Towards a deep learning approach for classifying response to treatment in glioblastomas

Ana Matoso[1*], Catarina Passarinho[1], Marta 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]Learning Health, Hospital da Luz, Lisbon, Portugal

*anamatoso@tecnico.ulisboa.pt

Introduction

Glioma

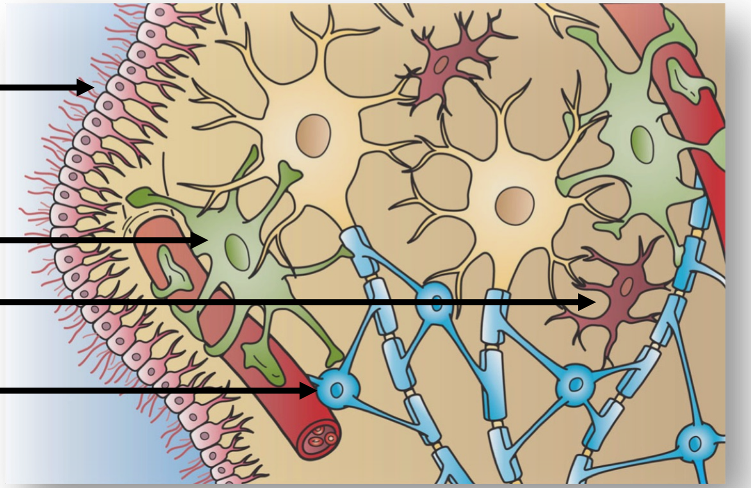
Glial cells Tumor

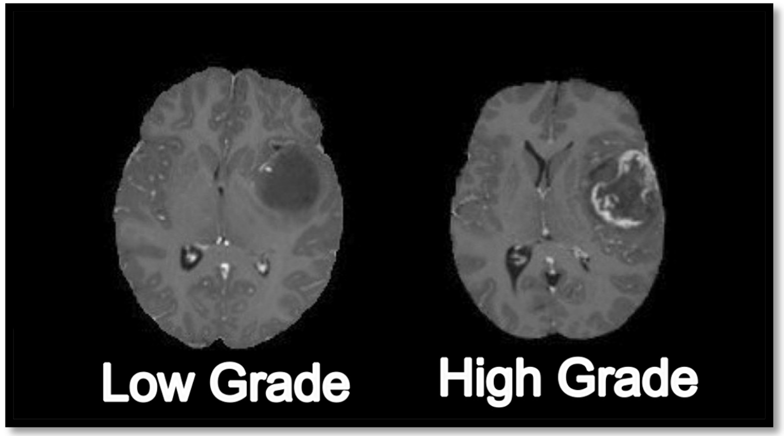
Glioma incidence: ~5/100 000 per year

Gliomas
- Grade I, Grade II — Low Grade
- Grade III, Grade IV — High Grade

worse outcome

20% survival after 2 years





Larjavaara, S., et al., Neuro-oncol., 2007; Ho, V. K. Y., et al., Eur. J. Cancer, 2014; Wen, P. Y., et al., J. Clin. Oncol., 2010

Introduction

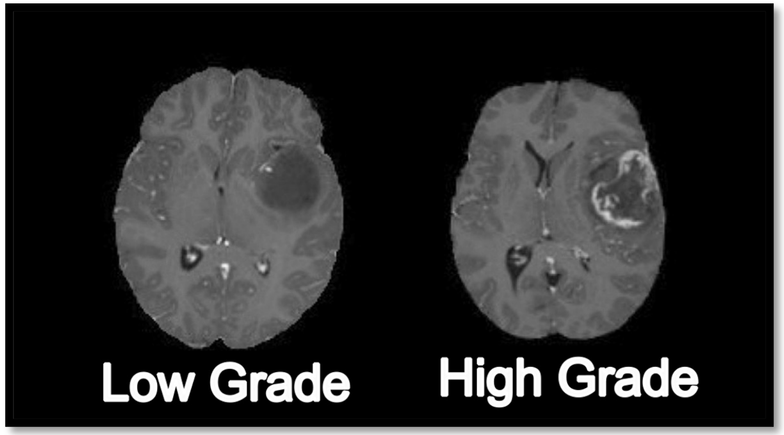
Glioma

Glial cells | Tumor

Glioma incidence: ~5/100 000 per year

Gliomas
- Grade I, Grade II — Low Grade
- Grade III, Grade IV — High Grade

worse outcome



Response Assessment in Neuro-Oncology (RANO) criteria
- Complete Response
- Partial Response
- Stable Disease
- Progressive Disease

	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*

Larjavaara, S., et al., Neuro-oncol., 2007; Ho, V. K. Y., et al., Eur. J. Cancer, 2014; Wen, P. Y., et al., J. Clin. Oncol., 2010

Introduction

Glioma

Glial cells Tumor

Glioma incidence: ~5/100 000 per year

Gliomas
- Grade I, Grade II — Low Grade
- Grade III, Grade IV — High Grade

worse outcome

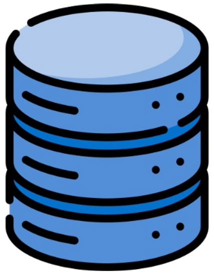
Response Assessment in Neuro-Oncology (RANO) criteria
- Complete Response
- Partial Response
- Stable Disease
- Progressive Disease

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*

Larjavaara, S., et al., Neuro-oncol., 2007; Ho, V. K. Y., et al., Eur. J. Cancer, 2014; Wen, P. Y., et al., J. Clin. Oncol., 2010

Methods – Data



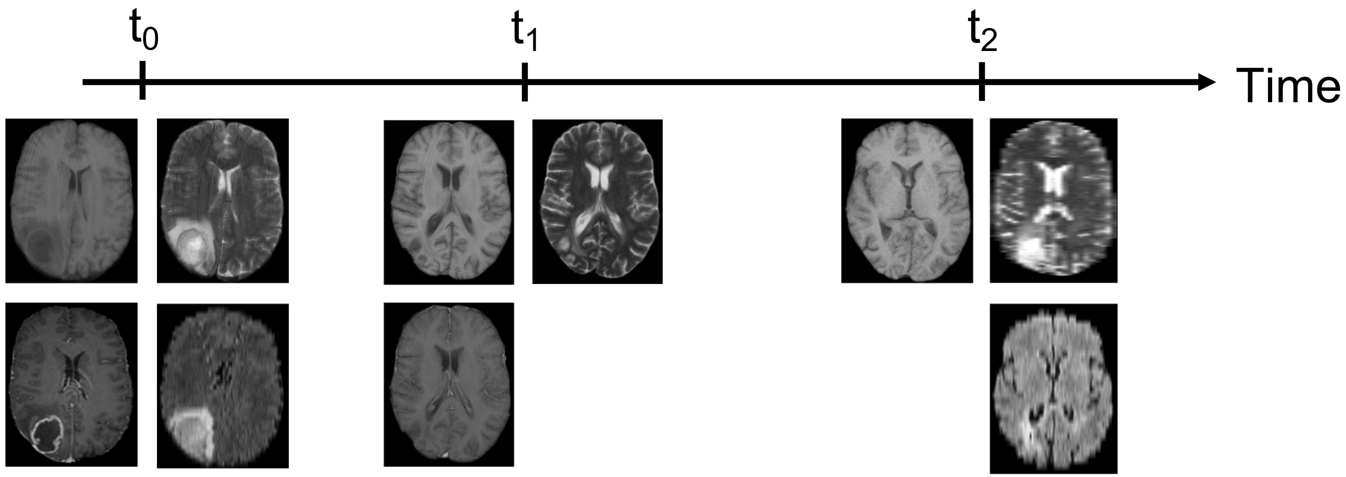
LUMIERE longitudinal dataset

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

638 timepoints

91 patients

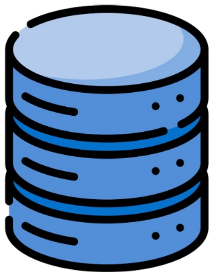
Class	Prevalence
Progressive Disease (PD)	67%
Stable Disease (SD)	20%
Progressive Response (PR)	6%
Complete Response (CR)	7%



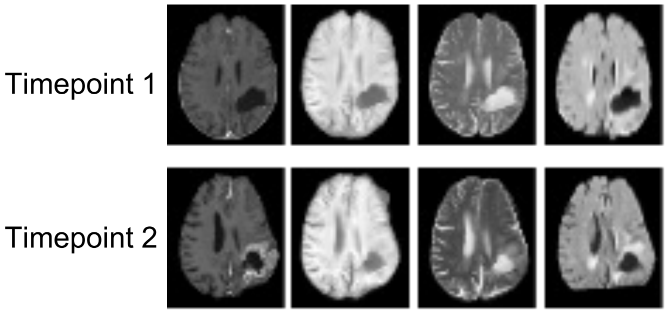
Suter, Y., et al., Scientific Data Data, 2022

Methods – Pipeline

LUMIERE dataset



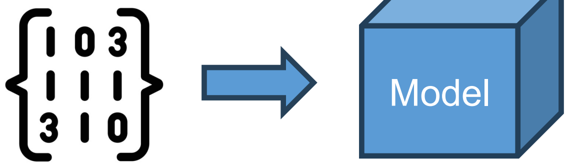
Data



5-fold Cross Validation
80/20 Stratified Split

Model Training

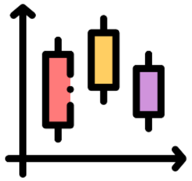
Weight Initialization



Training Setup:
- 100 epochs maximum
- Cross Entropy loss
- AdamW optimizer
- LR = 1e-4
- Patience = 10

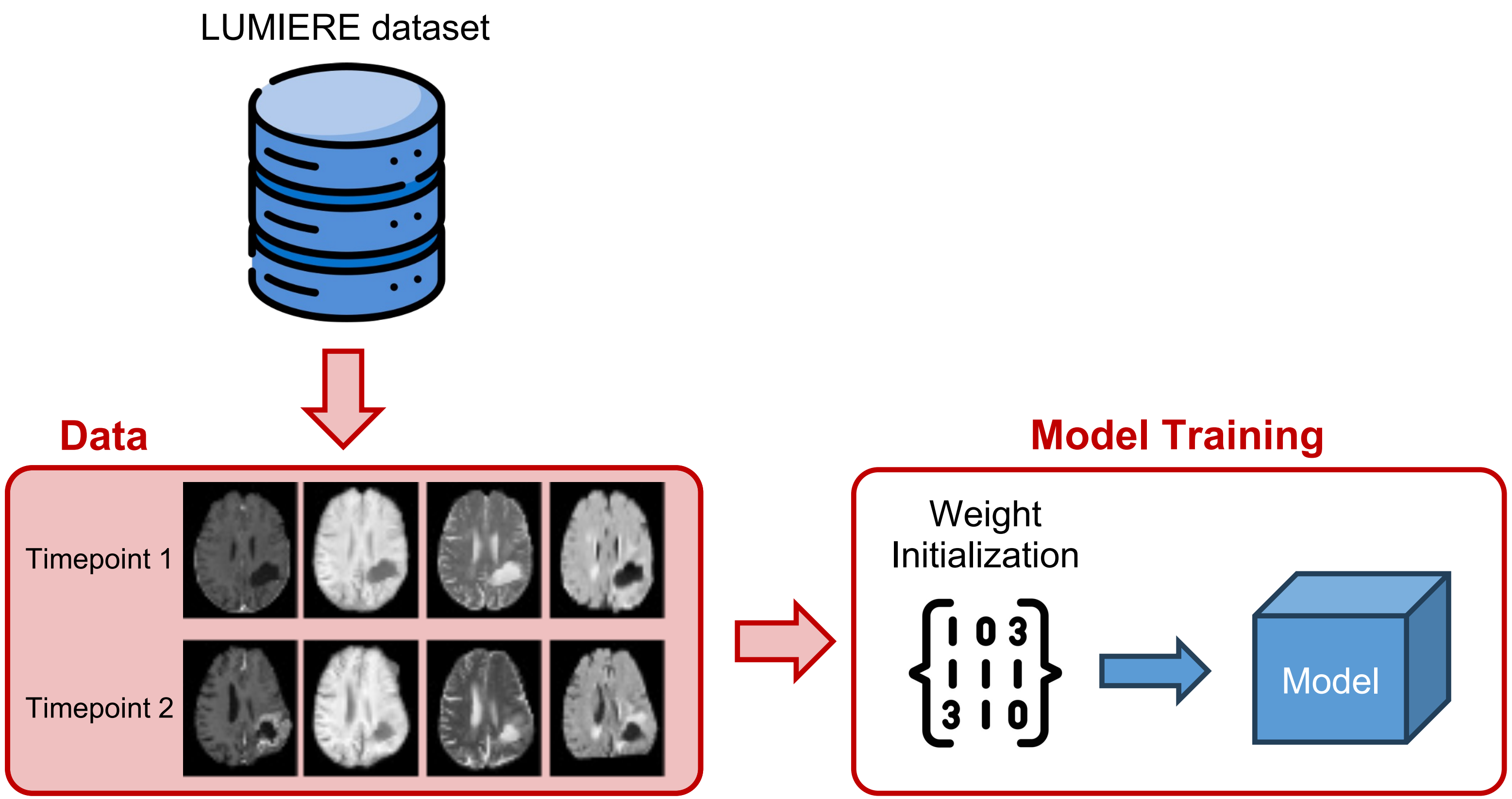
Model Testing

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



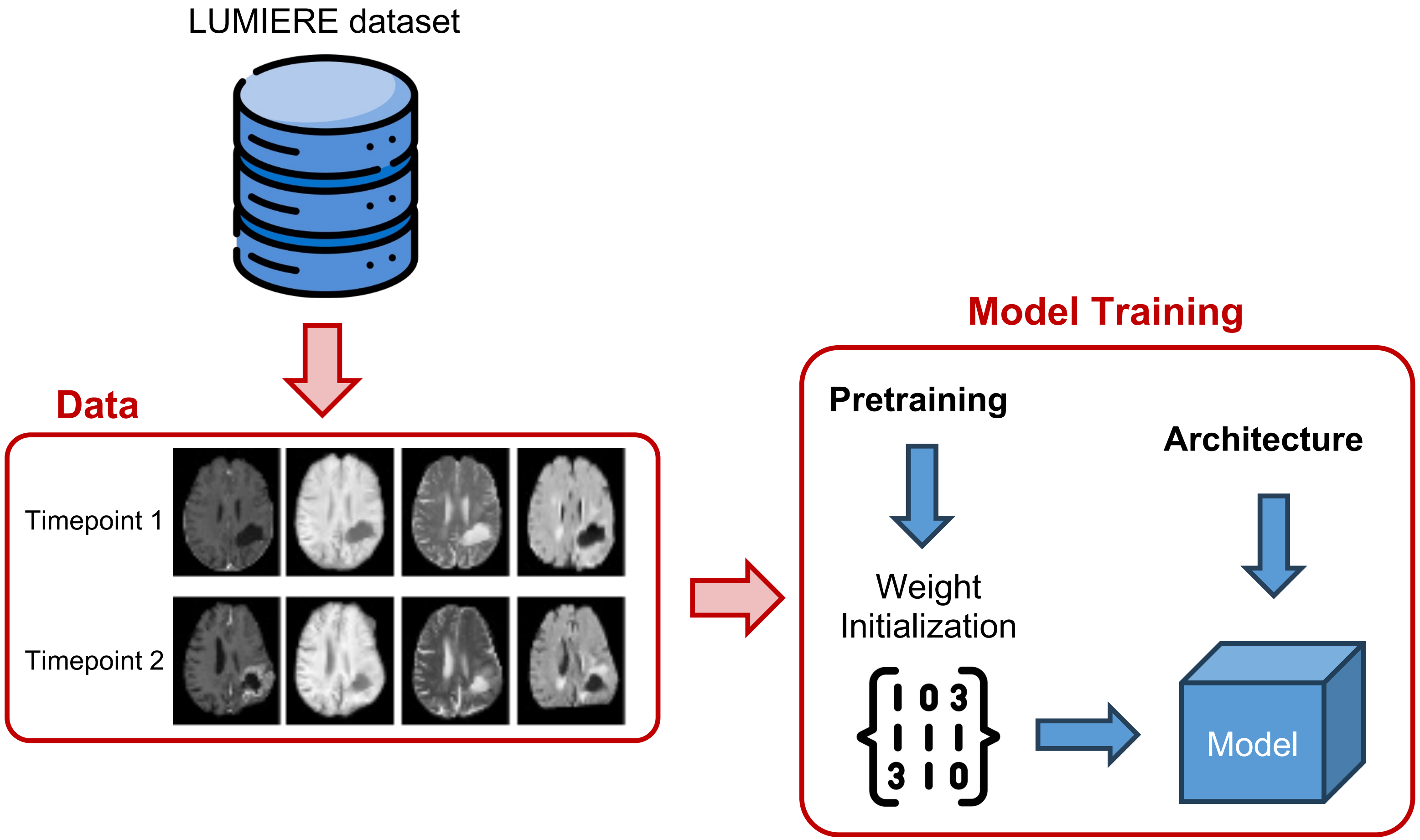
MONAI PyTorch

Methods – Tested Approaches



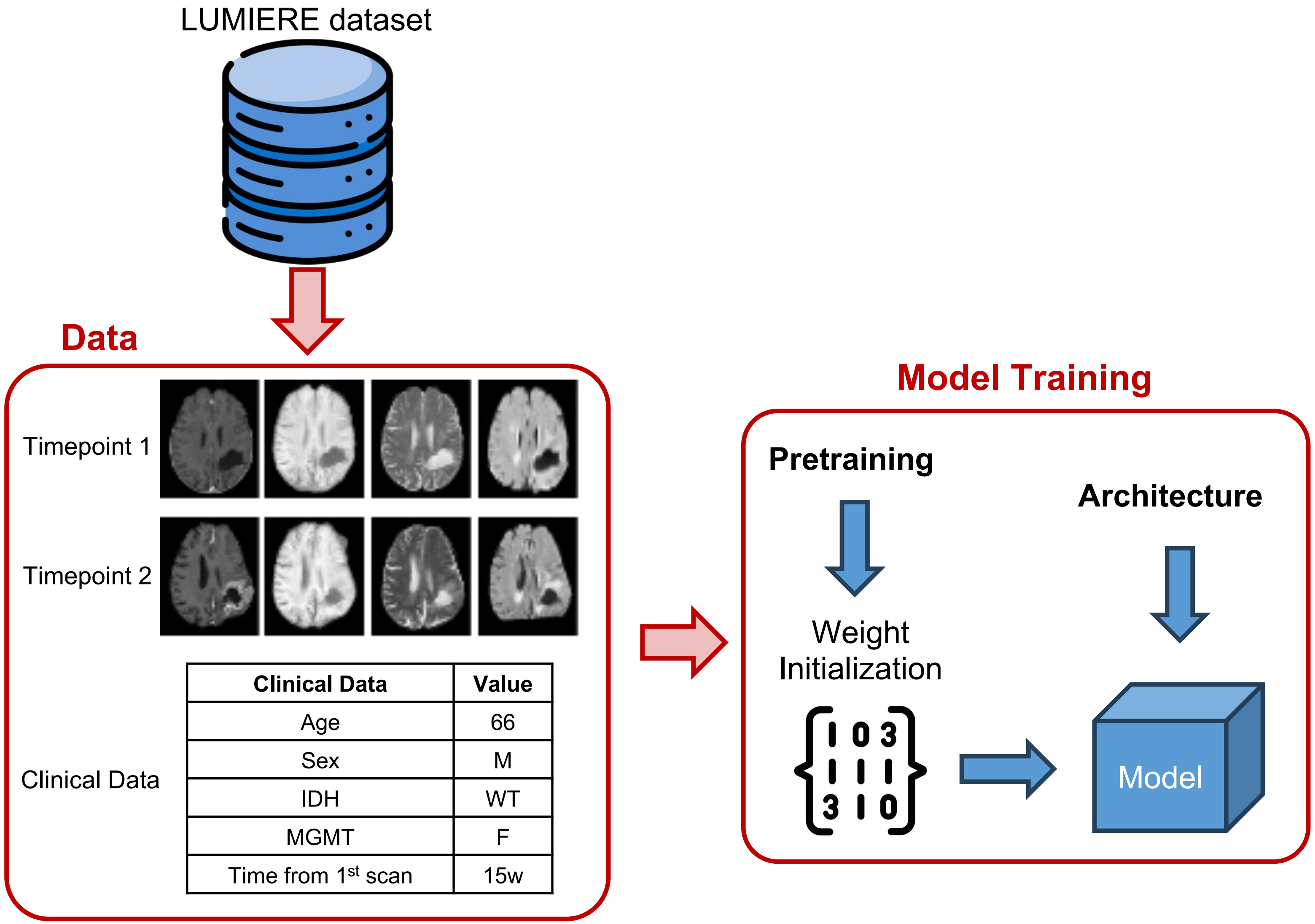
1. Subtraction of Timepoints
2. Combination of modalities

Methods – Tested Approaches



1. Subtraction of Timepoints
2. Combination of modalities
3. Model Architectures
4. Pretraining

Methods – Tested Approaches

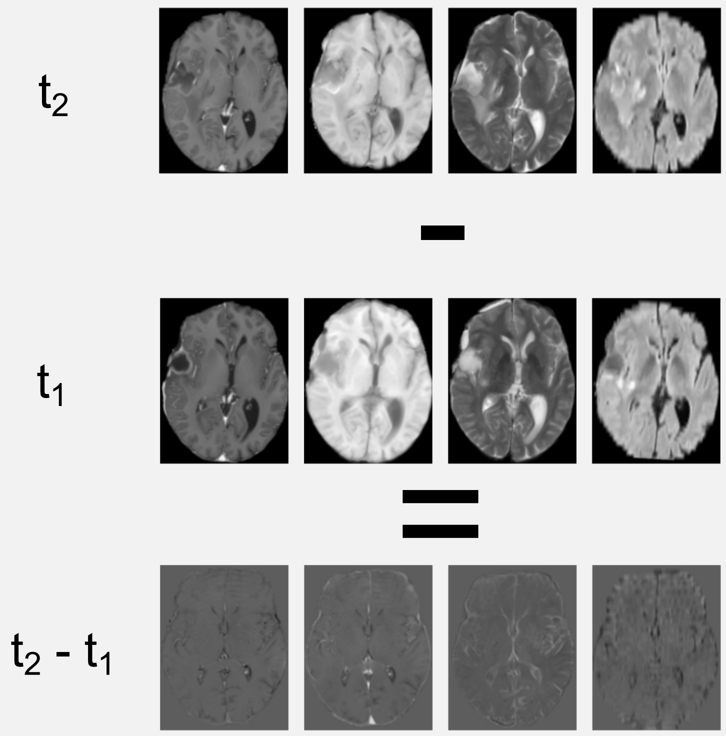


Clinical Data	Value
Age	66
Sex	M
IDH	WT
MGMT	F
Time from 1st scan	15w

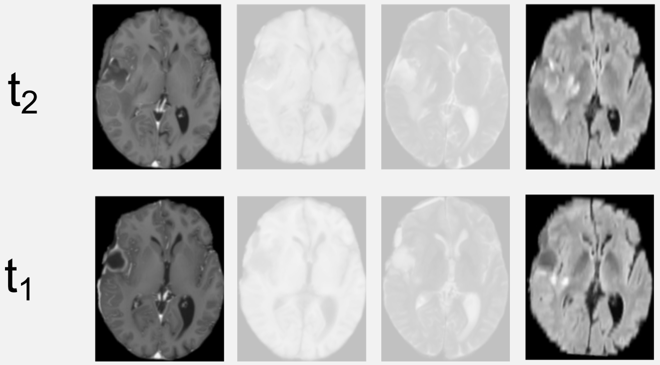
1. Subtraction of Timepoints
2. Combination of modalities
3. Model Architectures
4. Pretraining
5. Clinical Data

Methods – Tested Approaches

1. Subtraction of timepoints



2. Combinations of modalities



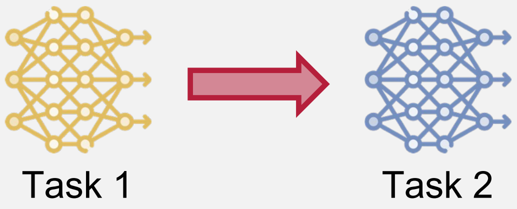
Combination of Modalities	Size of Dataset
CT1+T1+T2+FLAIR	337
CT1+FLAIR	344
T1+T2+FLAIR	338
CT1	355
T1+FLAIR	338

3. Model Architectures

- Densenets:
  - Densenet 121
  - Densenet 169
  - Densenet 264
- Vision Transformer
- Alexnet3D

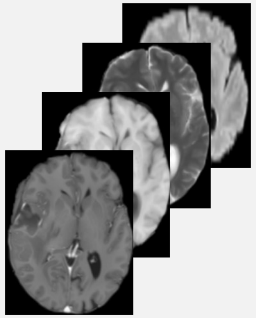
Methods – Tested Approaches

4. Pretraining



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

5. Use of Clinical Data

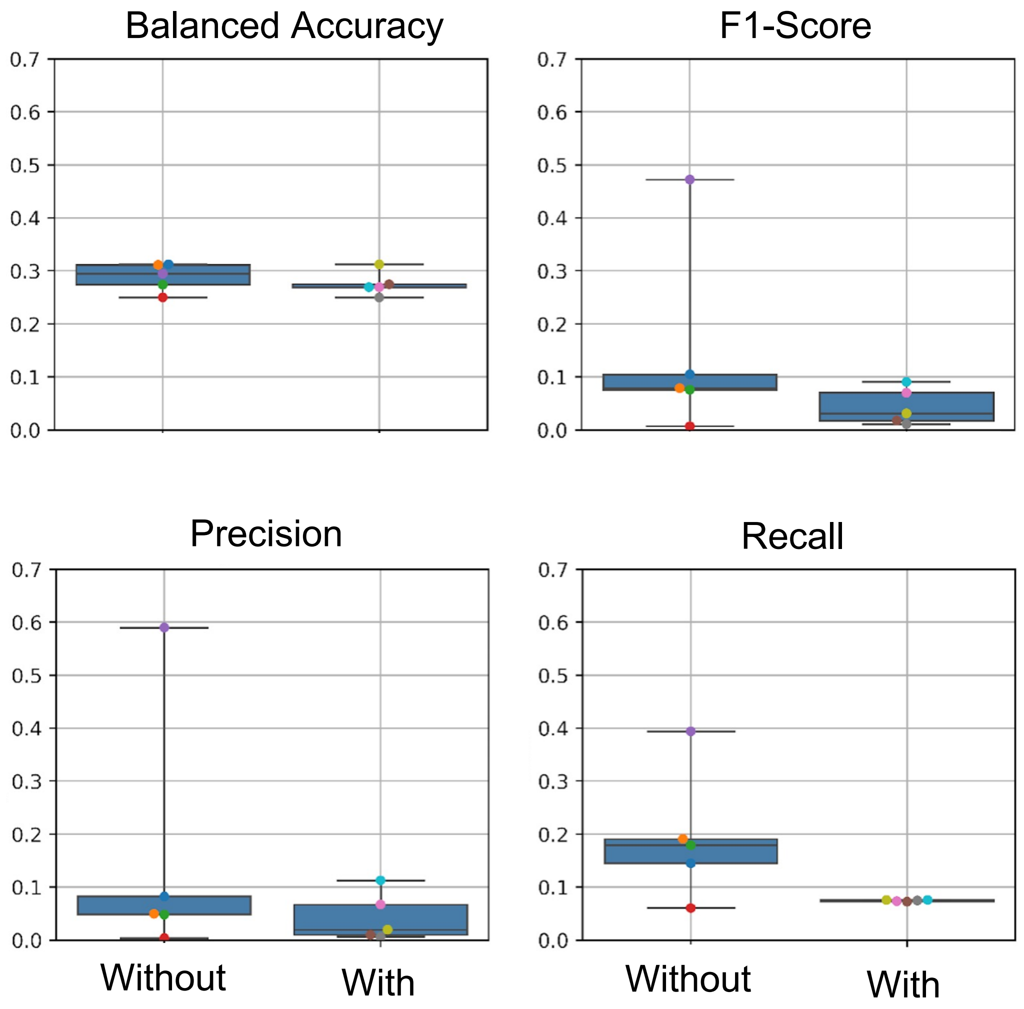


+

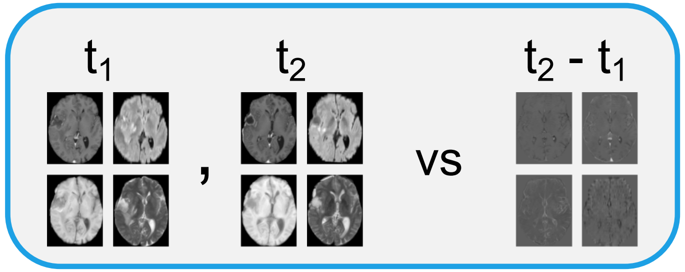
Clinical Data	Value
Age	66
Sex	M
IDH	WT
MGMT	F
Time from 1st scan	15w

Approaches will be tested sequencially

Results – Subtraction



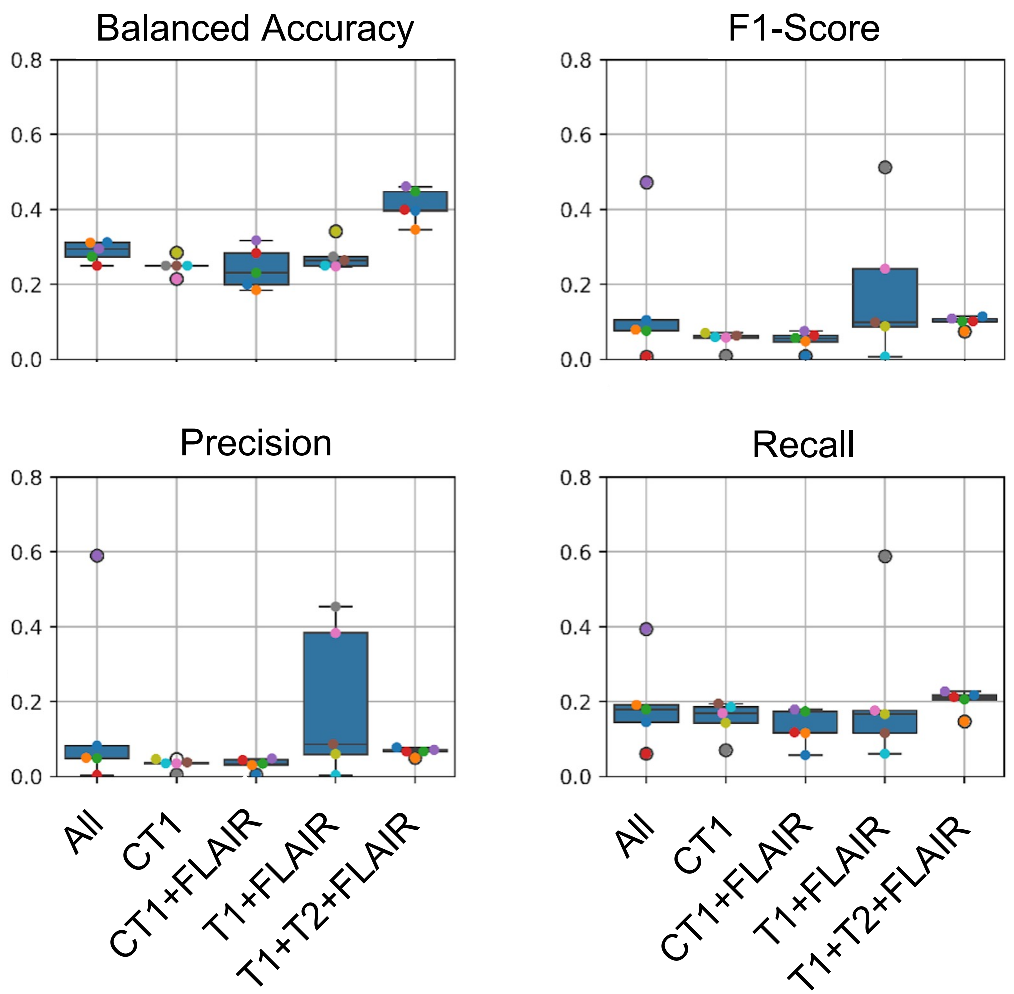
Approach



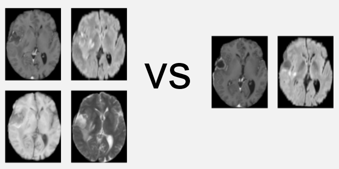
- Similar BA
- Slight decrease in Recall and Precision
- Decrease in F1-Score

→ No subtraction was done in the next stages

Results – Modalities



Approach



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

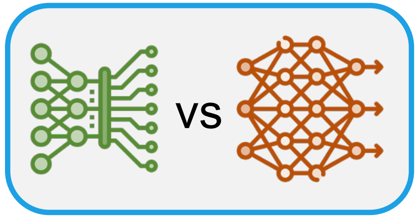
⇩

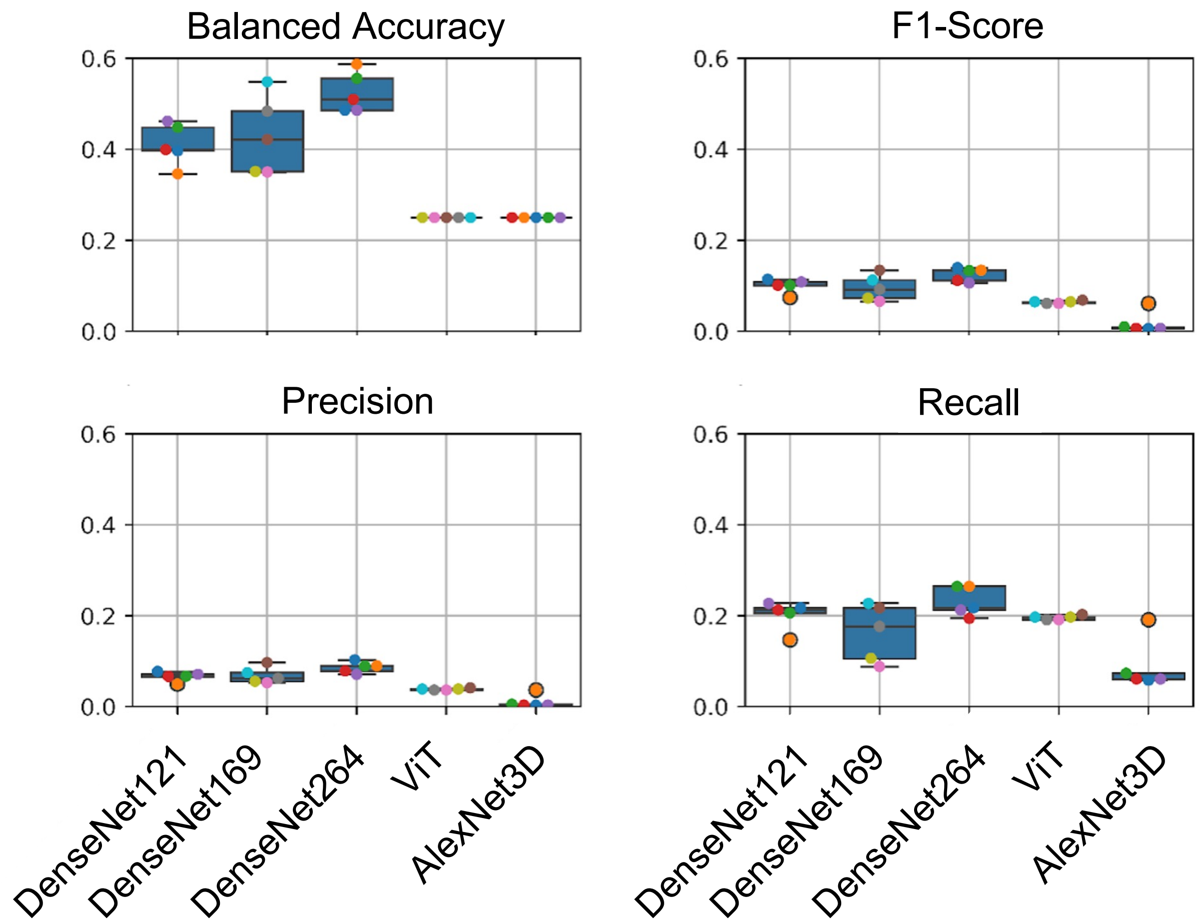
- Increased F1 Score in T1+FLAIR

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

Results – Architectures

Approach

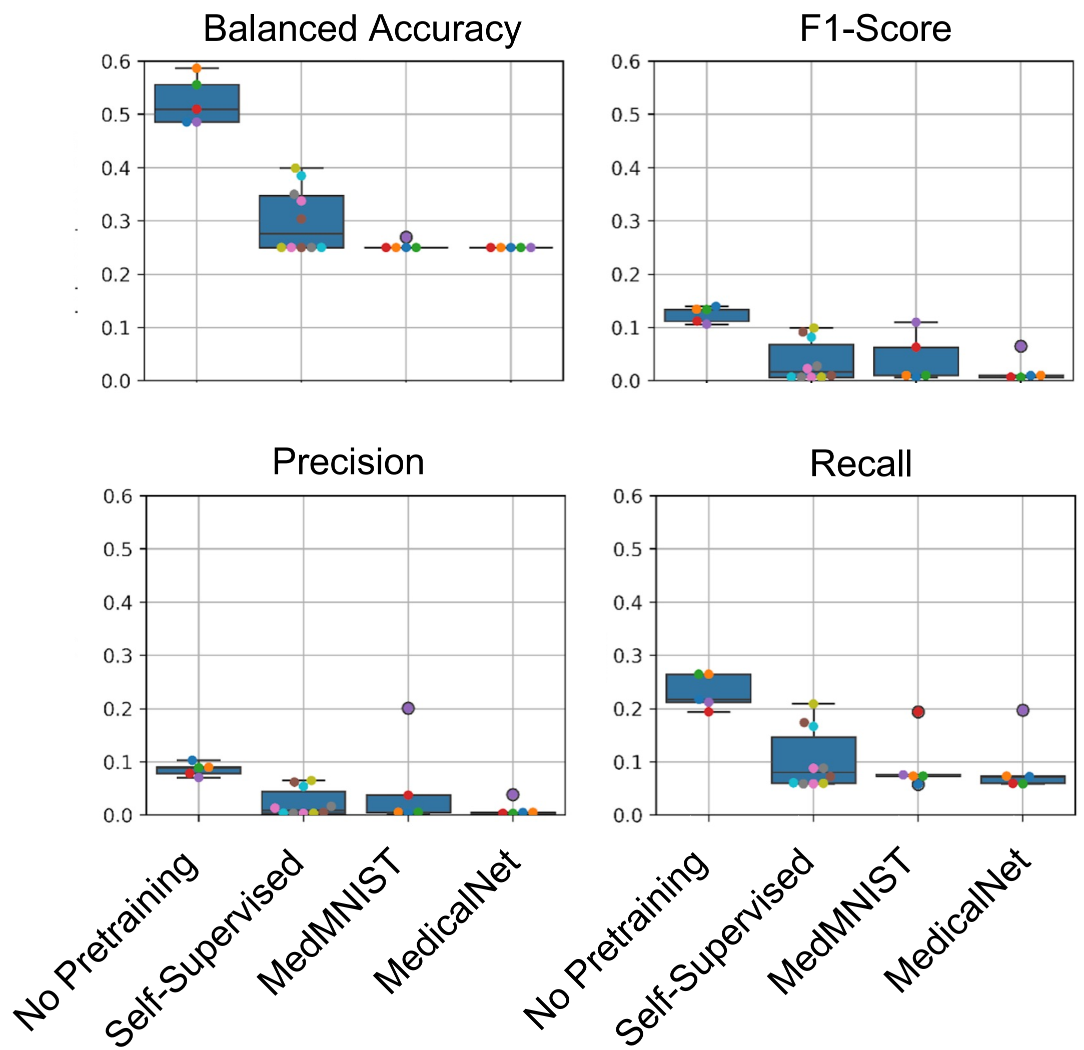




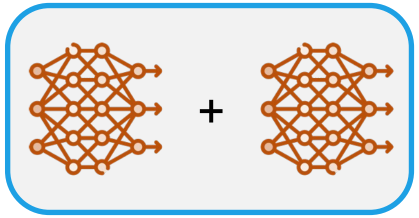
- DenseNets performed better than ViT and AlexNet3D
- More complex DenseNets improve performance

⟶ DenseNet264 has overall better performance

Results – Pretraining



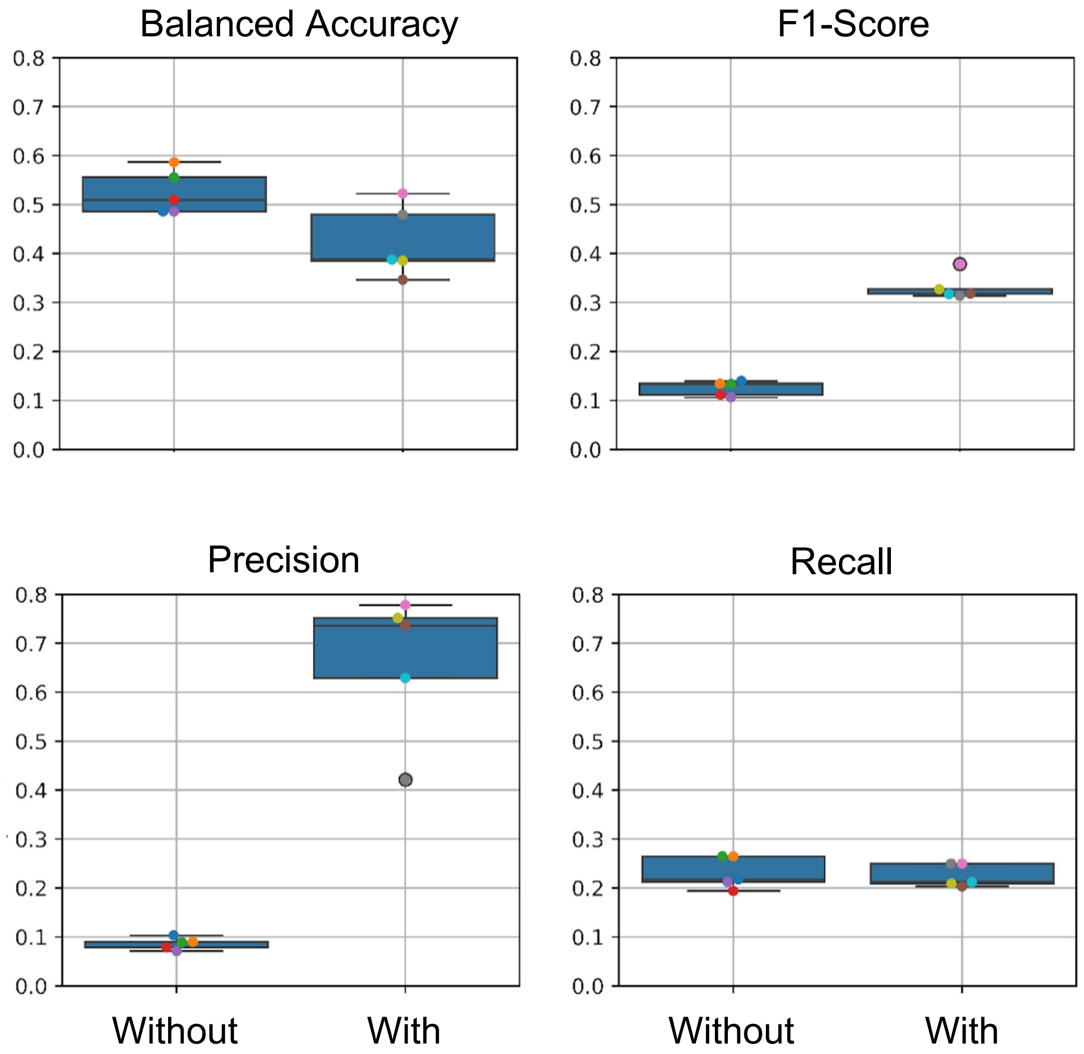
Approach



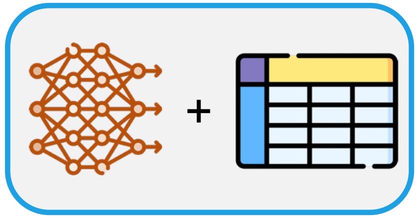
None of the pretraining options improved the results over doing no pretraining

⟶ No pretraining was done

Results – Clinical Data



Approach

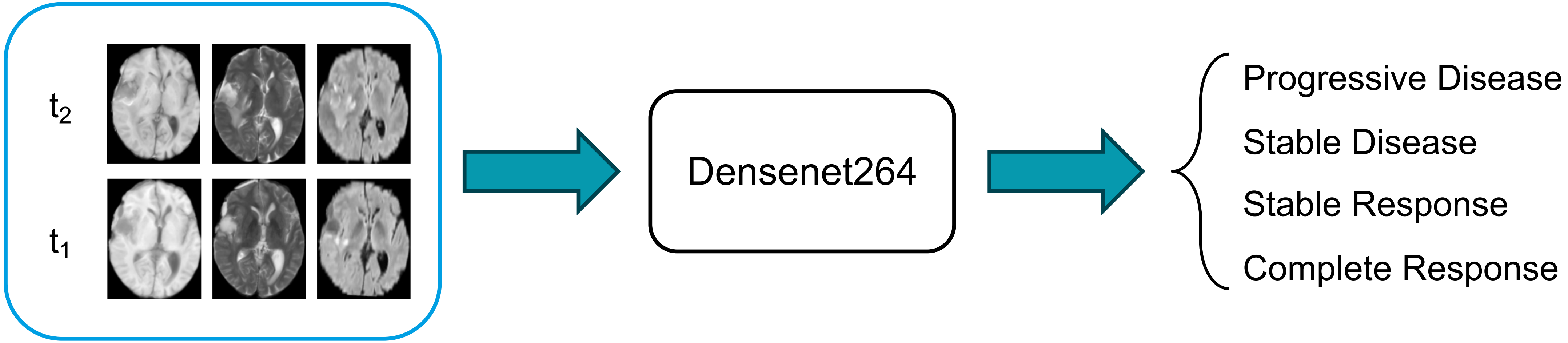


- BA is higher when clinical data is not used
- Using Clinical Data improves Precision
- Increased F1-Score when using Clinical Data

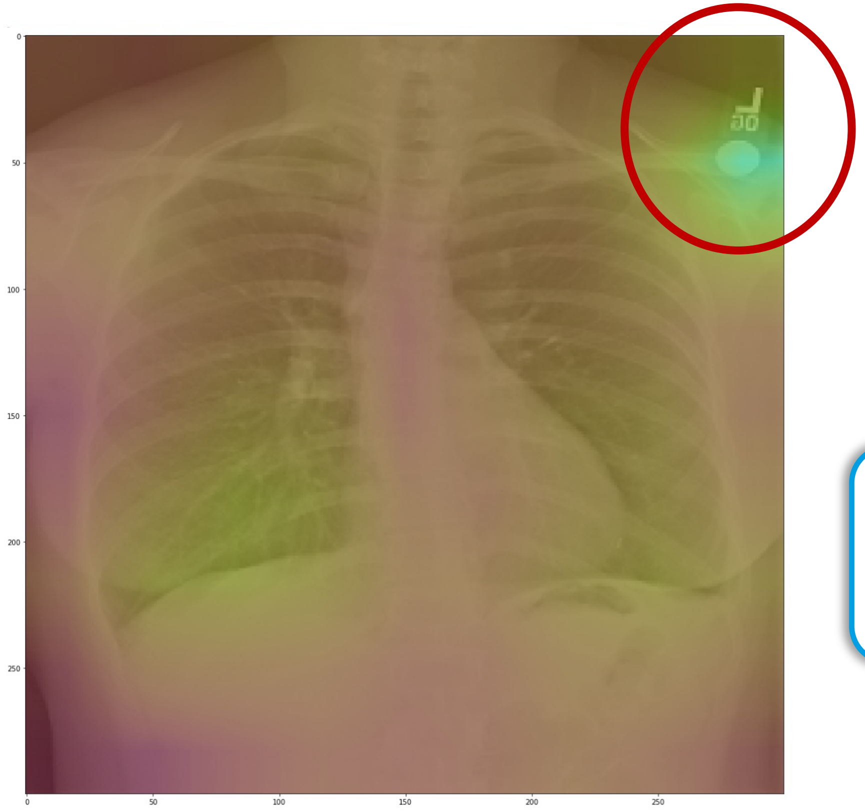
→ Clinical Data was not inputted

Best Results

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



Methods – Explainability



Burned-in Annotations

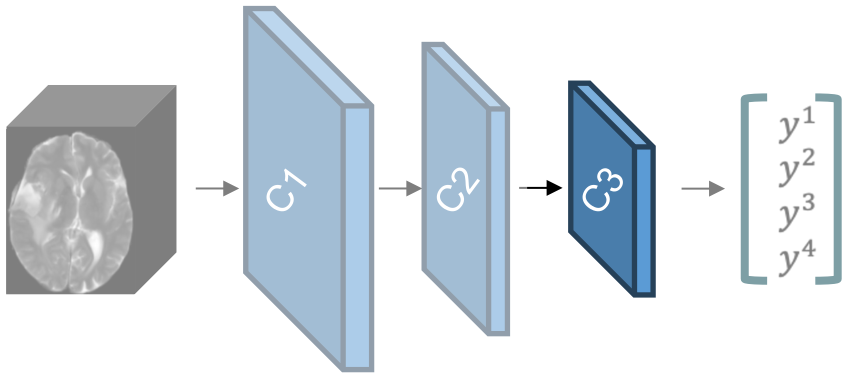
Important to check impactful regions for classification

Teixeira, L. O., Pereira, R. M., Bertolini, D., Oliveira, L. S., Nanni, L., Cavalcanti, G. D. C., & Costa, Y. M. G. (2021). Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. Sensors, 21(21), 7116.

Methods – Explainability

Class Activation Maps

⟶ Last convolutional layer



with: Grad-Cam package

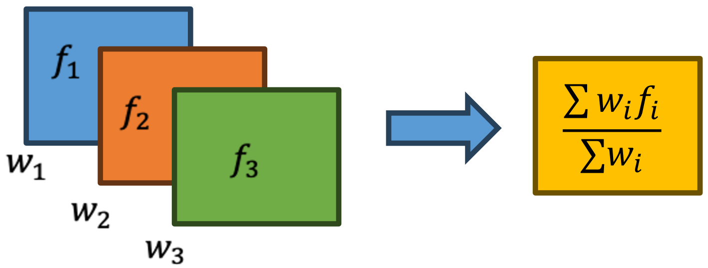
Saliency Maps

Methods – Explainability

Class Activation Maps

⟶ Last convolutional layer

⟶ Weighted Average of Feature Maps by the gradients



where $w_i = ReLU\left(k\frac{\partial y^c}{\partial f_i}\right)$

with: Grad-Cam package

Saliency Maps

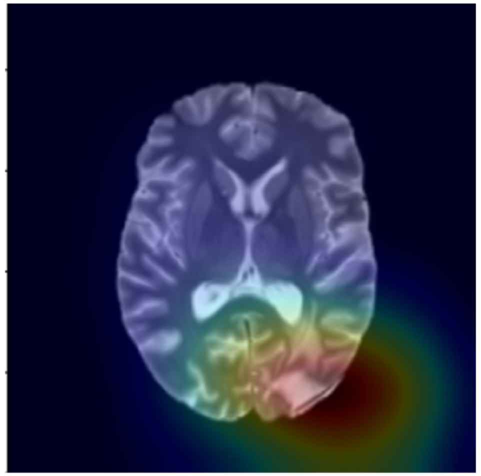
Methods – Explainability

Class Activation Maps

⟶ Last convolutional layer

⟶ Weighted Average of Feature Maps by the gradients

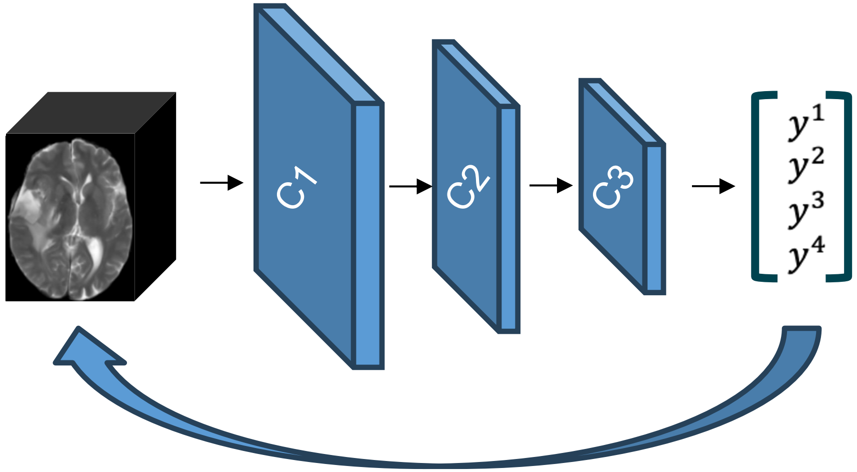
⟶ Coarse heatmap



with: Grad-Cam package

Saliency Maps

⟶ Gradients with respect to inputs



Captum

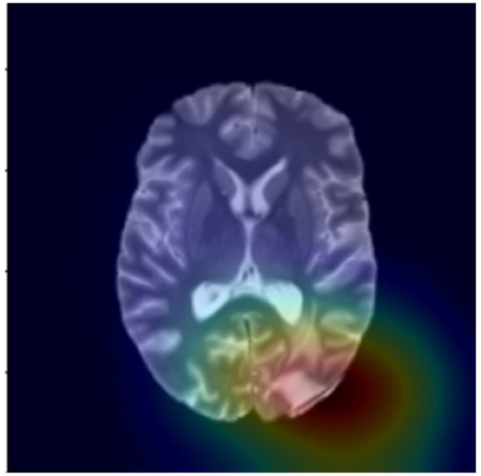
Methods – Explainability

Class Activation Maps

⟶ Last convolutional layer

⟶ Weighted Average of Feature Maps by the gradients

⟶ Coarse heatmap

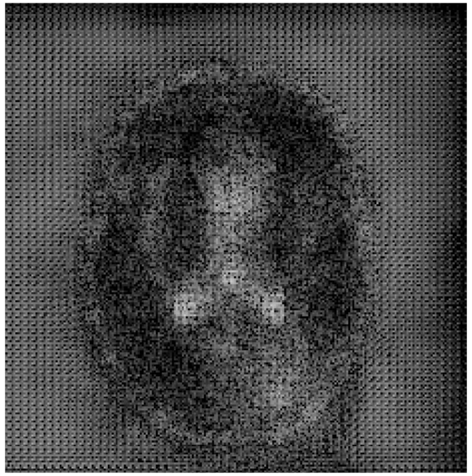


with: Grad-Cam package

Saliency Maps

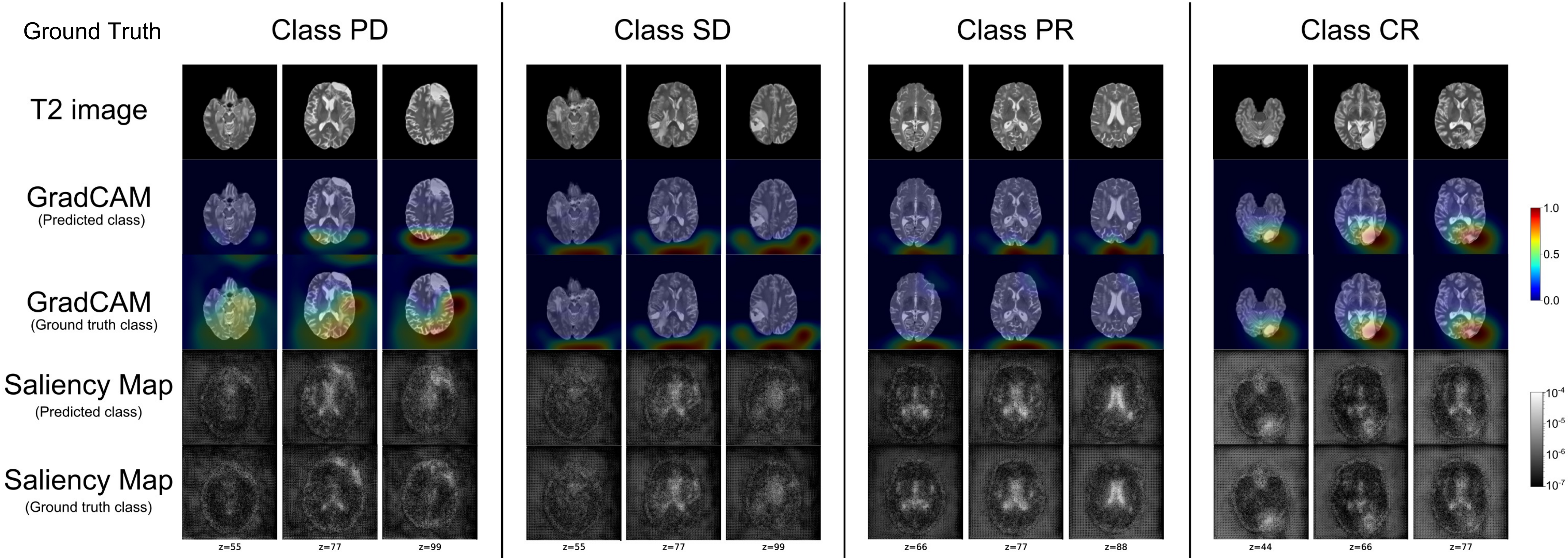
⟶ Gradients with respect to inputs

⟶ Granular impact of input



Captum

Results – Explainability



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

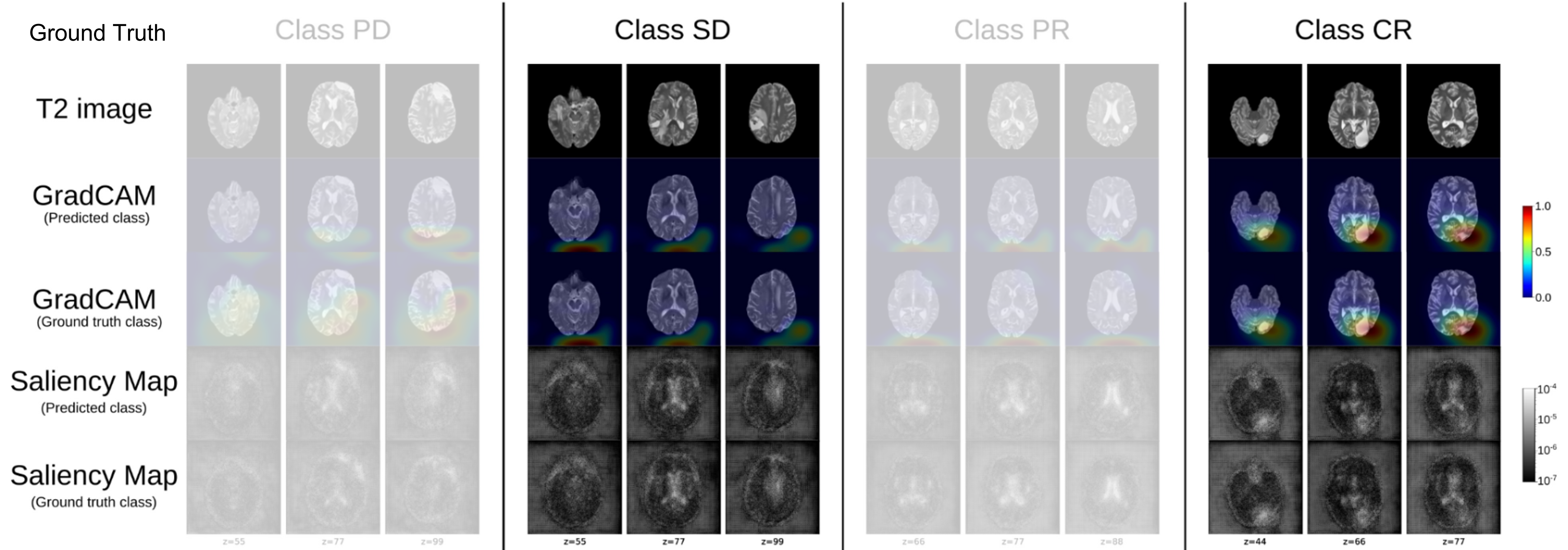
Classes	PD	SD	PR	CR
Probability (%)	19.3	**45.3**	21.3	14

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

- ➢ Tumor is not highlighted in some cases

Results – Explainability



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 (%)	17.4	43.2	15	24.3

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

➢ Correct prediction with unexpected highlighted region

➢ High probability of being CR

Conclusion

Models tested have poor performance

Complex problem

Small dataset size hinders learning

Test other approaches to increase performance

Need for Open Access Datasets

Importance of Explainability in Healthcare

Acknowledgements

LaSEEB

Fundação para a Ciência e a Tecnologia

Grant: 2023.03810.BDANA

LARSyS Laboratory of Robotics and Engineering Systems

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

José Maria Moreira from Learning Health