



OPEN Decision level scheme for fusing multiomics and histology slide images using deep neural network for tumor prognosis prediction

Tingting Zhao¹, Yongyong Ren^{2,3}, Hui Lu^{1,2,3,4} & Yan Kong^{1,2}✉

Molecular biostatistical workflows in oncology often rely on predictive models that use multimodal data. Advances in deep learning and artificial intelligence technologies have enabled the multimodal fusion of large volumes of multimodal data. Here, we presented a decision level multimodal data fusion framework for integrating multiomics and pathological tissue slide images for prognosis prediction. Our approach established the spatial map of instances by connecting the neighboring nuclei in space and calculated the characteristic tensor via graph convolution layers for the input pathological tissue slide images. Global Average Pooling was applied to align and normalize the feature tensors from pathological images and the multiomics data, enabling seamless integration. We tested our proposed approach using Breast Invasive Carcinoma data and Non-Small Cell Lung Cancer data from the Cancer Genome Atlas, which contains paired whole-slide images, transcriptome data, genotype, epigenetic, and survival information. In a 10-fold cross-validation, the comparison results demonstrated that the multimodal fusion paradigm improves outcome predictions from single modal data alone with the average C-index increasing from 0.61 to 0.52 to 0.75 and 0.67 for breast cancer and non-small cell lung cancer cohort, respectively. The proposed decision level multimodal data fusion framework is expected to provide insights and technical methodologies for the follow-up studies.

Keywords Multimodal data fusion, Pathological slide images, Deep learning, Multiomics, Prognosis

Cancer is caused by aberrant cell proliferation in human body and is defined by hallmark histopathology, genomic, and genotypic heterogeneity in the tumor micro-environment that contribute to variability in treatment response and survival rate^{1,2}. Advances in sequence technologies and bio-informatics have sped up the widespread of multiomics data in cancer research for biomarkers, prognosis, and therapeutic response prediction^{3,4}. Moreover, the digital pathology and deep convolution neural networks have promoted the potential applications to analyze gigapixel whole slide images for lesion detection, benign and malignant discrimination, and prognosis prediction^{5,6}. The integration of multimodal data presents more opportunities on precision oncology and provides possibilities to increase the robustness and precision of diagnostic and predictive models⁷. Recent developments in bioinformatics platforms facilitate the comprehensive analysis of multiomics data, bridging the gap between genomic, transcriptomic, and imaging modalities to provide deeper insights into tumor heterogeneity and treatment response⁸. Such advancements bring AI closer to mimicking human intelligence in clinical decision-making⁹.

In the field of prognostic prediction, data-driven approaches have shown remarkable power in informing clinical decisions¹⁰. Combining whole-slide pathological images with genomic data has been shown to enhance the accuracy of prognostic predictions, as this integration can provide complementary information^{11–13}. However, the current research on the integration of pathological images and genomic data remains insufficient, particularly in the in-depth exploration of multiomics data, with most studies limited to one or two types of omics data (e.g., mRNA)¹³.

¹Center for Biomedical Informatics, Shanghai Engineering Research Center for Big Data in Pediatric Precision Medicine, Shanghai Children's Hospital, Shanghai 200240, China. ²SJTU-Yale Joint Center for Biostatistics and Data Science, Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai, China. ³Institute of Bioinformatics, Shanghai Academy of Experimental Medicine, Shanghai, China. ⁴State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ✉email: kongyan@sjtu.edu.cn

In this paper, we proposed a decision level multimodal data fusion framework^{14,15} for fusing imaging features from whole slide images and molecular features from multiomics data which consist of expression data, methylation data, and the mutation data. To verify the improvement of model genuinely results from the fusion features, deep-learning-based prognostic prediction models were built for the single-modal data, ensuring consistency with the neural network architecture embedded in the multimodal fusion framework prior to the fusion module. Compared with the models based on either histology information or omic information, the proposed multimodal fusion network achieved the best performance when tested on both The Cancer Genome Atlas (TCGA)¹⁶ Breast Invasive Carcinoma (BRCA) data¹⁷ and Non-Small Cell Lung Cancer (NSCLC) data¹⁸. Fusing the spatial morphological information from pathological whole slides with the multiomics provides the possibility to better understand the associations between phenotype and genotype and to further explore the capabilities of image-omic assays for tumor diagnosis, treatment, prognosis, and drug resistance prediction.

Results

Network implementation details

All the models were implemented using the Python PyTorch (version 1.4.0) deep learning framework and trained using two Tesla V100 GPUs on a single node. Due to GPU memory constraints, we used stochastic gradient descent with a batch size of 20. The initial learning rate was set to 0.01 and then decreased by a factor of 10 every 10 epochs. The weights were initialized from a Gaussian distribution with a mean and standard deviation of 0 and 0.01, respectively. Models for single modal data were trained from scratch for a total of 50 epochs since there is no prior knowledge. Moreover, rather than training the multimodal fusion network end-to-end, we initialized the model using weights from the single-modal network. All models were trained and tested on the same configurations and run for about 10–16 h.

Typical features from multimodal data

For the gene expression data, 1340 genes were selected using a cut-off of $|\log_2(\text{Fold Change})| > 1$ and false discovery rate (FDR) < 0.05 . For the somatic mutation data, 4518 sites with high mutation frequency (appearing in over 80% of samples) were chosen. For the methylation data, 9305 sites that were hypermethylated or hypomethylated were selected using a cut-off of $|\log_2(\text{Fold Change of M-value})| > 2$ and FDR < 0.01 . After intersecting the samples across data modalities, the multiomics data matrix was updated with 679 patients and the updated dimensions were as follows: $679 \times 1,340$ for expression data, $679 \times 4,518$ for mutation data, and $679 \times 9,305$ for methylation data. As shown in Fig. 1A, we got a multiomics feature-based tensor for each sample.

We used the hematoxylin and eosin (H&E) stained imaging as the second modality of data. Considering the large dimensions of the whole slide images, we selected 50 slide tiles of 1000×1000 pixels for each patient, resulting in a total of 33,950 image patches, along with the corresponding cellular spatial graph data. As shown in Fig. 1a and b, we finally got an image features-based tensor for each sample.

The fusion tensor was calculated using the tensor fusion method and was considered as the input for Coxnet for prognosis analysis (Fig. 2). The data mentioned here all pertain to the BRCA cohort, and the details of the samples and features used in this study were listed in Table 1.

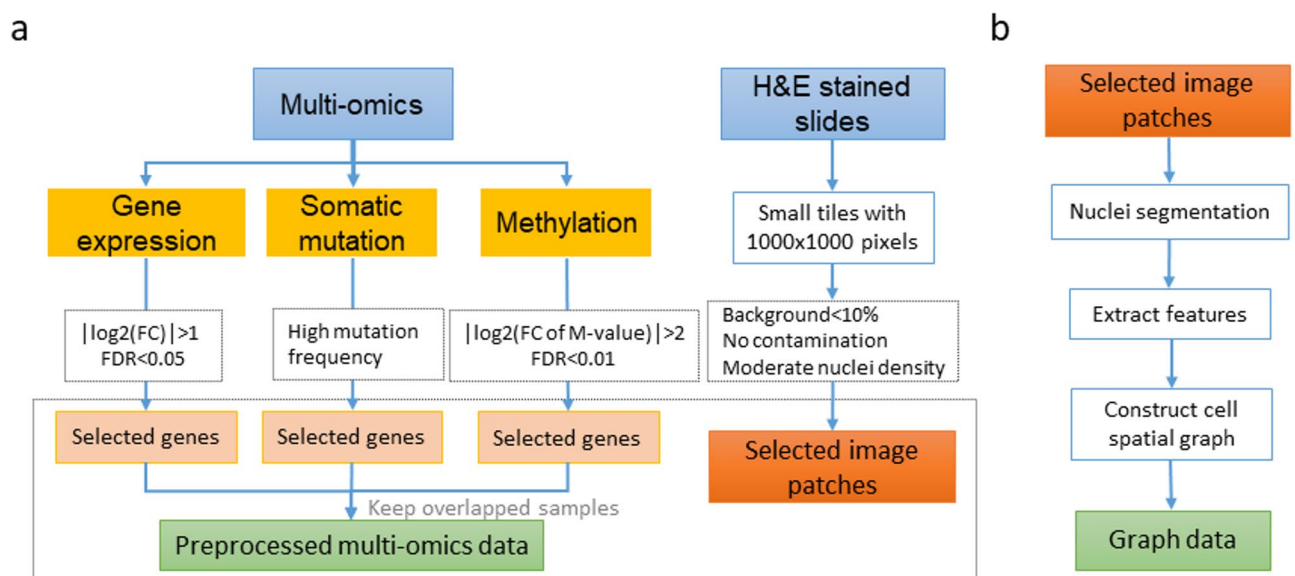


Fig. 1. Details of feature extraction steps for different modal data. **(a)** Pre-process workflow of different modal data. **(b)** The steps of generating nuclei spatial graph data.

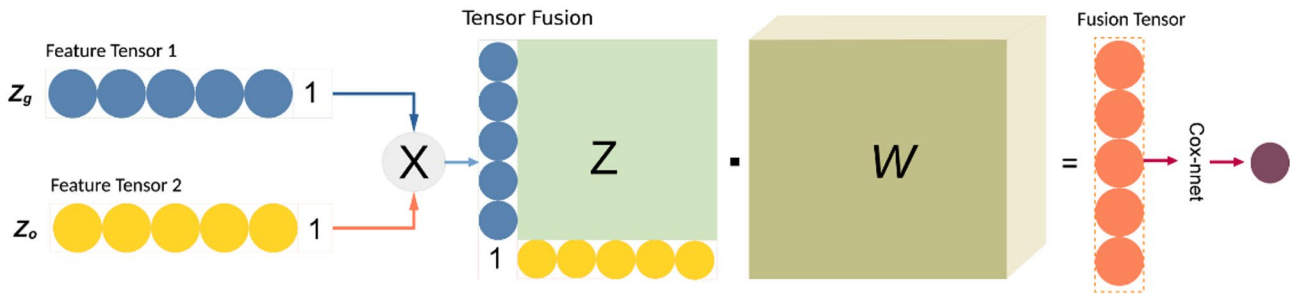


Fig. 2. Tensor fusion method used in this study.

Cancer	Data type	Raw tumor samples	Raw normal samples	Filter condition	Selected samples	Selected features
BRCA	RNA-seq	1071	99	$ \log_2FC > 2$ and $FDR < 0.05$	679	1340
	Methylation site	774	81	$ \log_2FC \text{ of M-value} > 2$ and $pvalue < 0.05$	679	9305
	Somatic mutation	989	–	Appeared in over 80% samples	679	4518
	Graphs of tissue slide	–	–	–	679×50	1041 (17 + 1024)
LUAD	RNA-seq	510	58	$ \log_2FC > 2$ and $FDR < 0.05$	443	2500
	Methylation site	451	32	$ \log_2FC \text{ of M-value} > 2$ and $pvalue < 0.05$	443	8743
	Somatic mutation	567	–	appeared in over 80% samples	443	1577
	Graphs of tissue slide	–	–	–	443×50	1041
LUSC	RNA-seq	496	49	$ \log_2FC > 2$ and $FDR < 0.05$	359	2500
	Methylation site	365	41	$ \log_2FC \text{ of M-value} > 2$ and $pvalue < 0.05$	359	8743
	Somatic mutation	485	–	appeared in over 80% samples	359	1577
	Graphs of tissue slide	–	–	–	359×50	1041

Table 1. The data summary before and after pre-processing.

Model performance comparison

Different prognostic prediction models were developed using only multiomics data, only slide image data, and the fusion data. In order to avoid overfitting, the model was trained and tested using 10 fold cross-validation. The predicted hazard ratios on 10 independent test datasets were added and averaged. We also built a baseline model using the cox regression function based on clinical characteristics, including clinical grade and tumor stage. To assess the prognostic accuracy of the prognostic prediction model using multimodal data, we performed the assessments using Kaplan-Meier method both on the BRAC and NSCLC datasets. Patients were stratified into high- or low-risk group according to the predicted survival time using the median value as the cutoff. The results demonstrated that the two groups were more distinct with the model that used the fusion data (Fig. 3).

Concordance index (C-index) for the prognostic prediction models were calculated and all the deep learning-based models outperformed the baseline model (see Tables 2 and 3). The average C-index for prognosis prediction of the multimodal fusion data model was higher than the single-modal data model for both NSCLC and BRCA dataset. Moreover, area under the curve (AUC) of the multimodal fusion model in the prediction of five-year disease-specific outcome was significantly higher than the multiomics-only model (Venkatraman’s p-value was 0.00113 for BRCA and 0.01267 for NSCLC) and pathology-only model (Venkatraman’s p-value was 0.00251 for BRCA and 0.00096 for NSCLC). Taken together, our proposed multimodal data fusion framework was significantly better in discriminating low-risk from high-risk patients when applied to both the BRCA (log-rank p-value < 0.05) and NSCLC (log-rank p-value < 0.01) datasets. We further demonstrate the extensibility and generalization performance of our fusion method through testing on 10 independent cross-validation test sets.

In multivariable analyses, the Fusion Net score remained a robust independent predictor of survival after adjusting for clinical covariates. In the BRCA cohort, the Fusion Net score demonstrated a strong independent association with survival outcomes in multivariable Cox regression (HR=4.24, 95% CI: 3.04–5.91, $p < 0.001$), adjusted for age and tumor stage (Supplementary Table). While advanced tumor stages (e.g., Stage IV, HR=6.27) showed expected high mortality risks, the Fusion Net score provided additional prognostic discrimination beyond traditional staging. In the NSCLC cohort, the Fusion Net score demonstrated a strong independent association with survival outcomes in multivariable Cox regression (HR=1.75, 95% CI: 1.57–1.95, $p < 0.001$), adjusted for age, tumor stage and gender (Supplementary Table).

Discussion

Computational oncology is experiencing unprecedented challenges and opportunities due to the accumulated multimodal data and the emerging deep learning based method^{19,20}. A major goal of computational oncology

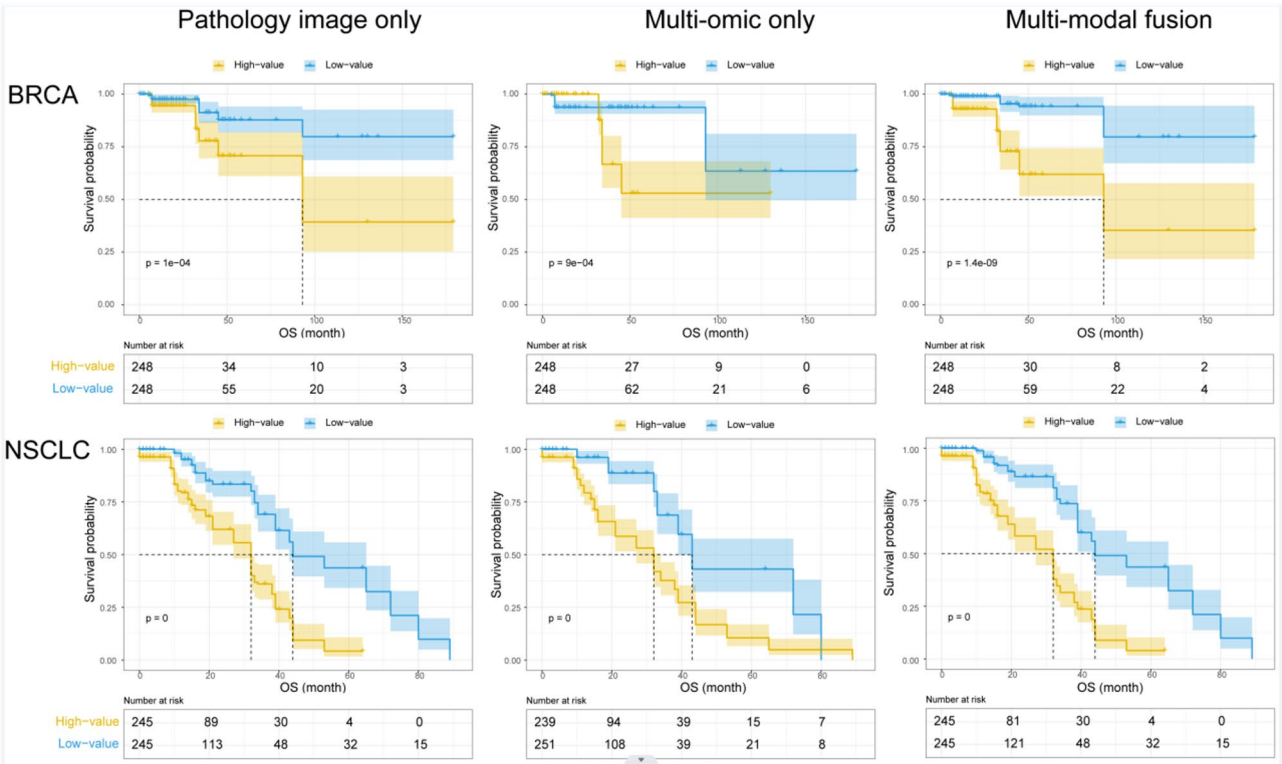


Fig. 3. Kaplan-Meier curves for different risk groups predicted by 3 models.

Model Name	C-index
Baseline COX (clinical)	0.586±0.043
Graph Convolution Net (only graph)	0.610±0.129
Multimomics Neural Net (only omics)	0.696±0.153
Fusion Net	0.753±0.067

Table 2. Comparison of prognosis between 4 models for BRCA.

Model name	C-index
Baseline COX (clinical)	0.553±0.107
Graph Convolution Net (only graph)	0.515±0.0143
Multimomics Neural Net (only multiple omics)	0.647±0.0840
Fusion Net	0.669±0.0673

Table 3. Comparison of prognosis between 4 models for NSCLC.

is to predict the future states of tumors such as prognosis of patients²¹ lymphatic metastasis²² and the recurrent status²³. More generally, the aim is to link data characteristics with tumor state prediction or clinical outcomes. Although multimodal data are widely used in various oncology research areas^{24,25} few studies have applied them together to address clinical prediction problems especially using both multimomics and image data.

Images and omics data are two different modal data that represent the genotype and phenotype levels, respectively. Combining imaging and omics data is not a trivial task, as these data are typically designed and generated for different purposes. Modality-specific characteristics and complementary features provide additional information and hold great potential for enhancing the power of statistical models. Therefore, fusing omics and images approaches is essential for combining the macro-level structural and functional data embedded in the biomedical images, with micro-level molecular signatures in omics, thereby providing a coherent, accurate, and reliable source of information. Integration of Omics and biomedical images is an emerging interdisciplinary field aimed at development of methods for integration of biomedical images and omics data. Some studies have suggested several prognostic risk factors, including differentially expressed genes, somatic mutation

sites²⁶, abnormal methylation modification sites²⁷ and the morphological features of tumor cells²⁸. In our study, we proposed a decision level multimodal fusion framework for integrating multiomics data with pathological tissue slide images. The multiomics data we used included RNA-Seq expression, somatic mutations, and methylation profiles which covered genomic, transcriptomic and epigenetic information. With prognosis prediction as the goal, we compared the performances of different models. Our proposed multimodal fusion network achieved better performance on both independent datasets. The results demonstrated that the modal complementary could improve the prediction performance. Furthermore, multimodal data fusion methodology could also be helpful in the tumor mechanism studies, such as genotype-phenotype relationships, invasive analyses, and the identification of new biomedical markers. The fusion strategy holds promise for improving current preventive and predictive disease models, enabling earlier detection, more precise diagnosis, and tailored treatment. Therefore, multimodal data fusion will be fundamental for the development of a personalized medicine approach, extremely beneficial for the patients as well as for the whole healthcare system. Our proposed decision level multimodal fusion framework is hoped to provide insights and technically methodologies for the follow-up studies that seed to merge heterogeneous data in disease mechanisms.

Despite these advantages, it is always rare to obtain all modal data at the same time. According to our comparison results, although multimodal fusion could improve the prediction power, the performance of the single-modal model is also acceptable (Fig. 3; Tables 2 and 3). In our study, we trained the multimodal fusion network on BRCA and NSCLC datasets separately. Following this framework, one can conduct training steps on any dataset to obtain a pretrained weight. To solve the problem of missing modalities while leveraging multimodal fusion, we suggest using the pretrained feature tensor for the missing modality.

For future works, it is attractive to incorporate the structured text data as another modal data, especially the records of medical care and related information, into our framework. In terms of the technical approach, the temporal point process (TPP) could be an effective tool for incorporating certain priors, such as self-exciting patterns, into the record sequences^{29,30}. In particular, the deep point process has emerged as a promising topic by combining the interpretability of classic Bayesian TPP and high capacity of neural models. It offers a pathway to integrate with our deep learning based multimodal learning and analysis framework.

Methods

Data description

The multimodal dataset used in this study are sourced from TCGA. We downloaded the multiomics data and H&E stained pathological tissue slide images of BRCA and NSCLC patients from the website. The multiomics data we used included RNA-Seq expression, somatic mutations, and methylation profiles. To construct the typical feature matrix within the matched samples, we use different pre-process pipelines for the multiomics data and the pathological tissue slide images (Fig. 1a). A total of 679 BRCA patients and 802 NSCLC patients were selected in this study, with the NSCLC dataset comprising two parts: 443 cases of Lung Adenocarcinoma (LUAD) and 359 cases of Lung Squamous Cell Carcinoma (LUSC). The correspondingly clinical pathological characteristics were also used, including age, sex, smoking history, stage, pathological subtype, the last follow-up days, and overall survival (OS) days.

For the RNA-Seq expression profiles, we extracted the mRNA expression matrix and calculate the different expression levels of genes between the tumor and adjacent normal tissues using the edgeR package³¹. The expression matrix of significantly differentially expressed genes of tumor tissues were retained as the first omic data. For the mutation data, we downloaded the publicly available masked somatic mutation results generated by mutect2³² software and arranged them into the mutation frequency matrix according to gene names. And for the methylation data, the beta-value matrix was downloaded as the source data, which was then transformed into an M-value matrix as suggested by Du et al.³³. We performed the T-test and retained those significant methylated tags and their methylation profiles on tumor tissues. Patients who have all three types of omics data were conserved.

Due to the large size of high-resolution tissue slide images and the limitations of GPU memory, these images cannot be used directly. We randomly cropped 50 tiles with 1000 × 1000 pixel size for each patient using the OpenSlide package of python³⁴. During the process of selection, we set some filtering criteria including background area less than 10%, without obvious contamination, and with moderate nuclei density. Small tiles were preserved if they met the criteria.

The proposed fusion network

In this work, we proposed a decision level multimodal data fusion framework for multiomics and pathological tissue slide images for survival outcome prediction. Our proposed method models feature interactions across modalities by using the tensor fusion strategy and assess the expressiveness of each representation before fusion. The workflow of the decision level fusion framework is depicted in Fig. 4.

The tensor representations of pathological slide images and multiomics data are processed in parallel. Prior to the Tensor fusion module, the feature tensor of slide images was obtained via Graph Convolutional Network (GCN), and the feature tensor of the multiomics data was calculated using multi-layer perception. The Tensor fusion module will generate a fused tensor conducted as the input of the Cox-nnet.

Prognosis prediction based on H&E slide images

The spatial information of cells in the tumor microenvironment, as observed in histopathology slide images, is critical for subtyping and prognosis prediction. Unlike feature representations from whole slide images using Convolutional Neural Networks, graph representations explicitly capture only specific preassigned features of each nucleus, which can be scaled to cover large regions of tissue. To construct graphs containing the micro-environment information, we first performed nuclei instance segmentation in the histopathology region of

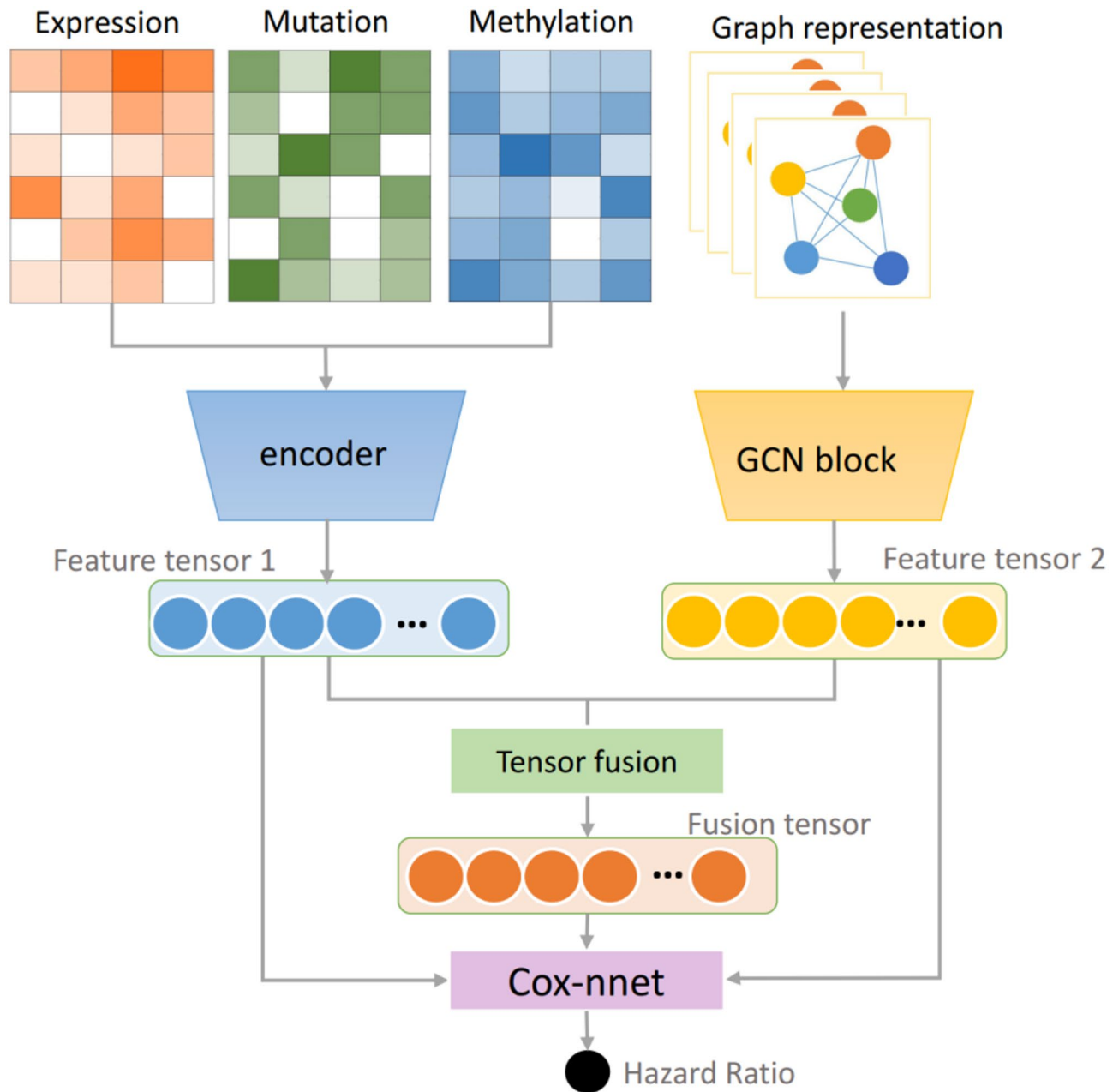


Fig. 4. Overview of our proposed Decision level fusion strategy for multimodal data. The details of the GCN block is explained in Fig. 5.

interest to define the graph nodes. Next, the adjacent nuclei were connected using the K-Nearest Neighbors to represent the edges among the graph nodes. In addition, the attributions for each nucleus we defined were captured. Finally, GCN was used for learning the robust representation of the entire graph for prognosis prediction.

Nuclei segmentation and feature extraction

We tried to focus on utilizing the spatial information and image features of the nuclei. The nuclei instances were segmented firstly using TSN-Unet⁶. And 17 morphological characteristics were extracted for each nucleus using the Python OpenCV package, including area, minor axis length, major axis length, radio of major inner circle, average values of RGB channels, average gray value, range of gray value, value range of RGB channels, the standard deviation of values from RGB channels, the standard deviation of gray value, the distance from the nearest cell. Besides the measurable morphological features, we also used the contrastive predictive coding³⁵ to extract another 1024 features from 64×64 pixel size region centered around each nucleus. Next, we built the cell graph with the edge set generated by connecting the adjacent 5 neighbor cells using the K-Nearest Neighbors

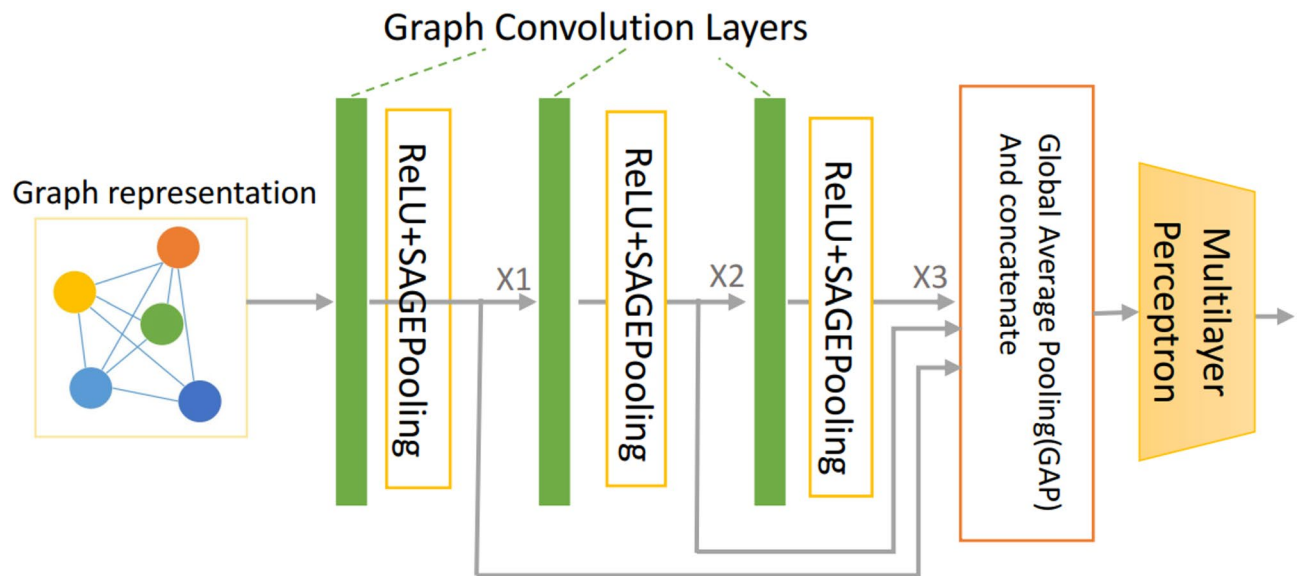


Fig. 5. The detailed architecture of the proposed GCN for the spatial graphs of nuclei. The input is the original spatial graph of nuclei. The aggregate and combine functions are adopted from the GraphSAGE architecture. The self-attention pooling strategy, SAGPooling is adopted to encode the hierarchical structure of nuclei graphs.

algorithm from the FLANN library. Finally, the graph structured data were constructed with each node represent a nuclei and the previous defined features were treated as the attributes of the node (Fig. 1b).

GCN for spatial graphs of nuclei in tissue slide images

In this study, we introduced a three-layer GCN. The GraphSAGE architecture³⁶ was used in the aggregate and combine functions. The hierarchical structure of nuclei graphs was encoded by using the self-attention pooling strategy (SAGPooling)³⁶ which is a hierarchical pooling method that performs local pooling operations of node embeddings in a graph.

As shown in Fig. 5, the input was the original spatial graph of nuclei. Three graph convolution layers were used in our proposed method, each employing ReLU as the activation function, followed by a single SAGPooling layer. During the training steps, the feature maps from one batch will be transformed into a vector representation. All three vectors generated from the 3 graph convolution layers correspondingly were concatenated. Then, a Global Average Pooling layer followed by a fully connected module was sequentially added to the network.

Prognosis prediction model based on nuclei Spatial graphs

The cox-nnet module was employed to calculate the Hazard Ratio for prognosis measurement. We got the feature tensor for slide imaging from the output of the fully connect layers, which we named Z_g . The cox-nnet module³⁷ uses the partial log-likelihood function [44] to calculate the loss, which is defined as:

$$\text{cost} = \sum (\theta_i - \log \sum e^{\theta_i * R}) * C$$

where R indicates the adjacent function and is defined as:

$$R_{ij} = \begin{cases} 1, & t_i \leq t_j \\ 0, & t_i > t_j \end{cases}$$

While θ is the predicted hazard ratio of the patient, and C represents the survival status.

Prognosis prediction based on multiomics data

Due to the difference in dimension and representation forms, the multiomics data used in this study were initially processed separately. At this stage, we employed a multi-layer perceptron to extract their feature representations. Then, the feature maps generated from three parallel multi-layer perceptrons were integrated via vector computation since they have the same dimension. The integrated tensor was also the feature tensor of the multiomics data, which we named Z_o . Finally, the multiomics integrated tensor was sent into the cox-nnet to predict the hazard ratio.

Prognosis prediction based on multimodal fusion data

We introduced the tensor fusion method³⁸ in this study to intergrate two different modalities: multiomics data and the cell spatial graph derived from the tissue slide image. According to the separate process of the

multiomics data and the tissue slide image data, we obtained the feature tensors for each modal- Z_g (dimension was 128) and Z_o (dimension was 32). The first step for the tensor fusion algorithm was to construct the input tensor Z ($Z \in \mathbb{R}^{d_1, d_2}$), which was generated by the cross product of Z_g and Z_o after adding 1 for each other. Then a vector representation (H_{fusion}) was produced through a fully connected layer (Fig. 2).

$$H_{\text{fusion}} = g(z; W, b) = W \cdot Z + b$$

where $h, b \in \mathbb{R}^{d_h}$, W denotes the weight matrix, and b is the deviant vector. Since Z is the two-order tensor, W is a tensor of order 3. The dimension of W is $d_1 \times d_2 \times d_h$, where d_h represents the size of the output layer and d_h equals 1 in the prognosis prediction task.

The fused tensor would be considered as the input of cox-nnet to predict the hazard ratio. We also used the partial log-likelihood function as the loss function.

Evaluation metrics for models

We evaluated our method using two types of metrics. The commonly used log-rank p-value of Cox-PH regression was utilized as the first evaluation metric. Kaplan-Meier survival curves for two risk groups were plotted, and the log-rank p-value was calculated to assess the differences between these curves.

We also employed C-index as the second evaluation metric. The C-index is calculated based on Harrell's C statistics and can be seen as the ratio of all pairs of individuals whose predicted survival time are ranked correctly^{39,40}. A model is considered good when C-index score around 0.7, whereas a score around 0.5 means random background.

To assess the independent prognostic value of the Fusion Net score beyond standard clinical variables, multivariable Cox proportional hazards regression models were constructed. For the BRCA dataset, the model incorporated the Fusion Net score, age, and tumor stage. For the NSCLC dataset, the model included the Fusion Net score, age, gender, and tumor stage. Proportional hazards assumptions were verified using Schoenfeld residuals, and sensitivity analyses using stratified models were performed for covariates violating the assumptions.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Received: 11 September 2024; Accepted: 30 June 2025

Published online: 15 July 2025

References

- Antonelli, L., Guarracino, M. R., Maddalena, L. & Sangiovanni, M. Integrating imaging and omics data: A review. *Biomed. Signal Process. Control.* **52**, 264–280. <https://doi.org/10.1016/j.bspc.2019.04.032> (2019).
- Chen, R. J. et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell.* **40**, 865–878. <https://doi.org/10.1016/j.ccell.2022.07.004> (2022). <https://doi.org/10.1016/j.ccell.2022.07.004>
- Gong, P. et al. Multi-omics integration method based on attention deep learning network for biomedical data classification. *Comput. Methods Programs Biomed.* **231**, 107377. <https://doi.org/10.1016/j.cmpb.2023.107377> (2023).
- Boroń, D. et al. Recent multiomics approaches in endometrial cancer. *Int. J. Mol. Sci.* **23** <https://doi.org/10.3390/ijms23031237> (2022).
- Jiao, Y., Li, J., Qian, C. & Fei, S. Deep learning-based tumor microenvironment analysis in colon adenocarcinoma histopathological whole-slide images. *Comput. Methods Programs Biomed.* **204**, 106047. <https://doi.org/10.1016/j.cmpb.2021.106047> (2021).
- Kong, Y., Genchev, G. Z., Wang, X., Zhao, H. & Lu, H. Nuclear segmentation in histopathological images using two-stage stacked U-Nets with attention mechanism. *Front. Bioeng. Biotechnol.* **8**, 573866. <https://doi.org/10.3389/fbioe.2020.573866> (2020).
- Stahlschmidt, S. R., Ulfenborg, B. & Synnergren, J. Multimodal deep learning for biomedical data fusion: a review. *Brief. Bioinform.* **23** <https://doi.org/10.1093/bib/bbab569> (2022).
- Berrar, D. & Dubitzky, W. Deep learning in bioinformatics and biomedicine. *Brief. Bioinform.* **22**, 1513–1514. <https://doi.org/10.1093/bib/bbab087> (2021).
- Ren, Y. et al. BMAP: a comprehensive and reproducible biomedical data analysis platform. *bioRxiv*, 2024.2007.2015.603507 <https://doi.org/10.1101/2024.07.15.603507> (2024).
- Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416. <https://doi.org/10.1016/j.cell.2018.02.052> (2018). .e411.
- Lu, Z., Xu, S., Shao, W., Wu, Y. & Huang, K. Deep-learning-based characterization of tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data. *JCO Clin. Cancer Inf.* **4**, 480–490 (2020).
- Binder, A., Bockmayr, M., Hgele, M., Wienert, S. & Klauschen, F. Morphological and molecular breast cancer profiling through explainable machine learning. *Nat. Mach. Intell.* 1–12 (2021).
- Shao, W. et al. Characterizing the survival-associated interactions between tumor-infiltrating lymphocytes and tumors from pathological images and multi-omics data. *IEEE Trans. Med. Imaging.* **42**, 3025–3035. <https://doi.org/10.1109/tmi.2023.3274652> (2023).
- Roheda, S., Krim, H., Luo, Z. Q. & Wu, T. In *26th European Signal Processing Conference (EUSIPCO)*. (2018).
- Veeramachaneni, K., Osadciw, L., Ross, A. & Srinivas, N. Decision-level fusion strategies for correlated biometric classifiers. *IEEE* (2008).
- Tomczak, K., Czerwinska, P. & Wiznerowicz, M. The Cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. / Współczesna Onkologia.* **19**, A68–77 (2015).
- Network, T. C. G. A. Comprehensive molecular portraits of human breast tumors. *Nature* **490**, 61–70 (2012).
- Chang, J. T., Lee, Y. M. & Huang, R. S. The impact of the Cancer Genome Atlas on lung cancer. *Transl. Res.* **166**, 568–585. <https://doi.org/10.1016/j.trsl.2015.08.001> (2015).
- Azuaje, F. Artificial intelligence for precision oncology: beyond patient stratification. *Npj Precis. Oncol.* **3** (2019).
- Luchini, C., Pea, A. & Scarpa, A. Artificial intelligence in oncology: current applications and future perspectives. *Br. J. Cancer.* **126**, 4–9 (2021).

21. Wang, Y. et al. Deep learning based time-to-event analysis with PET, CT and joint PET/CT for head and neck cancer prognosis. *Comput. Methods Programs Biomed.* **222**, 106948. <https://doi.org/10.1016/j.cmpb.2022.106948> (2022).
22. Zou, Y. et al. Extreme gradient boosting model to assess risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: individual prediction using SHapley additive explanations. *Comput. Methods Programs Biomed.* **225**, 107038. <https://doi.org/10.1016/j.cmpb.2022.107038> (2022).
23. Rueda, O. M. et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* **567**, 399–404. <https://doi.org/10.1038/s41586-019-1007-8> (2019).
24. Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell.* **40**, 1095–1110. <https://doi.org/10.1016/j.ccell.2022.09.012> (2022).
25. Lucieri, A. et al. ExAID: A multimodal explanation framework for computer-aided diagnosis of skin lesions. *Comput. Methods Programs Biomed.* **215**, 106620. <https://doi.org/10.1016/j.cmpb.2022.106620> (2022).
26. Tashakori, M. et al. TP53 copy number and protein expression inform mutation status across risk categories in acute myeloid leukemia. *Blood* **140**, 58–72. <https://doi.org/10.1182/blood.2021013983> (2022).
27. Liang, L. et al. Plasma CfDNA methylation markers for the detection and prognosis of ovarian cancer. *EBioMedicine* **83**, 104222. <https://doi.org/10.1016/j.ebiom.2022.104222> (2022).
28. Xiao, X., Wang, Z., Kong, Y. & Lu, H. Deep learning-based morphological feature analysis and the prognostic association study in colon adenocarcinoma histopathological images. *Front. Oncol.* **13**, 1081529. <https://doi.org/10.3389/fonc.2023.1081529> (2023).
29. Li, S. et al. In *International Conference on Learning Representations*.
30. Xiao, S. et al. Learning time series associated event sequences with recurrent point process networks. *IEEE Trans. Neural Netw. Learn. Syst.* 1–13 (2019).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. <https://doi.org/10.1093/bioinformatics/btp616> (2010).
32. Pei, S., Liu, T., Ren, X., Li, W. & Xie, Z. Benchmarking variant callers in next-generation and third-generation sequencing analysis. *Brief. Bioinform.* (2020).
33. Du, P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* **11**, 587. <https://doi.org/10.1186/1471-2105-11-587> (2010).
34. Goode, A., Gilbert, B., Harkes, J., Jukic, D. & Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Inform.* **4** (2013).
35. Lu, M. Y., Chen, R. J., Wang, J., Dillon, D. & Mahmood, F. Semi-Supervised Histology Classification using Deep Multiple Instance Learning and Contrastive Predictive Coding. (2019).
36. Hamilton, W. L., Ying, R. & Leskovec, J. Inductive Representation Learning on Large Graphs. (2017).
37. Travers, C., Xun, Z., Garmire, L. X. & Florian, M. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comp. Biol.* **14**, e1006076 (2018).
38. Zadeh, A., Chen, M., Poria, S., Cambria, E. & Morency, L. P. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. (2017).
39. Harrell, F. E., Lee, K. L. & Mark, D. B. *Prognostic/Clinical Prediction Models: Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors* (Tutorials in Biostatistics, 2005).
40. Raykar, V. C., Steck, H., Krishnapuram, B., Dehing-Oberije, C. & Lambin, P. In *Conference on Neural Information Processing Systems*.

Author contributions

Conceptualization, T.Z., H. L. and Y.K.; methodology, H. L. and Y.K.; formal analysis, T.Z. and Y.K.; investigation, T.Z. and Y.K.; data curation, T.Z., Y.R. and Y.K.; writing—T.Z., Y.R. and Y.K.; writing—review and editing, H. L. and Y.K.; All authors read and approved the final manuscript.

Funding

This work is supported by National Key R&D Program of China (Grant No. 2022YFC2703103), the Science and Technology Commission of Shanghai Municipality (Grant No. 23JS1400700), the Science and Technology Innovation Key R&D Program of Chongqing (Grant No. CSTB2024TIAD-STX0006), SJTU-Yale Collaborative Research Seed Fund and the Neil Shen's SJTU Medical Research Fund.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-09869-0>.

Correspondence and requests for materials should be addressed to Y.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025