Diagnosing TMJ Arthritis

Team:
Kaylie Zhu, Symbolic Systems
Huaiyang Zhong, MS&E
Nicolas Bievre, Stats
Eric Loreaux, BioE

Mentors:
Prof Anand Rajaraman
Prof Jeffrey Ullman

Problem Description

Context

- Juvenile idiopathic arthritis is a chronic inflammatory disease that affects children
- It can present in the the jaw/temporomandibular joint (TMJ) - painful & disfiguring
- Early detection via MRI and subsequent rapid treatment is critical
- **Problem:**
    - Large degree of variability in diagnosis amongst radiologists.
    - Very limited dataset (123 patients)

Project statement

- **Goal**: Accurately diagnose TMJ arthritis from MRI scans using neural networks
- **Data**: MRI scans of patients from the Bristol Hospital UK

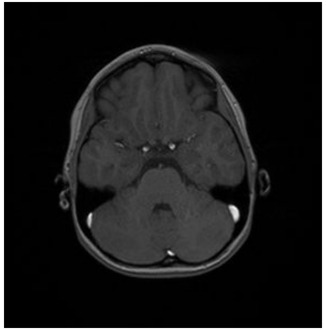
Data Processing

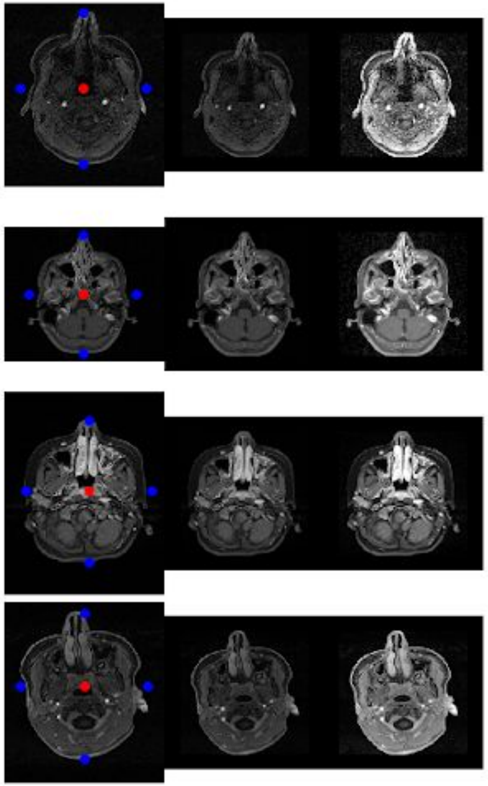
123 Patients

41- | 82+

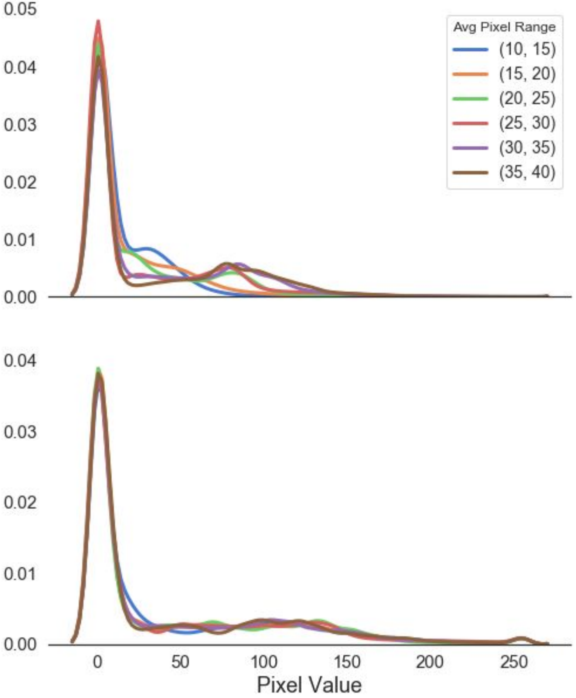
4,561 MRI Slices

Axial Orientation



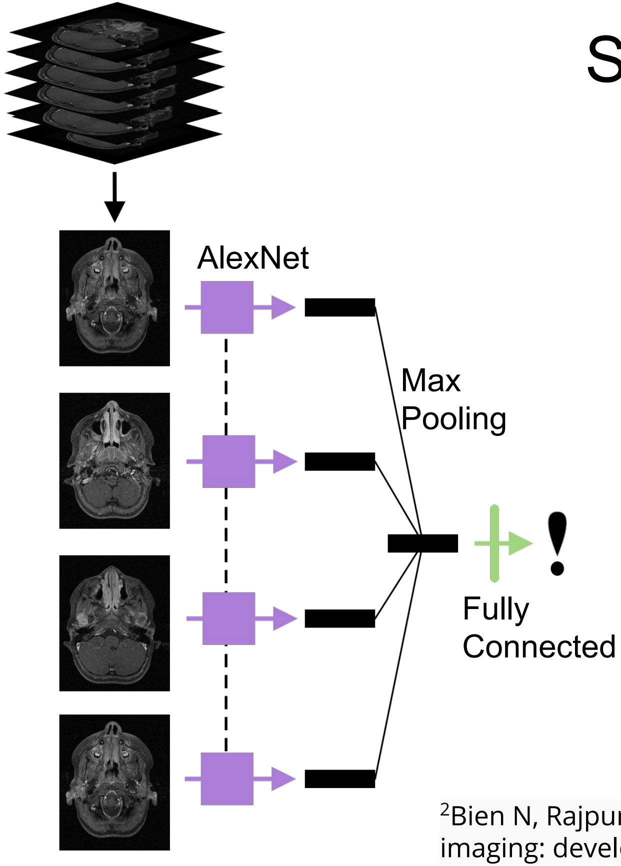


Resizing & Histogram Normalization[1]



[1] Nyu, Laszlo G. and Udupa, Jayaram K. On standardizing the MR image intensity scale. Magnetic Resonance in Medicine, 42:1072–1081, 1999.

Experimental Methodology - Series-Level Approach



Series-Level Approach

- All slices are stacked into one series and treated as a single training example
- Train the model to generate a single diagnosis

Pro: Fully end-to-end
Con: Requires a large number of patients

Example: MRNet[2] (2018)

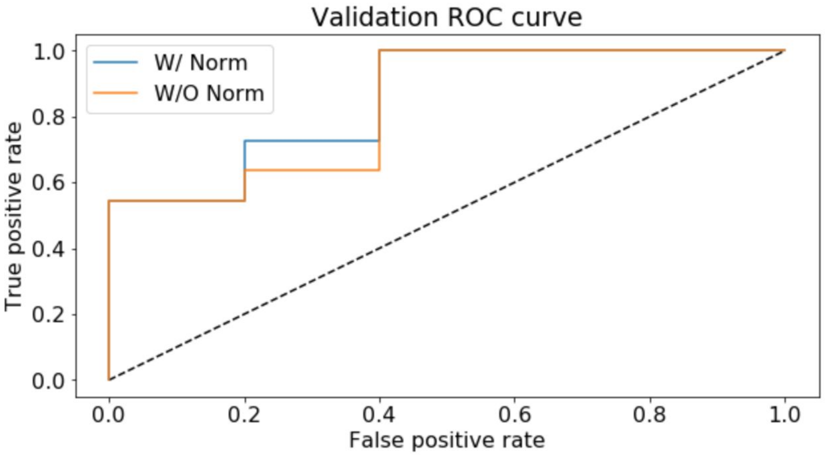
- Combine AlexNet activations across all MRI slices (Global MaxPool)
- Single fully connected layer
- Trained on ~1000 patients

[2]Bien N, Rajpurkar P, Ball RL et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Med2018;15(11):e1002699.

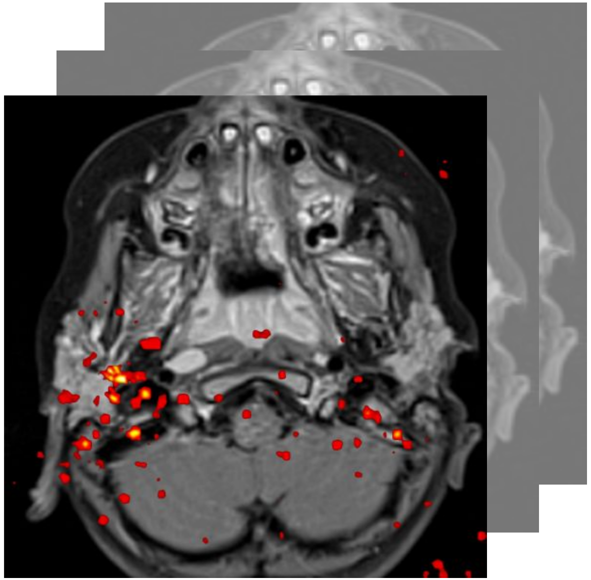
Results - Series-Level Approach

Quantitative Results

Model: AlexNet	Accuracy		F1		AUC	
	Train	**Val**	Train	**Val**	Train	**Val**
w/o Normalization	0.72	**0.875**	0.75	**0.91**	0.81	0.83
w/ Normalization	0.83	0.87	0.84	**0.91**	0.85	**0.85**



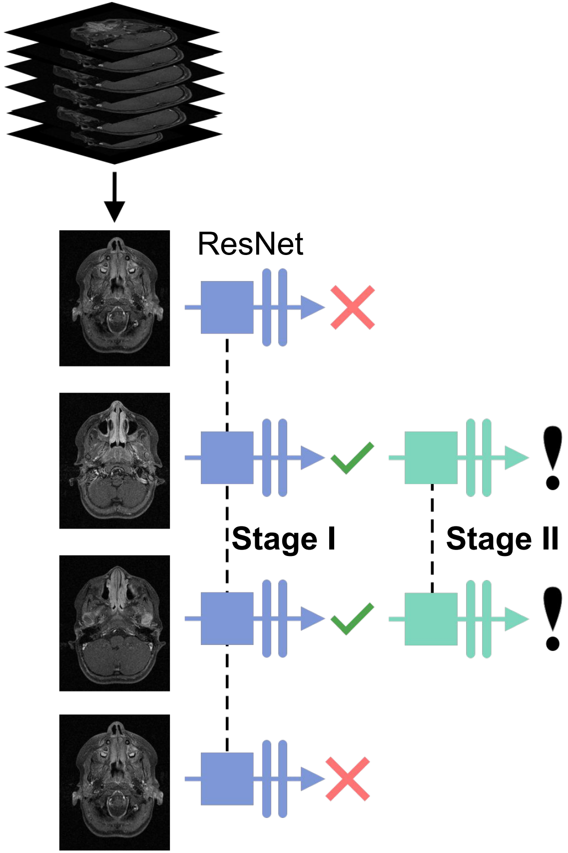
Qualitative Results



Saliency map

Patient with an abnormal left, normal right

Experimental Methodology - Slice-Level Approach



MRI Slice-Level Approach

- Treat each MRI slice as a single training example
- Train the models to generate slice-level votes
- Use a custom voting scheme to generate diagnosis

Pro: Requires fewer patients to train

- Many more disease applications

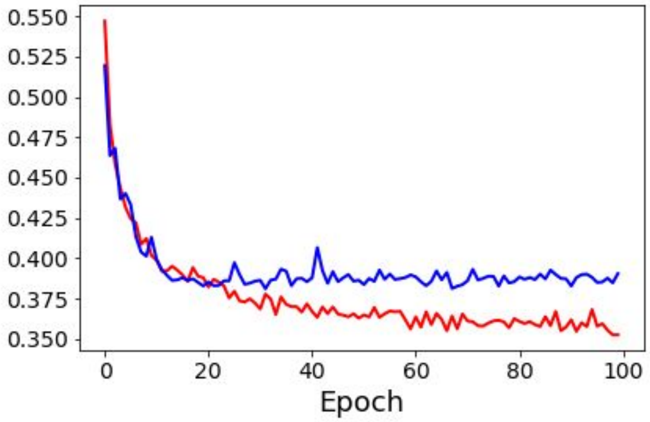
Con: Cannot train end-to-end

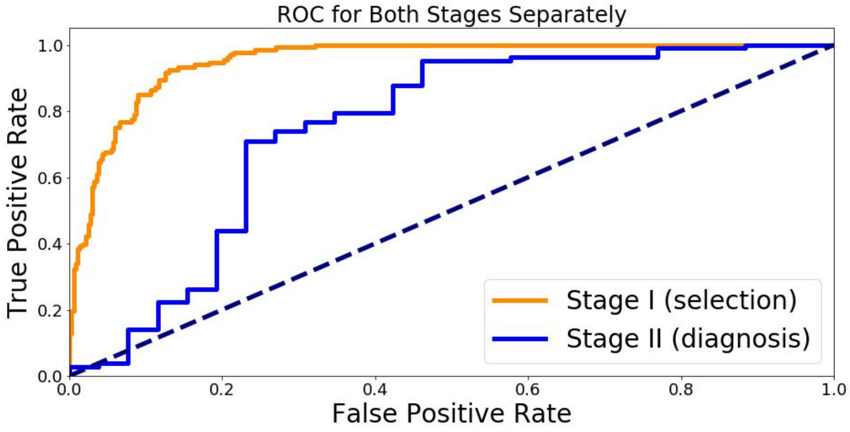
Our implementation

- Train first network to detect diagnostic relevance of each slice
- Train second network to make slice-level diagnosis
- Try combining models for multi-task approach

Results - Slice-Level Approach

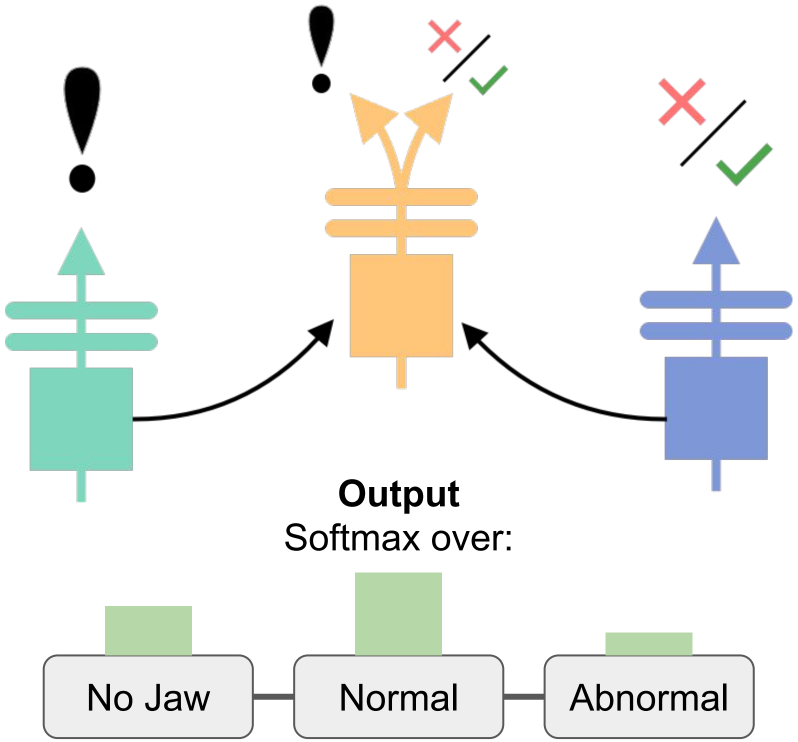
Stage	Model	Criteria	Accuracy		F1		AUC	
			Train	**Val**	Train	**Val**	Train	**Val**
I	**ResNet18**	Best F1	0.855	**0.898**	0.689	0.816	0.917	**.952**
		Best AUC	0.855	**0.898**	0.689	0.816	0.917	**.952**
II	**ResNet34**	Best F1	0.660	**0.872**	0.765	**0.923**	0.702	0.746
		Best AUC	0.735	0.805	0.805	0.874	0.786	**0.748**





MRI Slice-Level Results - Multitask

Goal: Combine both tasks into a single model



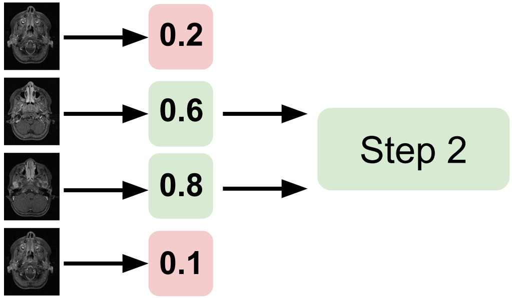
Quantitative Results

Model	Accuracy		F1	
	Train	**Val**	Train	**Val**
Resnet18	.82	.77	.73	.70
Resnet34	.83	.75	.74	.70
Resnet50	.87	.79	.76	.72
Densenet121	.93	.83	.78	.74
PNASNet-5-Large	.91	**.84**	.79	**.745**

Voting Scheme

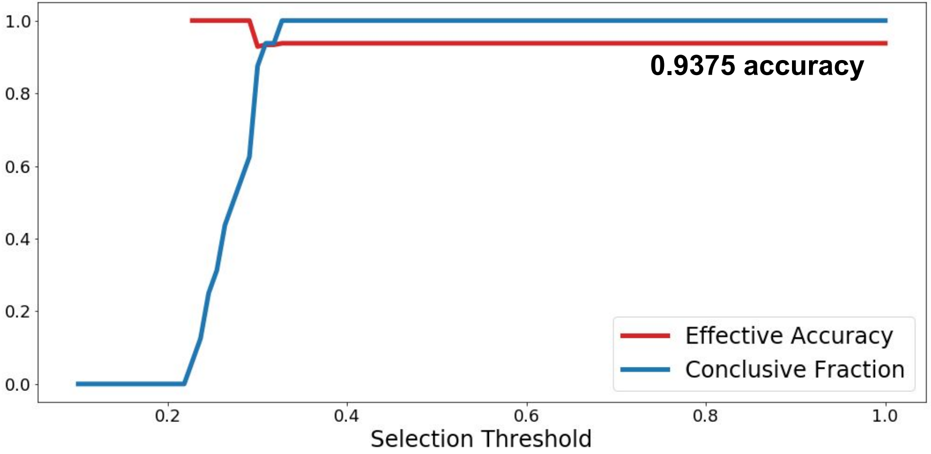
How to combine slice-level model outputs into a diagnosis?

Step 1: Select slices by diagnostic relevance using a selection threshold on Stage I output



Step 2: Combine Stage II outputs for selected slices using voting scheme to diagnose

Multitask model w/ global max voting scheme



Potential voting schemes:

- Global max
- Weight predictions by relevance

Future Works

- Continue tuning the models
  - Directly tune F1/AUC via alternative loss functions[3]
  - Tune voting schemes and parameters
- Introduce more complex disease labels
  - Left/Right specific
  - Mild/moderate/severe
- Incorporate other MRI imaging orientations and modalities
  - Coronal/Sagittal orientations
  - T1/T2/T1 fs/T2 fs modalities
- Incorporate expert radiologists into analysis
  - Quantify current expert biases
  - Assess differential performance using model as decision support

[3]Eban et al. Scalable Learning of Non-Decomposable Objectives. AISTATS, **54**. 2017.

Thank you