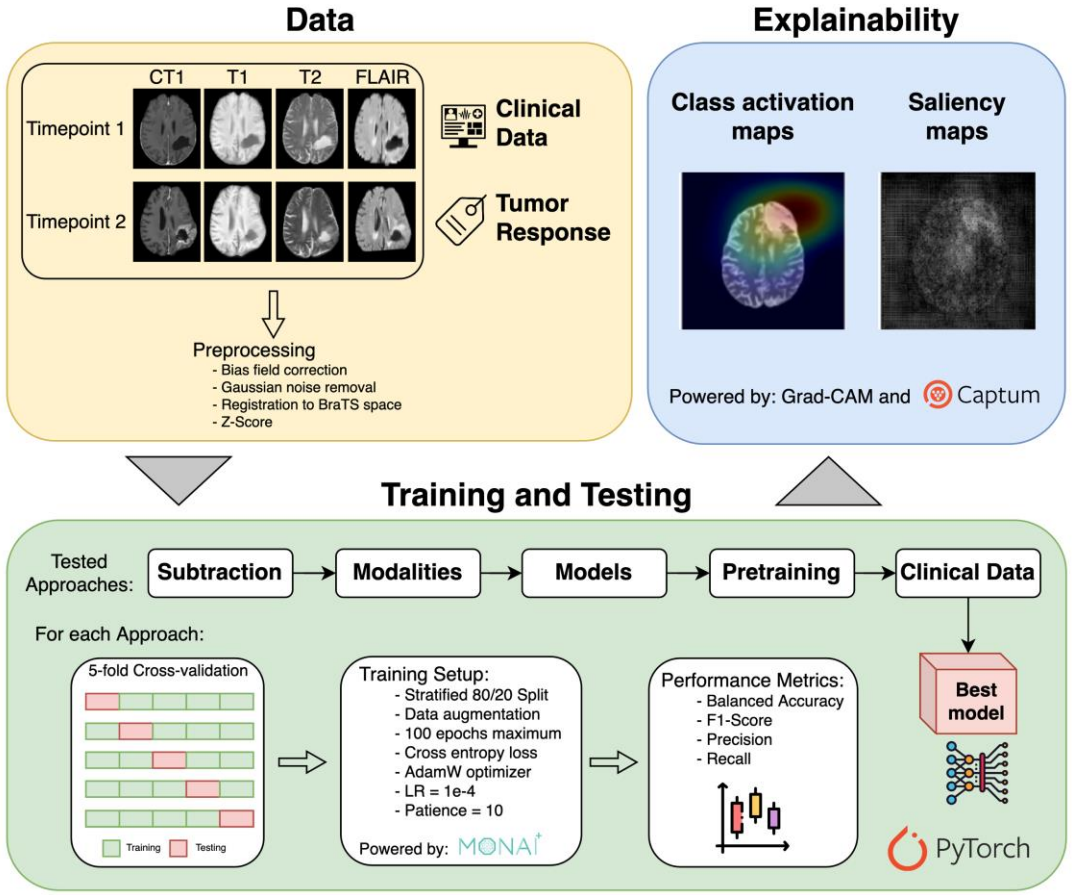
Institute for Systems and Robotics Lisboa
LaSEEB Evolutionary Systems Biomedical Engineering
HOSPITAL DA LUZ LEARNING HEALTH TRAINING, RESEARCH & INNOVATION CENTER
TÉCNICO LISBOA

Prediction of treatment response in a longitudinal glioblastoma dataset using deep learning

Ana Matoso, Catarina Passarinho, Marta P. Loureiro, José Maria Moreira, Patrícia Figueiredo, Rita G. Nunes



GOAL: To analyse and compare different **Deep Learning** approaches for **RANO criteria classification** based on two consecutive MRI acquisitions

Check out the preprint

Prediction of treatment response in a longitudinal glioblastoma dataset using deep learning

Ana Matoso[1*], Catarina Passarinho[1], Marta P. Loureiro[1], José Maria Moreira[2], Patrícia Figueiredo[1], Rita G. Nunes[1]

[1]Institute for Systems and Robotics – Lisboa and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Portugal;

[2]Hospital da Luz Learning Health, Luz Saúde, Lisboa, Portugal

*anamatoso@tecnico.ulisboa.pt

Declaration of Financial Interests or Relationships

Speaker Name: Ana Matoso

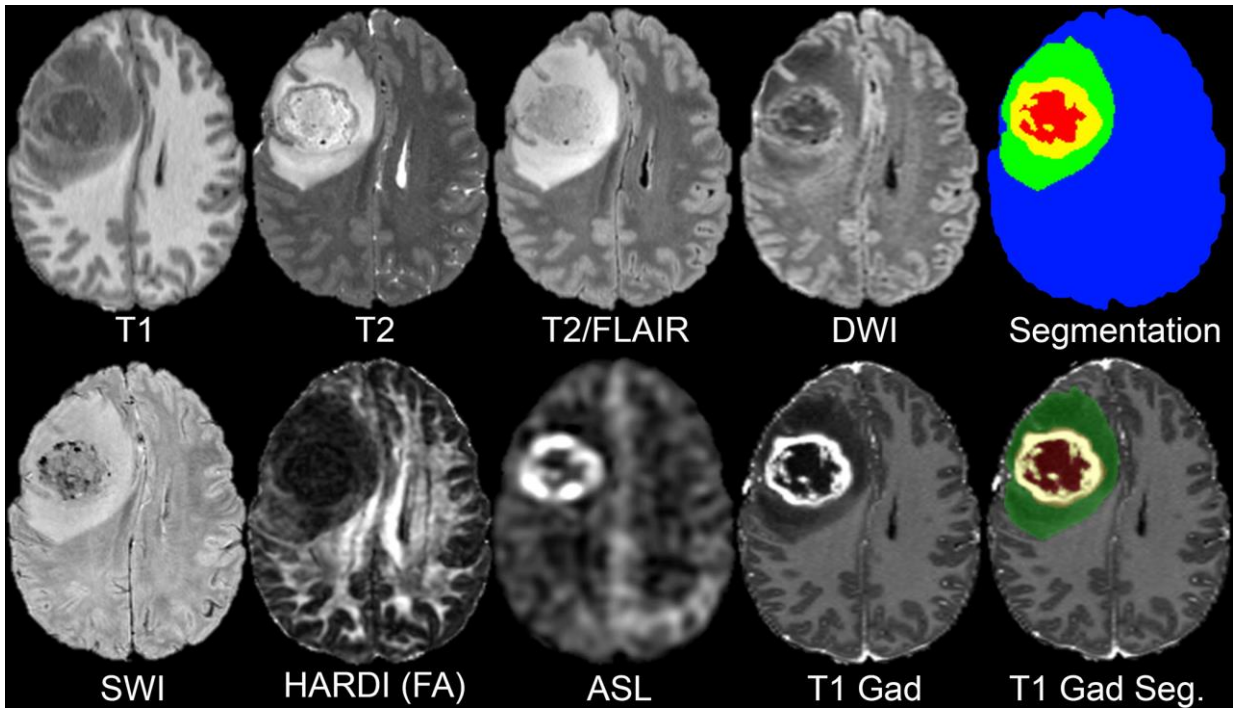
I have no financial interests or relationships to disclose with regard to the subject matter of this presentation.

Introduction

Glioblastoma

Glial cells Tumor

- Prevalence: ~3/100 000 per year
- Poor prognosis
  - Average survival: 9 months
  - 41% survival after 1 years
  - 13% survival after 2 years
- Frequent MRI scans to assess treatment



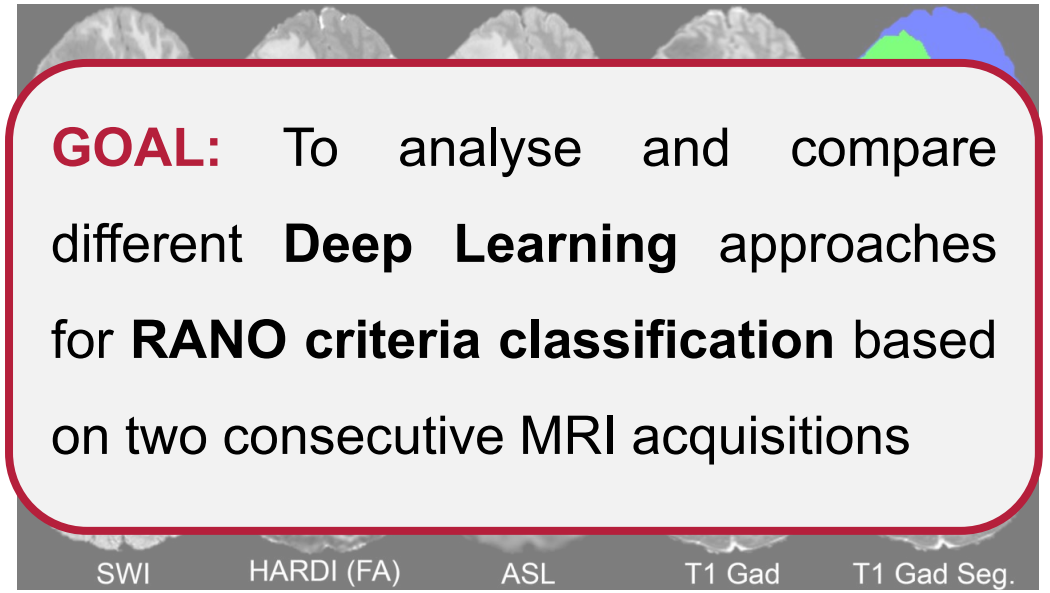
Glioblastoma on different MRI modalities

Nicholas F. Brown, et al. *Cancers*. 2022; Quinn T. Ostrom, et al. *Neuro-Oncology*, 2015; Evan Calabrese, et al. *Radiology: Artificial Intelligence* 2022.

Motivation

Response Assessment in Neuro-Oncology (RANO) criteria

- Complete Response ☺
- Partial Response
- Stable Disease
- Progressive Disease ☹



GOAL: To analyse and compare different **Deep Learning** approaches for **RANO criteria classification** based on two consecutive MRI acquisitions

	Complete Response	Partial Response	Stable Disease	Progressive Disease[a]
T1-Gd+	None	≥50% ↓	<50% ↓– <25% ↑	≥25% ↑*
T2/FLAIR	Stable or ↓	Stable or ↓	Stable or ↓	↑*
New lesion	None	None	None	Present*
Corticosteroids	None	Stable or ↓	Stable or ↓	NA
Clinical status	Stable or ↑	Stable or ↑	Stable or ↑	↓*
Requirement for response	All	All	All	Any*

Evan Calabrese, et al. *Radiology: Artificial Intelligence* 2022; Patrick Y. Wen, et al., *Journal of Clinical Oncology* 2010

Methods – Data

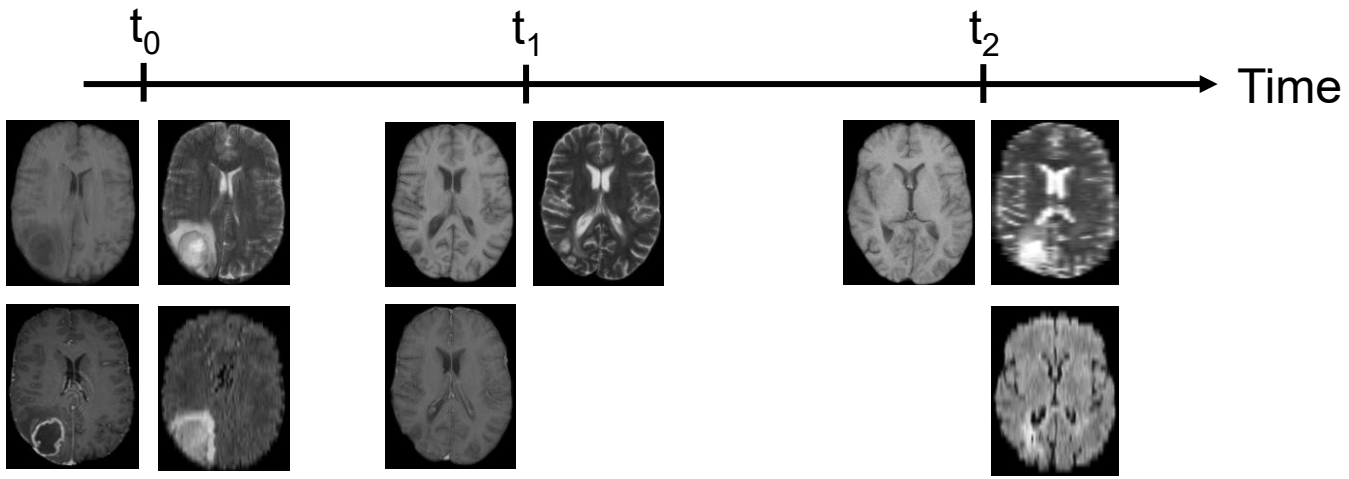
LUMIERE longitudinal dataset

- T1w
- CT1 (T1w contrast enhanced)
- T2w
- FLAIR
- Clinical Data
- RANO classification

638 timepoints

91 patients

Class	Prevalence
Progressive Disease (PD)	67%
Stable Disease (SD)	20%
Progressive Response (PR)	6%
Complete Response (CR)	7%

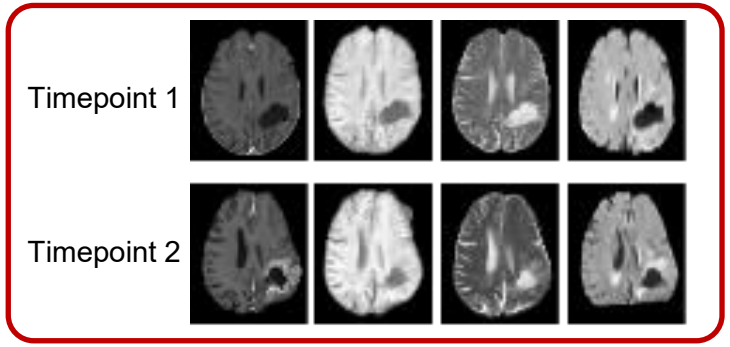


Yannick Suter, et al., Scientific Data Data, 2022

Methods – Pipeline

LUMIERE dataset

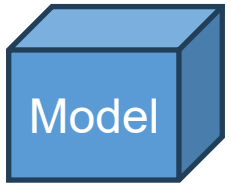
Data



5-fold Cross Validation
80/20 Stratified Split

Model Training

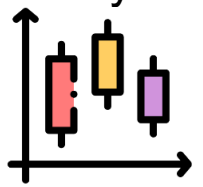
Weight Initialization



Training Setup:
- 100 Epochs
- Cross Entropy Loss
- AdamW Optimizer
- LR = 1e-4
- Patience = 10

Model Testing

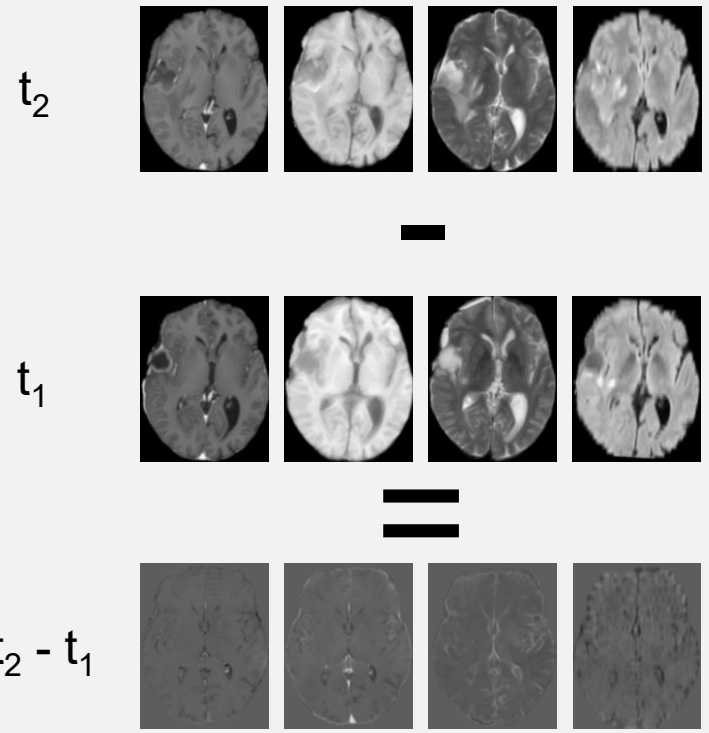
Performance Metrics:
- Balanced Accuracy
- F1-Score
- Precision
- Recall



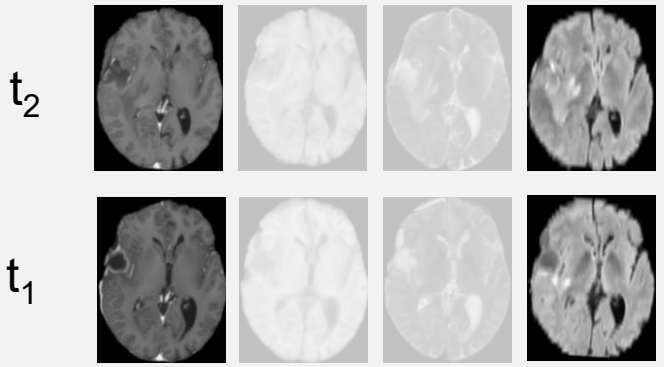
MONAI PyTorch

Methods – Tested Approaches

1. Subtraction of timepoints



2. Combinations of modalities



Combination of Modalities	Size of Dataset
CT1+T1+T2+FLAIR	337
CT1+FLAIR	344
T1+T2+FLAIR	338
CT1	355
T1+FLAIR	338

3. Model Architectures

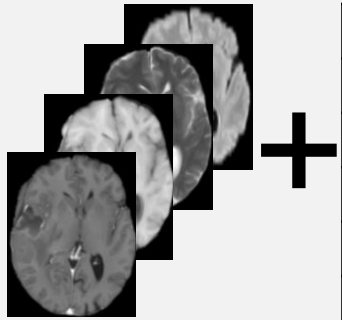
- DenseNets:
  - DenseNet 121
  - DenseNet 169
  - DenseNet 264
- Vision Transformer (ViT)
- AlexNet3D

Methods – Tested Approaches

4. Pretraining

- Self-Supervised Rotation Classifier
- MedMNIST Organ Classifier
- MedicalNet Segmentation Encoder

5. Use of Clinical Data



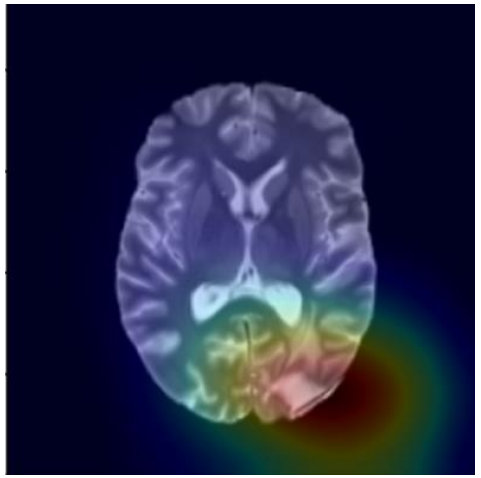
Clinical Data	Value
Age	66
Sex	M
IDH	WT
MGMT	F
Time from 1st scan	15w

Methods – Explainability

Class Activation Maps

⟶ Weighted Average of Feature Maps by the gradients

⟶ Coarse heatmap

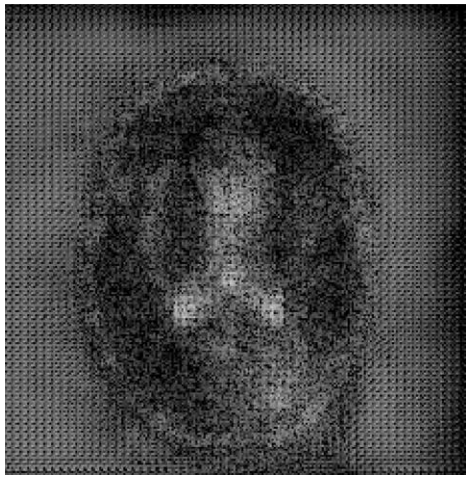


with: Grad-Cam package

Saliency Maps

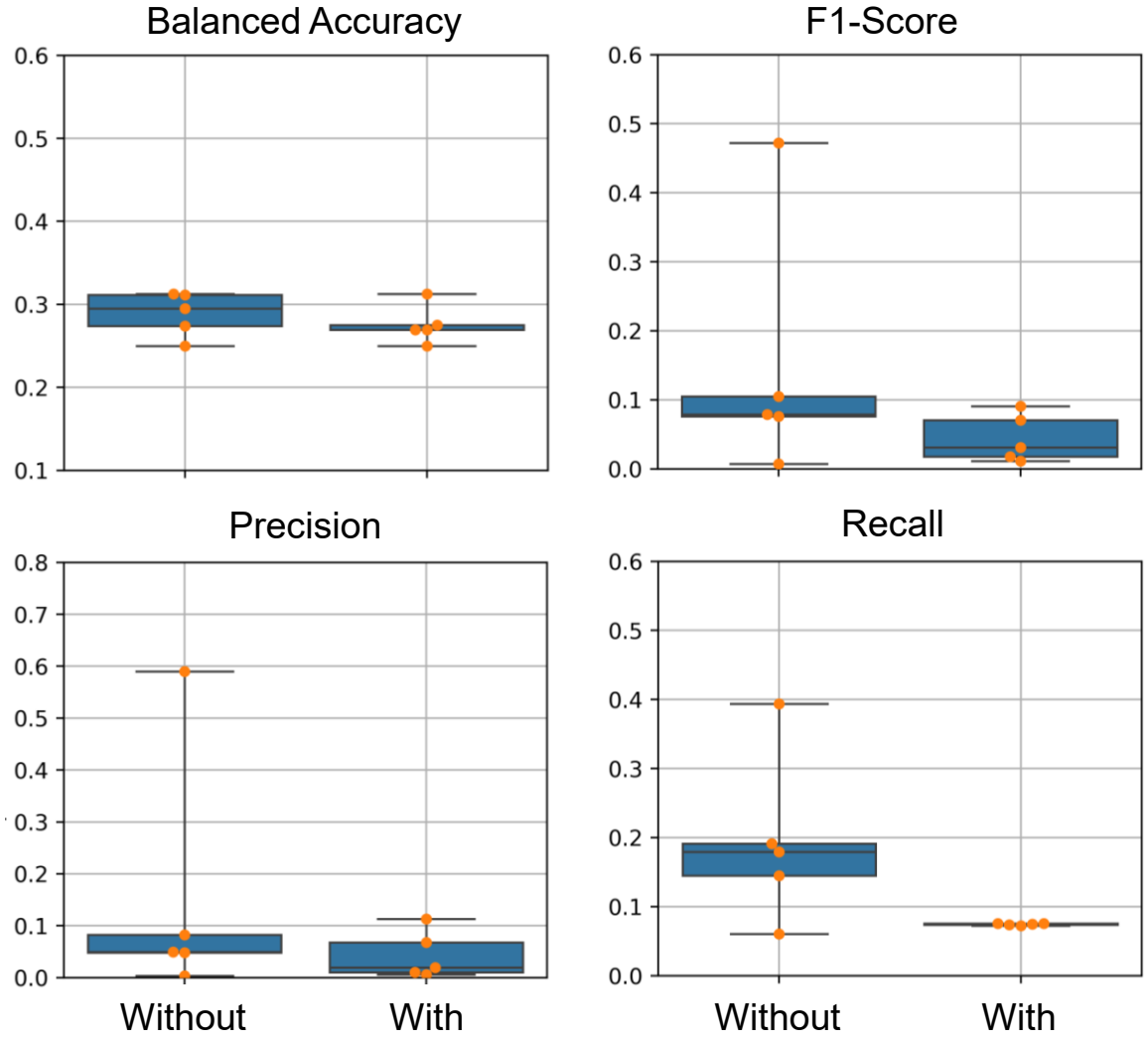
⟶ Gradients with respect to inputs

⟶ Granular impact of input

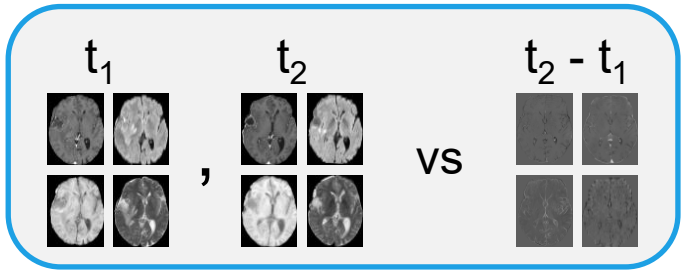


Captum

Results – Subtraction



Approach



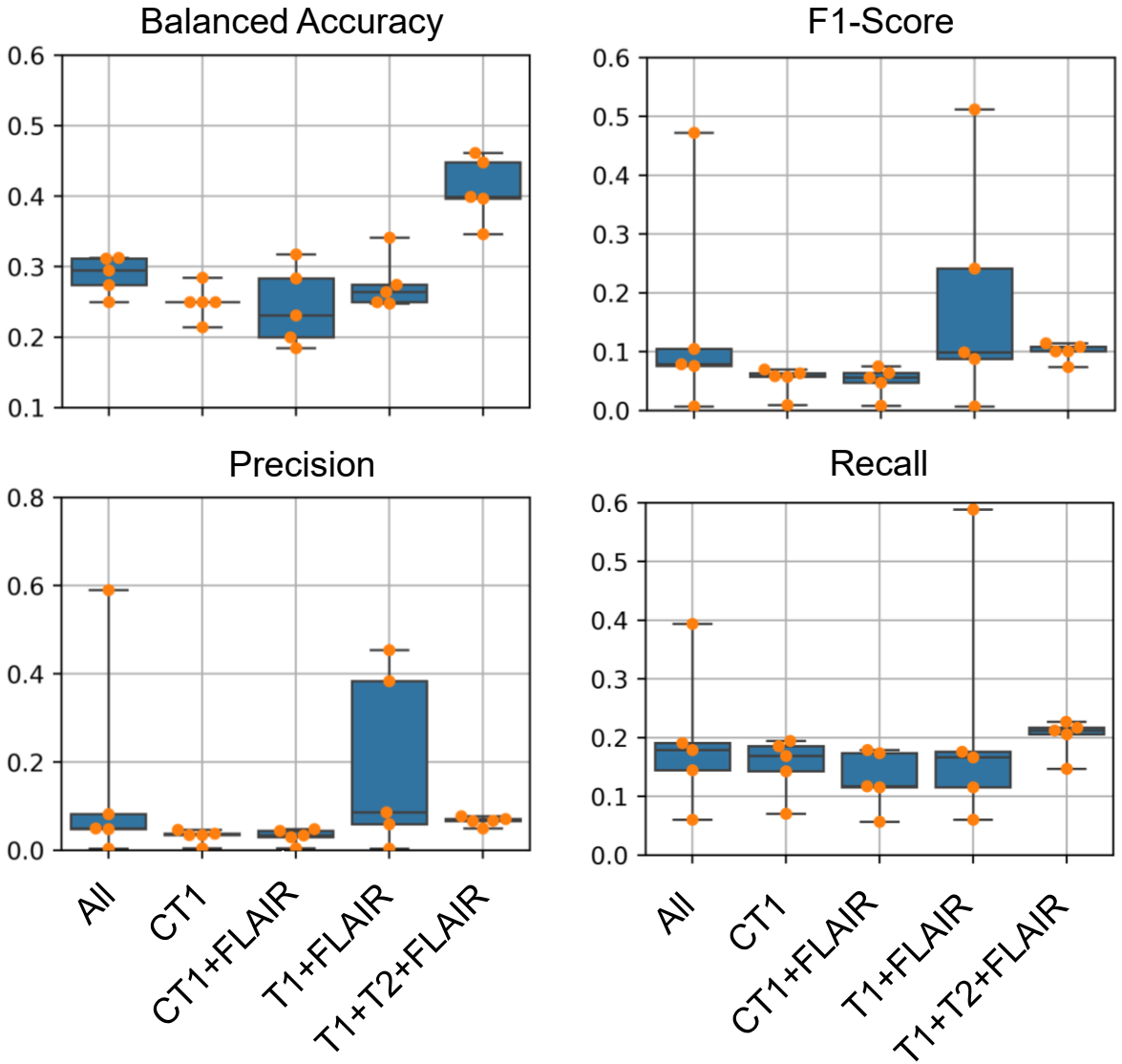
- Similar BA
- Slight decrease in Recall and Precision

⇩

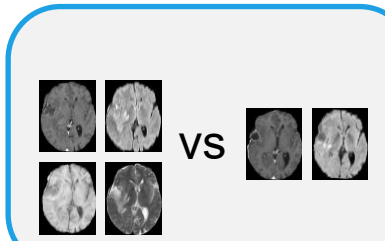
- Decrease in F1-Score

⟶ No subtraction was done in the next stages

Results – Modalities



Approach



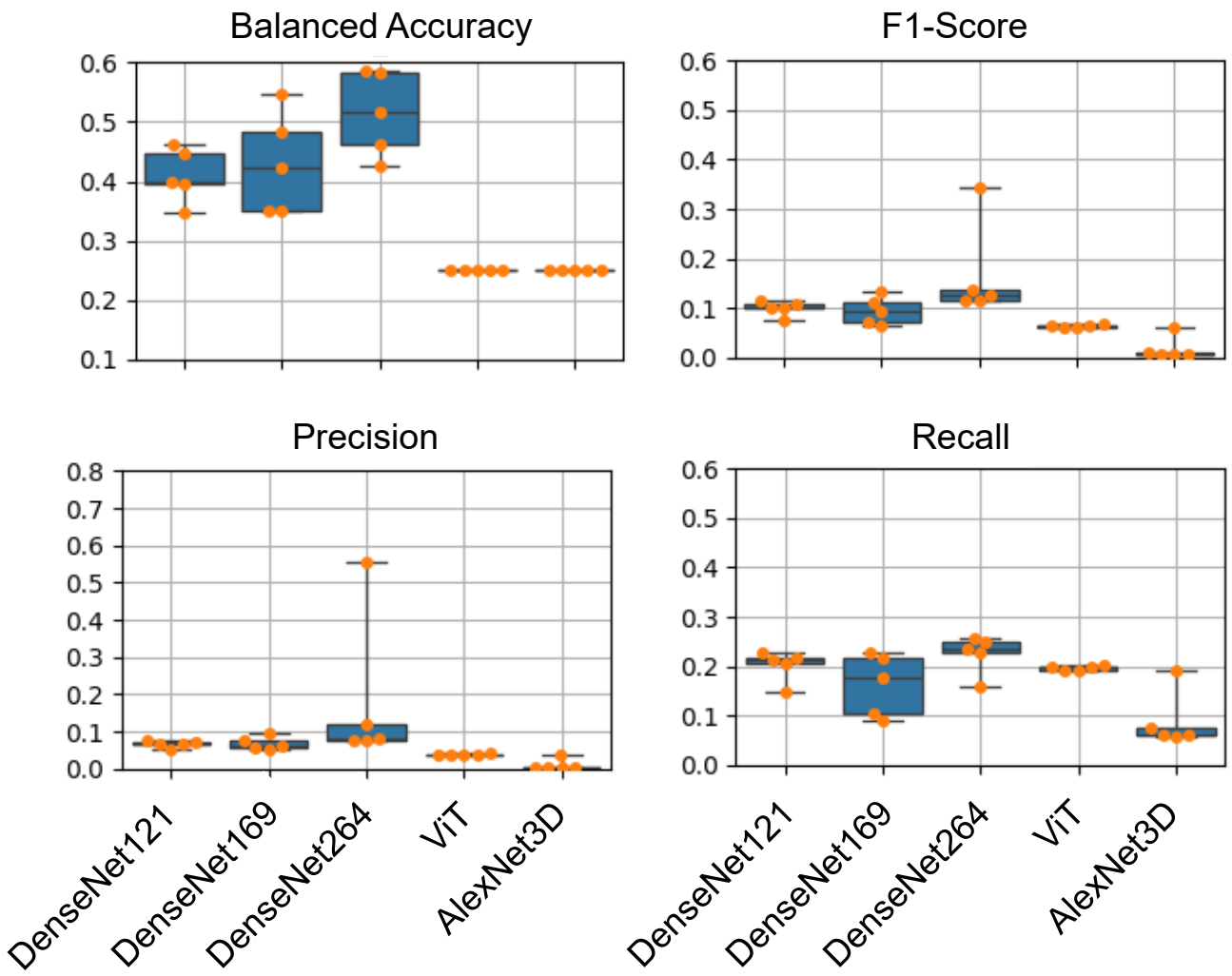
- Higher BA in T1+T2+FLAIR
- Higher Precision in T1+FLAIR

⇩

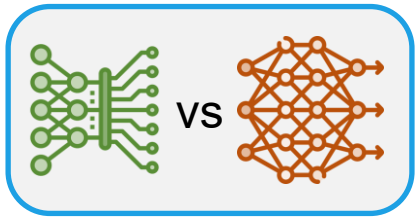
- Increased F1 Score in T1+FLAIR

⟶ The combination that uses T1 + T2 + FLAIR was used henceforth

Results – Architectures



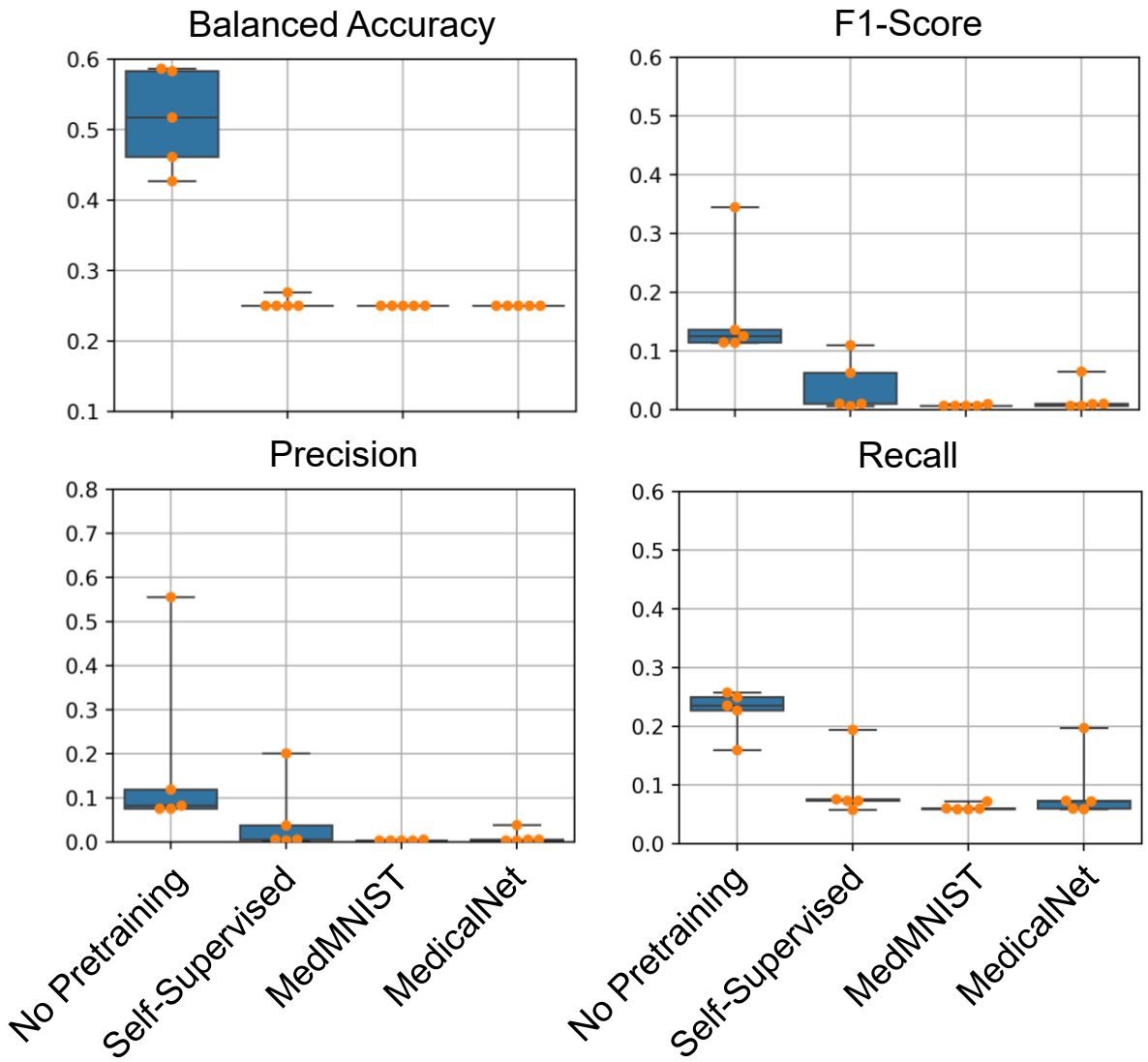
Approach



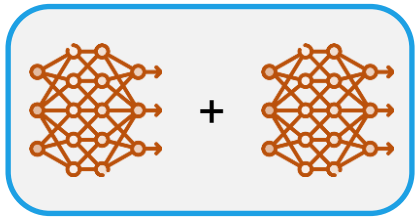
- DenseNets performed better than ViT and AlexNet3D
- More complex DenseNets improve performance

⟶ DenseNet264 has overall better performance

Results – Pretraining



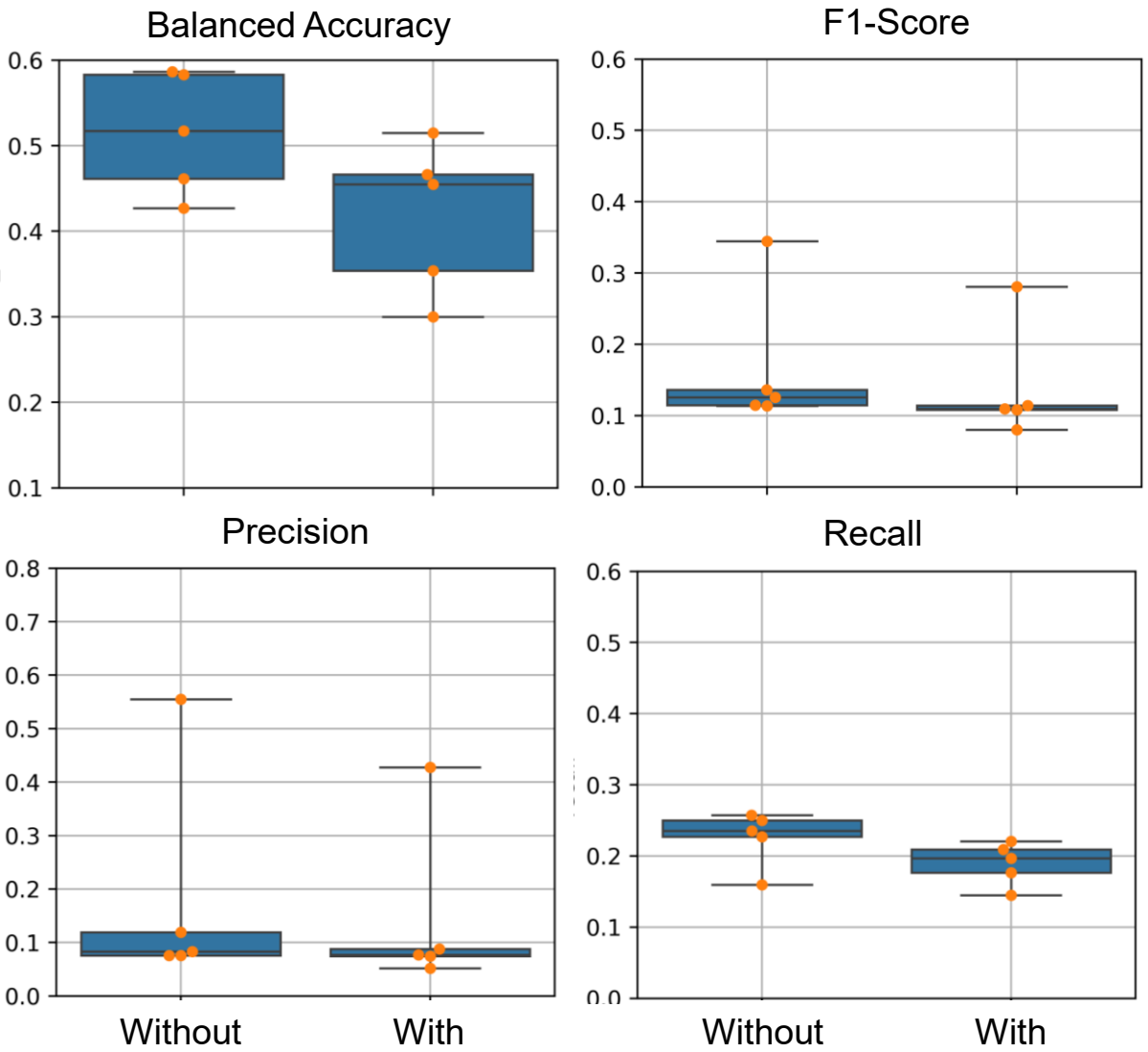
Approach



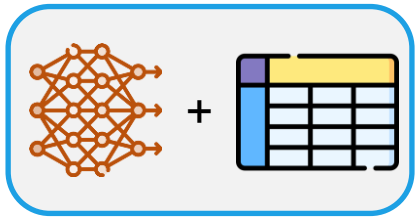
None of the pretraining options improved the results over doing no pretraining

→ No pretraining was done

Results – Clinical Data



Approach

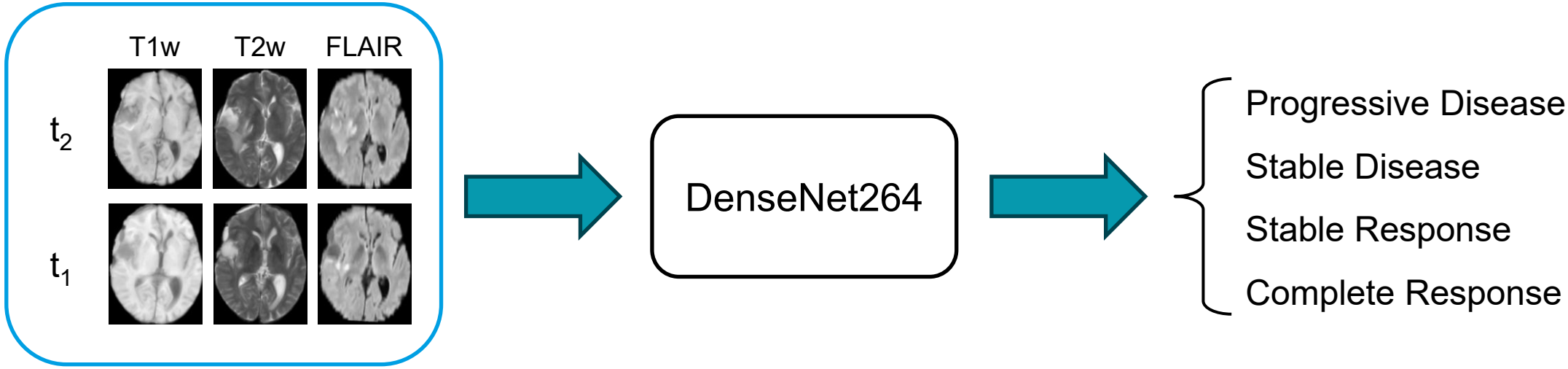


- BA is higher when clinical data is not used

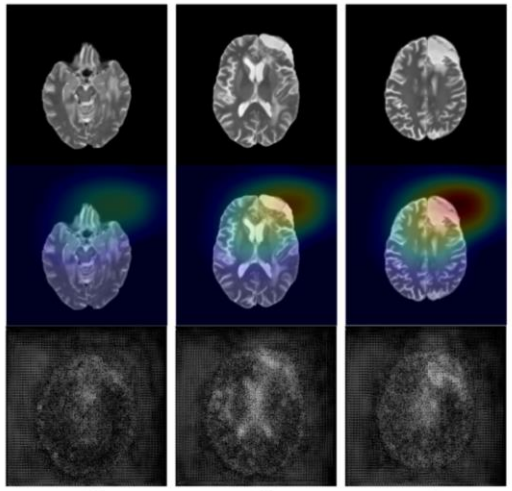
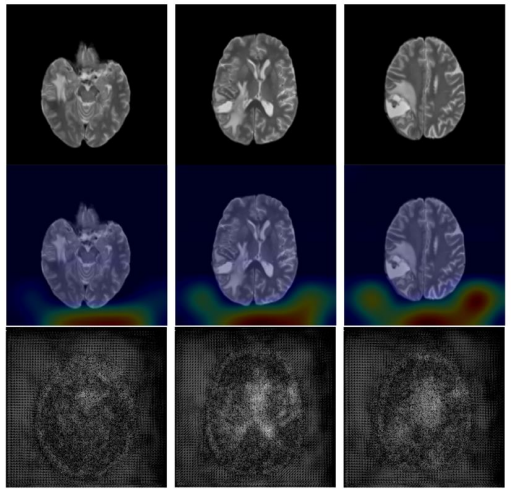
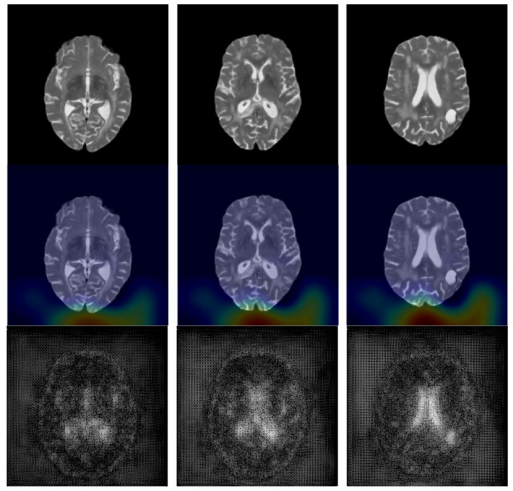
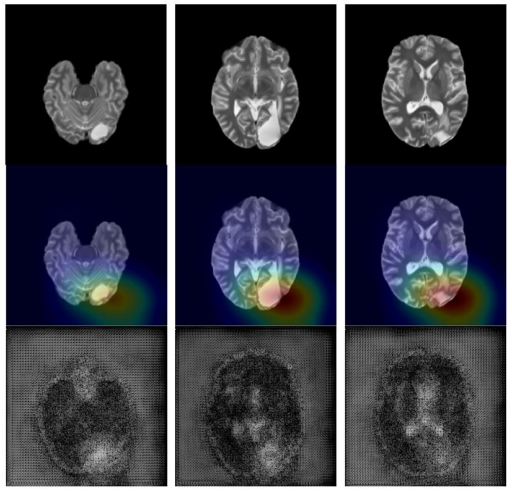
→ Clinical Data was not inputted

Best Results

No subtraction of timepoints | T1+T2+FLAIR | DenseNet264 | No pretraining | No Clinical Data Inputted



Results – Explainability

	Class PD	Class SD	Class PR	Class CR
Ground Truth				
T2 image				
Grad-CAM (Predicted class)				
Saliency Map (Predicted class)				
	z=55 z=77 z=99	z=55 z=77 z=99	z=66 z=77 z=88	z=44 z=66 z=77
Probability **Predicted** | Ground Truth	Classes: PD SD PR CR Probability (%): 17.1 22.6 16.6 **43.7**	Classes: PD SD PR CR Probability (%): 18.6 **45** 16.8 19.6	Classes: PD SD PR CR Probability (%): 19.3 **45.3** 21.3 14	Classes: PD SD PR CR Probability (%): 8.7 13.8 8.3 **69.2**

PD=Progressive Disease; SD=Stable Disease; PR=Progressive Response; CR=Complete Response

Results – Explainability

Ground Truth

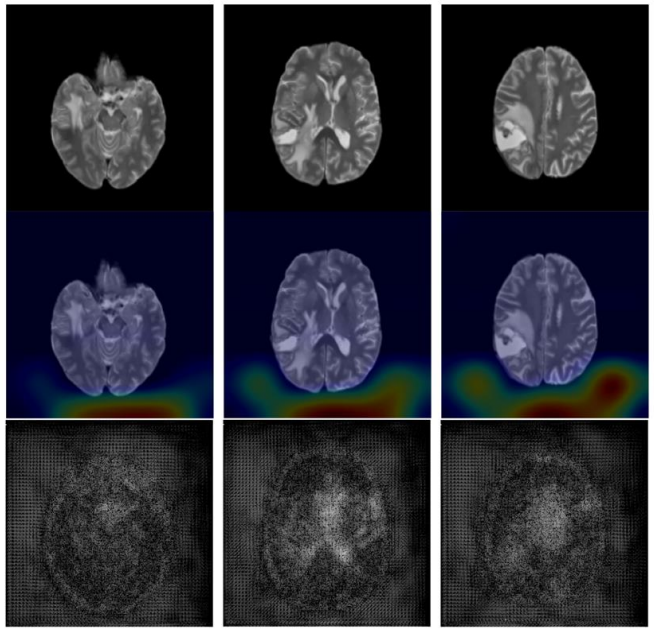
Class SD

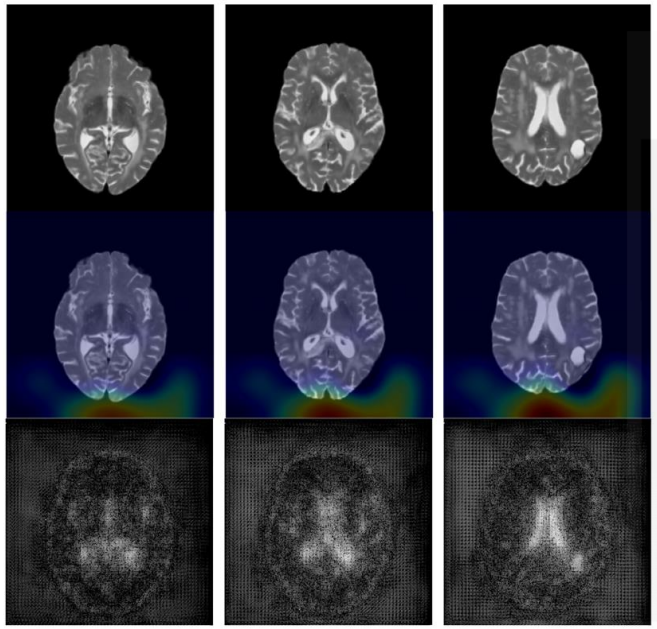
Class PR

T2 image

Grad-CAM
(Predicted class)

Saliency Map
(Predicted class)





Probability
Predicted | Ground Truth

Classes	PD	SD	PR	CR
Probability (%)	18.6	**45**	16.8	19.6

Classes	PD	SD	PR	CR
Probability (%)	19.3	**45.3**	21.3	14

PD=Progressive Disease; SD=Stable Disease;
PR=Progressive Response; CR=Complete Response

- Tumor is not highlighted in some cases

Results – Explainability

Ground Truth

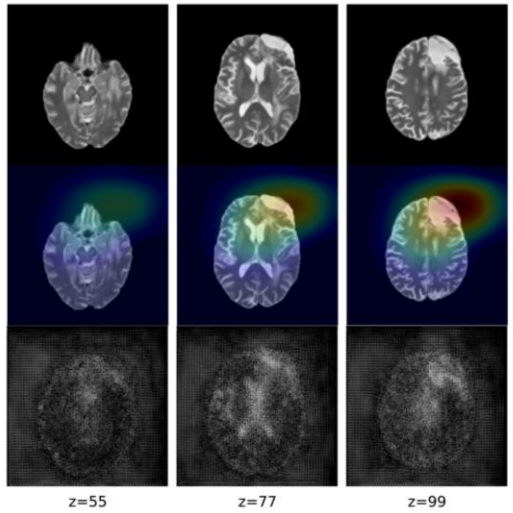
T2 image

Grad-CAM (Predicted class)

Saliency Map (Predicted class)

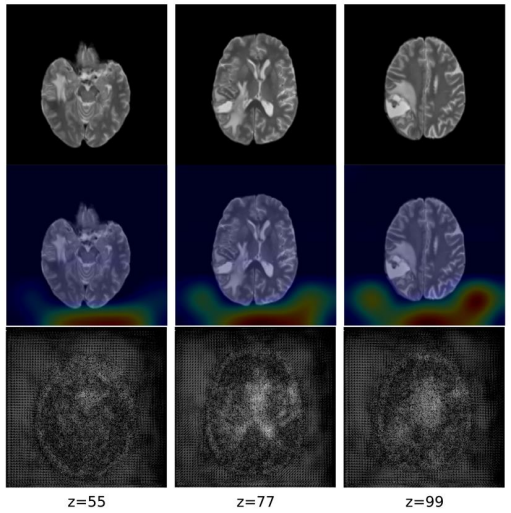
Probability
Predicted | Ground Truth

Class PD



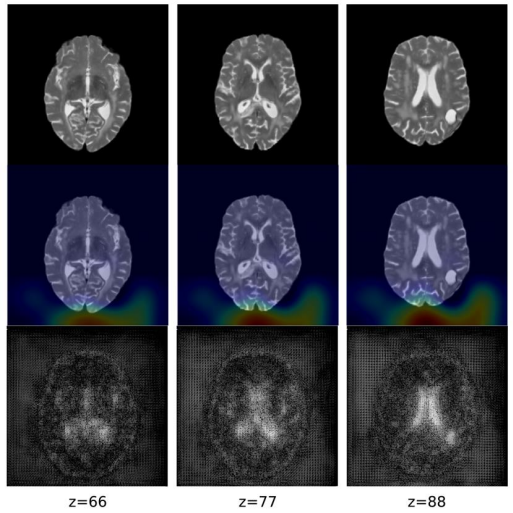
Classes	PD	SD	PR	CR
Probability (%)	17.1	22.6	16.6	**43.7**

Class SD



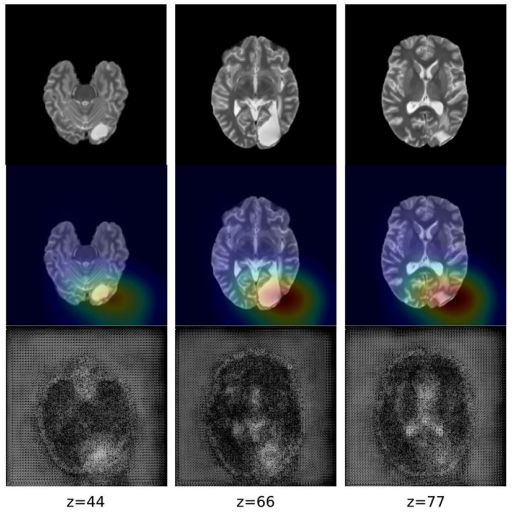
Classes	PD	SD	PR	CR
Probability (%)	18.6	**45**	16.8	19.6

Class PR



Classes	PD	SD	PR	CR
Probability (%)	19.3	**45.3**	21.3	14

Class CR



Classes	PD	SD	PR	CR
Probability (%)	8.7	13.8	8.3	**69.2**

PD=Progressive Disease; SD=Stable Disease; PR=Progressive Response; CR=Complete Response

Results – Explainability

Ground Truth

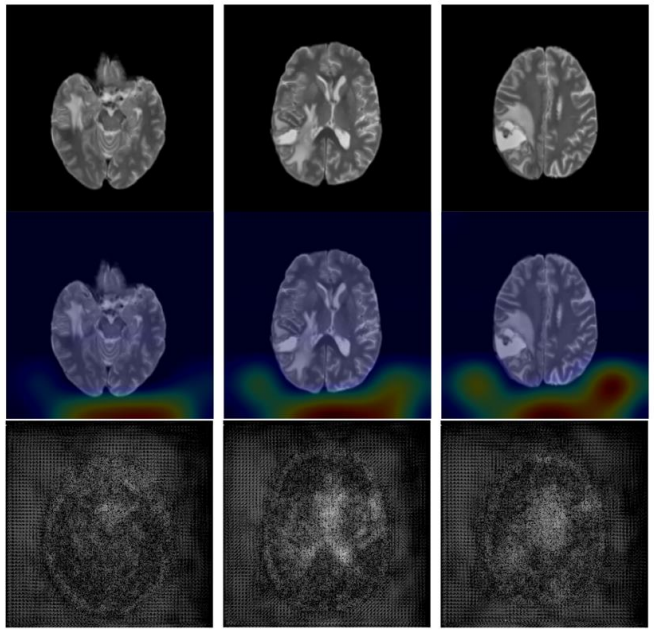
Class SD

Class CR

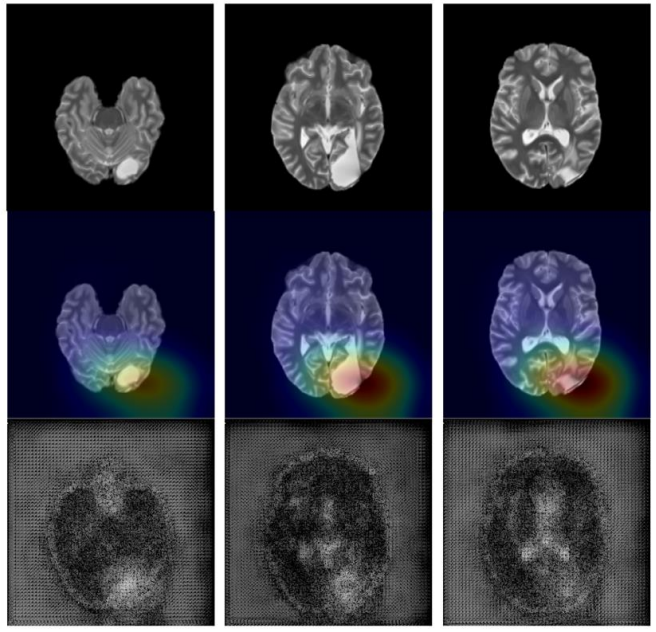
T2 image

Grad-CAM
(Predicted class)

Saliency Map
(Predicted class)



z=55 z=77 z=99



z=44 z=66 z=77

Probability
Predicted | Ground Truth

Classes	PD	SD	PR	CR
Probability (%)	18.6	**45**	16.8	19.6

Classes	PD	SD	PR	CR
Probability (%)	8.7	13.8	8.3	**69.2**

PD=Progressive Disease; SD=Stable Disease;
PR=Progressive Response; CR=Complete Response

- Correct prediction with unexpected highlighted region
- High probability of being CR

Conclusion

- Models tested have poor performance
- Test other approaches to increase performance
- Complex problem
- Need for Open Access Datasets
- Small dataset size hinders learning
- Importance of Explainability in Healthcare

Check out the preprint

Acknowledgements

LaSEEB

José Maria Moreira from Learning Health

Grant: 2023.03810.BDANA

Grants' DOI:
10.54499/LA/P/0083/2020,
10.54499/UIDP/50009/2020, and
10.54499/UIDB/50009/2020