

Department of Physics



**UNIVERSITY OF
CAMBRIDGE**
Cavendish Laboratory

Pathomic Fusion Executive Summary

by

Deyi Zeng

June 2024

Supervised by

Dr Mireia Crispin-Ortuzar, Dr Zeyu Gao

Contents

1	Introduction	2
2	Task	2
3	Methodology	2
4	Sample Aggregation	2
5	Improvement by Pre-aggregation Trimodal Fusion Network	2
6	Conclusion	4
A	Figures and Tables	7
B	Generative Tools	13

1 Introduction

Currently, medical diagnosis relies heavily on the experience and knowledge of physicians, introducing subjectivity and leading to incorrect predictions. This underscores the importance of using statistical or deep learning models to develop objective tools that can evaluate the survival and risk conditions of patients. This project is particularly interesting and important because it aims to create such objective models, potentially improving the accuracy and reliability of objective medical diagnoses.

With the development of artificial intelligence and machine learning, different types of neural networks are extensively used in making cancer and tumor prognosis and diagnosis, based on the genomic profile and histologic tissues of patients. However, there is limited research related to making predictions by combining the genomic profile and the histologic tissues. *Our primary objective* in this project is to reproduce the results from [1] and compare the performance of classical uni-modal networks with state-of-the-art multi-modal fusion networks in making predictions. Another key contribution of us is the proposal of a pre-aggregation mechanism to capture the relationships between different histology information of the same patients.

Also, the performance of networks in stratifying the patients survival time is discussed. The Integrated Gradients (IG) [2] and Grad-CAM [3] are utilized to interpret the models.

2 Task

The datasets used in this project are The Cancer Genome Atlas - Glioblastoma Multiforme Lower Grade Glioma (TCGA-GBMLGG) and Kidney Renal Clear Cell Carcinoma (TCGA-KIRC). These datasets contain survival time, censor status, WHO grades, genomic profiles, and pathological slides. Also, cell graphs are created from the pathological slides, as the inputs of graph neural network.

There are two tasks. The survival outcome prediction focuses on predicting patient hazards, while the WHO Glioma Grade Classification predicts the probability of patients belonging to each grade.

3 Methodology

The goal of this project is to train a trimodal fusion network (Figure 1) that integrates genomic profiles, ROI images, and cell graphs to predict patient hazards for survival outcome prediction and classify patients' grades for WHO grade classification.

To train a large network from scratch can be time-consuming and inefficient. To address this, three unimodal networks are initially trained. The pathology-based CNN excels at implicitly extracting morphology features for prediction. The graph-based GCN utilizes explicitly and manually extracted features for prediction. The gene-based SNN makes predictions based on the genomic profiles of patients. Each of these well-trained networks extracts a 32-length feature vector.

A gating-based attention mechanism [4] is implemented to allow the interaction between feature vectors and reduce the model sensitivity to noisy data. This is followed by a Kronecker product, enabling a second interaction between features to explicitly captures interactions between features.

The omic, graph, and path features are input into a gating-based attention network, followed by the application of the Kronecker Product. Subsequently, an encoder flattens and transforms these high-dimensional features into length-64 vectors. Finally, a classifier is employed.

4 Sample Aggregation

Since a single patient may have multiple samples in the dataset, predictions for different samples from the same patient are aggregated as follows: the predicted hazards for the same patient are aggregated by taking the mean, while the predicted probabilities of grades are aggregated by taking the maximum.

5 Improvement by Pre-aggregation Trimodal Fusion Network

As introduced in the last section, the mean and maximum aggregation of multiple samples from the same patient is used in the original design. However, this ignores the correlations between different ROIs and cell graphs for the same patients during training.

To address this, we designed a pre-aggregation network (Figure 3) that packs all ROIs for the same patient into one sample, rather than splitting them into several samples as in the original design.

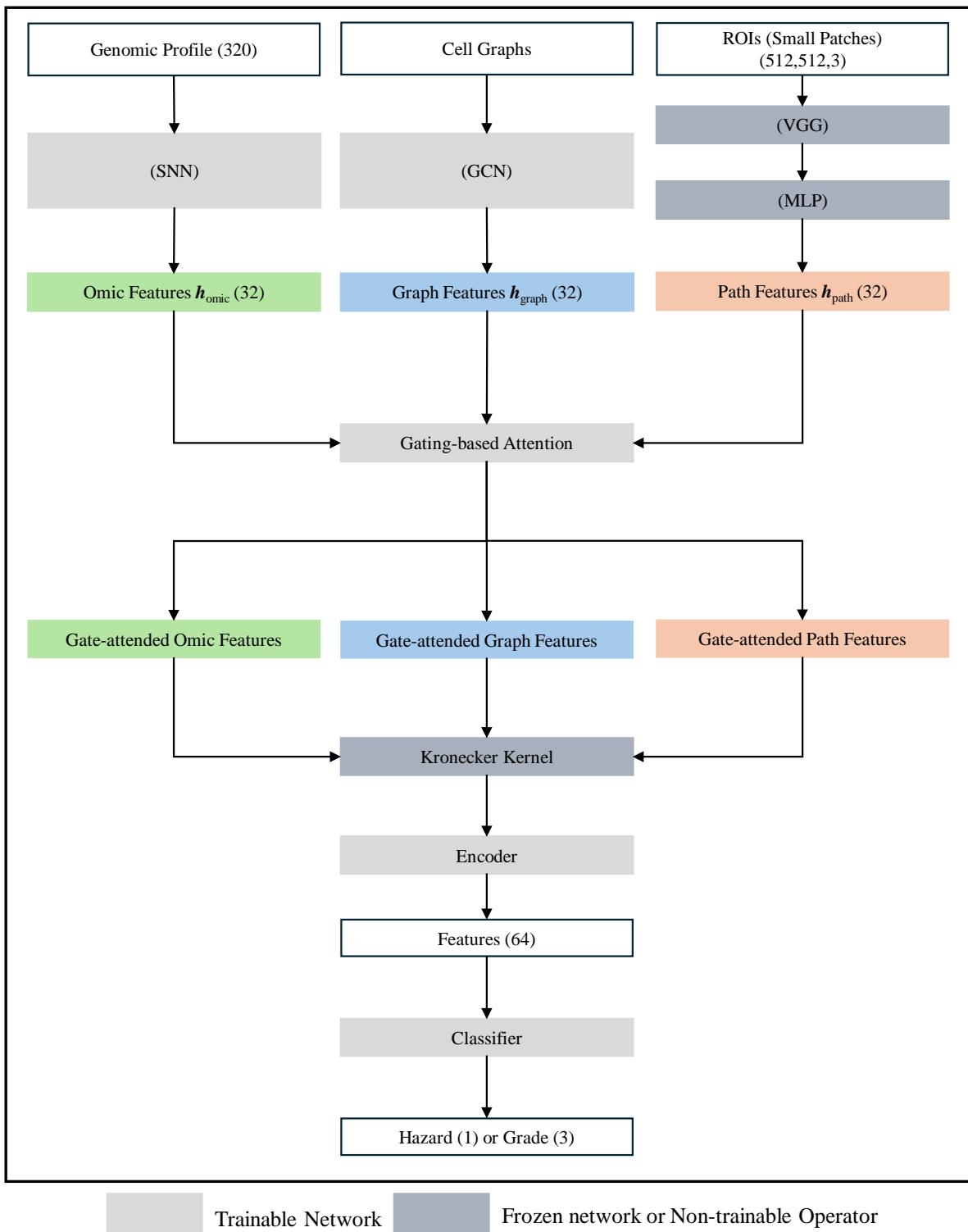


Figure 1: Trimodal Fusion Network for TCGA-GBMLGG and TCGA-KIRC (Training Version), freeze all blocks to make predictions or testing

Pre-aggregation involves concatenating the feature vectors, as illustrated in Figure 2, followed by feeding them into the attention network.

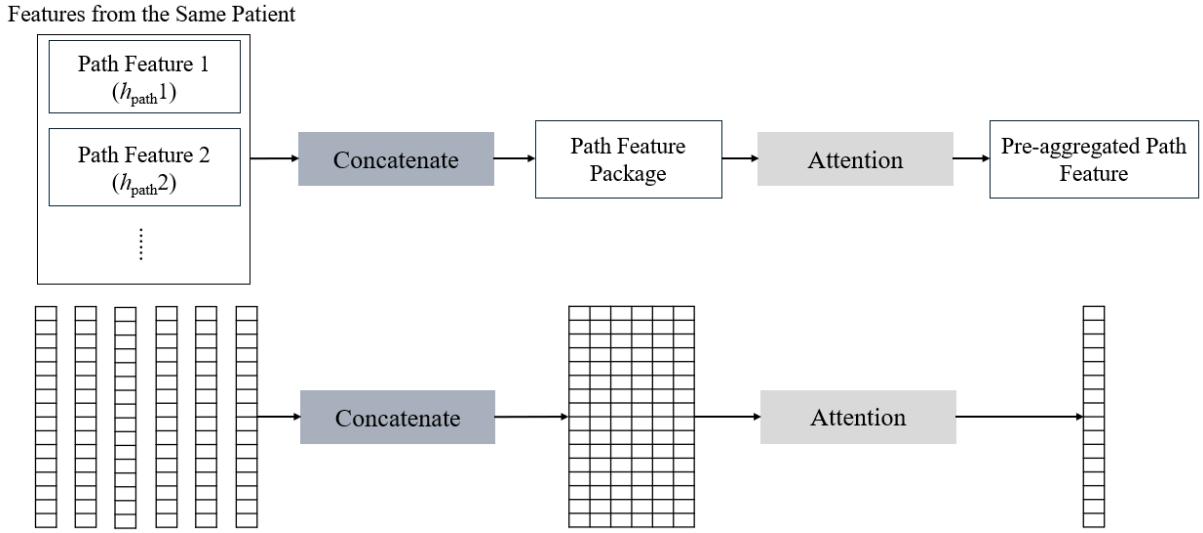


Figure 2: Pre-Aggregation

The pre-aggregation network structure is displayed in Figure 3, which is constructed by inserting another attention network between GCN/CNN and Gating-based attention to the original trimodal fusion Network in Figure 1.

6 Conclusion

In this project, the original paper [1] is reproduced and our results are similar to the original one.

In WHO Grade Classification of TCGA-GBMLGG task, the trimodal fusion network ($\text{CNN} \otimes \text{GCN} \otimes \text{SNN}$) ranks the highest in AUC, AP, F1-score, and F1 Grade II, Grade III, and Grade IV scores (Table 1).

In survival outcome prediction of TCGA-GBMLGG task (Table 2), the trimodal fusion network ($\text{CNN} \otimes \text{GCN} \otimes \text{SNN}$) by original paper ranks the third highest with 0.8254 C -Index . The pre-aggregate design improves the C -Index to the second highest 0.8318 ($(\cup \text{CNN}) \otimes \text{SNN}$) and highest 0.8361 ($(\cup \text{CNN}) \otimes (\cup \text{GCN}) \otimes \text{SNN}$). The stratification effect of ($\text{CNN} \otimes \text{GCN} \otimes \text{SNN}$) and ($(\cup \text{CNN}) \otimes (\cup \text{GCN}) \otimes \text{SNN}$) on low-to-high hazard patients are better than the other deep learning models, and on mid-to-high hazard patients are better than CNN model. According to IG (Figure 4), the IDH mutation and PTEN are the most influential genes. The Grad-CAM (Figure 6) tells that the CNN network pays the highest attention to the nucleoli and then the surrounding fried-eggs-like patches.

In survival outcome prediction of TCGA-KIRC task (Table 3), the bimodal fusion network ($\text{CNN} \otimes \text{SNN}$) ranks the highest with 0.7187 C -Index, followed by trimodal fusion network ($\text{CNN} \otimes \text{GCN} \otimes \text{SNN}$) with 0.7144 C -Index. The trimodal fusion ($\text{CNN} \otimes \text{GCN} \otimes \text{SNN}$) has the best low-to-high stratification performance among all models. According to IG (Figure 5), the CYP3A7 and DDX43 are the most influential genes. The Grad-CAM (Figure 7)tells that the CNN network pays the highest attention to the nucleoli and the edges of cells.

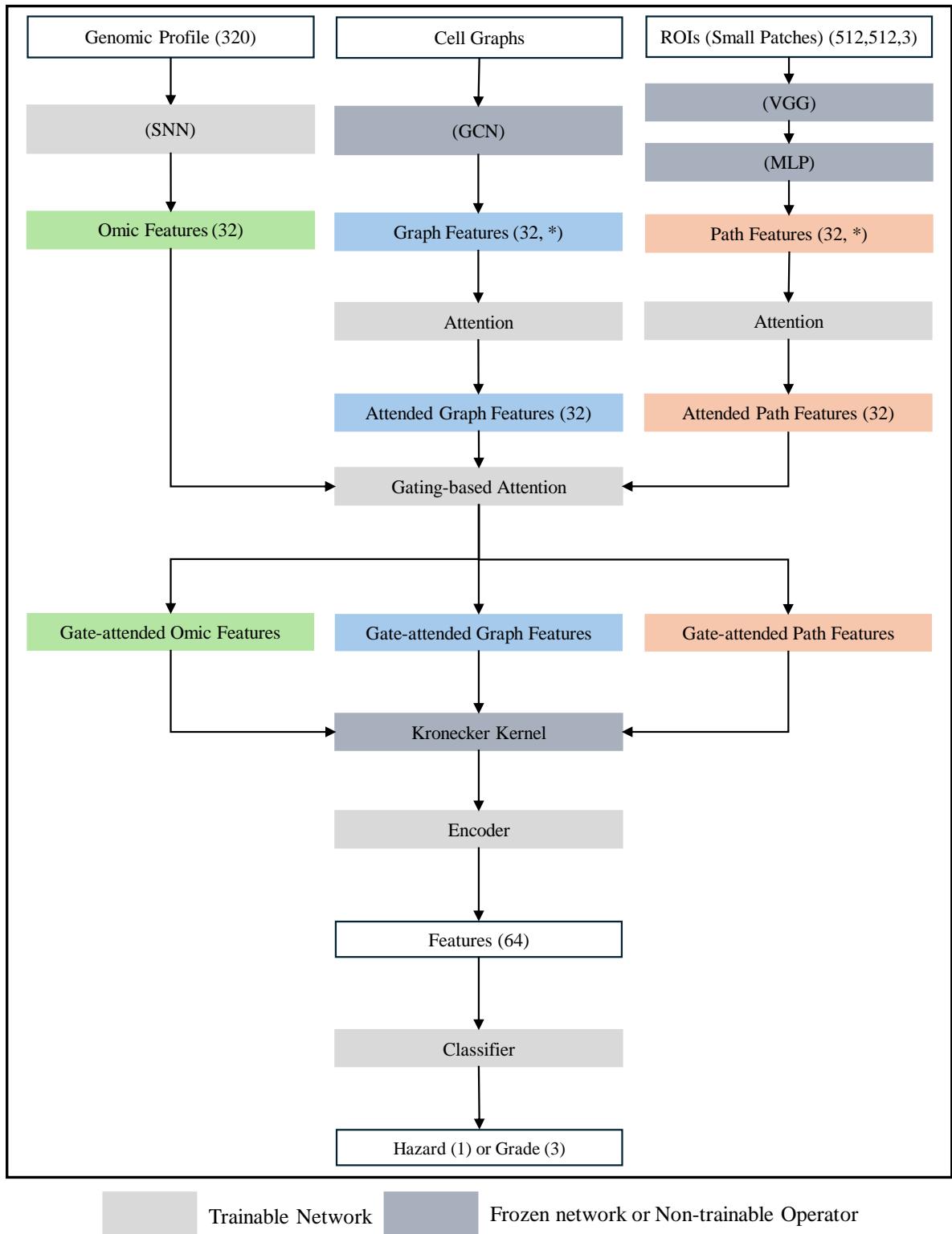


Figure 3: Pre-aggregation Tri-fusion Network (Training Version), freeze all blocks to make predictions or testing

References

- [1] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 41(4):757–770, 2020.
- [2] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.
- [3] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [4] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

A Figures and Tables

Table 1: Performance Metrics of WHO Grade Classification Task in TCGA-GBMLGG

Method	AUC	AP	F1-score	F1 Grade IV
Genomic SNN	0.8522 ± 0.012	0.7287 ± 0.018	0.6503 ± 0.016	0.8541 ± 0.018
Genomic (SNN \otimes SNN)	0.8529 ± 0.012	0.7295 ± 0.019	0.6503 ± 0.018	0.8552 ± 0.017
Histology GCN	0.8471 ± 0.014	0.7627 ± 0.017	0.6493 ± 0.024	0.8148 ± 0.027
Histology (GCN \otimes GCN)	0.8498 ± 0.011	0.7649 ± 0.016	0.6492 ± 0.015	0.8155 ± 0.019
Histology CNN	0.8823 ± 0.007	0.7791 ± 0.020	0.7152 ± 0.022	0.8799 ± 0.017
Histology (CNN \otimes CNN)	0.8875 ± 0.008	0.8083 ± 0.014	0.7176 ± 0.017	0.8734 ± 0.014
Pathomic F. (CNN \otimes SNN)	0.9050 ± 0.009	0.8338 ± 0.016	0.7330 ± 0.019	0.9141 ± 0.013
Pathomic F. (GCN \otimes SNN)	0.9018 ± 0.011	0.8243 ± 0.018	0.7318 ± 0.020	0.9127 ± 0.017
Pathomic F. (CNN \otimes GCN \otimes SNN)	0.9088 ± 0.010	0.8360 ± 0.018	0.7459 ± 0.021	0.9235 ± 0.013

Table 2: Concordance Index of Survival Outcome Prediction in TCGA-GBMLGG

Model	C-Index
Cox (Age+Gender)	0.7316 ± 0.012
Cox (Subtype)	0.7595 ± 0.011
Cox (Grade)	0.7379 ± 0.013
Cox (Grade+Subtype)	0.7769 ± 0.013
Genomic SNN	0.8024 ± 0.017
Genomic (SNN \otimes SNN)	0.7848 ± 0.012
Histology GCN	0.7478 ± 0.019
Histology (GCN \otimes GCN)	0.7374 ± 0.026
Histology CNN	0.7955 ± 0.016
Histology (CNN \otimes CNN)	0.7930 ± 0.014
Pathomic F. (CNN \otimes SNN)	0.8205 ± 0.011
Pathomic F. (GCN \otimes SNN)	0.8126 ± 0.014
Pathomic F. (CNN \otimes GCN \otimes SNN)	0.8254 ± 0.013
Pathomic F. ((\cup CNN) \otimes SNN)	0.8318 ± 0.011
Pathomic F. ((\cup GCN) \otimes SNN)	0.8239 ± 0.016
Pathomic F. ((\cup CNN) \otimes (\cup GCN) \otimes SNN)	0.8361 ± 0.015

Table 3: Concordance Index of Survival Outcome Prediction in TCGA-KIRC

Model	<i>C</i> -Index
Cox (Age+Gender)	0.6300 ± 0.0240
Cox (Grade)	0.6754 ± 0.0360
Genomic SNN	0.6828 ± 0.0260
Genomic (SNN \otimes SNN)	0.6820 ± 0.0270
Histology GCN	0.6403 ± 0.0300
Histology (GCN \otimes GCN)	0.6382 ± 0.0320
Histology CNN	0.6628 ± 0.0210
Histology (CNN \otimes CNN)	0.6706 ± 0.0230
Pathomic F. (CNN \otimes SNN)	0.7187 ± 0.0270
Pathomic F. (GCN \otimes SNN)	0.6884 ± 0.0250
Pathomic F. (CNN \otimes GCN \otimes SNN)	0.7144 ± 0.0280

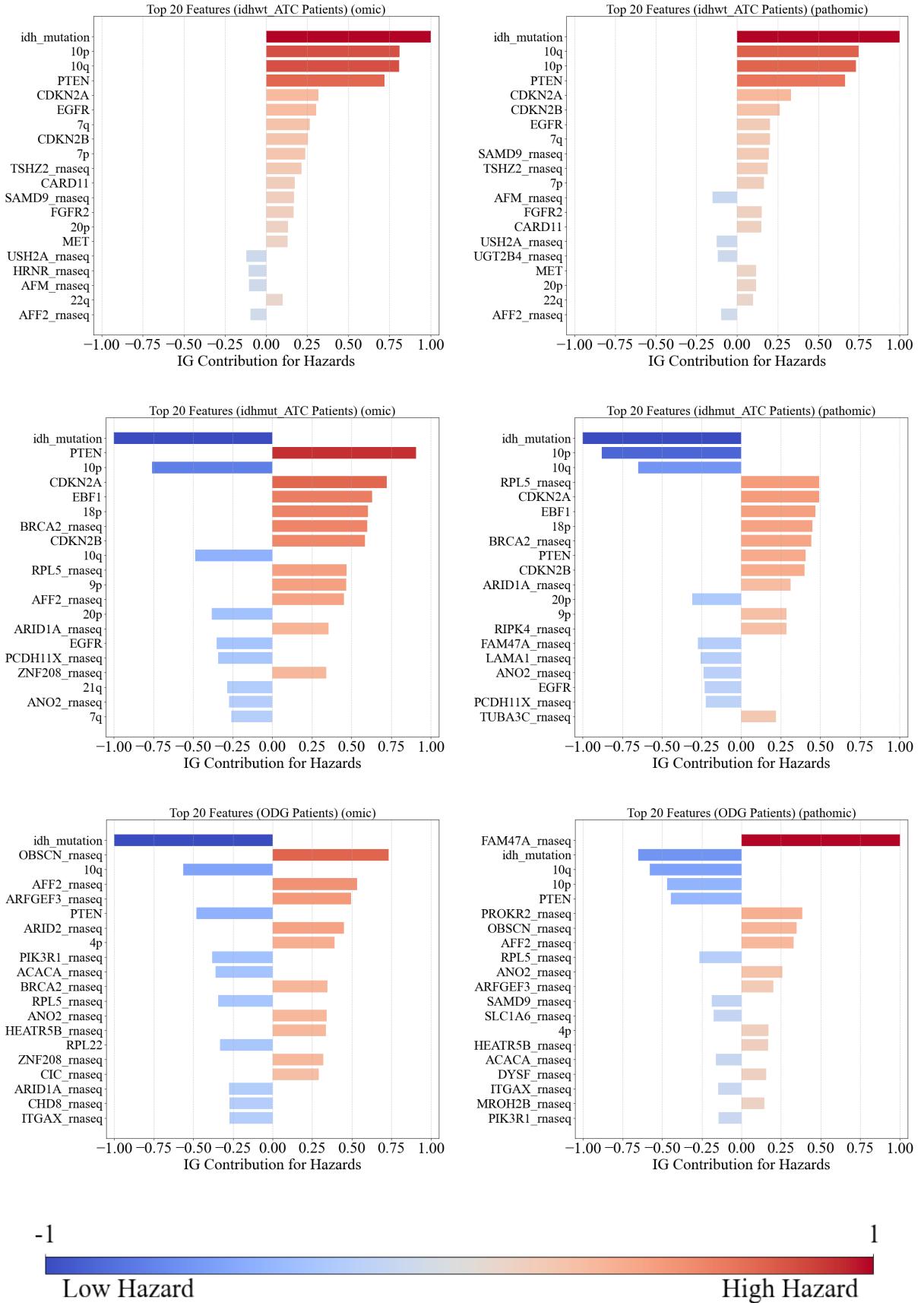


Figure 4: Comparison of Uni-modal (Omic, left) and the Multi-modal (Pathomic, right) IG Interpretability in TCGA-GBMLGG: The IDH mutation, PTEN are most influential genes among all 320 genes

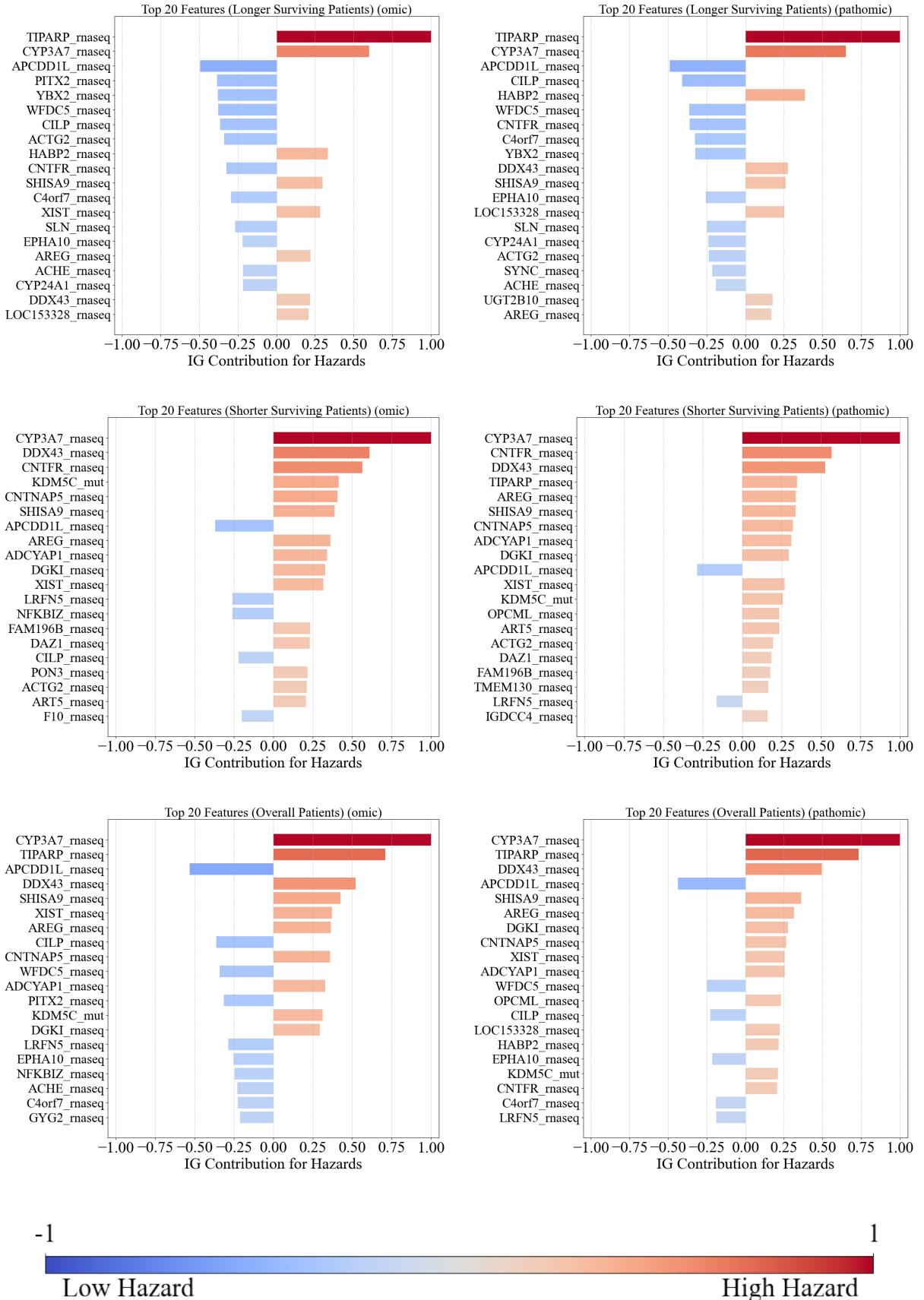


Figure 5: Comparison of Uni-modal (Omic, left) and the Multi-modal (Pathomic, right) IG Interpretability in TCGA-KIRC: The CYP3A7, DDX43 are most influential genes among all 362 genes

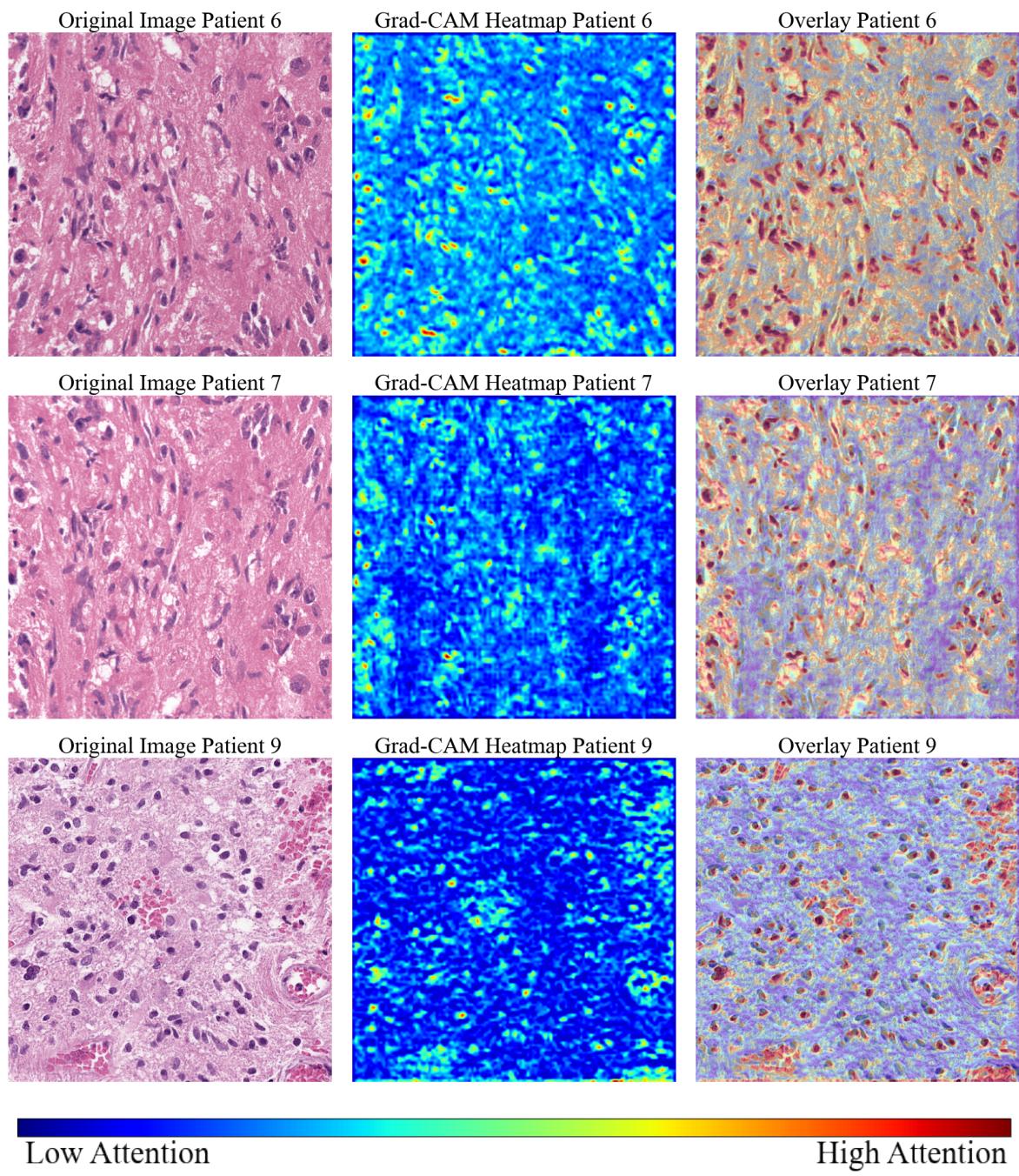


Figure 6: Grad-CAM of 4 Patients in TCGA-GBMLGG

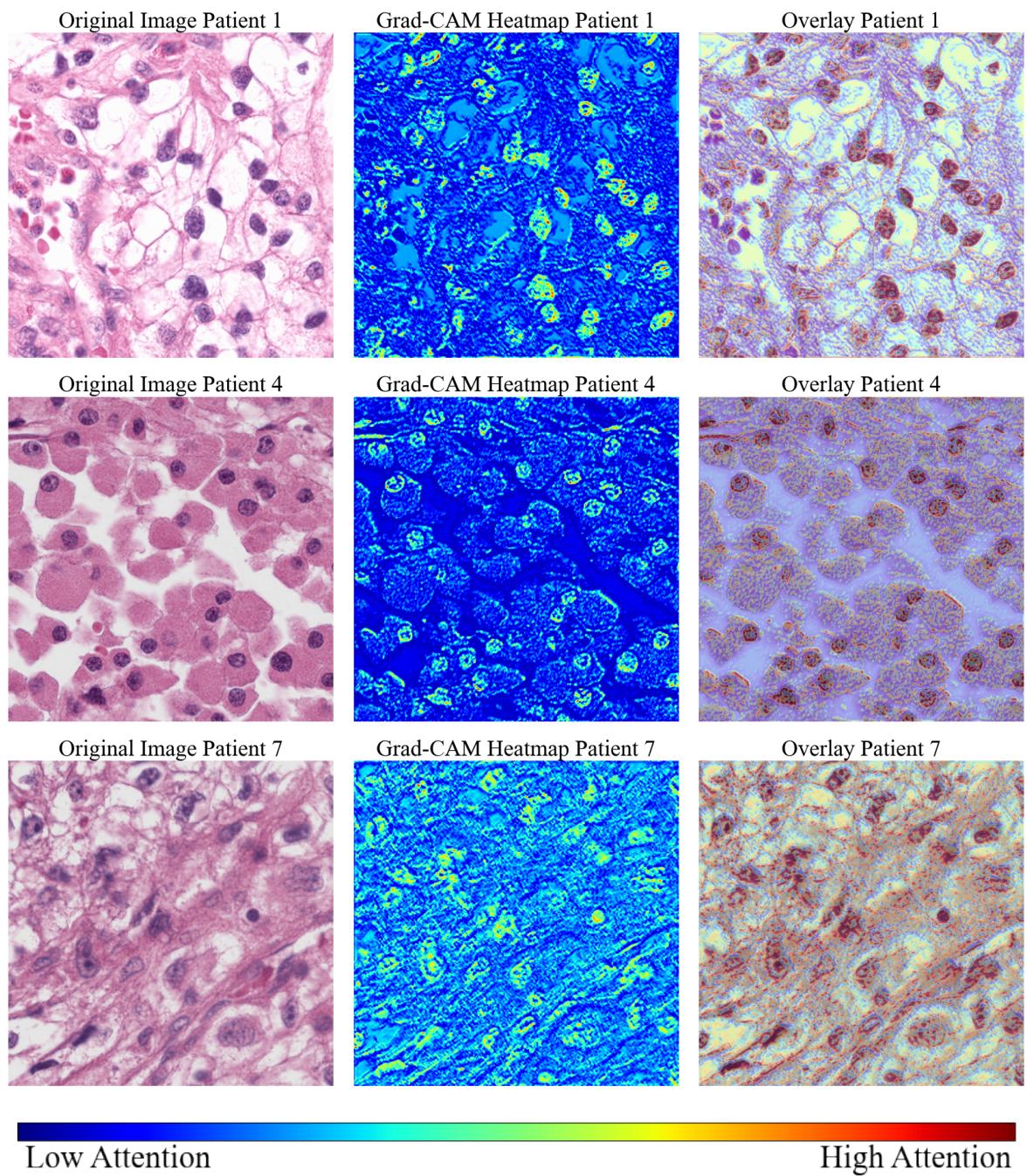


Figure 7: Grad-CAM of 4 Patients in TCGA-KIRC

B Generative Tools

The following are generated or revised by Chatgpt: 1. GradCAM classes in Gradcam_path_GBMGLL.py 2. GradCAM classes in Gradcam_path_KIRC.py 3. The pre-aggregated Bimodal and Trimodal Fusion in networks.py incorporates the Self Attention structure from ChatGPT and the original network structure. We integrated the concept and architecture from ChatGPT to establish a foundational structure, which we subsequently adapted by adjusting parameters and hyperparameters to suit our specific needs. 4. The Chatgpt is applied to polish the sentences, detect grammatical errors.