University of Cambridge

MPhil in Data Intensive Science

DiS MPhil Project 19

Executive Summary for 19 Pathomic Fusion: an interpretable attention-based framework that integrates genomics and imaging to predict cancer outcomes

Shizhe Xu
Email: sx263@cam.ac.uk
Submission Deadline: 30 June 2024
Total Word Count: 872

Executive Summary

Most deep learning-based approaches for cancer research rely solely on histology or genomics without leveraging complementary information from both sources. This project report aims to summarise, reproduce and improve the work in the paper published in 2020, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis", in which the novel fusion approach combines histopathology images, cell graph and genomic features for survival outcome prediction and grade classification. This multimodal fusion approach not only outperforms Cox proportion hazard models and unimodal deep networks trained separately on histology and genomic features, but also offers better interpretability of the correlation between gene expression and tissue morphology. In other words, the deep learning-based multimodal fusion model can provide additional clinical insights by incorporating heterogeneous data sources such as diagnostic slides and single cell sequencing. Based on our findings from ablation experiments, we conclude that pathomic fusion can achieve a higher Concordance Index (fraction of all pairs of samples whose predicted survival times are correctly ordered among all uncensored samples) and significantly improve image-omic-based patient stratification. In addition, its multimodal interpretability can identify new integrative biomarkers of diagnostic, prognostic, and therapeutic relevance.

To reproduce the results from the original paper, we train each unimodal network, including convolutional neural networks (CNN), graph convolutional networks (GCN), and self-normalising networks (SNN) to generate corresponding feature representations, which can be used for multimodal fusion. We also thoroughly examine data preparation steps such as cell graph reconstruction and genomic data alignment. For the novel fusion approach, we implement a gating-based mechanism that controls the expressiveness of features of each modality before computing the Kronecker product of feature representations. Our reproduced results validate performance improvement in multimodal networks (CNN+GCN+SNN) in terms of a higher Concordance index in survival analysis and fine-grained patient stratification. In addition to conducting quantitative analysis to compare model performance and different fusion strategies, we also apply the Grad-CAM and Integrated Gradients methods to generate local explanation of image, cell graph, and genomic modalities for individual patients, as well as global explanation of genomic modality across molecular subtypes. All evaluation plots in the original paper have been reproduced and improved in our report.

To obtain a comprehensive overview and deep understanding of multimodal networks, we start from scratch rather than using the provided data split. We create our own data splits for a 15-fold Monte Carlo cross-validation. The cell graph .pt files have also been reconstructed using nuclei segmentation and image regions of interest (ROIs). Several modifications have been made to update the original code, such as optimising the network architecture and the gating-attention mechanism, to achieve more accurate results in survival analysis and grade classification. Since the local and global explanation and visualisation code has not been provided, we explicitly rewrite and implement it to observe localisation of tumour cellularity and feature attribution shifts when conditioning on morphological features.

Potential improvements and novel methods are discussed at the end of this report. In the histology CNN, the VGG19 architecture and its layers are explicitly detailed. This can be optimised by importing pretrained models directly from 'torchvision.models' to create more structured code. We also leverage visual-language foundation models, specifically CONCH (CONtrastive learning from Captions for Histopathology), to offer a new perspective on image feature extraction. The VGG encoder in the original code can be replaced by

the CONCH encoder, which extracts the image embeddings effectively and produces new unimodal feature representations used for our multilinear fusion. CONCH also reduces the risk of data contamination without using public histology slides from platforms such as TCGA and GTEX.

In the histology GCN, nuclei segmentation is essential before performing unsupervised cell feature extraction and cell graph reconstruction. The original paper used the cyclic GAN and conditional GAN for segmentation. With advancements in transformer-based networks, we propose a vision transformer called CellViT, which can distinguish overlapping nuclei in segmentation and automatically identify cell types such as tumour, stroma, and necrosis.

In visualising and interpreting multimodal pathomic fusion, we explore alternative algorithms to Grad-CAM. Specifically, we compare the outputs of Grad-CAM++, EigenCAM, and LayerCAM with those of Grad-CAM. These variants might reveal missing details in intercellular interactions and the microvascular patterns surrounding each cell.

In addition to fusing feature representations from CNN/VGG, GCN and SNN as mentioned in the original paper, we extract additional feature representations using CONCH. We develop our own code to create data splits that store both CNN and CONCH features, and run the multilinear/quadrilinear fusion CNN+GCN+SNN+CONCH instead of the trilinear fusion CNN+GCN+SNN. By incorporating four unimodal feature representations, our fusion model achieves superior performance, yielding the highest Concordance index in the survival outcome analysis.

We have attempted to replace the multilinear perceptrons (MLPs) in the original code with Kolmogorov-Arnold networks (KANs), which have activation functions on edges instead of nodes. Moreover, every weight parameter is replaced with a univariate function parameterised as a B-spline. We have found that KANs deliver slightly better results when the input size of genomic features is relatively small. KANs might outperform traditional MLPs when training and testing with fewer parameters (smaller sample size), and it is feasible to test KANs with our fusion strategy in the near future.

The code, pretrained models, and checkpoints are available in our GitLab repository.