

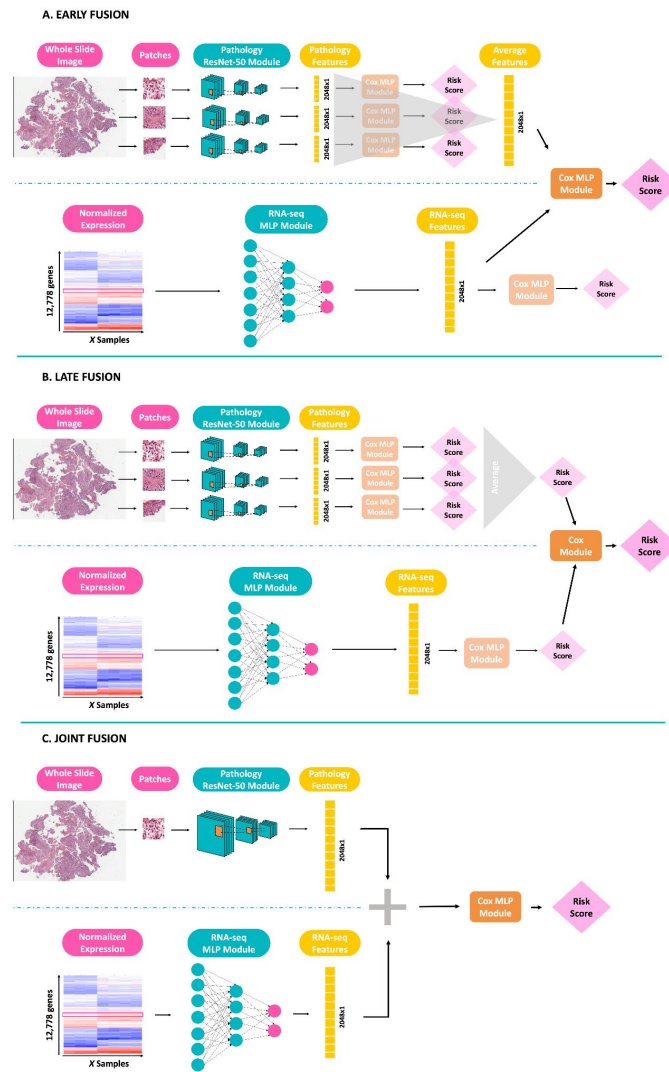
Multimodal Representation Learning for Healthcare integrating Imaging and Genomic data

Ekin Beyazit and Koushik Howlader

Goal

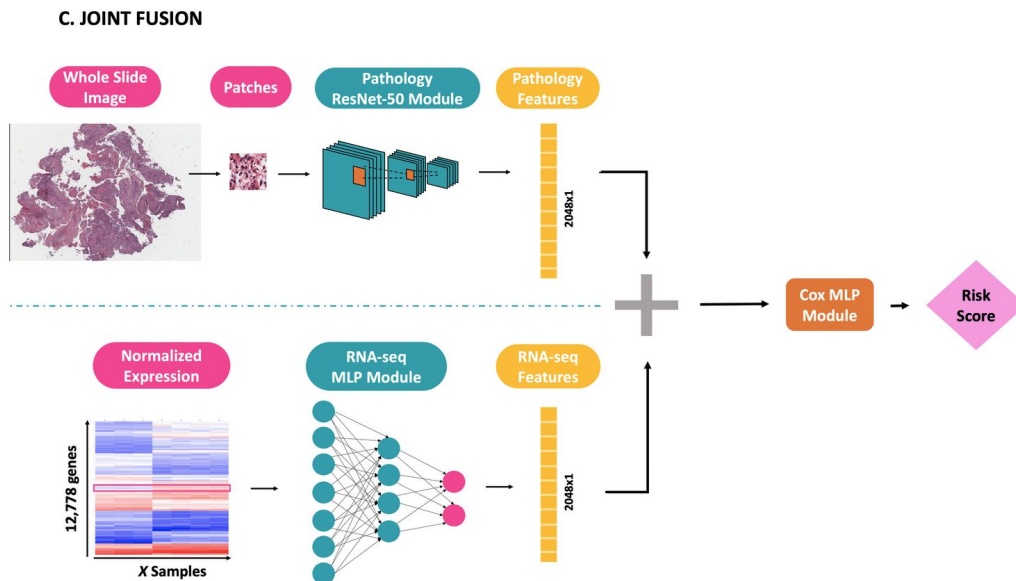
We tried to advance the existing methods in multimodal brain tumor survival prediction.

- We did this by using cross attention while aggregating patch features, doing a weighted average (whereas the paper is only averaging predictions, or features depending on the fusion type).

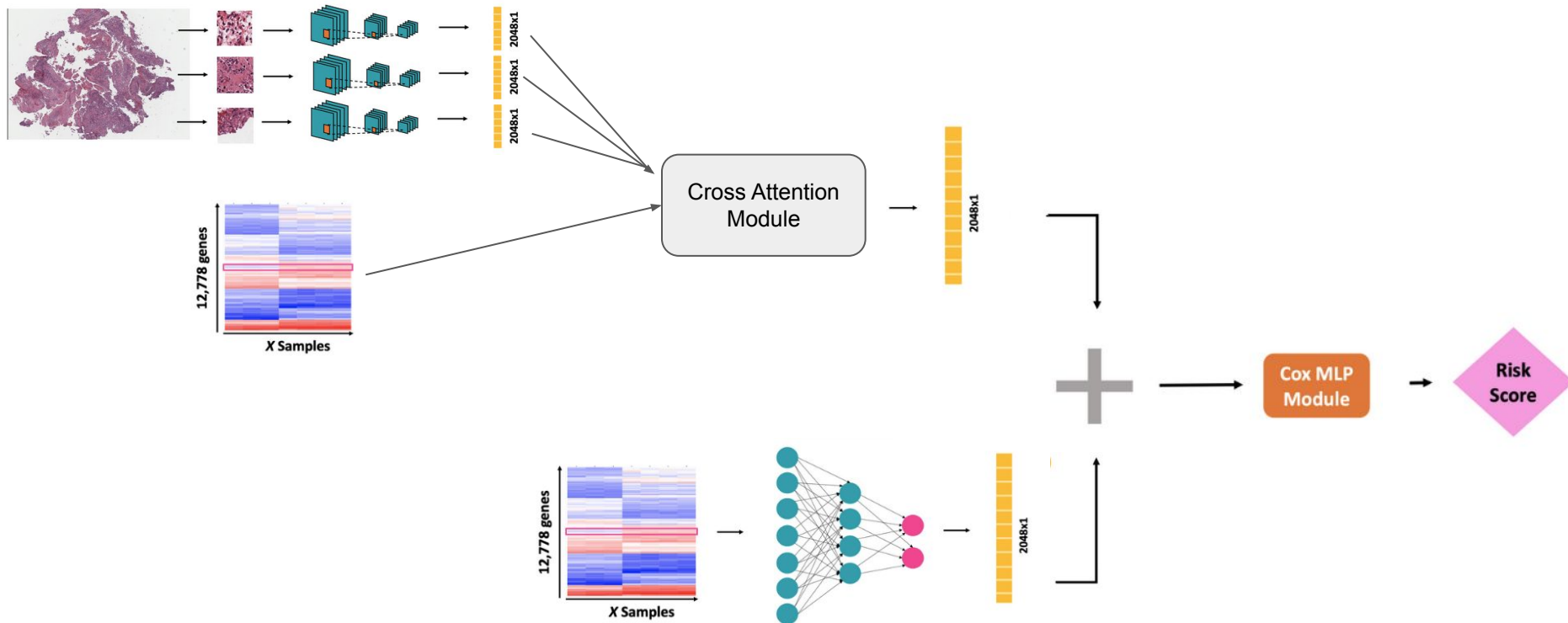


Approach

- We tackled the joint fusion method proposed in the paper “Multimodal data fusion of adult and pediatric brain tumors with deep learning”.



Approach



Approach

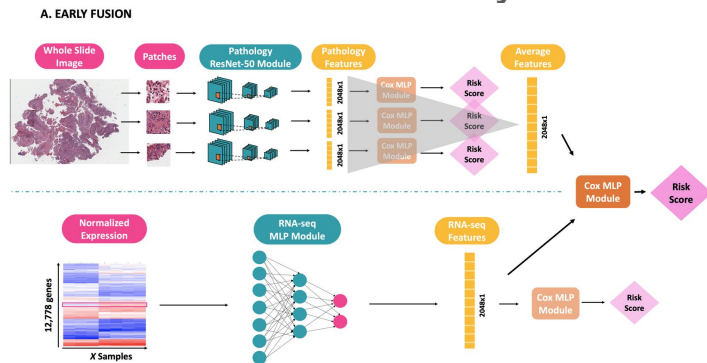
- We use the feature vectors as key and value vectors, while using the RNA sequence as a query vector.
- These operations essentially apply weights to the feature vectors, that are then averaged.
- We do not interfere with the rest of the architecture.

```
class TanhCrossAttention(nn.Module):
    def __init__(self, dim=2048):
        super(TanhCrossAttention, self).__init__()
        self.dim = dim
        self.gene_linear = nn.Linear(11047, dim, bias=False) # query vector coming from RNA sequence
        self.linear = nn.Linear(dim, dim, bias=False)

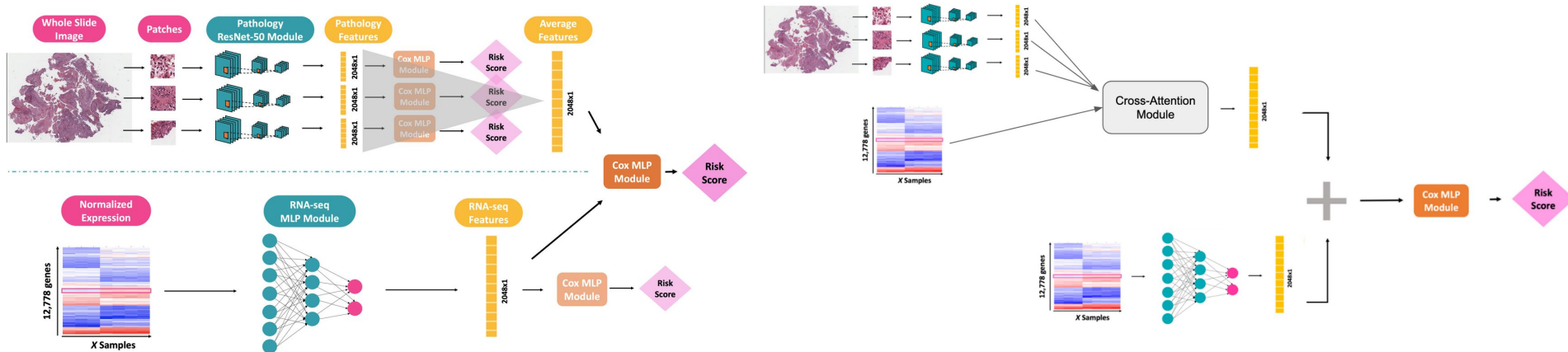
    def forward(self, x, gene_expression):
        gene_projection = self.gene_linear(gene_expression)
        logits = torch.tanh(self.linear(x)).matmul(gene_projection.unsqueeze(-1))
        attention_weights = torch.nn.functional.softmax(logits, dim=1)
        out = x * attention_weights * x.shape[1]
        return out, attention_weights
```

Challenges

- Doing multiple forward passes per sample seems to be messing with the loss, causing very different loss values based on bag_size.
- The baseline model proposed in the paper has very high dropout rates, which prevent our model from learning.
- The models take very long to train, and are very resource heavy in general, which makes it harder to do parameter tuning.
- Our approach may be more in line with early fusion rather than joint fusion.



A. EARLY FUSION



Dataset

Adult glioma cohort - (Brain Cancer Cell) - TCGA

Data Type:

- WSI (Whole Slide Image) - 844
- Gene Expression Data - 507

One patient can have multiple WSI

Results: by WSI and Case ID (Patient)

Due to resource constraints we consider 250 WSI (randomly).

For each WSI, prepared 2000 patches.

Data Split:

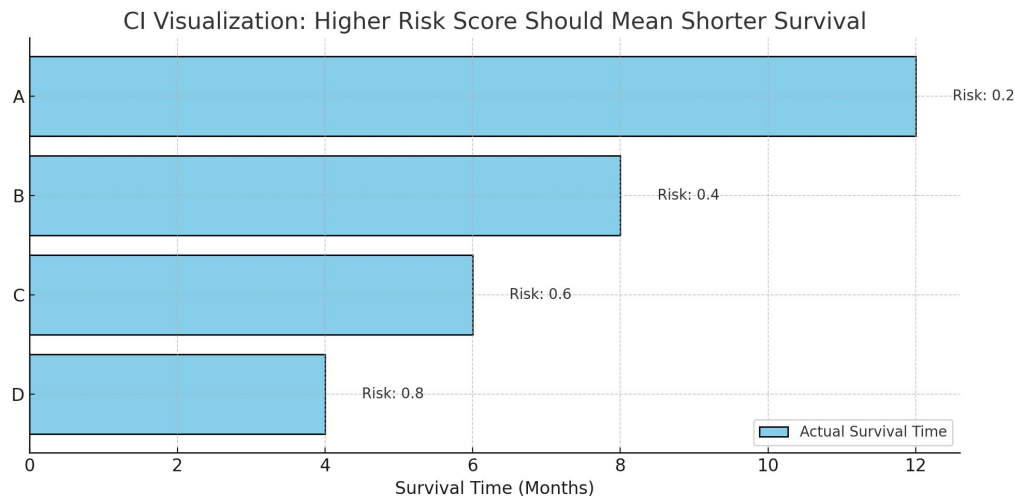
- Train - 70%
- Test: 20 %
- Validation: 10%

Evaluation Metric

CI (Concordance Index)

How well the predicted risk scores align with the actual survival outcomes.

- $CI = 1.0 \rightarrow$ perfect prediction
- $CI > 0.5 \rightarrow$ Better prediction
- $CI < 0.5 \rightarrow$ worse than random



Preliminary Results

Batch size	Epochs	Bag size	Dropout	Learning rate	CI
128	19,18,8,5,16	32	0	0.001, 0.001	WSI: 0.62 Case:0.5806
128	20,20,13,20,7	32	0	0.01, 0.001	WSI:0.55 Case:0.51
256	20,6,19,2,16	16	0	0.001, 0.001	WSI:0.61 Case:0.56

Baseline

Batch size	Epochs	Bag size	Dropout	CI
256	17,20,18,17,15	1	0.8	WSI: 0.656 Case: 0.610
128	20,18,18,16,9	1	0.8	WSI: 0.659 Case: 0.621
128	17,5,8,11,12	1	0.5	WSI: 0.659 Case: 0.620

Comparison

	WSI		Case	
Batch Size	MBS	Ours	MBS	Ours
128	0.659	0.62	0.621	0.5806

Challenges

- Resources
- We didn't perform with higher BagSize

Future Work

- Working with early fusion rather than joint fusion, as that is more aligned with our current approach.
- Trying to normalize the loss values such that they are comparable between different experiments.
- Doing proper grid-search to find the best parameters for both models.
- Displaying attention values for clarity.
- Using other kind of aggregation methods after weighting (such as pooling).

References

Steyaert, S., Qiu, Y.L., Zheng, Y. *et al.* Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. *Commun Med* 3, 44 (2023).

Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Journal of the American Medical Association* 1982; 247(18):2543--2546.

THANK YOU