Project 19: Glioma classification and biopsy guidance with multimodal MRI and Deep Learning

Paper critical review

Nhat Le

March 2022

1. Project Goals

Our project focuses on develop a deep learning method for classification tasks including glioma grading and assessment of tumor recurrence for post-treatment patients using their multimodal MRI scans. We also put an emphasis on the analysis of using additional images of Amide Proton Transfer-weighted (APTw), an MR sequence developed in Johns Hopkins, besides the conventional MR images in such an automated pipeline. While APTw imaging has shown clinical values in aforementioned tasks, using it effectively with deep learning faces challenges from small dataset and other intrinsic features of the sequence.

2. Paper Selection

The paper we review in this summary is “Learning-based analysis of amide proton transfer-weighted MRI to identify tumor progression in patients with post-treatment malignant gliomas” by Guo et al., which is the last study from Dr. Jiang research group to tackle similar problems with the subset of the dataset we are using for this project. The paper is recently submitted and under review, and it provides very pertinent information for our project. Due to certain reasons, the current publicly available version of this paper does not have the figures and tables shown in this critical review, which we get directly from the authors.

3. Critical Review

a. Summary of problem and key results

This paper develops and verify a deep learning method to identify tumor progression in post-treatment glioma patients using conventional MR and APTw images as input. The key results are:

- A convolutional neural network (CNN) could learn from the data to predict tumor recurrence at slice-level classification
- Additional modules such as a learnable subtraction layer, hierarchical classifications, long short-term memory (LSTM) helps improve the performance and extend the task to patient-level classification (AUC increased from 0.85 to 0.90 for slice-level)
- Using APTw images improve the performance both at slice-level and scan-level classification (AUC increased from 0.88 to 0.90 for slice-level with $p<0.001$and from 0.85 to 0.90 for scan-level with $p<0.05$)
- Using class activation map could broadly highlight the tumor regions

b. Significance of key results

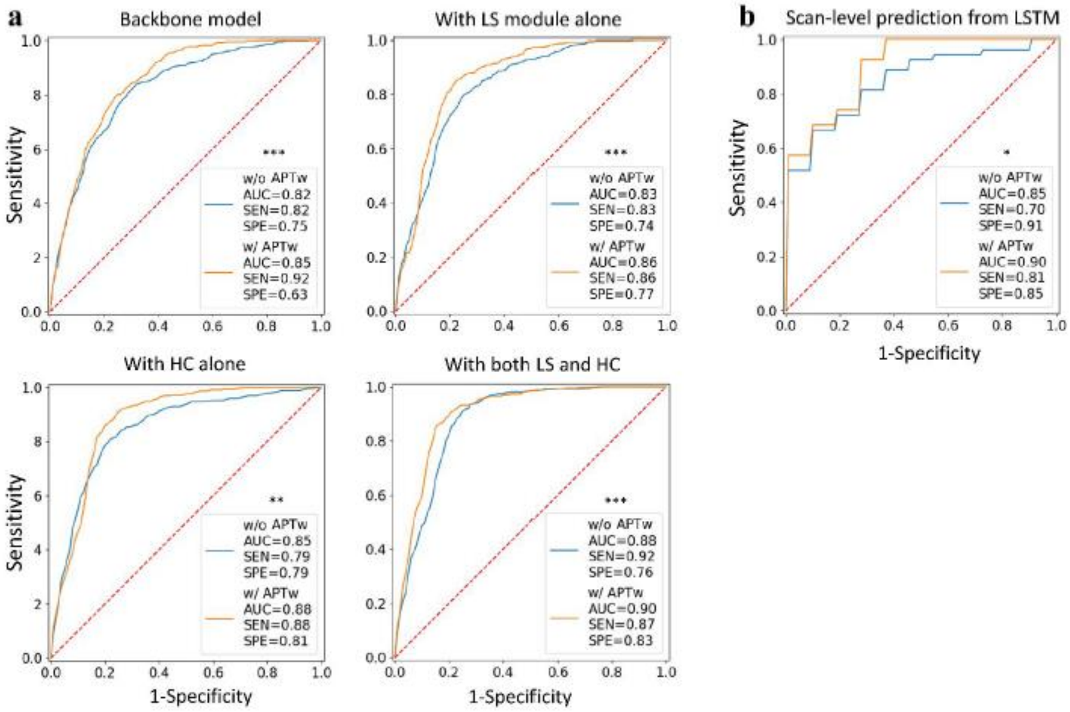
This paper is the first time APTw imaging is used in a deep learning pipeline to predict tumor recurrence and the result is promising. It also presents meaningful ways of augmenting an existing CNN to improve the performance when training with a dataset of small/medium size.

Table 2: Comparisons of performances with different MRI sequence data as input for the slice-level classification in the backbone model

Data Input					Diagnostic Performances			
APTw	GdT$_1$w	T$_1$w	T$_2$w	FLAIR	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	P value#
		√	√	√	0.77 (0.70, 0.81)	0.61 (0.53, 0.69)	0.81 (0.77, 0.85)	-
	√	√	√	√	0.82 (0.79, 0.87)	0.82 (0.75, 0.88)	0.75 (0.70, 0.80)	$P < 0.001$
√		√	√	√	0.84 (0.81, 0.89)	0.87 (0.81, 0.92)	0.68 (0.62, 0.73)	$P < 0.001$
√	√	√	√	√	0.85 (0.82, 0.89)	0.92 (0.87, 0.96)	0.63 (0.58, 0.69)	$P < 0.001$

Note. √ indicates that the input instance contains the corresponding MR sequence.

P value for the ROC curve comparison with respect to the input of T$_1$w, T$_2$w and FLAIR data.

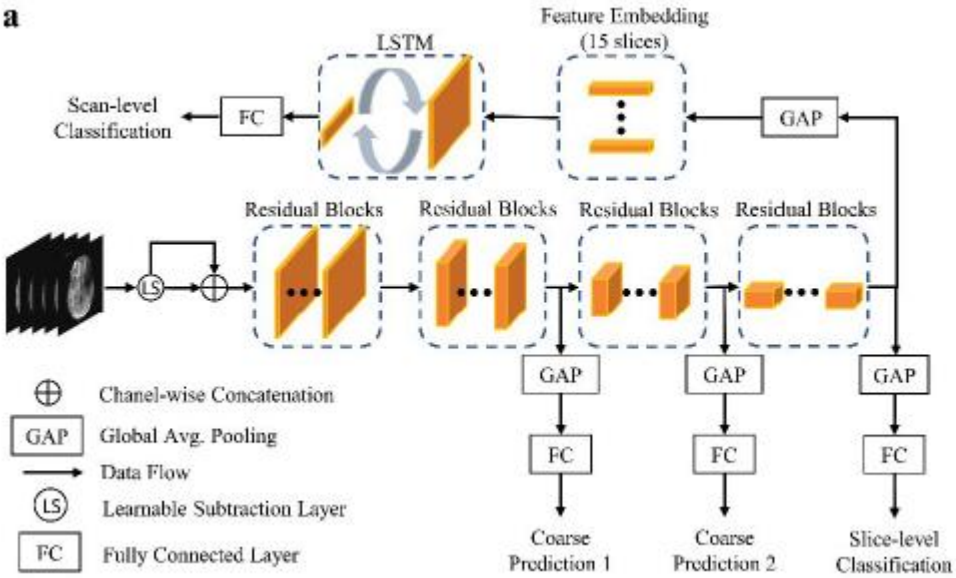


c. Experimental methods

- Overview:

The scan pixel resolution is 256x256x15 for each sequence. Different sequences are co-registered together during the preprocessing steps. For each slice, a 256x256 image from each sequence will be fed into corresponding input channel of the deep learning model.

This is an overview of the key CNN-LSTM architecture in this paper:

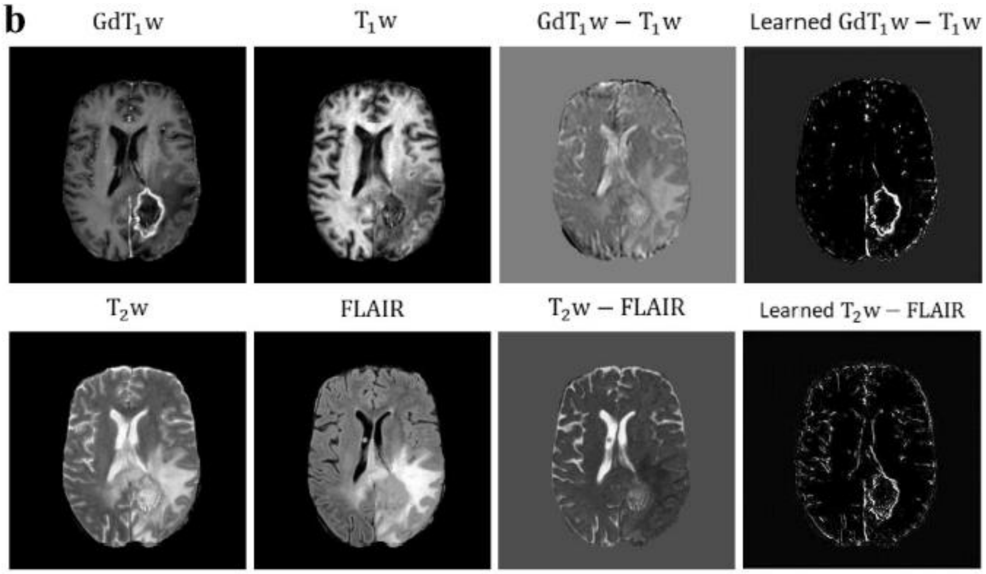


Overall, the "residual blocks" represent the well-known ResNet-18 backbone, which acts as feature extractor in this combined network. They are trained using hierarchical slice-level classification tasks to predict abnormal against normal slices, tumor against no-tumor slices, and recurrent tumor against non-recurrence. This training paradigm improves the overall performance by increasing gradient flows. The extracted feature vector for each slice in a scan is then stacked together as input to train the LSTM which performs the scan-level classification.

- Learnable subtract module

A learnable subtraction module learns to calculate the 'soft' difference between closely related MR sequences to provide additional input channel to the CNN backbone, including between T1w and GdT1w and between FLAIR and T2w images. It highlights the abnormal regions caused by tumors or the post-treatment cavity. It is illustrated in the following visualization:

b



- Ablation study:

To evaluate the additional benefits of certain image modality and network component, the author performs an ablation study with the schemes and results shown in the table and graph in the result section above.

d. Our assessment
    - Strength:

    This paper has developed a good preprocessing pipeline for the dataset to make it suitable for deep learning usage. For the model architecture, it gives motivation behind each of the key designs and conducts a thorough ablation study to show prove these points. The author also mentions common algorithms in similar medical imaging problems such as the 3D CNN family and clear explanations of why they will not work well with our dataset. This helps us avoid certain pitfalls in conceptualizing our improved methods.

    Furthermore, the slice-level classification method with hierarchical classification (HC) paradigm in the paper gives us a straightforward way to do our analysis of the effectiveness of using multiple APTw images in deep learning pipeline.

    - Weakness:

    There are a few things we think that the author could have done differently in this study. First, the decision to use chronological order to split the dataset into train, validation, and test sets was not explained in the paper. This prevents the use of k-fold cross validation method, which is commonly used in this kind of study to evaluate the performance of learned model. Secondly, there is a significant lack of documentation in the paper on the details of implementation and training of the deep learning models. As we get access to the source code (not publicly available), we can point out certain details. For example, the author pretrains the CNN on the slice-level classification tasks first, loads the pretrained weights to the CNN backbone in the combined CNN-LSTM, and freezes them while training the combined one on the patient-level classification tasks. Another example is the absence of the list of data augmentation techniques used in training, and some of which are not trivial. Without details like this or the source code being made public, the level of performance obtained in this paper will not be reproducible.

- Possible next steps for this work:

Our project extends from this paper and uses their architecture as one of the models for baseline evaluation. The main idea of using CNN to extract embedded features and using LSTM to stack them up could be transferred to more recent and powerful deep learning architecture for sequence modeling, such as the family of (vision) transformer model. We also plan to use the extra number of APTw images that is available but not used in this paper to propose meaningful methods of improving its suitability for deep learning methods.

4. Conclusions

This paper provides a validated deep learning method for the glioma classification task of tumor recurrence versus treatment effects. The key contributions of the model designed in the paper come from both deep learning and medical imaging expertise. The author insights of the problem and the dataset are very valuable for us in our project.

Reference

Guo, Pengfei and Unberath, Mathias and Heo, Hye-Young and Eberhardt, Charles and Lim, Michael and Blakeley, Jaishri and Jiang, Shanshan, Learning-Based Analysis of Amide Proton Transfer-Weighted MRI to Identify Tumor Progression in Patients with Post-Treatment Malignant Gliomas. Available at SSRN: https://ssrn.com/abstract=4049653 or http://dx.doi.org/10.2139/ssrn.4049653