

# Pathomic Fusion

Cailley Factor

June 2024

# 1 Abstract

This project sought to replicate and extend on Chen et al.’s paper ‘Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis’, hiterto referred to as the original paper [1]. The original paper developed a novel framework for integrating histology, genomic, and transcriptomic data for cancer grading and prognosis. The approach was validated using data from the The Cancer Genome Atlas for glioma and kidney clear cell renal cell carcinoma (ccRCC). We extend this by incorporating clinical data for ccRCC to enhance patient stratification. Various neural network architectures were used to extract unimodal feature representations: both convolutional neural networks and graph convolutional networks for histology, and self-normalizing networks for genomic, transcriptomic, and clinical data. These representations were then fused using a gating-based attention mechanism to control the expressiveness of different features followed by a Kronecker product for the bimodal and trimodal networks. As in the original paper, the pathgraphomic fusion model (CNN+GCN+SNN) achieved the highest performance in glioma grade classification across most metrics. For survival prediction, the pathomic fusion model (CNN+SNN) outperformed individual and other fusion models with the pathgraphomic fusion model also showing high performance. For ccRCC, the clinical SNN model showed superior performance to the other networks and impressive patient stratification, however the fusion models with the clinical data performed substantially worse than the clinical SNN model alone. This replication highlights the effectiveness of multimodal fusion approaches with a Kronecker product in cancer prognosis and patient stratification, useful for important problems in medicine such as personalised treatment. It demonstrates the utility of clinical data for cancer survival analysis, although refinements are needed for better integration of this data in multimodal fusion approaches.

# Contents

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>4</b>
2.1	Cancer Pathology and Modalities of Diagnosis . . . . .	4
2.2	Patient Stratification . . . . .	4
2.3	Graph Convolutional Networks for Cancer Outcome Prediction . . . . .	5
2.4	Multimodal Fusion Strategy . . . . .	5
<b>3</b>	<b>Methodology</b>	<b>6</b>
3.1	Data . . . . .	6
3.2	Model Architecture & Training . . . . .	6
3.2.1	Extracting Features from H&E Histology Images using a Convolutional Neural Network	6
3.2.2	Extracting Cell Morphometric and Graph Features using a Graphical Convolutional Neural Network . . . . .	6
3.2.3	Extracting Features from Molecular Profiles using a Self-Normalizing Network . . . .	8
3.2.4	Utilising Clinical Data . . . . .	9
3.2.5	Gating-Based Attention and Multimodal Tensor Fusion . . . . .	9
3.3	Evaluation & Interpretability . . . . .	10
<b>4</b>	<b>Results and Discussion</b>	<b>11</b>
4.1	Glioma Grade Analysis . . . . .	11
4.2	Glioma Survival Analysis . . . . .	12
4.3	Clear Cell Renal Cell Carcinoma Survival Analysis with Clinical Data . . . . .	15
4.4	Conclusion . . . . .	18
<b>5</b>	<b>Appendix</b>	<b>19</b>
5.1	Patient Inclusion . . . . .	19
5.2	Data for Each Model . . . . .	19
5.3	Cox Model . . . . .	19
5.4	Network Training . . . . .	20
5.5	Additional Hazard Plots . . . . .	24

## 2 Introduction

### 2.1 Cancer Pathology and Modalities of Diagnosis

Cancer is a complex disease characterized by hallmarks including sustained proliferative signaling, evasion of growth suppressors, resistance to cell death, enabling replication immortality, induction of angiogenesis, and activation of invasion and metastasis [2]. These biological processes underlie the heterogeneous nature of the tumour microenvironment and their effects be seen at different biological scales [2]. This necessitates a multimodal approach to diagnosis and treatment that integrates histology, genomics, and transcriptomics to provide a comprehensive understanding of the disease and to stratify patients for, e.g., personalised treatment.

The analysis of histology slides by highly trained pathologists is the cornerstone of the diagnosis of neoplasia and the grade classification of cancers [3]. Histology involves analysing tissue slides from a patient’s biopsy and often involves Hematoxylin and Eosin (H&E) staining for improved contrast [3]. With the advent of digital pathology, histology slides provide rich information about the tumour microenvironment, ripe for machine learning approaches [4]. Important features relating to the hallmarks can be extracted from these slides including cell density, nuclear morphology, mitotic count, and the presence of necrosis or microvascular proliferation, which are critical for accurate diagnosis and prognosis [5]. Genomic information such as data about mutation status and copy number variation can provide information about DNA alterations that drive tumourigenesis [6]. Mutations can lead to cancer progression by activating oncogenes (e.g., MYC, FGFR2) which can promote cell proliferation or inactivating tumor suppressor genes (e.g., BRCA2 and PTEN) which normally regulate cell growth [7]. Changes in copy number variation, involving gains or losses of DNA segments, can lead to the amplification of oncogenes or deletion of tumor suppressor genes, further driving tumourigenesis [6]. Transcriptomic analysis, such as mRNA sequencing evaluates RNA expression levels to understand the functional consequences of these genetic changes [6].

### 2.2 Patient Stratification

Effective patient stratification is crucial for evaluating cancer prognosis and improving treatment outcomes. Brain tumours are classified according to the cell from which they originate [8]. Gliomas are a general term to describe brain tumours from glial cells, which include astrocytes, oligodendrocytes, and microglial cells, and support neurons rather than directly participating in synaptic interactions and electrical signalling [8, 9]. The World Health Organization (WHO) revised their classification system in 2021 for gliomas to include molecular markers for categorizing tumors as these generate a stronger correlation with tumor biology, behavior, and patient prognosis compared to the older system of grading based solely on histology [10, 11]. The paradigm for glioma classification involves stratifying diffuse gliomas based on the following characteristics: glial cell type, IDH1 mutation status, 1p19q chromosome codeletion status (the short arm of chromosome 1 and the long arm of chromosome 19) [10]. These guidelines have created three categories of molecular subtype for gliomas which are heavily linked to survival: 1) IDH-wildtype astrocytomas (IDHwt ATCs), 2) IDH-mutant astrocytomas (IDHmut ATCs), and 3) IDH-mutant and 1p/19q-codeleted oligodendrogliomas (ODG). Grade II, III, and IV gliomas have a five-year survival rate of around 40-50%, 25-30%, and 5-10%, respectively [12]. Accordingly, IDHwt ATCs (mostly WHO grades III and IV) have been shown to have worse patient survival outcomes compared with IDHmut ATCs (WHO Grades II, III, and IV) and ODGs (mostly WHO Grades II and III) [10].

Kidney Renal Cell Carcinoma (RCC) is the most common malignancy of the kidneys and comprises approximately 90% of kidney cancer [13]. It originates from epithelial cells in the renal tubular epithelium, and includes three common subtypes: clear cell renal cell carcinoma (ccRCC), which accounts for 75-80% of RCC, papillary RCC, and chromophore RCC [13]. Methods for staging ccRCC are poor and suffer from high variability between observers [14]. The Fuhrman Grading System based on the appearance of cancer cells under a microscope is the most widely used metric for ccRCC prognosis, although it still suffers from some interpretation difficulties and poor reproducibility [15]. The Fuhrman Grading System creates four classifications based on nuclear size, shape, and nucleolar prominence from G1 to G4 [14]. The five year survival rate for ccRCC is 50-69% longer than for gliomas [16]. For both diseases, there is a need for finer patient stratification beyond grade status and molecular subtype which could be improved using more rich histology and genomic features, as well as multimodal fusion strategies.

## 2.3 Graph Convolutional Networks for Cancer Outcome Prediction

New approaches are needed for extracting the rich information from histology slides. Although several papers have utilised graph convolutional networks for histology classification in breast, colon, and prostate cancer, this paper was novel in using graph convolutional networks (GCNs) for survival outcome prediction [17, 18, 19]. Graph convolutional networks (GCNs) are advantageous in that they can explicitly capture information regarding cell-to-cell interactions and cell spatial organization in the tissue. A graph  $G$  is represented as  $G = (V, E)$  where  $V$  is a set of vertices and  $E$  is a set of nodes, cells in the case of the histology images [20]. This representation allows GCNs to model the complex spatial organization of cells and interactions between neighboring cells within the tissue. Unlike traditional CNNs which operate on regular grid structures, GCNs generalize convolutional operations to irregular and sparse graph structures [20]. GCNs utilise convolutional operations to learn features by aggregating information from neighboring nodes (cells) in the graph, taking into account their spatial proximity and connectivity [20]. This enables GCNs to capture features that reflect the local and global spatial relationships between cells in the histology images. GCNs extract features which are different to CNNs, as they dynamically adjust receptive fields based on the connectivity of nodes (cells), whereas CNNs process images based on fixed receptive fields [20]. This adaptability allows GCNs to handle varying scales of spatial relationships and capture higher-order dependencies.

## 2.4 Multimodal Fusion Strategy

Multimodal fusion strategies have been explored to more depth in computer vision and natural language processing tasks, than in biomedical data science. A lot of the work on the multimodal fusion of histology and genomics in cancer has previously focused on establishing correlations between histology and genomics [21, 22, 23, 24]. However, some work has been done on fusion for supervised learning tasks, and these approaches break down into three main categories: early, late, and intermediate fusion [25]. Early fusion involves integrating information from all data modalities through deep or handcrafted features before feeding it into one model for a prediction, through varying integration operations from vector concatenation to bilinear pooling (Kronecker product) [25]. Late fusion involves training a separate model for each modality and coming up with predictions from the unimodal models then aggregating the predictions by methods such as majority voting or learned models such as a multilayer perceptron [25]. Prior to this paper, most existing work in the space involved late fusion, such as vector concatenation of histology and genomic features [25, 26]. Intermediate fusion occur when the loss from the multimodal model propagates back to most of the layers to iteratively improve the feature representations and the data can be fused gradually or at a single level [25]. Another form of intermediate fusion is guided fusion whereby the model uses information from one modality to guide feature extraction from another modality with mechanisms such as attention [25, 27, 28]. The multimodal fusion approach employed in the original paper is similar to intermediate fusion, except in regards to the histology CNN network, which resembles early fusion [1].

### 3 Methodology

As this paper first sought to replicate the results of the original paper, much of the methodology is adopted from it [1]. A substantial amount of the code base was missing or not able to be run and needed to be amended or coded from scratch. Additionally, there were several discrepancies between the code base provided and the paper’s description, as well as changes that were made to improve runtime speed, which may account for any discrepancy between the methodology description below for the replication and that of the paper.

#### 3.1 Data

The TCGA contains paired whole-slide image, genomic, transcriptomic and clinical data with ground truth survival and histology grade labels [29]. Genomic, transcriptomic, clinical, and histology data were extracted from datasets of The Cancer Genome Atlas (TCGA) as described in more depth in the appendix 5.1 and 5.2. For grade prediction for glioma, the TCGA-GBMLGG dataset was used and only mutation status and copy number variation (CNV) were included, as the inclusion of RNA-Seq data led to the overtraining of the model, while RNA-seq data was also included for survival analysis. Data from TCGA-KIRC was used for the ccRCC data. As an extension to the original paper, clinical data was utilised for some of the ccRCC models. Clinical data was selected to add to the ccRCC versus the glioma dataset as the ccRCC dataset had a substantially greater variety of clinical datatypes. The following clinical datatypes were included: diagnosis age, sex, Buffa hypoxia score, prior diagnosis, neoplasm disease stage, metastasis stage, and primary lymph node presentation. The original paper utilised a 15-fold cross-validation with 80% training and 20% testing (split by TCGA ID).

#### 3.2 Model Architecture & Training

For TCGA-GBMLGG, histology features were extracted for the unimodal networks using convolutional neural networks (CNN) and graph-based features were extracted using graph convolutional networks (GCN). Genomic and transcriptomic features were learned using a self-normalizing network (SNN) to prevent overfitting. For the TCGA-KIRC dataset, the GCN architecture was not trained in contrast to the original paper. As an extension, clinical data was utilised in a unimodal network, as well as for trimodal and bimodal fusion strategies. The bimodal and trimodal fusion networks utilised attention-based gating of the unimodal features to control their expressiveness followed by a Kronecker product to capture a joint representation of the features. For survival outcome prediction, the networks were supervised by the Cox partial log likelihood loss (appendix 5.3). For grade classification, they are supervised by the cross entropy loss. A summary of the methodology for the trifusion networks is shown in figure 1.

##### 3.2.1 Extracting Features from H&E Histology Images using a Convolutional Neural Network

To capture the features from histology tissue, a convolutional neural network (CNN) was trained on 512 x 512 regions-of-interest (ROI) at 20x magnification. CNNs can capture important information about cancer biology, such as high cellularity, nuclear atypia, the proliferation of microvasculature [26]. The architecture of the CNN is VGG19, a popular CNN architecture with 16 convolutional, 3 fully connected and 5 max pooling layers. Drop-out probabilities of 0.25 were applied after the first two fully connected layers and 0.05 after the last fully connected layer. Pre-existing weights trained on ImageNet were utilised. From this architecture, an embedding  $e_p \in \mathbb{R}^{32 \times 1}$  was extracted.

##### 3.2.2 Extracting Cell Morphometric and Graph Features using a Graphical Convolutional Neural Network

Although GCNs provide some similar information to CNNs regarding morphological features, a big difference between GCNs and CNNs is that GCNs require the creation of a graph structure to capture relationships and dependencies, whereas CNNs operate directly on the image data. Nuclei segmentation was used to define the set of nodes (cells) in the graph. An adjacency matrix was created using a k-nearest neighbours algorithm. Both handcrafted and deep features were extracted for each cell. As the nuclei segmentation masks were missing from the data, the precomputed graph information was utilised to train the GCN.

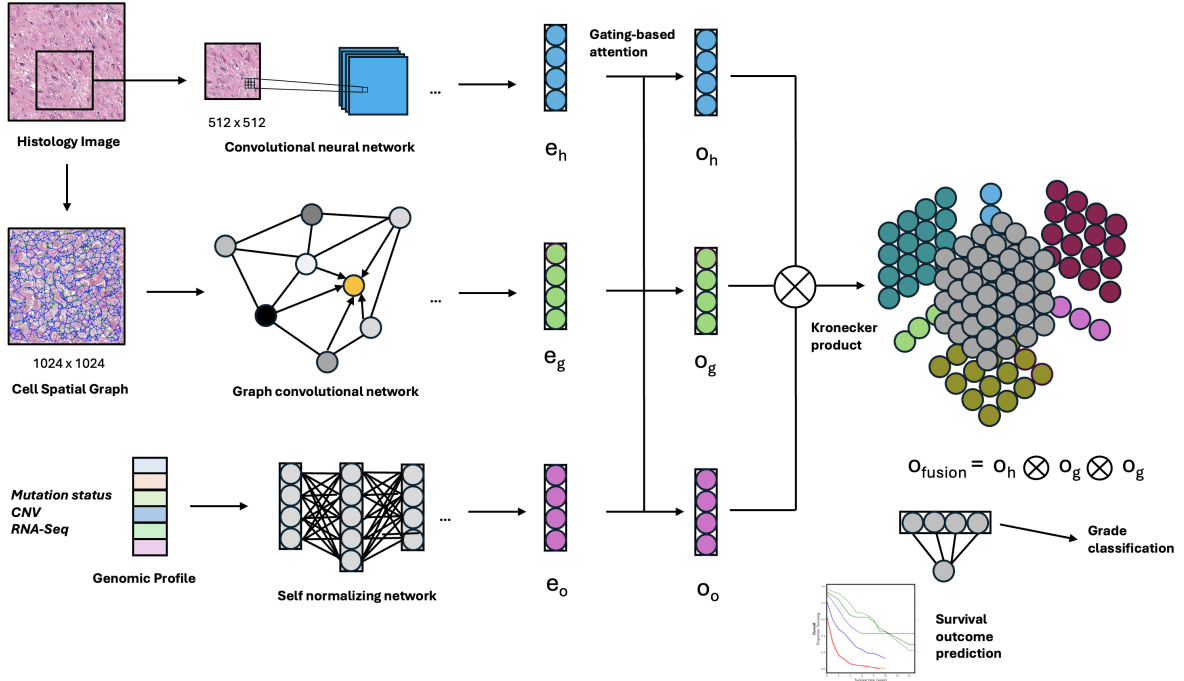


Figure 1: Summary of the methodology for pathgraphomic fusion (CNN+SNN+GCN) modified from a figure from the original paper. Unimodal features are extracted using a histology CNN, histology GCN, and genomic SNN. For the trimodal networks, the features are fused together using gating-based attention followed by a Kronecker product and then used for survival outcome prediction or grade classification.

Accurate segmentation of nuclei is crucial for identifying abnormal cell features such as nuclear atypia and dense tumor cellularity, which are indicative of cancer progression [30]. A conditional generative adversarial network (cGAN) was utilised to enhance semantic segmentation accuracy without relying on manually engineered loss functions, such as pixel-wise loss, which can often lead to errors such as merging nuclei [31]. The cGAN comprises a generator and a discriminator which engage in a competitive game where the generator  $S$  learns to produce realistic segmentation masks  $S(n)$  from histology images ( $n$ ), while  $D$  distinguishes between real pairs ( $n, m$ ) and generated pairs ( $n, S(n)$ ) [31]. The model is trained using a combined L1 loss and discriminative loss, resulting in equation (3), where the generator is trying to minimise the loss and the discriminator is trying to maximise the loss [31].

$$\mathcal{L}_{\text{GAN}}(S, D_M) = \mathbb{E}_{m, n \sim p_{\text{data}}(m, n)}[\log D_M(m, n)] + \mathbb{E}_{n \sim p_{\text{data}}(n)}[\log(1 - D_M(m, S(n)))] \quad (1)$$

$$\mathcal{L}_1(S) = \mathbb{E}_{m, n \sim p_{\text{data}}(m, n)}[\|m - S(n)\|_1] \quad (2)$$

$$\arg \min_S \max_{D_M} \mathcal{L}_{\text{GAN}}(S, D_M) + \mathcal{L}_1(S). \quad (3)$$

K-nearest neighbours was then used to determine the set of edges connecting nodes in the graph with a parameter of  $k=5$ . The adjacency matrix was defined based on this algorithm using the FLANN library [32].

$$A_{ij} = \begin{cases} 1 & \text{if } j \in \text{KNN}(i) \text{ and } D(i, j) < d \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Cell features were then extracted both manually and using contrastive predicted coding (CPC) [33]. For each cell, eight contour features were extracted from the segmentation results (major axis length, minor axis length, angular orientation, eccentricity, roundness, area, and solidity).  $64 \times 64$  image crops were created around each centroid of the contour and four additional texture features were extracted from gray-level co-occurrence matrices (dissimilarity, homogeneity, angular second moment, and energy). The types of features which are able to be identified from this information include atypia, nuclear pleomorphism, and hyperchromatism. CPC was utilized to extract 1024-dimensional features from the  $64 \times 64$  crops, shared among different portions of a  $256 \times 256$  image crop. This enables the extraction of features shared among different portions of the cell such as intercellular interactions and the microvasculature networks around each cell. The feature from the CPC are then combined with the handcrafted features to make the cell graph.

GCNs generate feature representations for each node by aggregating feature vectors from neighboring nodes and this enables the pooling of all node features to form a comprehensive graph representation, useful for the paper’s tasks of classification or survival prediction. The paper utilised ‘GraphSAGE’s ‘AGGREGATE’ and ‘COMBINE’ functions for updating node representations and utilises the ‘SAGPOOL’ self-attention pooling method to capture the hierarchical structure of cell graphs, culminating in a pooled feature vector that serves as input for Pathomic Fusion taken from a paper by Lee et al. [34, 35]. Overall, the GCN comprised 3 ‘GraphSAGE’ and ‘SAGPOOL’ self-attention pooling layers with dimension 128, followed by two linear layers of size 128 and 32. From this architecture, an embedding  $e_g \in \mathbb{R}^{32 \times 1}$  was extracted.

### 3.2.3 Extracting Features from Molecular Profiles using a Self-Normalizing Network

Feedforward neural networks have shared weights and tend to be prone to overfitting, especially in the case of the high-dimensional genomic data compared with the number of samples [20]. The genomic SNN consists of 4 consecutive blocks of fully-connected layers with dimensions [64, 48, 32, 32], scaled exponential linear units (SeLU) activation, and Alpha Dropout. A SeLU was selected as the activation as it causes outputs after every layer to tend towards a zero mean and unit variance. By randomly setting a fraction of the input units to a specific alpha value, Alpha Dropout helps prevent overfitting, similar to traditional dropout. This ensures that the model does not rely too heavily on specific neurons and generalizes better to new data, as well as ensuring that the self-normalizing properties of SeLU activation functions are preserved in order to help maintain the network’s stability and performance. From this architecture, an embedding  $e_i \in \mathbb{R}^{32 \times 1}$  was extracted.



### 3.2.4 Utilising Clinical Data

The clinical data for ccRCC was used to replace the graph data for the multimodal fusion networks. The same SNN network as the genomic network was utilised for the clinical network, i.e., a self-normalizing network with the same dimensions and number of fully-connected layers, and same activation and dropout. From this architecture, an embedding  $e_c \in \mathbb{R}^{3 \times 1}$  was extracted.

### 3.2.5 Gating-Based Attention and Multimodal Tensor Fusion

Graph, pathology and genomic data each have very different formats with information held in a graph structure with nodes and edges, a two-dimensional grid, and a one-dimensional vectors of covariates, respectively. A key motivation of multimodal fusion is to stratify patients into more precise subtypes, for example, for treatment groups, as features from one modality have different clinical significance when conditioned upon another modality. The unimodal feature representations described in the previous sections were combined using gating-based attention followed by a Kronecker product. Figure 2 shows the method of trifusion and bifusion employed in the paper.

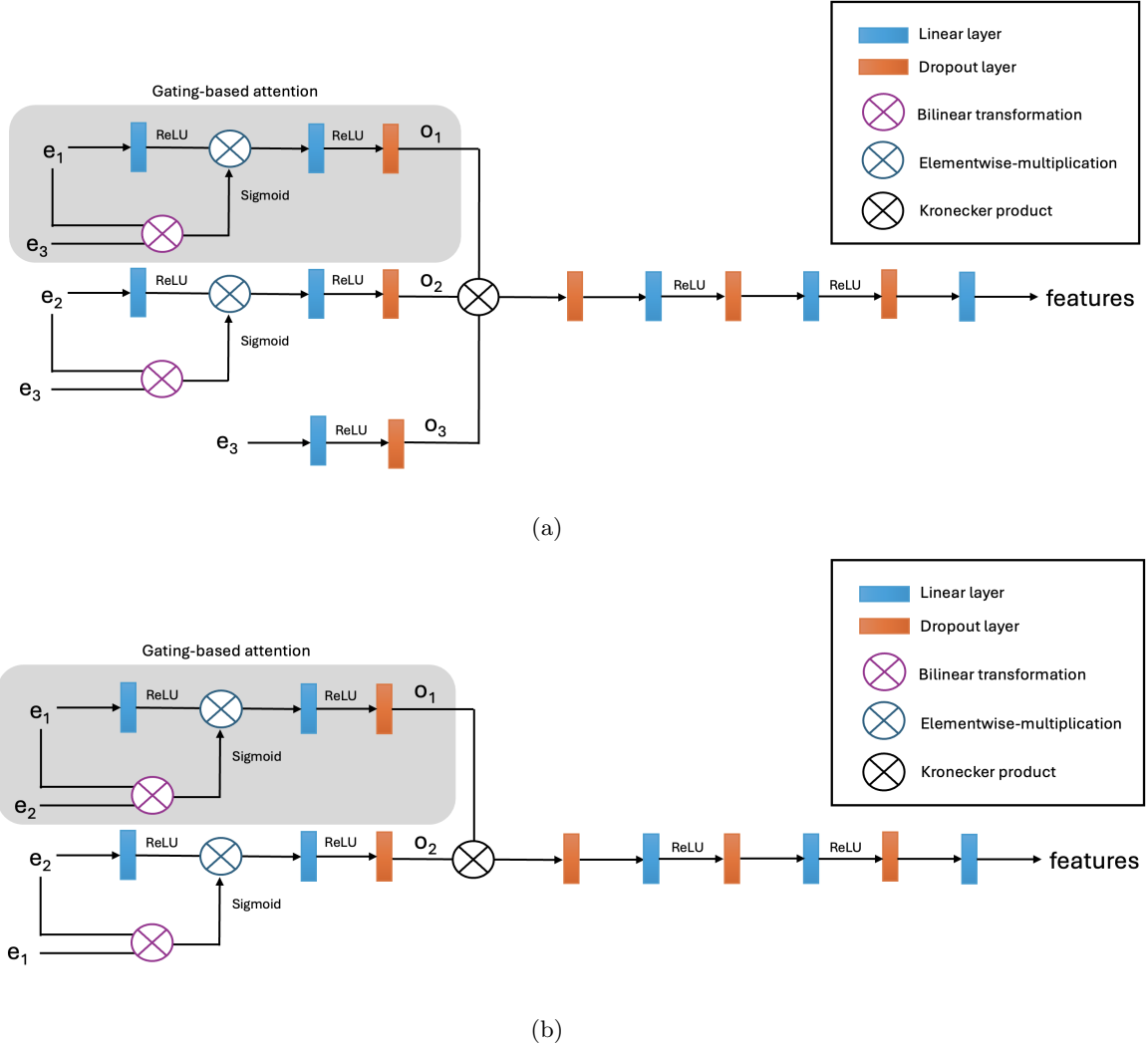


Figure 2: Comparison of gating-based attention followed by a Kronecker product for the embeddings for (a) the trifusion networks and (b) the bifusion networks.

Gating-based attention serves to reduce the feature space by reducing collinearities between unimodal

features before calculating the Kronecker product. In gating-based attention, a bilinear transformation is computed between the feature of interest and the feature that is attending to it. The feature of interest and the result of the previous computation are then multiplied elementwise before being passed through more linear layers prior to computing a Kronecker product. For survival analysis, e1 and e2 of figure 2a represent the unimodal embeddings from the histology and graph networks (or the clinical network in the case of TCGA-KIRC), respectively, and e3 represents the features extracted from the genomic network. Therefore, the genomic features are used to control the expressiveness of the other networks. For grade classification for TCGA-GBMLGG, e1 and e2 represent the embeddings from the genomic and graph networks, respectively, and e3 represents the embeddings from the histology network. In other words, the histology features are used to control the expressiveness of the other features before computing the Kronecker product for grade analysis.

The bifusion mechanism employed by the paper involves gating-based attention of the two modalities of interest with respect to each other before computing a Kronecker product, as shown in figure 2b. The Kronecker product models feature interactions across the unimodal features that would not otherwise be captured. The Kronecker product computes a matrix outer product of the unimodal feature vectors of the model. To maintain unimodal and bimodal feature interactions while computing trimodal interactions, a value of 1 was appended to each unimodal feature representation. The equation for the Kronecker product in the trifusion model is as follows:

$$\mathbf{o}_{\text{fusion}} = \begin{bmatrix} \mathbf{o}_1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{o}_2 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{o}_3 \\ 1 \end{bmatrix}$$

For TCGA-GBMLGG, trifusion was performed with the genomic, graph, and histology features. Whereas for TCGA-KIRC, trifusion was performed with the genomic, and histology features, as well as with the addition of novel clinical features.

### 3.3 Evaluation & Interpretability

The original paper utilised a 15-fold cross-validation, however, due to the long training time, the replication for TCGA-GBMLGG was performed on five of the fifteen folds and the replication for TCGA-KIRC was performed on four of the fifteen folds. Standard statistical metrics were used to evaluate the models. For grade classification, the standard metrics of Area Under the Curve (AUC), Average Precision (AP), F1-Score (microaveraged across all classes), F1-Score (WHO Grade IV only), and ROC curves were utilised.

For survival outcome prediction, the standard metric of the Concordance Index (c-index) was utilised, which assesses the ability of a model to correctly predict the order of time-to-event outcomes, such as the time until a particular event occurs, in this case, death [36]. The c-index ranges from 0.5 to 1.0, where a value of 0.5 suggests that the model’s predictions are no better than random chance, and a value of 1.0 indicates perfect predictive accuracy. The metric compares pairs of individuals within the dataset. For each pair where one individual experiences the event (death) before the other, the metric involves checking if the model predicts a higher risk for the individual who experienced the event first. The c-index is calculated as the proportion of all usable pairs where the predictions and outcomes are concordant.

Patients were stratified by changing their hazard ratios into percentiles. For TCGA-GBMLGG, the predicted hazards were separated using percentiles of 33-66-100 to plot Kaplan-Meier curves. For TCGA-KIRC, the predicted hazards were separated by percentiles into 25-50-75-100 or 50-100. Kaplan-Meier estimates were used to visualise the efficacy of the models at stratifying patients. Long rank tests were performed to determine the statistical significance of patient stratification.

Integrated gradients (IG) from ‘Captum’ is a gradient-based feature attribution method [37]. The integrated gradients method calculates the gradients of the input tensor  $x$  across different scales against a baseline  $x_i$  (zero-scaled). The Gauss-Legendre quadrature is then used to approximate the integral of gradients.

$$\text{IG}_i(x) ::= (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha$$

This method was used to attribute the predictions of the networks to the inputs for the genomic features of the genomic SNN network for TCGA-GBMLGG and for the clinical features for the clinical SNN network for TCGA-KIRC.

## 4 Results and Discussion

### 4.1 Glioma Grade Analysis

The loss curves for training the models for grade classification and survival analysis for the TCGA-GBMLGG datasets can be found in the appendix 5.4. The performance metrics for the models for the grade analysis are displayed in table 1 and table 2. Table 1 displays the mean Area Under the Curve (AUC) score and Average Precision (AP) over the five folds. Table 2 displays F1-scores micro-averaged across all classes (Micro) and for WHO Grade IV class only. From these values, it can be seen that the pathgraphomic fusion (CNN+SNN+GCN) model has the highest classification performance for all metrics other than AP, where pathomic fusion (CNN+SNN) outperformed pathgraphomic fusion. This is exactly in line with the original paper which exhibited the same pattern. However, the standard deviations are relatively higher than in the original paper and thus, the results for pathgraphomic fusion (CNN+SNN+GCN) were very similar to the performance of the other fusion types. The larger standard deviations are due to a difference in methodology, whereby the original paper used the max softmax activation score from overlapping  $512 \times 512$  patches to determine class in the evaluation. In this methodology, only one random  $512 \times 512$  patch per  $1024 \times 1024$  histology image was utilised for the evaluation. As in the original paper, multiple slides from one patient were aggregated using the max softmax activation score. Among the unimodal networks, the histology CNN network outperformed the other unimodal networks, while the graph GCN network outperformed the genomic SNN network. This is in-line with the fact that grade analysis is based on histology analysis by a trained pathologist. The histology CNN network even outperformed the graphomic (GCN+SNN) fusion. The AP results appeared to be slightly lower than the metrics in the original paper, while the other metrics were very similar values. As the replication was only performed on the first five folds, this could have also accounted for small differences between the results.

Model	AUC	AP
Genomic SNN	0.807571 $\pm$ 0.0217	0.609871 $\pm$ 0.0553
Graph GCN	0.857788 $\pm$ 0.0124	0.706830 $\pm$ 0.0211
Histology CNN	0.873109 $\pm$ 0.0288	0.732673 $\pm$ 0.0499
Graphomic Fusion (GCN+SNN)	0.867079 $\pm$ 0.0244	0.705593 $\pm$ 0.0593
Pathomic Fusion (CNN+SNN)	0.893656 $\pm$ 0.0210	0.750632 $\pm$ 0.0422
Pathgraph Fusion (CNN+GCN)	0.889795 $\pm$ 0.0168	<b>0.764240 <math>\pm</math> 0.0288</b>
Pathgraphomic Fusion (CNN+GCN+SNN)	<b>0.899006 <math>\pm</math> 0.0322</b>	0.756393 $\pm$ 0.0720

Table 1: Mean AUC and AP metrics with their standard deviations for the different models for grade classification of TCGA-GBMLGG across five folds.

Model	F1-Score (Micro)	F1-Score (Grade IV)
Genomic SNN	0.656763 $\pm$ 0.0165	0.860153 $\pm$ 0.0124
Graph GCN	0.666179 $\pm$ 0.0171	0.872645 $\pm$ 0.0184
Histology CNN	0.704426 $\pm$ 0.0417	0.890018 $\pm$ 0.0526
Graphomic Fusion (GCN+SNN)	0.715734 $\pm$ 0.0476	0.896489 $\pm$ 0.0278
Pathomic Fusion (CNN+SNN)	0.746046 $\pm$ 0.0321	0.919581 $\pm$ 0.0187
Pathgraph Fusion (CNN+GCN)	0.724476 $\pm$ 0.0206	0.908947 $\pm$ 0.0127
Pathgraphomic Fusion (CNN+GCN+SNN)	<b>0.746208 <math>\pm</math> 0.0564</b>	<b>0.922600 <math>\pm</math> 0.0275</b>

Table 2: Mean F1-Score metrics with their standard deviations for the different models for grade classification of TCGA-GBMLGG across five folds.

The F1-scores on the WHO grade IV data only were notably higher than the microaveraged F1-scores. Receiver operating characteristic (ROC) curves were also plotted in figure 3 to display the differences between the pathgraphomic fusion model (CNN+SNN+GCN) and the histology CNN model. Comparing the AUC values in the figure, it can be seen that pathgraphomic fusion outperforms the histology CNN across all the grades as well as overall, though the large standard deviation results in a substantial overlap between

the performances. This is in-line with the original paper where pathgraphomic fusion achieves higher AUCs than the histology network across all grades and overall.

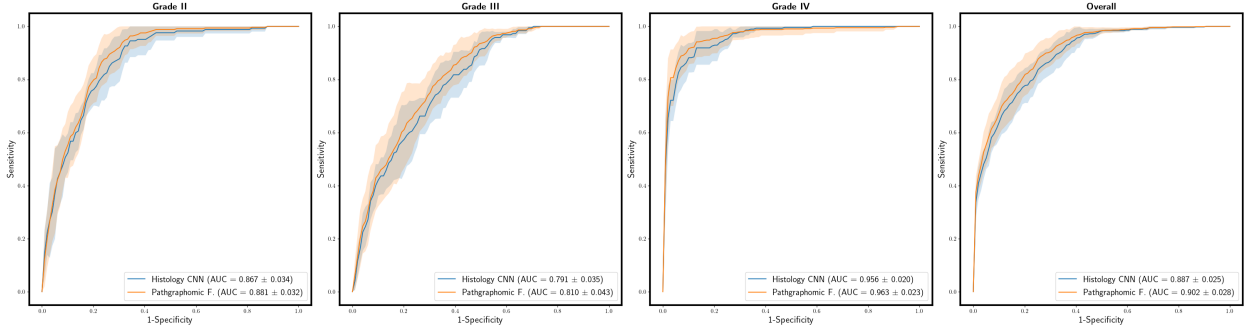


Figure 3: ROC curves by WHO grade and overall comparing the histology CNN model to the pathgraphomic fusion model (CNN+GCN+SNN) for grade prediction of TCGA-GBMLGG data aggregated for five folds.

## 4.2 Glioma Survival Analysis

For survival analysis, the TCGA-GBMLGG networks were also trained on five folds. The c-index was the primary metric utilised in the paper to evaluate the performance of the survival analysis networks. Evaluated on the five folds, the pathgraphomic model (CNN+GCN+SNN) performed worse than the pathomic fusion model (CNN+SNN). This is unlike the original paper where the pathgraphomic model (CNN+GCN+SNN) performed the best. This is likely due to a difference in the methodology of the replication compared to the original paper where the mean of hazard predictions were computed from 9 overlapping  $512 \times 512$  image crops across each histology region of interest belonging to the patient. In this replication, only one random  $512 \times 512$  image crop per patient was utilised for evaluation with the hazards from multiple regions of interest aggregated based on their mean. This may account for why the models all had slightly lower values than in the original paper with the genomic SNN having a quite similar value to the original paper, as it would not be impacted by this methodology difference. Additionally, the standard deviations were quite high compared with the paper, which could have also been explained by this. Notably in both the paper and the reproduction, the genomic SNN outperformed the other unimodal networks. In this reproduction, it outperformed the pathgraphomic fusion (CNN+GCN+SNN) network. This demonstrates the strength of genomic information in survival analysis, which could be in part due to the provision of information to determine molecular grade including IDH mutation and 1p19q co-deletion status, which are known to be important in survival prediction [10, 11].

Model	Mean C-index
Genomic SNN	$0.798098 \pm 0.052117$
Graph GCN	$0.703496 \pm 0.040011$
Histology CNN	$0.720992 \pm 0.022110$
Graphomic Fusion (GCN+SNN)	$0.787477 \pm 0.054785$
Pathomic Fusion (CNN+SNN)	<b><math>0.802934 \pm 0.043578</math></b>
Pathgraph Fusion (CNN+GCN)	$0.718602 \pm 0.023739$
Pathgraphomic Fusion (CNN+GCN+SNN)	$0.796474 \pm 0.052041$

Table 3: Mean C-index and standard deviation for different models for five folds for TCGA-GBMLGG survival analysis.

To further analyse the importance of the genomic features, integrated gradient attribution was conducted for fold k=1 to understand the gradient attribution of the features of the SNN genomic model. Integrated gradients were also used to analyse gradient attribution between the different histomolecular subtypes. Canonical oncogenes in glioma such as IDH, PTEN, CDKN2A and CDK2B all have high attribution across all

of the different molecular subtypes. IDH mutation status is related to the molecular subtyping of glioma and mutations are commonly associated with better prognosis and are a hallmark of lower-grade astrocytomas [10, 11]. PTEN mutations are linked to tumor suppression and are one of the most frequent mutations in human cancers, as PTEN is important in preventing uncontrolled cell migration, among other roles in genomic stability, cell migration, and apoptosis [38]. The long arm of chromosome 10 (10q) includes the PTEN gene, and loss of PTEN due to 10q deletion can lead to uncontrolled cell growth, in-line with its copy number variation appearing as an important feature for the integrated gradients attribution [39]. CDKN2A and CDKN2B both encode tumor suppressor proteins that play key roles in cell cycle function and thus their mutation can lead to cell cycle dysregulation [40]. MYC is another canonical oncogene and the attribution of MYC copy number variation appeared only in the integrated gradients for the astrocytomas, including IDHwt and IDHmut. MYC plays a pivotal role in cell cycle regulation, apoptosis, and cellular transformation, promoting cell proliferation and growth [41].

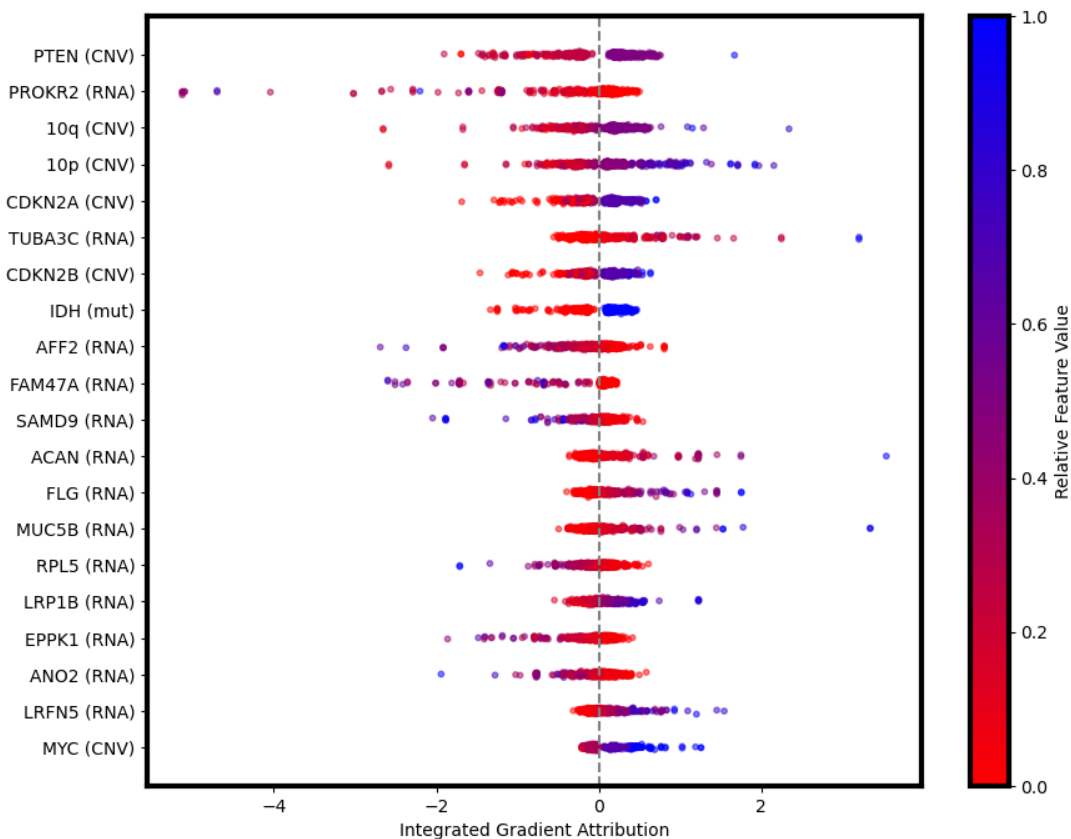
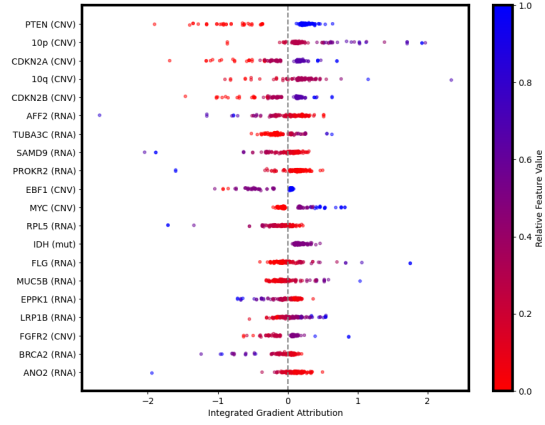
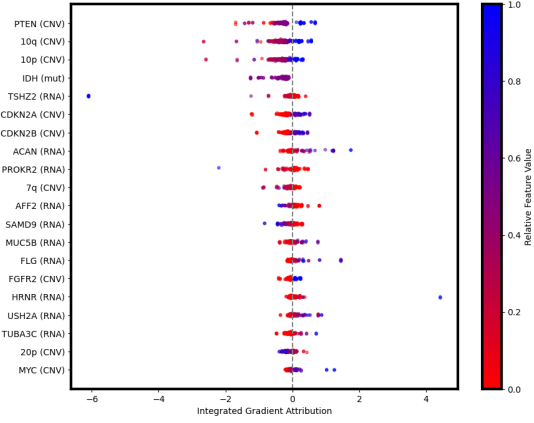


Figure 4: Integrated gradient attribution of the genomic features of the SNN genomic model for survival analysis of the first fold of the TCGA-GBMLGG dataset.

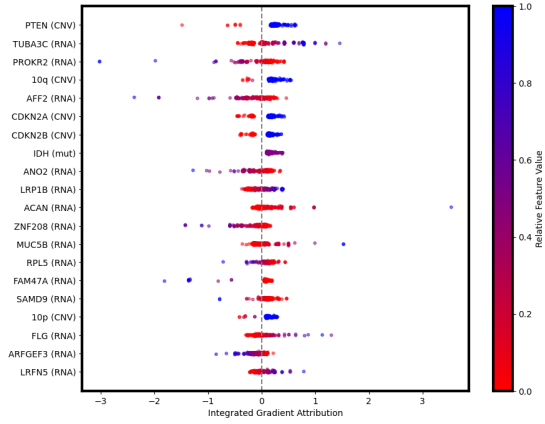
A key purpose of medical machine learning is to improve patient stratification, which is important for the ultimate goal of predicting treatment response. To stratify patients, Kaplan-Meier plots were used on the aggregated hazards between the five folds and log-rank test values were computed and added to the figures to assess the difference between the survival outcomes of the different survival groups, i.e., the adjacent curves. Figure 6 plots the survival outcomes by different stratification: true grade, molecular subtype, and the predictions of the histology CNN and the pathgraphomic fusion (CNN+GCN+SNN) network. The



(a)



(b)



(c)

Figure 5: Integrated gradient attribution of the genomic features of the SNN genomic model for survival analysis of the first fold of the TCGA-GBMLGG dataset by histomolecular subtype for (a) IDHwt Astrocytoma (b) IDHmut Astrocytoma and (c) Oligodendroglioma.

stratification by the histology CNN led to a very similar stratification to the stratification by true grade. On the other hand, the stratification by the pathgraphomic fusion network was more similar to the stratification by molecular subtype which highly correlates with survival outcomes. The middle percentile of hazards (33 to 66%) did not lead to a particularly substantial difference in stratification by pathgraphomic fusion, as shown by a very high p-value between the two curves. A breakdown of the Kaplan-Meier curves by histomolecular subtype of the hazard predictions of the different networks can be found in appendix 5.5. Swarm plots of the hazards are displayed in 7. It can be seen that the swarm plots for the pathgraphomic fusion (CNN+SNN+GCN) model result in a near bimodal separation for each of the different histomolecular subtypes and patients with grade IV cancer have much higher hazards than those with cancer of WHO grade II or III. The histology CNN network leads to a much more gradual spread of hazards for the grades. The raw hazards were plotted instead of the z-scoring done in the original paper. The patterns are similar to the original paper which also displayed a much more segregated clustering of hazard values at extremes for the pathgraphomic fusion model than for the histology CNN network.

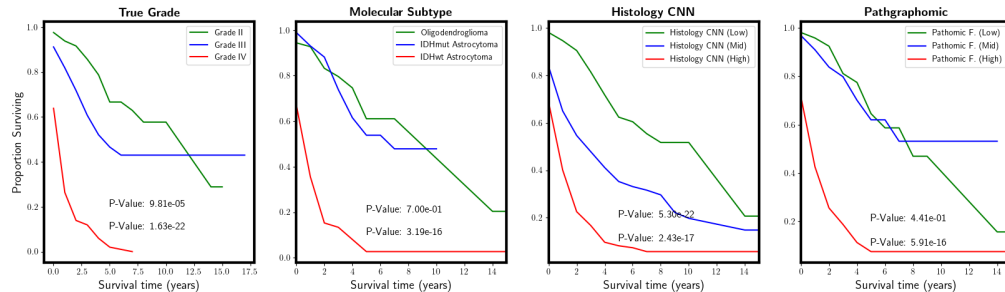


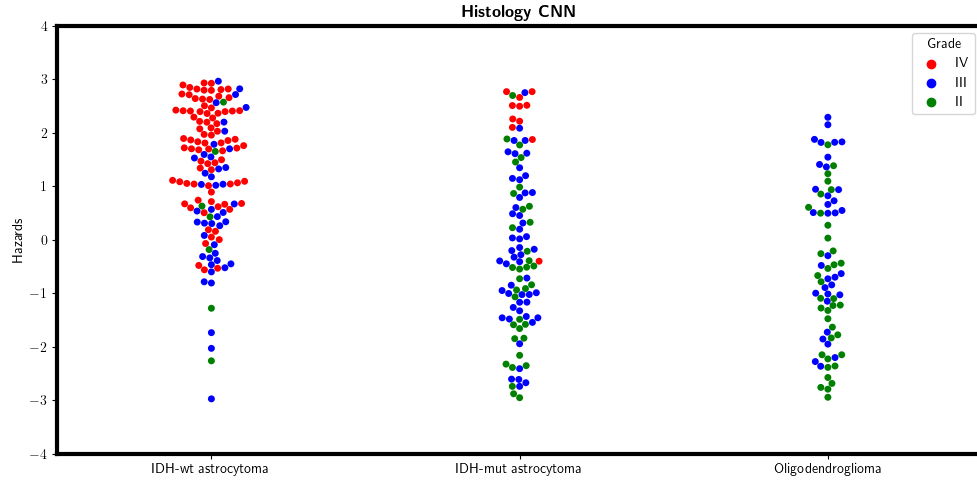
Figure 6: Kaplan-Meier curves to compare the performance of the histology CNN, histology GCN, genomic SNN, and pathgraphomic fusion (CNN+GCN+SNN) model for the TCGA-GBMLGG survival analysis.

### 4.3 Clear Cell Renal Cell Carcinoma Survival Analysis with Clinical Data

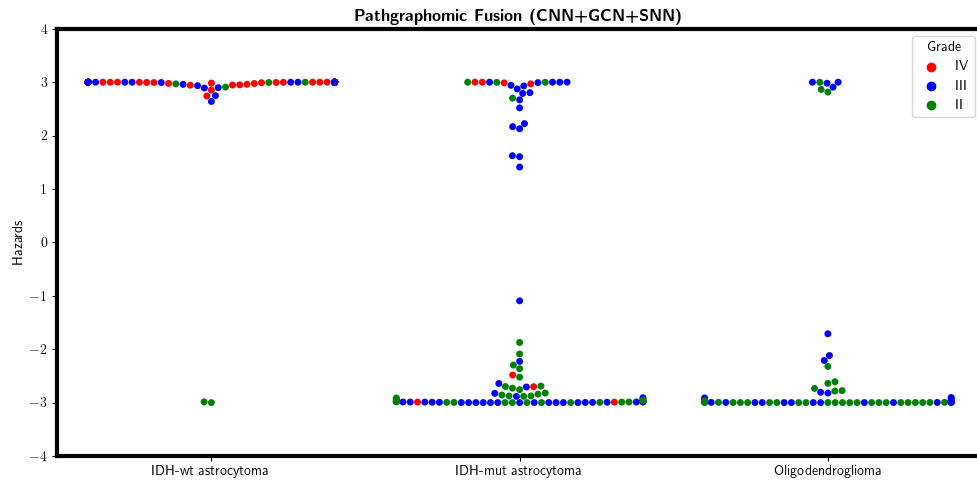
Survival analysis was also performed for TCGA-KIRC to analyse the performance of the fusion approach for another cancer type. As an extension, clinical data was added in addition to the histology and genomic data. C-indices were calculated for the models for the first four folds, as displayed in table 4. The clinical data was highly efficacious compared to the other modalities at predicting the survival outcomes. The other modalities had lower c-indices than in the original paper with the genomic SNN outperforming pathomic fusion (CNN+SNN). Again, the discrepancy in these results could be accounted for by the difference in evaluating patches for the histology regions of interest, as the genomic SNN had the most similar c-index to the original paper. The pathclinomic fusion (CNN+SNN+SNN) model performed worse than the pathomic (CNN+SNN) fusion model alone. The training and performance losses are displayed in appendix 5.4 and demonstrate overtraining for the histology CNN and genomic SNN which could account for the poor performance of the pathclinomic fusion model after fusion. Even in the original paper, the c-indices for TCGA-KIRC are substantially lower than for TCGA-GBMLGG.

Model	Mean C-index
Histology CNN	0.620536 $\pm$ 0.027413
Clinical SNN	<b>0.802990 <math>\pm</math> 0.030280</b>
Genomic SNN	0.652888 $\pm$ 0.061092
Pathomic Fusion (CNN+SNN)	0.648174 $\pm$ 0.053488
Clinomic Fusion (SNN+SNN)	0.609105 $\pm$ 0.044369
Pathclinomic Fusion (CNN+SNN+SNN)	0.634705 $\pm$ 0.041995

Table 4: Mean C-indices and standard deviation of four folds for different models for TCGA-KIRC survival analysis.



(a)



(b)

Figure 7: Swarm plots of the real hazard values for (a) the histology CNN and (b) the pathgraphomic fusion (CNN+GCN+SNN) model for survival analysis of the TCGA-GBMLGG dataset.



To analyse the effect of stratification, Kaplan-Meier curves were plotted by separating the hazards from the models by different percentiles. For the separation of hazard percentiles by 25%-50%-75%-100%, it can be clearly seen that the clinical SNN was the most efficacious model at stratifying the different patients. The pathclinomic fusion model (CNN+SNN+SNN) resulted in particularly poor separation of the hazard percentiles compared with the separation by Fuhrman grades (G1-G4). Stratification was also analysed using 50%-100% hazard percentiles, and again the clinical SNN resulted in the best stratification of patients, while the pathclinomic (CNN+SNN+SNN) fusion resulted in the worst patient stratification. Given the stellar performance of the clinical SNN model, integrated gradient attribution was conducted on this model to assess the attribution to the different features. It can be seen in figure 10 that the neoplasm disease stage was the most greatly attributed and reflects four clear categories corresponding to the four stages.

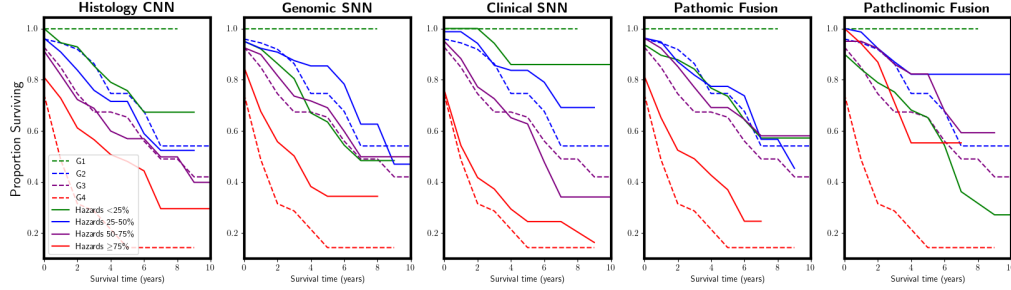


Figure 8: Kaplan Meier curves to compare the performance of the different models trained on the TCGA-KIRC dataset, stratifying the hazards by 25-50-75-100% percentiles, compared to stratifying by Fuhrman Grade (G1-G4).

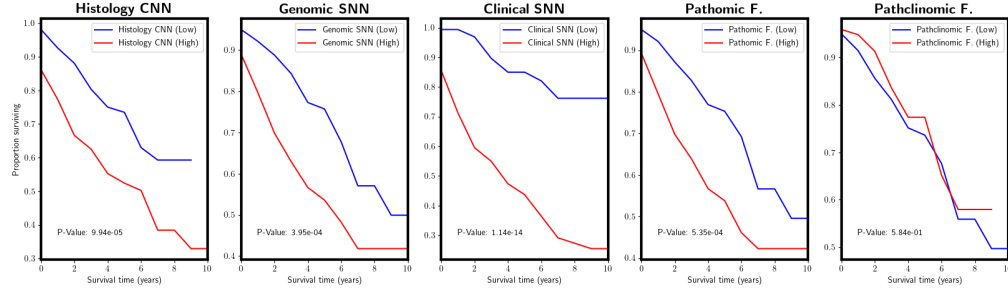


Figure 9: Kaplan Meier curves to compare the performance of the different models trained on the TCGA-KIRC dataset, stratifying the hazards by 50% and 100% percentiles, and with the log-rank test p-values.

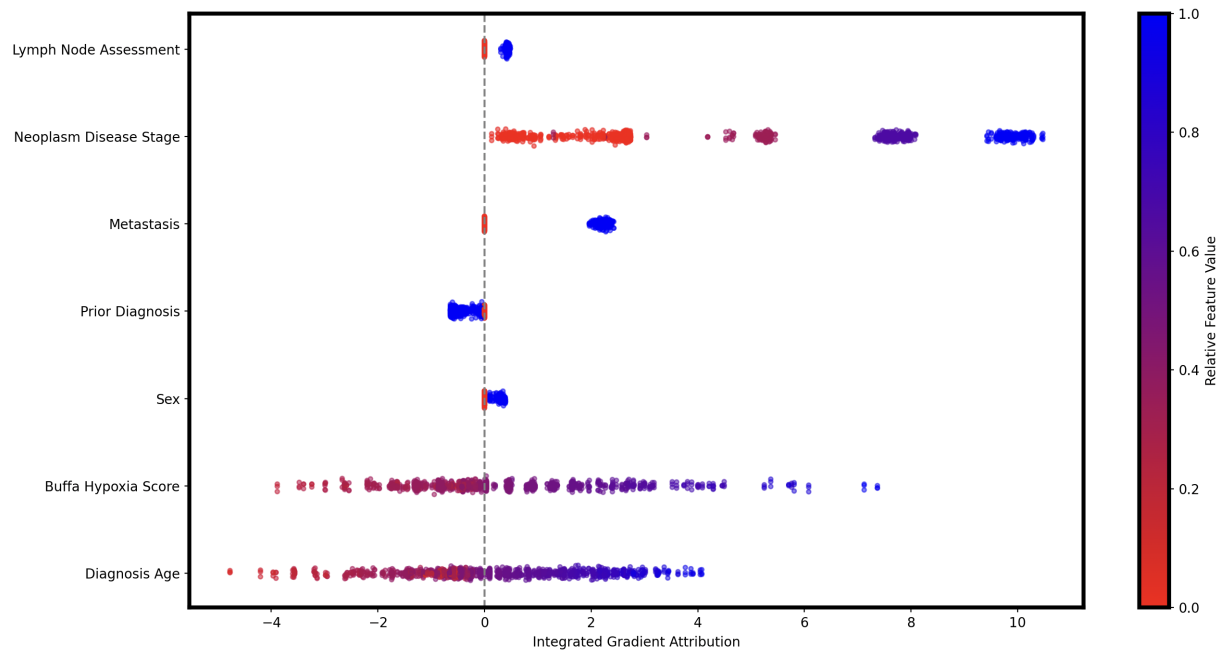


Figure 10: Integrated gradient attribution of the clinical features of the TCGA-KIRC clinical SNN model for fold k=1.

#### 4.4 Conclusion

Overall, the pathgraphomic fusion (CNN+GCN+SNN) model demonstrated great success at grade prediction for the TCGA-GBMLGG dataset, as in the original paper, outperforming the other networks across almost all performance metrics. It performed very well for the survival analysis of TCGA-GBMLGG; however, pathgraphomic fusion (CNN+GCN+SNN) had an overall average c-index of slightly less than both pathomic fusion (CNN+SNN) and the genomic SNN. Notably, this could be due to the differences in the evaluation methods on the histology ROIs in this reproduction and the original paper. The fusion models did not perform as well for the TCGA-KIRC dataset with the simple clinical model far surpassing their performance. The merit of clinical information in survival analysis is clear but alternative fusion approaches may need to be utilised to integrate the clinical data more effectively, such as concatenating the clinical embedding with the genomic embedding, rather than incorporating it in the Kronecker product as its own embedding. Overall, this study replicated and extended upon the findings of the original paper, confirming the efficacy of Kronecker product fusion approaches in cancer grade and survival analysis for TCGA-GBMLGG. The integration of histology, genomic, and clinical data through advanced neural network architectures and fusion strategies holds promise for more accurate and personalized cancer prognosis and treatment.

## 5 Appendix

### 5.1 Patient Inclusion

Malignant gliomas include both low-grade gliomas (LGG), which are classified as grades II-III according to the World Health Organisation, and glioblastoma (GBM), formerly called glioblastoma multiforme, which are grade IV gliomas [42]. The TCGA-GBMLGG dataset from the paper is a combination of two TCGA datasets of patients with GBM and LGG resulting in 769 patients. Not all genomic features were available for each of the patients and thus the fusion models were trained on different subsets of patients than the unimodal models for both survival analysis and grade classification. This is because 72 patients have missing molecular subtype (IDH mutation and 1p19q codeletion status), 33 patients have missing histology subtype and grade labels, and 256 patients have missing mRNA-Seq data. The KIRC dataset of the paper comprises 417 patients with ccRCC, but 389 patients had the relevant clinical data, so this subset of patients was utilised for the analysis.

### 5.2 Data for Each Model

For each patient, 1-3  $20 \times 1024 \times 1024$  histology regions of interest (0.5 $\mu$  / pixel) were each utilised as a singular data point in the cross-validation and the relevant genomic and clinical data were copied over. Aggregations by patient TCGA ID were utilised in the evaluation methodology to account for this. For survival analysis of the TCGA-GBMLGG dataset, mutation status, copy number variation (CNV), and RNA-Seq data were included resulting in 357 genomic and transcriptomic features. For grade prediction for TCGA-GBMLGG, only mutation status and copy number variation (CNV) were included, resulting in 80 features, as the inclusion of RNA-Seq data led to the over-training of the model. For TCGA-GBMLGG, the genes curated for the mutation and CNV data were based on the genomics features in a paper by Mobadersany et al., and included important genes such as EGFR and MYC [26]. A binary indication was utilised for the mutation status. CNV information measures amplified and deleted copies of genes and chromosome regions with amplified regions having positive values, and deleted regions having negative values. Bulk RNA-Seq measures the quantified bulk abundance of mRNA transcripts and the top 240 differentially expressed genes were included. For TCGA-KIRC, only survival analysis was performed. As with the TCGA-GBMLGG dataset, the RNASeq of the top 240 differentially expressed genes were included. The CNV information that was included comprised of the amplified/deleted genes with greater than 7% deletion or amplification, yielding 117 CNV features. This resulted in a total of 357 genomic and transcriptomic features for the TCGA-KIRC dataset.

### 5.3 Cox Model

The Cox proportional hazards model is a semi-parametric model used in survival analysis to assess the effect of several variables on the time a specified event (death) takes to happen [43]. The hazard function  $h(t|X)$  for an individual with covariates  $X$  represents the instantaneous rate of occurrence of the event at time  $t$ , given that the subject has survived up to time  $t$  and is given by:

$$h(t|X) = h_0(t) \exp(\beta^T X)$$

where  $h_0(t)$  is the baseline hazard function, which describes how the risk of an event changes over time in the absence of the covariates and  $\beta$  is the vector of regression coefficients. The Cox partial log likelihood loss is based on the ratio of the hazard functions, which are "proportional hazards", as their ratio is constant over time and only varies with the covariates. The Cox partial log likelihood loss can be used to estimate  $\beta$  and is as follows:

$$l(\beta, X) = - \sum_{i \in U} \left( X_i \beta - \log \sum_{j \in R_i} e^{X_j \beta} \right)$$

where  $U$  is the set of uncensored patients (those who experienced the event, i.e., death before last follow-up), and  $R_i$  is the set of patients whose time of death or last follow-up is later than  $i$ . In deep learning, features

from hidden layers can be used as covariates  $X$  in the Cox model.

## 5.4 Network Training

Figure 11 demonstrate different training and testing losses per epoch, as well as different grade accuracies, for grade prediction for TCGA-GBMLGG averaged over the first five folds for the different models. Figure 12 demonstrates different training and testing losses per epoch, as well as different survival accuracies and c-indices, for the different models of TCGA-GBMLGG averaged over the first five folds for the different models. Figure 13 demonstrates different training and testing losses per epoch, as well as the different survival accuracies and c-indices for the different models for TCGA-KIRC trained on the first four folds for the different models. Most of the training parameters used were taken from the ‘train\_opt.txt’ files found in the checkpoints file within the github repository associated with the paper. The Histology CNN was initialized using pretrained weights from ImageNet, followed by fine-tuning the network using a low learning rate of 0.0005, but the batch size was changed to 32 and the number of epochs set to 25 for faster training. Data augmentation of the form of random crops of  $512 \times 512$ , colour jittering, and random vertical and horizontal flips were performed on the  $1024 \times 1024$  images for training and testing. The histology GCN and genomic SNN were trained with a learning rate of 0.002 with a batch size of 32 and 64, respectively, for 50 epochs. The genomic network was regularised with L1 regularisation. For the fusion models with genomic information, L1 regularisation was used on just the genomic network. The networks were all trained with an Adam optimizer, dropout probability of  $p = 0.25$ , and a linearly decaying learning rate scheduler. For the fusion models, the genomic and graph networks were initially frozen and the final linear layers of the multimodal networks were trained utilizing a learning rate of 0.0001 and an Adam optimizer. At epoch 5, the genomic and graph networks were unfrozen, continuing the training for an additional 25 epochs with a learning rate of 0.0001.

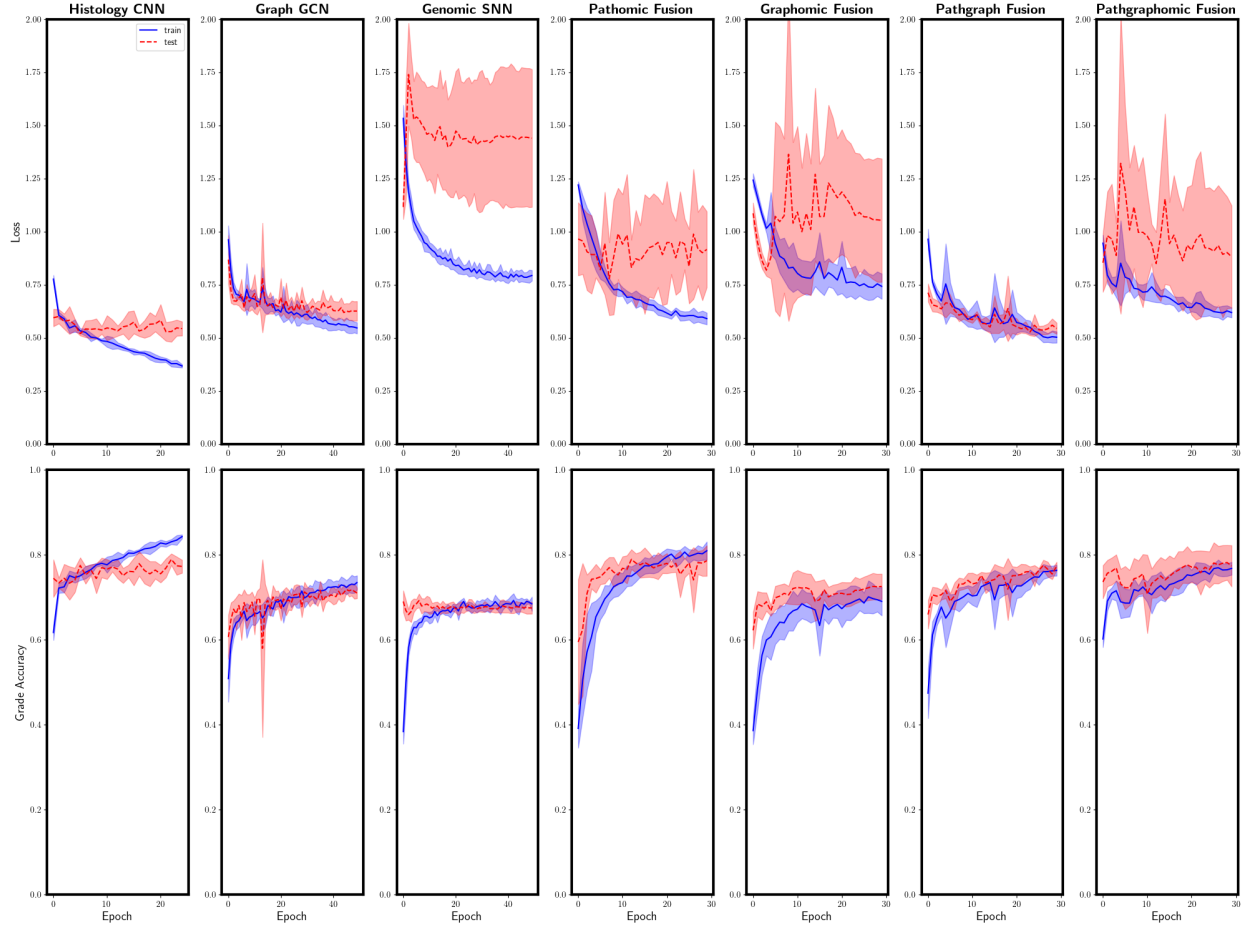


Figure 11: Plot of mean losses, and grade accuracies for the train and test set evaluated at each epoch during training of the TCGA-GBMLGG dataset for grade prediction over five folds.

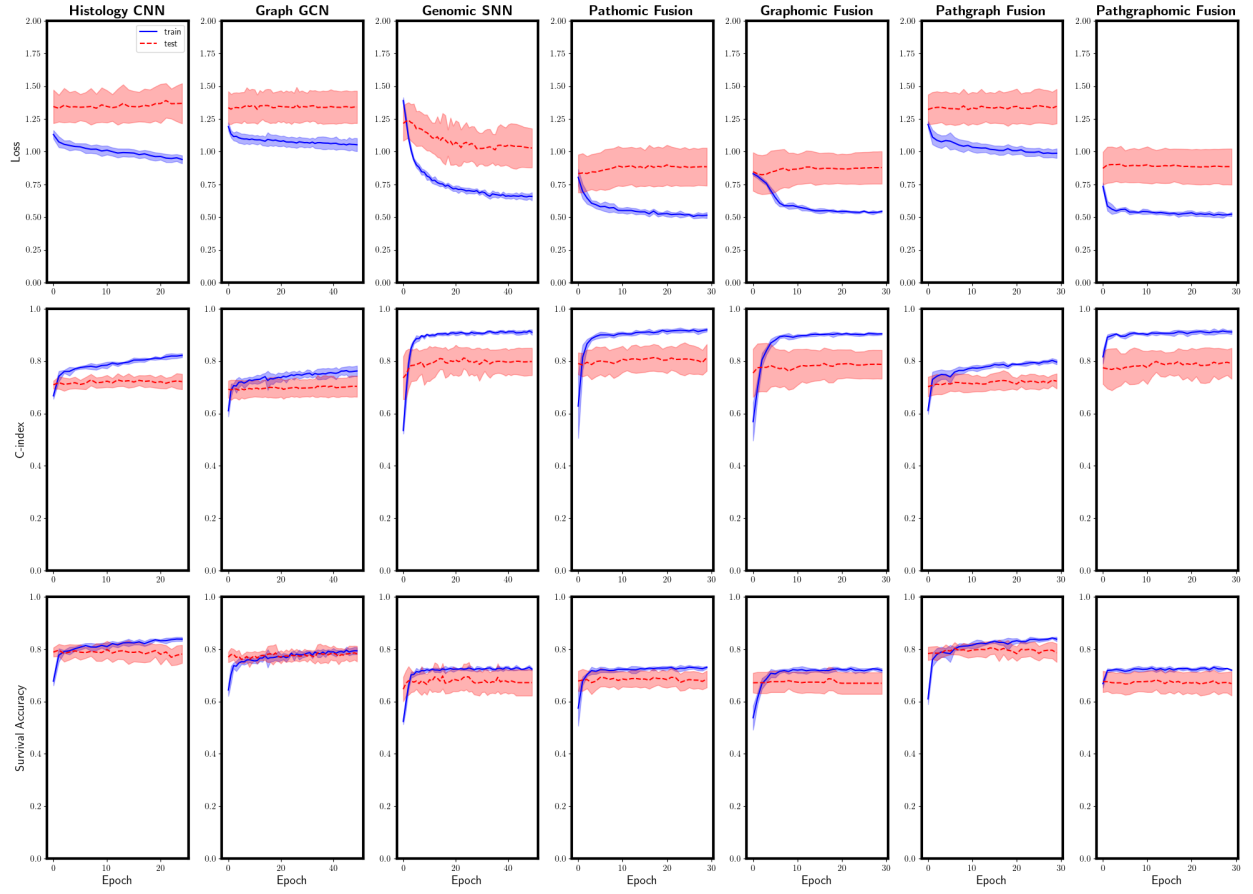


Figure 12: Plot of mean losses, c-indexes, and survival accuracies for the train and test set evaluated at each epoch during training of the TCGA-GBMLGG dataset for survival analysis over five folds.

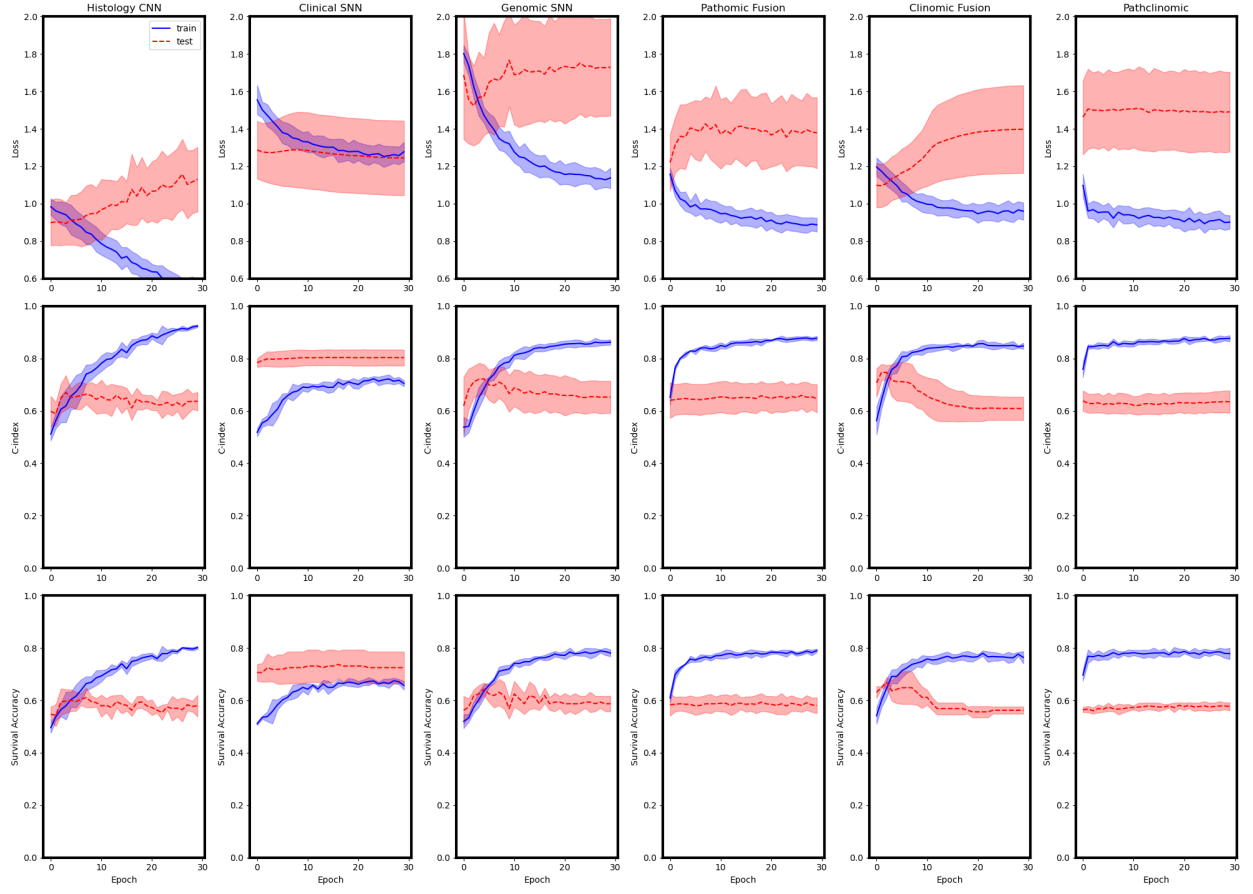


Figure 13: Training losses and other performance metrics for survival analysis of TCGA-KIRC averaged over the first four folds.

## 5.5 Additional Hazard Plots

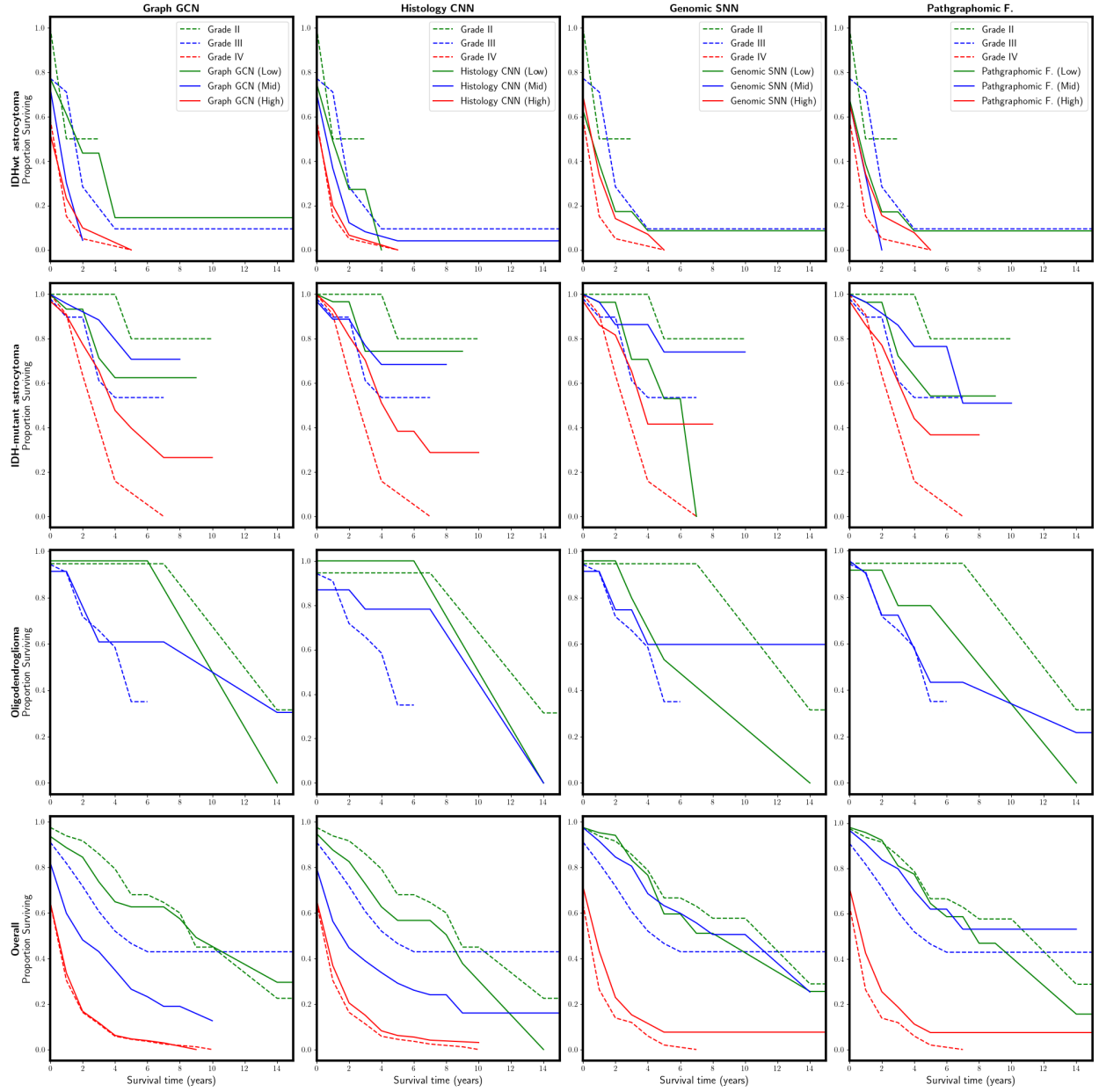


Figure 14: Kaplan-Meier comparative analysis of the histology CNN, histology GCN, genomic SNN, and pathgraphomic fusion for TCGA-GBMLGG survival analysis split by histomolecular subtype and overall.



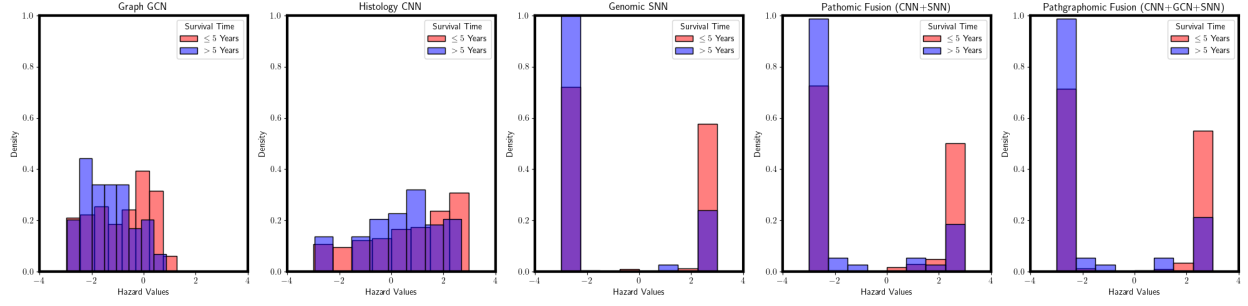


Figure 15: Histograms of the hazards from the pathgraphomic fusion (CNN+GCN+SNN) model for TCGA-GBMLGG survival analysis grouped by high survival time (greater than 5 years) and low survival time (less than five years).

## References

- [1] RJ Chen et al. “Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis”. In: *IEEE Trans Med Imaging* 41.4 (2022), pp. 757–770. DOI: 10.1109/TMI.2020.3021387.
- [2] Douglas Hanahan. “Hallmarks of Cancer: New Dimensions”. In: *Cancer Discovery* 12.1 (Jan. 2022), pp. 31–46. DOI: 10.1158/2159-8290.CD-21-1059.
- [3] Metin N Gurcan et al. “Histopathological image analysis: a review”. In: *IEEE Reviews in Biomedical Engineering* 2 (Oct. 2009). Epub 2009 Oct 30, pp. 147–171. DOI: 10.1109/RBME.2009.2034865.
- [4] Vaishali Baxi et al. “Digital pathology and artificial intelligence in translational medicine and clinical practice”. In: *Modern Pathology* 35.1 (2022), pp. 23–32. DOI: 10.1038/s41379-021-00919-2.
- [5] S. Zulkarnain et al. “Evaluation Study of Intraoperative Cytology Smear and Frozen Section of Glioma”. In: *Asian Pacific Journal of Cancer Prevention* 21.10 (Oct. 2020), pp. 3085–3091. DOI: 10.31557/APJCP.2020.21.10.3085.
- [6] S. Supplitt et al. “Current Achievements and Applications of Transcriptomics in Personalized Cancer Medicine”. In: *International Journal of Molecular Sciences* 22.3 (2021). Published 2021 Jan 31, p. 1422. DOI: 10.3390/ijms22031422.
- [7] Ph.D. Amy Y. Chow. “Cell Cycle Control by Oncogenes and Tumor Suppressors: Driving the Transformation of Normal Cells into Cancerous Cells”. In: *Nature Education* 3.9 (2010). Division of Tumor Cell Biology, Beckman Research Institute, City of Hope, p. 7.
- [8] W Aiman, DP Gasalberti, and A Rayi. *Low-Grade Gliomas*. StatPearls [Internet]. Updated 2023 May 6. Treasure Island (FL): StatPearls Publishing, 2024. URL: <https://www.ncbi.nlm.nih.gov/books/NBK560668/>.
- [9] D Purves, GJ Augustine, D Fitzpatrick, et al., eds. *Neuroscience*. 2nd. Sunderland (MA): Sinauer Associates, 2001. Chap. Neuroglial Cells. URL: <https://www.ncbi.nlm.nih.gov/books/NBK10869/>.
- [10] David N Louis et al. “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary”. In: *Neuro-oncology* 23.8 (2021), pp. 1231–1251. DOI: 10.1093/neuonc/noab106.
- [11] Yae Won Park et al. “The 2021 WHO Classification for Gliomas and Implications on Imaging Diagnosis: Part 1—Key Points of the Fifth Edition and Summary of Imaging Findings on Adult-Type Diffuse Gliomas”. In: *Journal of Magnetic Resonance Imaging* (2023). DOI: 10.1002/jmri.28743.
- [12] Brain Tumour Research. *Types of Brain Tumours: Glioma*. Accessed: 2024-07-03. 2023. URL: [https://braintumourresearch.org/pages/types-of-brain-tumours-glioma%5C#:~:text=Grade%5C%20II%5C%20gliomas%5C%20\(such%5C%20as,of%5C%20around%5C%205%5C%2D10%5C%25](https://braintumourresearch.org/pages/types-of-brain-tumours-glioma%5C#:~:text=Grade%5C%20II%5C%20gliomas%5C%20(such%5C%20as,of%5C%20around%5C%205%5C%2D10%5C%25).
- [13] J. Yang, K. Wang, and Z. Yang. “Treatment strategies for clear cell renal cell carcinoma: Past, present and future”. In: *Frontiers in Oncology* 13 (2023), p. 1133832. DOI: 10.3389/fonc.2023.1133832.
- [14] ZL Smith, EJ Pietzak, CK Meise, et al. “Simplification of the Fuhrman grading system for renal cell carcinoma”. In: *Canadian Journal of Urology* 22.6 (2015), pp. 8069–8073.
- [15] X.Y. Chen et al. “MRI-Based Grading of Clear Cell Renal Cell Carcinoma Using a Machine Learning Classifier”. In: *Frontiers in Oncology* 11 (2021). Published 2021 Oct 1, p. 708655. DOI: 10.3389/fonc.2021.708655.
- [16] National Cancer Institute. *Clear Cell Renal Cell Carcinoma*. Accessed: 2024-07-06. 2020. URL: <https://www.cancer.gov/pediatric-adult-rare-tumor/rare-tumors/rare-kidney-tumors/clear-cell-renal-cell-carcinoma#:~:text=The%205%2Dyear%20survival%20rate,survival%20rate%20is%20about%2010%25..>
- [17] S. Gadiya, D. Anand, and A. Sethi. “Histograms: Graphs in histopathology”. In: *arXiv preprint arXiv:1908.05020* (2019).
- [18] J. Wang et al. “Weakly supervised prostate TMA classification via graph convolutional networks”. In: *arXiv preprint arXiv:1910.13328* (2019).

- [19] Y. Zhou et al. “CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images”. In: *arXiv preprint arXiv:1909.01068* (2019).
- [20] Simon JD Prince. *Understanding Deep Learning*. University College London, 2019.
- [21] V Subramanian et al. “Correlating cellular features with gene expression using cca”. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 805–808.
- [22] I Carmichael and J Marron. “Joint and individual analysis of breast cancer histologic images and genomic covariates”. In: *arXiv preprint arXiv:1912.00434* (2019).
- [23] N Coudray et al. “Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning”. In: *Nature medicine* 24.10 (2018), p. 1559.
- [24] JN Kather et al. “Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer”. In: *Nature medicine* (2019), p. 1.
- [25] Jana Lipkova et al. “Artificial intelligence for multimodal data integration in oncology”. In: *Cancer Cell* 40.10 (Oct. 2022), pp. 1095–1110. DOI: 10.1016/j.ccell.2022.09.012.
- [26] P Mobadersany et al. “Predicting cancer outcomes from histology and genomics using convolutional networks”. In: *Proceedings of the National Academy of Sciences* 115.13 (Mar. 2018), E2970–E2979.
- [27] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. “Attention-based Deep Multiple Instance Learning”. In: *arXiv preprint arXiv:1802.04712* (2018). ICML 2018 paper. URL: <https://doi.org/10.48550/arXiv.1802.04712>.
- [28] R.J. Chen et al. “Multimodal co-attention transformer for survival prediction in gigapixel whole slide images”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 4015–4025.
- [29] Carolyn Hutter and Jean Claude Zenklusen. “The Cancer Genome Atlas: Creating lasting value beyond its data”. In: *Cell* 173.2 (2018), pp. 283–285. DOI: 10.1016/j.cell.2018.03.042.
- [30] Wolf-Dieter Heiss, Peter Raab, and Heinrich Lanfermann. “Multimodality Assessment of Brain Tumors and Tumor Recurrence”. In: *Journal of Nuclear Medicine* 52.10 (2011), pp. 1585–1600. ISSN: 0161-5505. DOI: 10.2967/jnumed.110.084210. eprint: <https://jnm.snmjournals.org/content/52/10/1585.full.pdf>. URL: <https://jnm.snmjournals.org/content/52/10/1585>.
- [31] F. Mahmood et al. “Adversarial training for multi-organ nuclei segmentation in computational pathology images”. In: *IEEE Transactions on Medical Imaging* (2018).
- [32] Marius Muja and David G Lowe. “Fast approximate nearest neighbors with automatic algorithm configuration”. In: *VISAPP (1)*. Vol. 2. Citeseer. 2009, pp. 331–340.
- [33] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. “Representation learning with contrastive predictive coding”. In: *arXiv preprint arXiv:1807.03748* (2018).
- [34] William L Hamilton, Zhitao Ying, and Jure Leskovec. “Inductive representation learning on large graphs”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 1024–1034.
- [35] J. Lee, I. Lee, and J. Kang. “Self-attention graph pooling”. In: *arXiv preprint arXiv:1904.08082* (2019).
- [36] Abdallah Alabdallah et al. “The Concordance Index Decomposition: A Measure for a Deeper Understanding of Survival Prediction Models”. In: *Center for Applied Intelligent Systems Research (CAISR), Halmstad University, Sweden* (2024).
- [37] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. “Axiomatic Attribution for Deep Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR. 2017, pp. 3319–3328.
- [38] Young-Ran Lee, Mei Chen, and Pier Paolo Pandolfi. “The functions and regulation of the PTEN tumour suppressor: new modes and prospects”. In: *Nature Reviews Molecular Cell Biology* 19.9 (2018), pp. 547–562. DOI: 10.1038/s41580-018-0015-0.
- [39] Hidde F van Thuijl, Ilari Scheinin, Daoud Sie, et al. “Spatial and temporal evolution of distal 10q deletion, a prognostically unfavorable event in diffuse low-grade gliomas”. In: *Genome biology* 15.9 (2014), p. 471. DOI: 10.1186/s13059-014-0471-6.

- [40] H Suzuki, X Zhou, J Yin, et al. “Intragenic mutations of CDKN2B and CDKN2A in primary human esophageal cancers”. In: *Human Molecular Genetics* 4.10 (1995), pp. 1883–1887. DOI: 10.1093/hmg/4.10.1883.
- [41] SE Ahmadi, S Rahimi, B Zarandi, et al. “MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies”. In: *Journal of Hematology & Oncology* 14 (2021), p. 121. DOI: 10.1186/s13045-021-01111-4.
- [42] H Zhang, B Zhang, W Pan, et al. “Preoperative Contrast-Enhanced MRI in Differentiating Glioblastoma From Low-Grade Gliomas in The Cancer Imaging Archive Database: A Proof-of-Concept Study”. In: *Frontiers in Oncology* 11 (Jan. 17, 2022), p. 761359. DOI: 10.3389/fonc.2021.761359.
- [43] John Fox and Sanford Weisberg. *Cox Proportional-Hazards Regression for Survival Data in R: An Appendix to An R Companion to Applied Regression, third edition*. Tech. rep. Last revision: 2023-01-31. Jan. 2023.