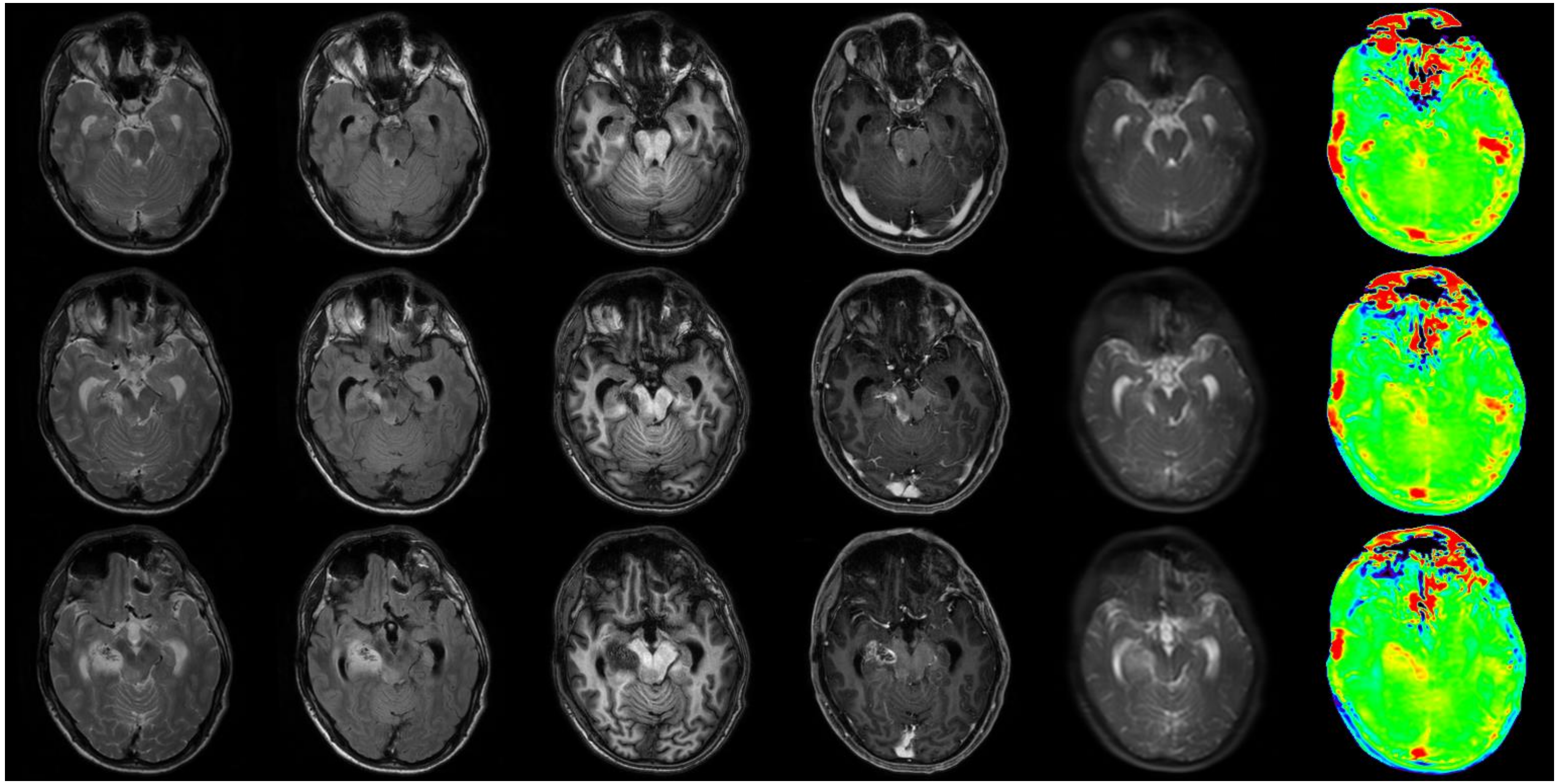
Glioma Classification and Biopsy Guidance with Multimodal MRI and Deep Learning

Checkpoint presentation

Team: Nhat Le

Mentors: Prof. Mathias Unberath, Prof. Shanshan Jiang

Background



- Common structural MRI sequences:
  - T2-weighted
  - T1-weighted
  - FLAIR
  - Gado-enhanced T1-weighted
- APT-weighted: a molecular MRI sequence developed by JHU
- Classification tasks:
  - Tumor grading
  - Tumor recurrence

Deliverables:

- Analysis of using average of multiple APT images
- Validated deep learning methods for classification tasks

Dataset

- Each scan
  - 15 slices
  - Slice resolution: 256x256
  - 5 modalities
  - Mostly preprocessed
- Focus on post-treatment dataset
  - Fine-tuning for tumor grading
  - All recurrent tumors are high grade

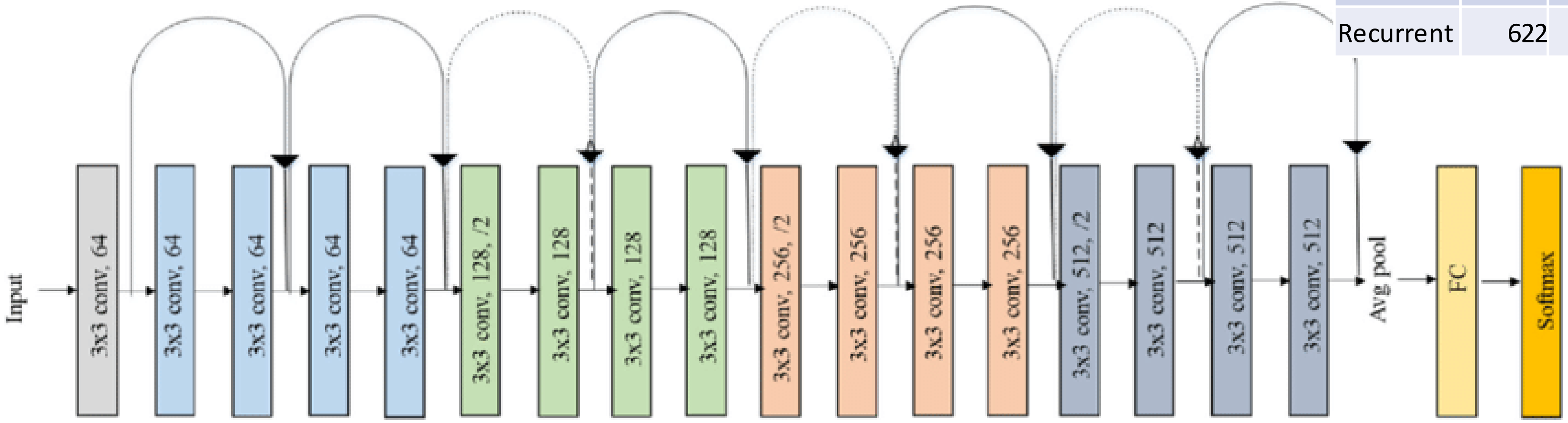
Type	**# scans**
Post-treatment tumor recurrence	103
Post-treatment treatment effects	61
Newly-diagnosed low-grade	18
Newly-diagnosed high-grade	20
Total	202

Annotation type	**# scans**
Scan-level	202
Slice-level	147 (all post-treatment)
Pixel-level	~140 (all post-treatment)

Analysis of using averaged APT images

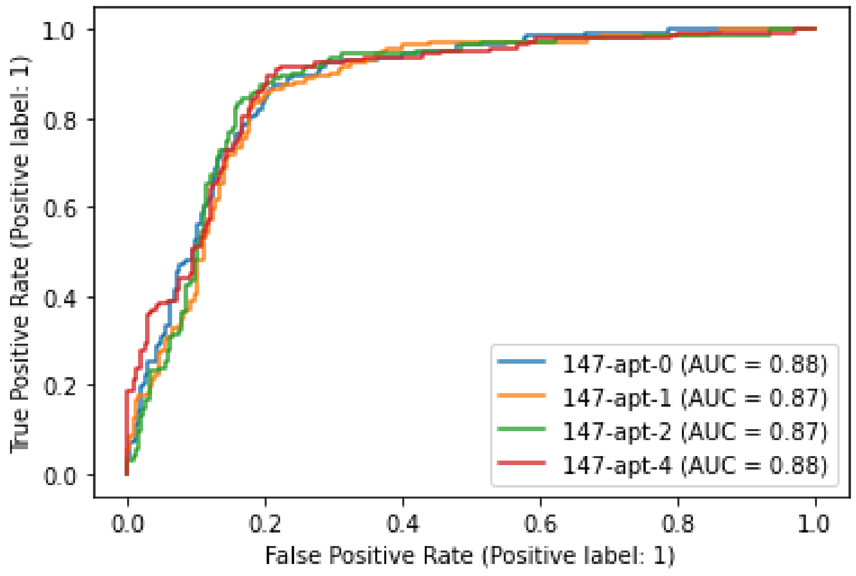
- Method: slice-level classification
- Model backbone: ResNet-18
- Hierarchical classification paradigm (from a previous study)
  - Normal vs. Abnormal
    - Abnormal: Tumor vs. No tumor
      - Tumor: Recurrent tumor vs. Non-recurrent
- APT images averaged after co-registration

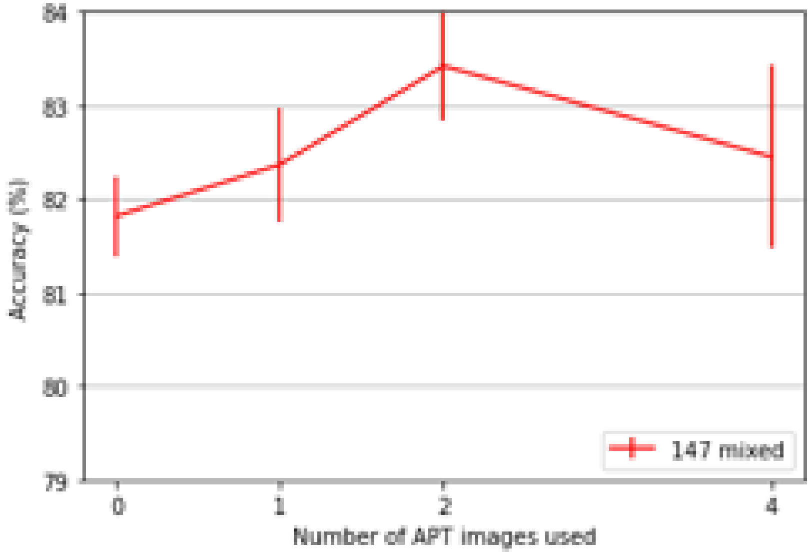
Type	Train	%	Val	%	Total	%		Notes
Scan-level	118	100	29	100	147	100		
Scan-TE	47	39.83	11	37.93	58	39.45		Treatment effect
Scan-TR	71	60.17	18	62.07	89	60.55		Tumor recurrence
Slice-level	1770	100	435	100	2205	100		
Label 0	333	18.81	83	19.08	416	18.87		
Label 1	622	35.14	143	32.87	765	34.69		
Label 2	494	27.91	106	24.37	600	27.21		
Label 3	321	18.14	103	23.68	424	19.23		
Normal	333	18.81	83	19.08	416	18.87		Label 0 presence
Abnormal	1437	81.19	352	80.92	1789	81.13		Label 1 or 2 or 3 presence
Tumor	943	53.26	246	56.55	1189	53.92		Label 1 or 3 presence
Recurrent	622	35.14	143	32.87	765	34.69		Label 1 presence



147-case mixed datasets

- 147 post-treatment cases with slice-level annotations
- 3 different versions
  - APT-1: single apt image for each slice
  - APT-2: average of 2 APT images if available
  - APT-4: average of 4 APT images if available
- Results
  - APT-4 likely suffered from inhomogeneous data
  - Using APT-2 is better than no APT (p=0.90)





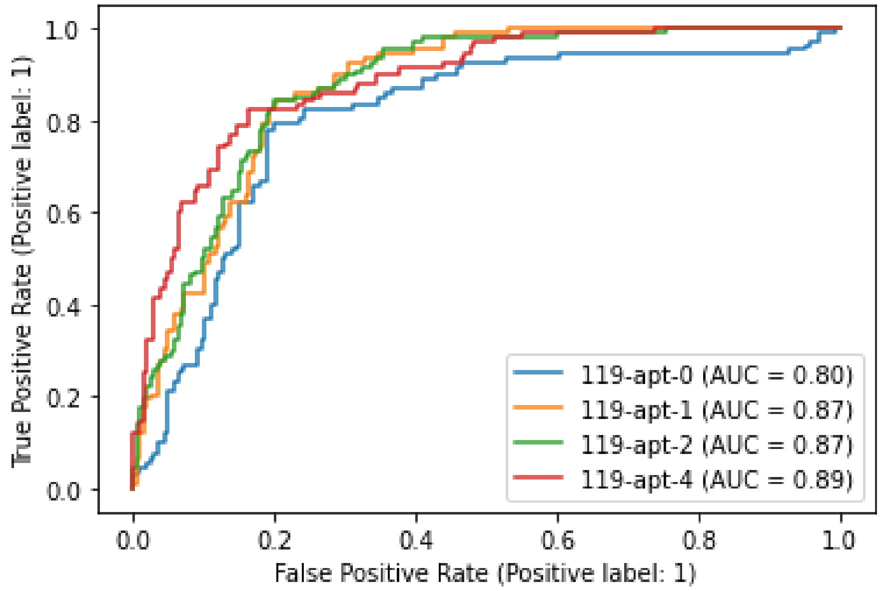
APT-1	Train	%	Val	%
apt-1	118	100	29	100
Total	118	100	29	100

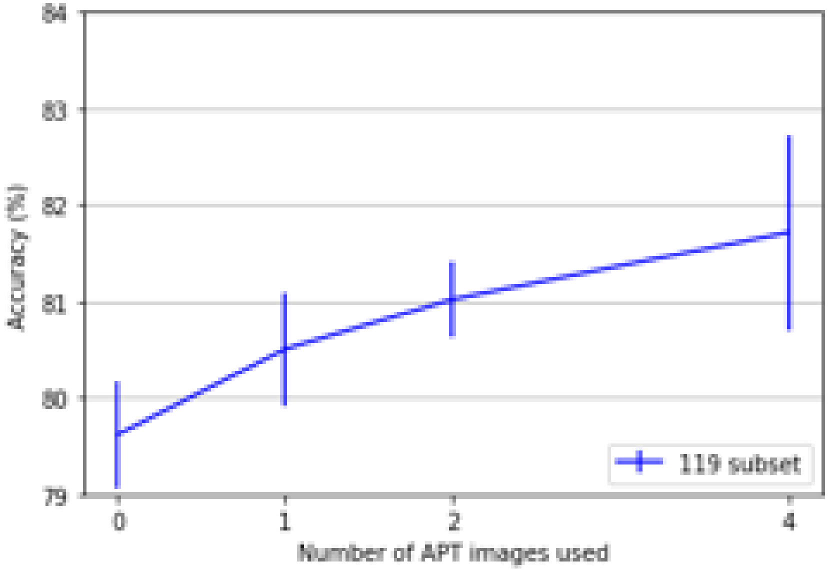
APT-2	Train	%	Val	%
apt-1	2	1.69	4	13.79
apt-2	116	98.31	25	86.21
Total	118	100	29	100

APT-4	Train	%	Val	%
apt-1	2	1.69	4	13.79
apt-2	18	15.25	4	13.79
apt-4	98	83.06	21	72.42
Total	118	100	29	100

119-case subset

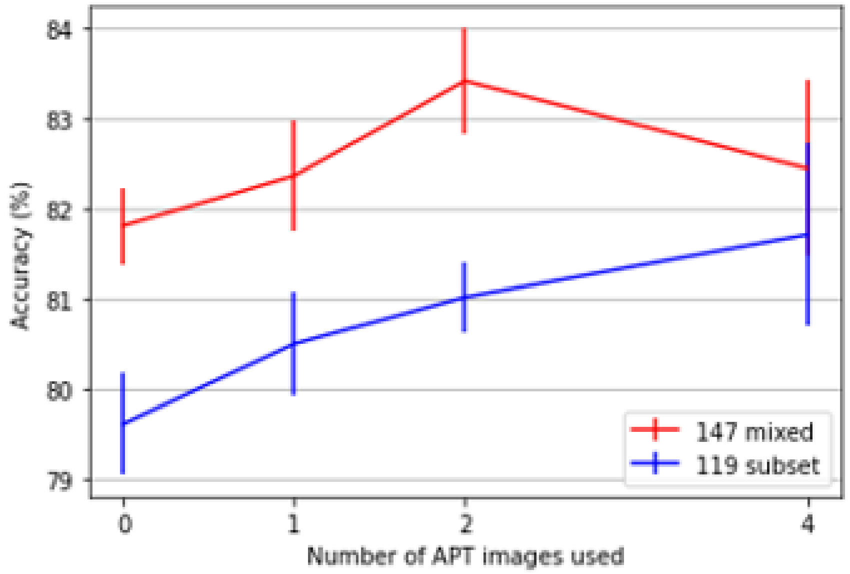
- All 119 cases have single, 2-average, and 4-average APT images
- Train and validation sets are subsets of those of 147-case dataset
- Results:
  - Expected increase in performance for both overall accuracy and AUC
  - Using 2-average is better than no APT (p = 0.90)





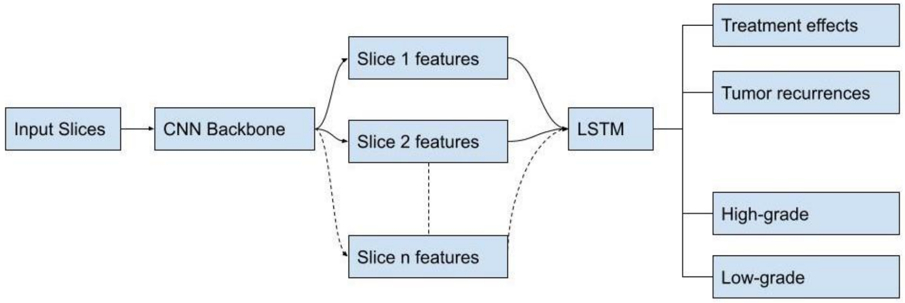
Quick Comparison

- Dataset:
    - 147 cases: 118 train – 29 val (4:1)
    - 119 cases: 98 train – 21 val (5.6:1)
- Need more homogeneous data and slice-level annotations to generalize results

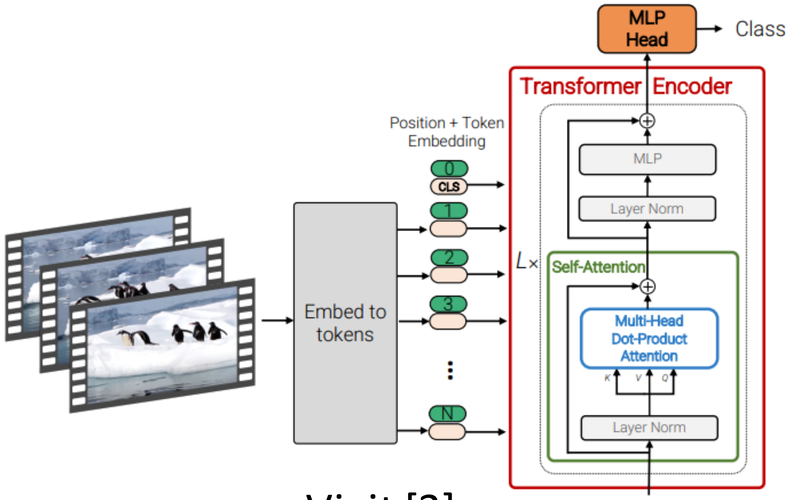


Baseline evaluation for scan-level classification

- Baseline 1: CNN-LSTM
  - Reproduced results of a previous study
    - CNN trained with hierarchical classification 147 cases
    - CNN output stacked as slice embedding
    - Accuracy: ~80%
  - Trained end-to-end from scratch using scan-level labels only
    - Low gradient flow to CNN
    - More overfitting on slightly larger dataset
    - Accuracy: 65 – 68%
- Baseline 2: Vivit
  - Trained end-to-end
    - Pretrained weights for ViT components
    - Overfitting
    - Accuracy: 65 – 70%



CNN-LSTM



Vivit [3]

Management

- Currently designing the improved method of augmenting vision transformer with memory
- ~1.5 weeks behind original schedule due to 2-week delay in annotations (expected 2/26 - received 3/10)