IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 August 2024, accepted 28 August 2024, date of publication 3 September 2024, date of current version 11 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3453930

RESEARCH ARTICLE

Survival Prediction for Non-Small Cell Lung Cancer Based on Multimodal Fusion and Deep Learning

XIAOPU MA[1], FEI NING[2], XIAOFENG XU[1], JIANGDAN SHAN[1], HE LI[1], XIAO TIAN[3], AND SHUAI LI[1]

[1]Henan Engineering Research Center of Intelligent Processing for Big Data of Digital Image, School of Computer Science and Technology, Nanyang Normal University, Nanyang 473061, China
[2]School of Life Sciences and Agricultural Engineering, Nanyang Normal University, Wolong, Nanyang 473061, China
[3]Nanyang Medical College, Wolong, Nanyang 473061, China

Corresponding author: Xiaofeng Xu (nanyang_xuxiaofeng@126.com)

This work was supported in part by Wolong Scholar Award Program Project in Nanyang Normal University and Nanyang Normal University Doctoral Research Fund under Grant 218299, in part by the National Natural Science Foundation of China Youth Fund under Grant 62002180, and in part by the 2022 Nanyang City Science and Technology Research Program Project under Grant KJGG105.

ABSTRACT Non-small cell Lung Cancer (NSCLC) is one of the most common types of lung cancer, accounting for approximately 80% to 85% of lung cancer cases, and survival prediction in this cancer is an essential medical task. Traditional survival prediction methods, which only rely on demographic and clinical variables, fail to fully characterize patients' pathological characteristics and clinical factors, thus limiting the prediction effect. With the development of medical imaging technology and genomic methods, survival prediction studies have ushered in new analytical perspectives in recent years. However, most of the existing advanced techniques only rely on one class and few classes of medical data, which do not comprehensively characterize patients' pathological features and clinical conditions, and similarly limit the predictive effect. To this end, in this paper a new method is proposed to solve this problem, that is, using flexible interpretable graph structure to fuse and model the multimodal data (including clinical data and CT images, etc.) of patients, so as to solve the problem of fragmentation and one-sidedness among multi-class data of survival prediction. A new multi-modal fusion graph convolutional network (FGCN) is proposed according to the characteristics of multi-modal graph data. The main body of the network structure is composed of a SAGE graph convolution layer with the self-attention mechanism, which accelerates the convergence of the model compared to the traditional one and ensures that the risk predicted by the model is as close to the actual situation as possible. More importantly, this study is the first to introduce the TopKPooling strategy in this model to address the problems of feature redundancy and excessive noise features arising from multimodal data fusion and to reduce the model complexity. A large number of ablation and comparison experiments and analyses on public non-small cell lung cancer datasets show that the method proposed in this paper achieves better prediction results, with a C-index value of 0.76, which exceeds the currently known advanced techniques, and fully proves the effectiveness and superiority of the method in this paper.

INDEX TERMS Survival prediction, multimodal fusion, CT, deep learning, graph convolutional networks.

I. INTRODUCTION

Non-Small Cell Lung Cancer (NSCLC) is one of the most common types of lung cancer, accounting for approximately 80% to 85% [1] of lung cancer cases. The widespread existence of this type of cancer poses a serious challenge to the medical community. Survival prediction helps doctors and patients to develop individualized treatment plans jointly. By knowing what to expect from patients, doctors can better weigh the pros and cons of different treatment options

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

to obtain the best treatment effect. For the treatment and management of NSCLC patients, good survival prediction is crucial. This predictive information can not only help patients better cope with the disease, but also help to allocate medical resources more effectively. However, survival prediction in NSCLC is not an easy task, and many complex factors, including clinical and pathologic differences between patients, the disease progression rate, and the diversity of therapeutic regimens, challenge this task. Inaccurate survival predictions may not only lead to inappropriate clinical decisions, but also reduce clinicians' confidence in communicating with patients [2]. Despite the many difficulties, patients' survival rates and quality of life are expected to improve further with medical research advancement and new treatment strategies.

Over the past decades, clinicians have relied on clinical covariates, biomarkers, and clinical experience for cancer survival prediction [3]. TNM [4] staging and gene expression profiling were widely used in almost all solid cancers to predict prognosis and assist in clinical treatment decisions. For example, researchers Cai et al. [5] constructed a model containing eight messenger RNA prognostic signatures for identifying potential molecular biomarkers in esophageal cancer patients suffering from lymph node metastasis (LNM) [6], [7]. The model proved to be a good indicator of prognosis and contained biomarkers such as VEGF [8], cyclin D1 [9], Ki-67 [10], and squamous cell carcinoma antigen [11] that play an essential prognostic role in patients with esophageal squamous cell carcinoma who underwent surgical resection, which was derived from pathological analysis of tumor samples. However, there is still a large group of physically unresectable, inoperable, or refusing to operate patients in the actual clinical setting, who are not amenable to biomarker testing, and thus, there is an urgent need for a non-contact method of predicting survival risk through medical imaging.

With the continuous advancement of computer vision [12] and medical imaging technologies [13] in recent years, clinicians are gradually recognizing the importance of using AI techniques as decision support tools for cancer survival prediction. Hundreds of prediction models for different cancer types have emerged as a result [14], [15]. These models are based on a variety of data sources, including gene expression data [16], clinical data [17], biomarkers from pathology samples, and radiomics features extracted from medical imaging [18], [19], [20]. Han et al. [21] designed a deep learning model combining deep genomics and radiomic features that could group patients with high-grade gliomas into long-term and short-term survivors. Howard et al. [22] evaluated whether a machine-learning model could identify patients with intermediate-risk head and neck squamous cell carcinomas and assess the effectiveness of chemotherapy for them. Zhong et al. [23] used multi-task deep learning radiomics and preoperative MR images to construct a diagnostic criterion for identifying the prognosis of patients under different treatment regimens, based on which the optimal treatment regimen was recommended. Deep features obtained from chest CT scans using a convolutional neural network were recently used to classify COVID-19 pneumonia infections with 95.1% accuracy [24]. In addition, Xie et al. [25] used subregional and whole-region radiomics features of planning CT to predict the overall survival of patients with esophageal squamous cell carcinoma who received concurrent radiotherapy. In 2022, Akbar et al. proposed a deep learning prediction model for anticancer peptides, cACP-DeepGram [26], which achieved a maximum accuracy of 96.94%. In 2023, Akbar et al. proposed machine learning models pAtbP-EnC [27] and AIPs-SnTCN [28] for anti-tuberculosis peptide prediction and anti-inflammatory peptide prediction, respectively, and the prediction accuracies reached 97.80% and 92.04%, which were superior to the existing methods and predictors at that time. In 2024, Akbar et al. proposed another new computational model called Deepstacked-AVPs [29] for accurate differentiation of antiviral peptides, which achieved a prediction accuracy of 96.60%, an area under the curve (AUC) of 0.98, and a precision-recall (PR) value of 0.97 when using the training samples.

As mentioned above, current studies have demonstrated the feasibility of survival prediction modeling and prognosis of numerous cancer patients from preprocessed measurement features of medical imaging [21], [22], [23], [24], [25], [26], [27], [28], [29], and can predict survival in patients who are difficult to perform invasive modalities, and contribute to helping physicians make clinical treatment decisions, set appropriate treatment goals, and alleviate the anxiety of patients and their relatives. However, the same problem faced is that survival prediction analysis is a multifactorial-based task, and relying on only one category and a few categories in medical data does not fully characterize patients' pathological features and clinical conditions, thus limiting the prediction effect. From previous studies and literature, there are few studies that utilize multimodal fusion of clinical data and CT images for survival prediction, therefore, it is particularly important to explore the relationship within and between the modalities of the two types of medical data in order to provide more detailed evidence-based support for survival prediction.

In this article, this study will use a combination of quantitative and qualitative methods to fully explore the hidden information in multimodal medical data and propose a new survival prediction model based on the COX proportional risk model [30], as well as conduct in-depth experiments and analyses on real data to verify the validity and stability of the proposed model.

The main contribution of this work can be summarized as follows.

1. Aiming at the problem that traditional survival prediction methods are split between modal survival data when analyzing, this paper innovatively adopts the multimodal feature fusion technique, which utilizes the graph structure to model the multimodal

fusion of patients' clinical data and cancer CT images. This method can retain the importance of each data feature factor and the internal correlation between data features, and successfully solves the problem that traditional data cannot comprehensively describe the patient's pathological characteristics and clinical factors.

2. Aiming at the characteristics of multimodal graph data, this study improve the existing GraphSAGE graph convolutional network and name it FGCN, aiming at more effective inter-modal and intra-modal interactions. The main structure of the model consists of a graph convolutional layer with a self-attention mechanism and a well-designed TopKPooling module, as well as a gradient-augmented loss function, which enables the structure to be further adapted to the multifactorial problem of survival prediction. A large number of ablation and comparison experiments and analyses on public non-small cell lung cancer datasets show that the method proposed in this paper achieves better prediction results, with a C-index value of 0.76.
3. Aiming at the issue of unfavorable influence of noise features on the prediction results in both survival data and multimodal graph data, this paper innovatively introduces the TopKPooling module to the graph convolution model. This strategy is firstly applied to the survival prediction study of multimodal fusion. The TopKPooling strategy is actually a pruning operation on the graph structure data, selecting the low-scoring ones to be eliminated, and then recomposing a new graph structure. By introducing the TopKPooling strategy, this study is able to better handle the fused multimodal data feature maps, fully explore the key features, reduce the computational complexity of the prediction network, prevent the overfitting of the prediction model, and then improve the expressive ability and prediction performance of the model.
4. Aiming at the problem that COX proportional risk regression suffers from a serious drop in the iterative gradient during the training process, this study designed a gradient-enhanced loss function. This loss function is an optimization and improvement of the original COX proportional risk regression loss function by updating the gradient and introducing a new weight factor α to control the degree of influence of gradient changes on the loss function. This method allows the model to converge to the optimal solution faster, making the predicted risk of the model closer to the actual observed risk.

II. RELATED WORKS

A. SURVIVAL PREDICTION

Over the past decades, clinicians have typically relied on clinical covariates and clinical experience for cancer survival prediction. Pu et al. [31] conducted a population-based retrospective study using data from the Surveillance Epidemiology and End Results (SEER) database to establish a risk and prognostic column chart for YBCLM to help clinicians accurately predict the occurrence and survival of LM in YBC. However, with the continuous advancement of computer vision and medical imaging technologies, clinicians have come to realize the importance of using artificial intelligence techniques as decision support tools for cancer survival prediction. As a result, significant progress has been made in survival analysis and survival prediction in recent years, with the emergence of hundreds of advanced predictive models for different cancer types. These models are based on a variety of modal data. To further improve the performance of AI-based survival prediction, several studies have begun to experiment with combining clinical records and genomic profiling data with medical imaging data. Syed and Ma [32]. Their brain atrophy diagnostic model uses novel deep learning and multivariate mathematical models to detect, segment and classify atrophic regions and predict lesions. In 2022, Lin et al. [33] from Xiamen University used deep learning to model the interactions between pathological features and clinical information of 285 esophageal cancer patients and used it for the survival prediction task. Their experimental results showed that the method achieved a C-index index of 0.72. Hou et al. [34] proposed a new hybrid graph convolutional network, HGCN, equipped with an online mask autoencoder for robust multimodal cancer survival prediction.

B. GRAPH CONVOLUTIONAL NETWORK

Graph Convolutional Network (GCN) is a deep learning model based on graph-structured data, which has been widely used in recent research. The main feature of GCN is the ability to propagate and aggregate node features over the graph, thus enabling effective learning and representation of graph data. The network uses the node's neighbor information and node features to update the representation of each node, thus encoding the information of the graph into the node's vector representation. Through multiple layers of graph convolutional layers, the GCN is able to gradually fuse information between different nodes to obtain richer and more complex graph representations.

Graph Convolutional Networks have significant advantages in the prediction task, GCN can effectively fuse the structural information of graph data and encode the global information of the graph into each node's representation through the connectivity relationship between nodes and feature propagation of neighboring nodes, for incomplete nodes in the graph data, GCN is also able to perform effective inference to fill in the missing values and make predictions. This makes GCN show good robustness in dealing with missing data, since GCN retains the structural information of the graph during node feature propagation, its prediction results have good interpretability and can reveal the correlation and influencing factors between nodes.

The advantages of Graph Convolutional Networks in prediction tasks make them a powerful tool for processing

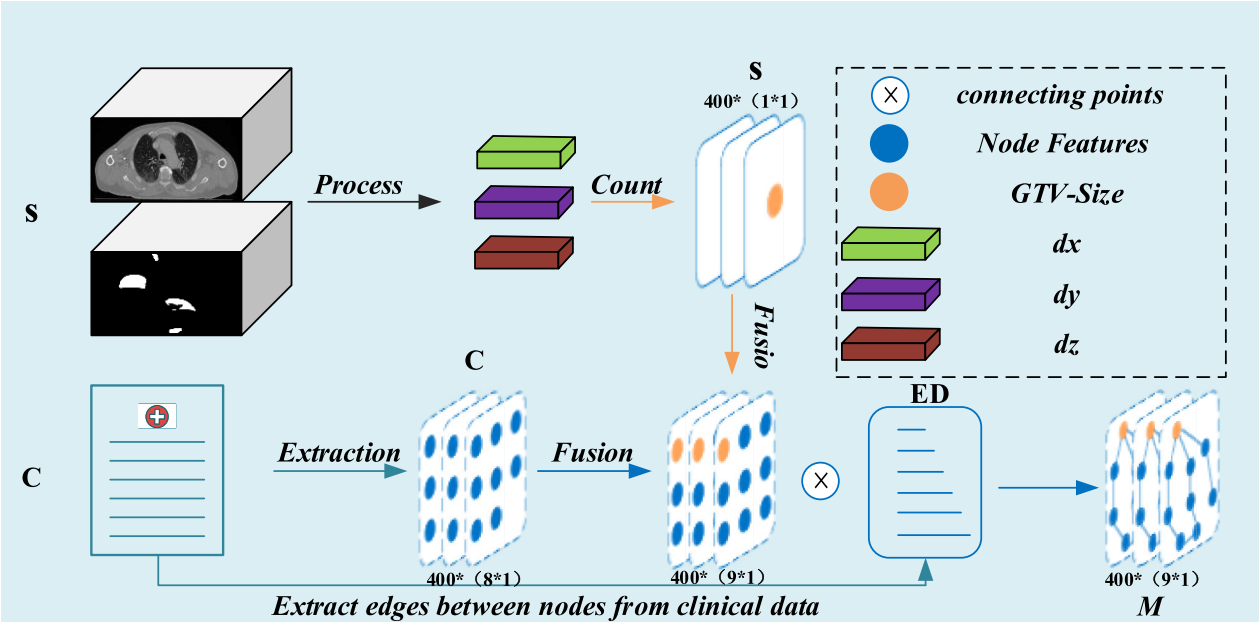

FIGURE 1. The Schematic of The Construction of Multimodal Information Graph. 1. for the CT image data s, this study extract its pixel spacing (dx, dy) and slice thickness (dz) from it, and combine it with the segmentation mask map to calculate the tumor volume GTV. 2. for the clinical feature data data c, this study normalize its Z-score as a way to avoid the influence of the scale difference between the features on the modeling, and at the same time, this study construct from the clinical data the feature-to-feature Edge relations ED={EB,EE} are constructed from the clinical data.3. S and C are used as the nodes of the graph structure data, and ED is used as the edge relations between the nodes to construct a complete multimodal information graph.

graph-structured data, which plays an important role especially in the fields of social network analysis, bioinformatics, drug prediction, and chemical molecule prediction. For example, Kipf and Welling [35] first proposed the concept of GCN and applied it to a semi-supervised chemical molecule classification task. Wu et al. [36] constructed a graph-convolutional testing framework called MoleculeNet for evaluating the performance of molecular machine learning algorithms in several chemical tasks, including drug prediction, molecular property prediction, etc.

C. MULTIMODAL FUSION

Multimodal fusion refers to the integration and fusion of information from different data sources or different modalities to obtain more comprehensive and accurate information. In research papers, multimodal fusion has become a research area of wide interest, and it has demonstrated excellent superiority in various fields. The superiority of the multimodal fusion strategy lies in its ability to make full use of the complementarity and correlation between different modal data, thus improving the characterization ability and information richness of the data. In the field of computer vision, multimodal fusion can combine multiple data sources such as images, text and speech for tasks such as image description generation and cross-modal retrieval. Liu et al. [37] proposed a method called multimodal mutual attention and multimodal mutual decoder to solve the reference image segmentation problem. In the field of medical imaging, multimodal fusion can fuse different medical image data (e.g., CT, MRI, PET, etc.) for disease diagnosis and prediction. Li et al. [38] used attention-guided deep supervisory nets for adaptive multimodal fusion to grade hepatocellular carcinoma. By taking full advantage of the correlation between different modal data, multimodal fusion can improve the efficiency of data processing and analysis, and bring more breakthroughs and advances in research and applications in various fields.

III. METHODOLOGY

In this paper, this study adopt an innovative approach that uses flexible interpretable graph structure to fuse and model the multimodal data (including clinical data and CT images, etc.) of patients, so as to solve the problem of fragmentation and one-sidedness among multi-class data of survival prediction. This study propose a novel Multi-modal fusion graph convolutional network approach (FGCN) for the characteristics of multimodal graph data, aiming at more effective inter- and intra-modal interactions. The network framework consists of three key components.

Firstly, this study design a method to fuse multi-modal data into graph-structured data. This method solves the problem of separating clinical data from image data in traditional data models, and can more flexibly capture the complex nonlinear relationship and interaction between data.

Secondly, this study also design a SAGA module with a self-attention mechanism [39], which is dominated by a SAGE graph convolutional layer [40] and innovatively introduces a TopKPooling [41], [42], [43] strategy, which successfully solves the problems of data redundancy and

excessive noise nodes that may be brought about by multimodal data fusion and retains the critical feature information while reduces the computational complexity of the model, improves the generalization ability of the model, and makes the model more lightweight.

Finally, this study design a gradient-enhanced loss function for the survival prediction network. The new loss function makes the predicted risk of the model as close as possible to the actual observed risk, thus better explaining and analyzing the survival probability of the sample.

A. THE CONSTRUCTION OF MULTIMODAL INFORMATION GRAPH

To address the research questions and objectives, this study adopted a feature-level extraction fusion approach when performing multimodal feature fusion [44]. This study believe that the multimodal feature data of a batch of patients should include at least two modalities M = {C, S}, i.e., the clinical feature data C and the CT pathology slice S, to characterize the patient's physical condition from the clinical level to the pathology level. To this end, this study designed a multimodal map network module to fuse the features contained in the clinical data and CT pathology maps into a multimodal feature map and use it as an input to the subsequent network model as shown in **Figure 1** in the following way.

For CT image data in dicom format, this study took the following steps to obtain the patient's tumor volume GTV [45]. Firstly, all dicom slices of each patient where loaded to make a complete CT pathology image, and their pixel spacing (dx, dy) and slice thickness (dz) were obtained. Then, the segmentation mask map corresponding to the patient, manually labeled and annotated by the radiation oncologist for the volume of the tumor nodule, is loaded and converted to a binary mask map (0 for the background and 1 for the target). Finally, the patient's tumor volume GTV was obtained by calculating the number of all pixels in the mask map with a pixel value of 1, multiplied by the pixel pitch and slice thickness. The formula is as follows:

$$GTV = All_{Pixel} \cdot dx \cdot dy \cdot dz \tag{1}$$

The resulting GTV for each patient can be represented as a 1*1 graph structure, and the set S of GTV graphs for 400 patients is 400*(1*1).

As for the clinical characteristics data (age, gender, clinical stages, medical history, survival time and status, etc.), this study quantitatively represented the characteristics other than the survival status with the Z-score standardization method. The formula is as follows:

$$\mathbf{Z} = (\mathbf{X} - \mu)/\sigma \tag{2}$$

This allows the scale of each feature to be scaled between 0 and 1, avoiding the impact of scale differences between features on modeling. This leads to the conclusion that the clinical characteristics of each patient can be represented as an 8*1 graph structure, while the graph set C of the clinical characteristics of 400 patients is 400*(8*1). At the same time, this study construct the edge index ED of the graph by the relationship between each feature in the clinical data. To establish the connection of the graph, this study use the indexes of the source node(EB) and the target node(EE). The edge of the graph institutional data can be represented as ED={EB, EE}. After the above preprocessing of the two parts of the data, new clinical data C and image feature data S are obtained respectively, when C={c1, c2, c3.....ci}, S={s1, s2, s3.....si} (i denotes the patient number). Next, put x={ci,si}, which can be abstractly understood as a concatenation operation of ci and si. At this point, a graph structure can be expressed as m={x, ED}, and the set of graphs of all patients M contains all patients m. This kind of graph structure contains both the CT image features of the patients as well as the clinical information features, which is used as the input of the subsequent model to realize the effect of using the two kinds of data at the same time.

B. THE MULTI-MODAL FUSION GRAPH CONVOLUTIONAL NETWORK

1) FGCN

In this study, this study aim to utilize multimodal information fused into graph-structured data to better capture the intrinsic connections between different feature nodes and obtain better prediction results. To achieve this goal, this study propose a framework called Multi-modal fusion graph convolutional network (FGCN), as shown in **Figure 2**, which consists of the following modules and methods.

First, this study map the node matrix x(9, 1) of the constructed multimodal graph structure data to an embedding layer matrix with a hidden dimension of 128 to obtain x0 (9, 128). Next, this study perform feature extraction and transfer of graph structure information through a combined SAGA module. In this study, due to the low complexity of the feature nodes and the real-time nature of clinical survival prediction, this study choose the SAGE graph convolution as the convolutional layer of SAGA. The SAGE graph convolution uses a learnable aggregation function to aggregate the features of the neighboring nodes and update the features of the target node through this function.

After the SAGE convolutional layer, this study adopt a TopKPooling strategy for the feature vector matrix x0 to filter out the noisy node features, thus obtaining the updated x0 (7,128). Then, this study introduce the self-attention mechanism as an additional information fusion layer and feedback layer, and finally, since this study need to extract the global features of the graph data, this study perform a global average pooling (GAP) operation on x0. to obtain a layer of convolutional global features x1. This study repeat the above convolutional and filtering operations twice to obtain x2 and x3, which represent global features on different convolutional scales, respectively. Specifically, it can be represented as:

$$\mathbf{x1} = \mathbf{gap}\,(\tau\,(\mathbf{E}\,(\mathbf{x0}))) \tag{3}$$

$$\mathbf{x2} = \mathbf{gap}\,(\tau\,(\tau\,(\mathbf{E}(\mathbf{x0})))) \tag{4}$$

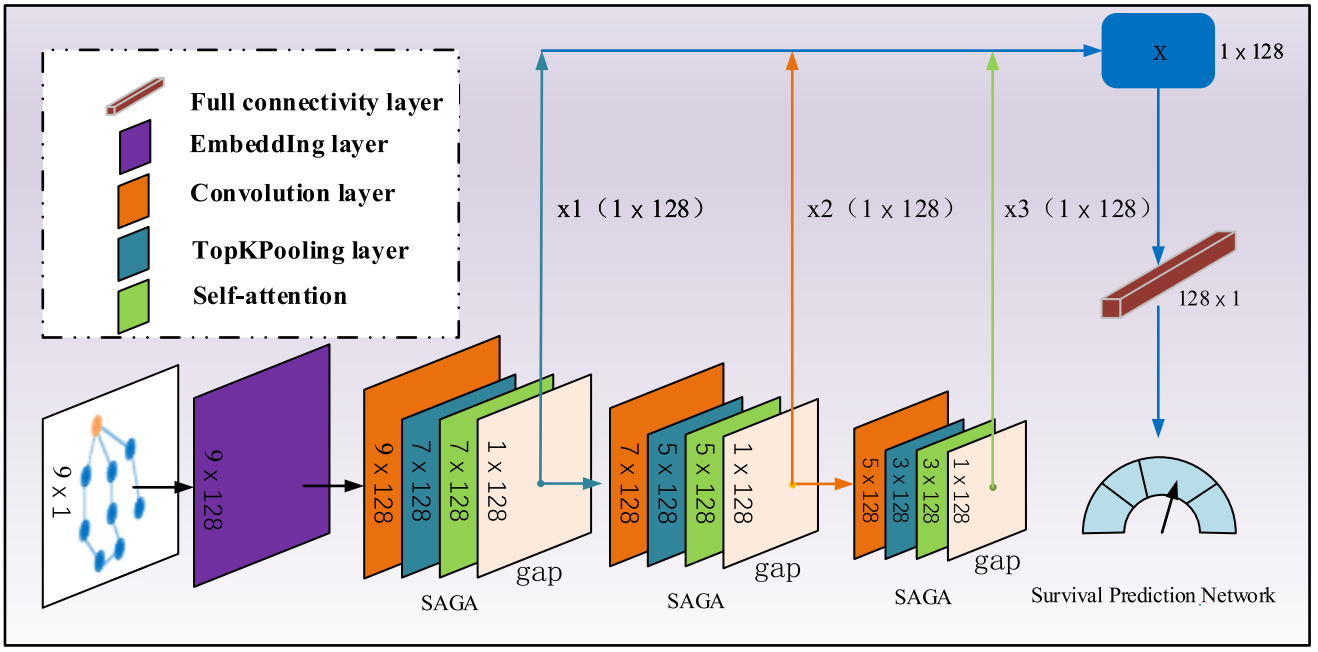

FIGURE 2. The Schematic of FGCN. 1. The node matrix x(9, 1) is mapped to an embedded row matrix E with hidden dimension 128 to obtain x0 (9, 128), where 9 is the number of feature nodes and 128 is a hyperparameter. 2. x0 is processed through the SAGA module, firstly, the features are extracted and passed through the SAGE convolutional layer; then the TopKPooling layer is utilized to interact with it and filter the features; then the self-attention mechanism is added as an additional information fusion layer and feedback layer; finally, a global average pooling layer is connected to extract the corresponding global features, and the output x1 is outputted and passed. 3. After three layers of SAGA module, three different scales of global features x1, x2, x3, respectively, and sum up the three global features to get the final global feature x. Finally, x is mapped to the label y through the fully connected layer to get the risk matrix β, and input into the Cox survival prediction module for the prediction of survival risk.

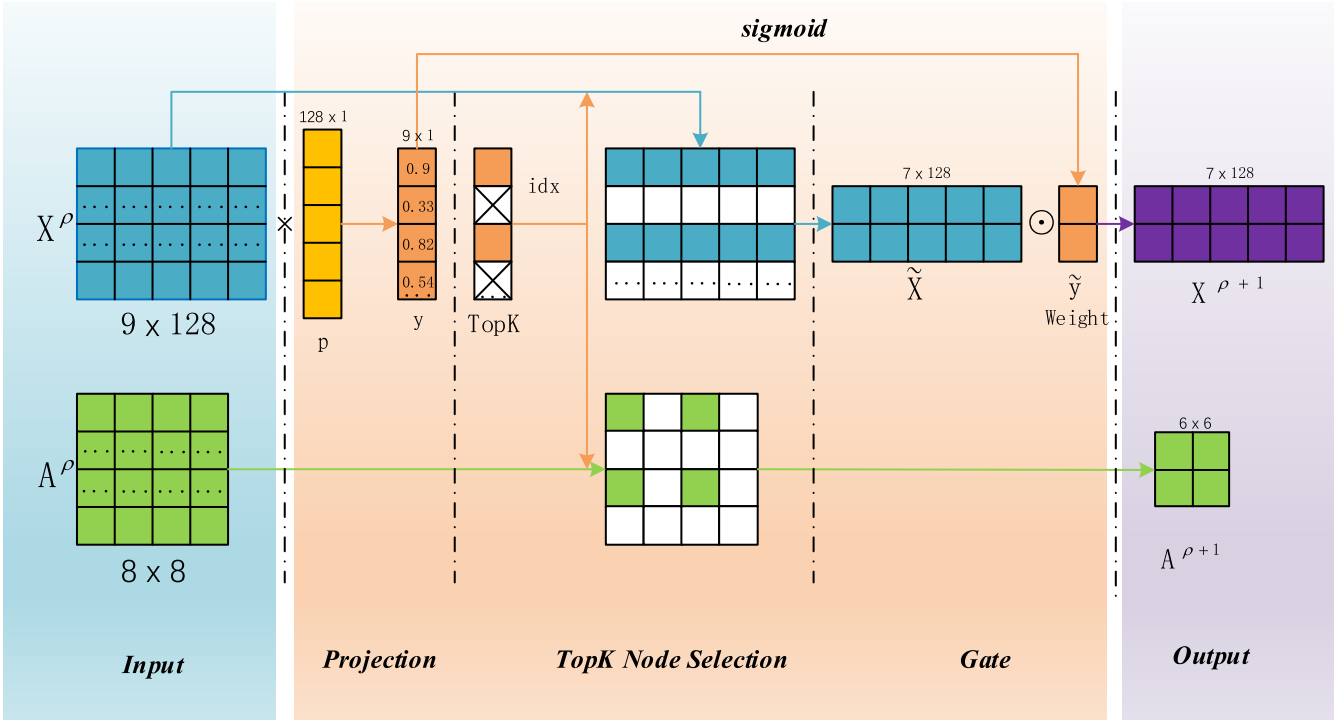

FIGURE 3. The Schematic of TopKPooling strategy. Input: when x0 is convolved by SAGE, a new feature mapping matrix X^{ρ} (9,128) and its adjacency matrix A^{ρ} (8,8) are obtained. Filtering: this study multiply X^{ρ} with a set of trainable weight parameters p(128,1) to obtain a corresponding score matrix y(9,1). Based on the score of y, this study choose to retain a number of K rows with the largest scores, in this experiment K=7. Next, this study map to X^{ρ} based on the retained 7 rows, which in turn yields the desired retained feature matrix $\tilde{X}$ (7,128). OUTPUT: Finally the filtered $\tilde{X}$ is weighted to get the final feature matrix $X^{\rho+1}$ (7,128). At the same time its adjacency matrix A^{ρ} is also updated in the same way to get the new adjacency matrix $A^{\rho+1}$.

$$\mathbf{x3} = \mathbf{gap}\left(\boldsymbol{\tau}\left(\boldsymbol{\tau}\left(\boldsymbol{\tau}\left(\mathbf{E}(\mathbf{x0})\right)\right)\right)\right) \tag{5}$$

$$\mathbf{x} = \mathbf{x1} + \mathbf{x2} + \mathbf{x3} \tag{6}$$

where E denotes the mapping operation on the input, and τ denotes the composite of the convolution and subsequent operations. In summary, this study obtain three global features x1, x2 and x3, and then sum them up to get the final global feature x. The global features learned by the model are mapped to the label y through the full connectivity layer to get the risk matrix β. After that, the result is inputted into the Cox survival prediction module for the prediction of survival risk.

2) TopKPooling STRATEGY

As shown in **Figure 3**, when dealing with the fused multimodal data feature maps, this study introduce an innovative TopKPooling strategy, which successfully solves the problems of feature redundancy and excessive noise features that may occur after multimodal data fusion. By adopting the TopKPooling strategy, this study are able to effectively reduce the dimensionality of the data features, which significantly reduces the discrete nature of the data features and further reduces the complexity of the model. The specific method is as follows.

In this study, this strategy is introduced for the first time in the study of survival prediction with multimodal fusion. The TopKPooling strategy is actually a pruning operation on the graph structure, selecting the low-scoring ones to be eliminated, and then recomposing them into a new graph structure. It retains the K nodes that are most similar to the maximum score point based on the scoring information of each node, while discarding other features that are not related to it. This approach allows us to focus on key features, avoiding the processing of excessive redundant features and noise, thus improving the processing efficiency and accuracy of the feature map. Compared to the method of pooling in CNN using null convolution, TopKPooling is more suitable for this kind of graph-structured data with the same density and graph convolutional networks, and the amount of computation is much smaller.

By introducing the TopKPooling strategy, this study is able to process the fused multimodal data feature maps better, fully explore the key features, reduce the computational complexity of the prediction network, prevent the overfitting of the prediction model, and then improve the expressive ability and prediction performance of the model.

Compared to some of the existing state-of-the-art pooling methods, such as the hierarchical pooling method Hierarchical Pooling [46], which is applicable to the aggregation of information at multiple levels and is not applicable to the aggregation of features among the information of a single patient. In addition, the most used state-of-the-art pooling method is Graph U-Net Pooling [47], which is based on graph structure, but it is suitable for segmentation of images and not for regression and survival prediction tasks. Since this study needs to explore the internal interrelationships and local relationships among the features of each patient, TopKPooling is more suitable for survival prediction tasks with multimodal fusion. In the subsequent experimental results and discussion section of this paper, the data in **Table 3** visualize the comparison of this strategy with other existing state-of-the-art strategies, further proving the positive effects and advantages of this strategy for the model.

3) GRADIENT-ENHANCED LOSS FUNCTION

In this study, this study used the Cox proportional risk regression model for the application of the survival prediction network module. First, this study extracted global features from the sample data through the FGCN module, and then added a fully connected layer to connect the global features and the risk coefficients. A single sample i has global characteristics x_i and event occurrence time ti, and a coefficient β exists. The model equation for Cox proportional risk regression can be expressed as:

$$\mathbf{h_i}\left(\boldsymbol{t}\right) = \mathbf{h_0}\left(\mathbf{t}\right) \cdot \mathbf{exp}\left(\boldsymbol{\beta} \cdot \boldsymbol{x_i}\right) \tag{7}$$

where $h_0\left(t\right)$ is the baseline risk function, which represents the risk when the covariate (global feature) x_i is 0, which is the same for all individuals, and the difference in risk between individuals is only caused by the difference in the global feature x_i. $h_i\left(t\right)$ is the risk function of sample i at time t, which represents the product of the probability of an event occurring at time t and the baseline risk function $h_0\left(t\right)$). The coefficient β is a parameter to be learned to measure the effect of the global feature x_i on the risk function. $exp(\beta \cdot x_i)$ denotes the variation of the effect of the global feature x_i on the risk function on an exponential scale.

Next, this study can use negative log-biased likelihood estimation as a loss function to minimize the model's error. Suppose this study have d event times t and d corresponding event state variables E, where $E_i = 1$ means the event occurred and $E_i = 0$ means the event did not occur. Then, the loss function of Cox proportional risk regression can be expressed as:

$$\mathcal{L}(\boldsymbol{\beta}) = -\sum\nolimits_{i=1}^{n}\left(\boldsymbol{E_i} \cdot \left(\boldsymbol{\beta} \cdot \boldsymbol{x_i} - \mathbf{log}\left(\sum\nolimits_{\boldsymbol{j} \epsilon \boldsymbol{R}(\boldsymbol{t_i})} \boldsymbol{e}^{\boldsymbol{\beta} \cdot \boldsymbol{x_j}}\right)\right)\right) \tag{8}$$

where $R(t_i)$ denotes the set of samples with events occurring before time, the risk set, and $log\left(\sum_{j \epsilon R(t_i)} e^{\beta \cdot x_j}\right)$ is the logarithm of the sum of the risk functions of samples with events occurring before time t_i.

To make the model's predicted risk as close as possible to the actual observed risk, this study need to optimize the loss function above to make the model converge faster to the optimal solution. To do this, this study use the gradient enhancement method for optimization. First, this study need to compute the gradient of the loss function with respect to the coefficient β. The gradient is obtained by taking the partial derivative with respect to β for the loss function $\mathcal{L}(\beta)$):

$$\eta = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = -\sum\nolimits_{i=1}^{n}\left(\boldsymbol{E_i} \cdot \boldsymbol{x_i} - \frac{\sum_{\boldsymbol{j} \epsilon \boldsymbol{R}(\boldsymbol{t_i})} \boldsymbol{E_j} \cdot \boldsymbol{x_j}}{\sum_{\boldsymbol{j} \epsilon \boldsymbol{R}(\boldsymbol{t_i})} \boldsymbol{e}^{\boldsymbol{\beta} \cdot \boldsymbol{x_j}}}\right) \tag{9}$$

Then, this study introduce the gradient information and add the gradient information as a weighting factor to the loss function, and update the loss function as follows:

$$\mathcal{L}(\boldsymbol{\beta}) = -\sum_{i=1}^{n} (E_i \cdot (\boldsymbol{\beta} \cdot \boldsymbol{x}_i - \log(\sum_{j \epsilon R(t_i)} e^{\boldsymbol{\beta} \cdot \boldsymbol{x}_j}))) + \boldsymbol{\alpha} \cdot \boldsymbol{\eta} \tag{10}$$

Here, $\boldsymbol{\alpha}$ is the weight factor of the gradient information, which is used to control the degree of influence of the gradient on the loss function. The value of $\boldsymbol{\alpha}$ can be adjusted according to the actual situation to balance the weights of the gradient and the loss function. During the training process, the gradient $\boldsymbol{\eta}$ is calculated by backpropagation and then added to the loss function to further adjust the model parameters. Similarly, in the subsequent Experimental Results and Discussion section of the paper, the data in **Table 4** visualize the improvement that the new loss function brings to the model.

TABLE 1. Distribution of features in the dataset.

Demographics	Values
Count	400
Survival Time（day）	
Range	10 -4328
Median	548
Gender	
Female	126
Male	274
Age（year）	
Range	33-91
Median	68
Overall Stage	
I	84
II	36
IIIa	110
IIIb	170
Status	
1	353
0	47

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET

This study performed a validation evaluation of the proposed method on Lung1 [48], a public dataset of non-small cell lung cancer in the TCIA [49] database. The dataset brings together available information on 400 patients with non-small cell lung cancer (NSCLC) from the Lung1 study published in *Nature Communications*. **Table 1** lists the details of the dataset. In addition, each patient's computed tomography scans included DICOM radiographic structure set (RTSTRUCT) and DICOM segmentation files (SEG), where the DICOM segmentation files were manual depictions of three-dimensional volumes of the primary tumor volume (''GTV-1'') and selected anatomical structures (lungs, heart, and esophagus) by radiation oncologists, and experienced radiologists were Review. 1 in the STATUS column of **Table 1** indicates that the death occurred, while 0 indicates that the death did not occur. IIIa and IIIb in **Table 1** denote stage III of the TNM staging of cancer and are both moderately advanced stages of cancer, but IIIb lesions are larger and may invade vital organs and blood vessels or metastasize to more distant lymph nodes and are usually more severe.

B. EXPERIMENTAL SETTINGS

Execution Details: The constructed graph-structured dataset is divided into training and test sets in the ratio of 8:2. The loss function of the model is used with ADAM optimizer, the batch size is set to 32 and the learning rate is set to 0.0005. The label y of the regression model is set to the survival state and the ratio K threshold for TopKPooling is set to 0.8 and 0.7. The model uses an ADAM optimizer with a batch size set to 32 and a learning rate set to 0.0005. An NVIDIA RTX 3090Ti graphics processor was used throughout the training process, and this study allowed early termination with the maximum number of training epochs set to 500. During the validation of the model, this study terminated training for optimization when the loss in the validation set could not be further reduced. In order to fully evaluate the performance of the proposed model and the comparison method, this study adopt a 5-fold cross-validation approach. Since there is an imbalance of positive and negative samples in the dataset, the problem is solved by using random oversampling, i.e., increasing the number of samples with a survival status category of 0 by replicating them. As illustrated in **Table 2,** there are additioanal paremeter settings for this model. In this study, gradient descent, early stopping method, and hyper-parameter tuning are utilized to inversely adjust and optimize the parameters.

Experimental Procedure: First, this study standardized eight features of the clinical data of 400 NSCLC patients. These features included patients' age, gender, clinical stages, medical history, and survival time. At the same time, this study calculated GTV (total tumor volume) from all CT sections and segmentation mask images of each patient. Next, this study fused these two kinds of data into a graph-structured dataset according to the multimodal graph structure construction method shown in **Figure 1**.

Next, the graph structure data of each patient is sequentially mapped to the corresponding feature matrix x(9,1) and embedded into the row matrix by hiding the dimension of 128, mapped to x0(9,128), at which point each feature row is equivalent to a row of 128-dimensional vectors. Then, the SAGE graph convolutional layer will update and convolve each row vector of x0. Immediately after the update, x0 goes through TopKPooling pooling to sieve out the noisy nodes, and at this point, x0 is (7,128). After that x0 will be updated by the self-attention layer and global average pooling to output a 128-dimensional row vector, which is the global feature x1. Similarly, two more rounds of this process at different scales yield x2, x3 respectively, while the global feature x is the sum of x1, x2, x3. Finally the output through the fully connected layer is a one-dimensional vector and mapped to the survival state label y to get the risk matrix β. Optimizing the survival function and the Cox predictive model through β and with the gradient augmented loss function makes the predicted risk

TABLE 2. Parameter list of the model.

FGCN	Parameter	Optimal value
SAGA	Learning rate	0.0005
	Activation function	ReLu，Sigmoid
	Dimensionality of the mapping layer	1,128
	Batchsize	32
	Epoch	500
	Graph Convolution Layer	3
	Dropout rate	0.2
	Optimizer	Adam
TopKPooling	Ratio K	0.8,0.7
Gradient-enhanced loss function	α	0.04

of the model as close as possible to the actual observed risk, thus better explaining the sample and analyzing the survival probability of the sample.

C. ASSESSMENT METRICS

First, this study use the consistency index (C-index), as assessment metrics. The C-index will compare the sequential relationship between the actual observed survival time and the predicted survival time, and calculate the degree of consistency between the predicted risk ordering and the actual observed risk ordering. The C-index is calculated as follows:

$$\mathbf{S} = \{(\mathbf{i}, \mathbf{j}) | \mathbf{1} \leq \mathbf{i} < \mathbf{j} \leq \mathbf{n}\} \tag{11}$$

$$\mathbf{M} = \{\mathbf{i} | \mathbf{1} \leq \mathbf{i} \leq \mathbf{n}, \boldsymbol{E_i} = \mathbf{1}\} \tag{12}$$

$$\mathbf{U} = \{(\mathbf{i}, \mathbf{j}) \in \mathbf{S} | \mathbf{i} \in \mathbf{M} \& \mathbf{j} \in \mathbf{M}\} \tag{13}$$

$$\mathbf{C} = \{(\mathbf{i}, \mathbf{j}) \in \mathbf{U} | \boldsymbol{t_i} > \boldsymbol{t_j} \& \boldsymbol{p_i} > \boldsymbol{p_j}\} \tag{14}$$

$$\textbf{C-index} = |\mathbf{C}|/|\mathbf{U}| \tag{15}$$

n is the total number of patient individuals, and S is the set of pairs, denoting the ensemble of all possible pairs of individuals. M is the set of individuals with event state 1. U is the set of useful pairs, denoting the set of pairs that satisfy the condition that the temporal state of both individuals in the pair of individuals is 1. C is the set of predicted consistent pairs, where t_i and t_j denote the actual observed survival times of individuals i and j, and p_i and p_j denote the predicted survival times of individuals i and j. The value of the C-index ranges from 0 to 1, where 0 indicates that the model predicts a completely wrong ordering and 1 indicates that the model predicts a completely correct ordering.

Second, this study assessed absolute prediction using the Brier score (Br), which represents the mean squared error between the observed survival state and the predicted probability of survival. Br is a value between 0 and 1, where 0 is the optimal value.

D. ABLATION EXPERIMENT

In order to evaluate the effectiveness of various pooling strategies, this study used a series of widely used and newer pooling methods and compared them with the TopKPooling strategy this study introduced. Keeping other experimental settings constant, this study compared and analyzed the model results produced by different pooling strategies as shown in **Table 3**. The C-index value of model without any pooling strategy, i.e., Baseline, is only 0.642, and with MaxPooling and MeanPooling strategies, the C-index is 0.625 and 0.656, respectively. And the C-index for Hierarchical Pooling is 0.672. whereas, with introduced TopKPooling strategy, the C-index of the model improves to 0.696. Similarly, the Br value is better than the other strategies. The experimental results show that the TopKPooling strategy is superior in performance and is able to better focus on the learning of key features compared to other pooling strategies, while effectively suppressing the influence of redundant features.

Additionally, for the task of prognosis prediction of non-small cell lung cancer (NSCLC) patients, this study conducted a series of ablation experiments on two critical components of the model, mainly focusing on the application of gradient-enhanced loss function and TopKPooling strategy. In this section, the results of these experiments will be elaborated in detail and analyzed as an example of ablation experiments on the model components in **Table 4**, in order to better reveal the enhancement effect of these methods on the model quality.

TABLE 3. Comparison of several pooling strategies.

Setting	C-index	Br
Baseline	0.642±0.03	0.287±0.005
MaxPooling	0.625±0.04	0.291±0.007
MeanPooling	0.656±0.03	0.240±0.009
Hierarchical Pooling	0.672±0.06	0.236±0.004
TopKPooling(ours)	**0.696±0.04**	**0.217±0.009**

The experimental results show that both of this study proposed methods have a positive impact on the prediction results. **Baseline 1** here refers to the base model, the FGCN that does not use the improved loss function as well as the TopKPooling pooling strategy. The data shows that using only the Gradient-enhanced loss function, the C-index of the model gains some improvement to 0.683. This demonstrates the advantages of the gradient-enhanced loss function in accelerating model convergence and reducing complexity. In addition, in order to profoundly investigate the role of the

TABLE 4. Ablation experiments of gradient-enhanced loss and TopKPooling.

Setting	Gradient-enhanced loss	TopKPooling	C-index	Br
Baseline 1	x	x	0.642±0.03	0.287±0.005
Baseline 2	√	x	0.683±0.05	0.236±0.006
Baseline 3	x	√	0.696±0.04	0.217±0.009
FGCN（ours）	√	√	**0.769±0.02**	**0.185±0.006**

TABLE 5. Comparison of FGCN results with other advanced models.

Type	Method	C-index	Br
Clinic	Lasso-Cox[50]	0.586±0.04	0.259±0.007
	Nomograms[31]	0.740±0.03	0.193±0.005
Machine Learning	DeepSurv[51]	0.658±0.06	0.220±0.004
Deep Learning	ResNet3D[52]	0.682±0.05	0.291±0.006
	GSCNN[53]	0.712±0.02	0.204±0.007
	UOCM[33]	0.724±0.04	0.196±0.007
	HGCN[34]	0.740±0.02	**0.177±0.005**
Our Methods	FGCN	**0.769±0.02**	0.185±0.006

TopKPooling strategy in model quality enhancement, based on the basic model, this study conducted a comparative study between Baseline 1 and Baseline 3 (using only the TopKPooling strategy). The experimental results show that the C-index of Baseline 3 is significantly improved to 0.696, while that of Baseline 1 is only 0.642. This demonstrates the effectiveness of the TopKPooling strategy in mining key features, which further improves the prediction quality of the model. The experimental results also show that with the simultaneous application of the gradient-enhanced loss function and the TopKPooling strategy, this study achieve a significant improvement of 12.7% in C-index relative to Baseline 1. On the other hand, this study also conducted an in-depth analysis of the Brier score, and the results show that both of the proposed methods positively impact the reduction of the Br score. This further validates the critical role of these two methods in model performance improvement.

In conclusion, through the systematic analysis of ablation experiments, the results clearly demonstrate that the introduction of the TopKPooling strategy and gradient-enhanced loss function has significant performance improvement for prognostic prediction of NSCLC patients, which provides a strong support for personalized treatment decision-making in the medical field.

Additionally, In order to evaluate the overall performance of the synthesized model, this study let its comparison with several state-of-the-art survival models and used a 5-fold cross-validation method, as shown in **Table 5**.

To ensure comparability of the comparison, this study applied the same preprocessing steps, such as normalization, to these methods and kept the other model settings unchanged. The data in **Table 5** show that this study proposed model obtains the highest C-index of 0.769 and the best prediction performance under the same conditions. Of particular note, the model improved the C-index by 18.3%, 2.9% and 11.1% compared to the clinical approach and DeepSurv, respectively, suggesting the limitations of traditional clinical approaches and relatively simple machine learning models in coping with the physiologic and lesion complexity of non-small cell lung cancer. In addition, **Table 5** shows that compared to other deep learning methods such as ResNet3D, GSCNN, UOCM, and HGCN, this study proposed FGCN model improves the performance by 8.7%, 5.7%, 4.5%, and 2.9%, respectively, and for the Br score, this study result of 0.185 is second only to the HGCN's 0.177. These encouraging results clearly demonstrate the superior performance of the FGCN model.

E. ANALYSIS

1) KAPLAN-MEIER SURVIVAL CURVES

In this section, this study will show the performance of FGCN in another essential task in the field of survival prediction by categorizing NSCLC patients into two subgroups based on this study predicted outcomes (risk coefficients). Specifically, this study use the median of the predicted outcomes as a threshold to categorize all patients into low-risk and high-risk groups. As shown in **Figure 4**, this study plotted the Kaplan-Meier survival curves [54], from which it can be seen that patients with longer survival time are classified as a low-risk group. In comparison, patients with shorter survival time are classified as a high-risk group. Suppose the two curves are close or even cross. In that case, it indicates that the model is unable to effectively differentiate between low-risk and high-risk patients, and vice versa, it indicates that the model's prediction results are better, and this method can vividly demonstrate the model's prediction ability.

For a comprehensive statistical analysis, this study assessed all data consistently. This study used the log-rank test to measure the difference between the two curves. P-values less than 0.05 indicate statistical significance, and p-values less than 0.01 indicate strong statistical significance. As shown in **Figure 4**, with the proposed TopKPooling strategy and gradient-enhanced loss function, this study

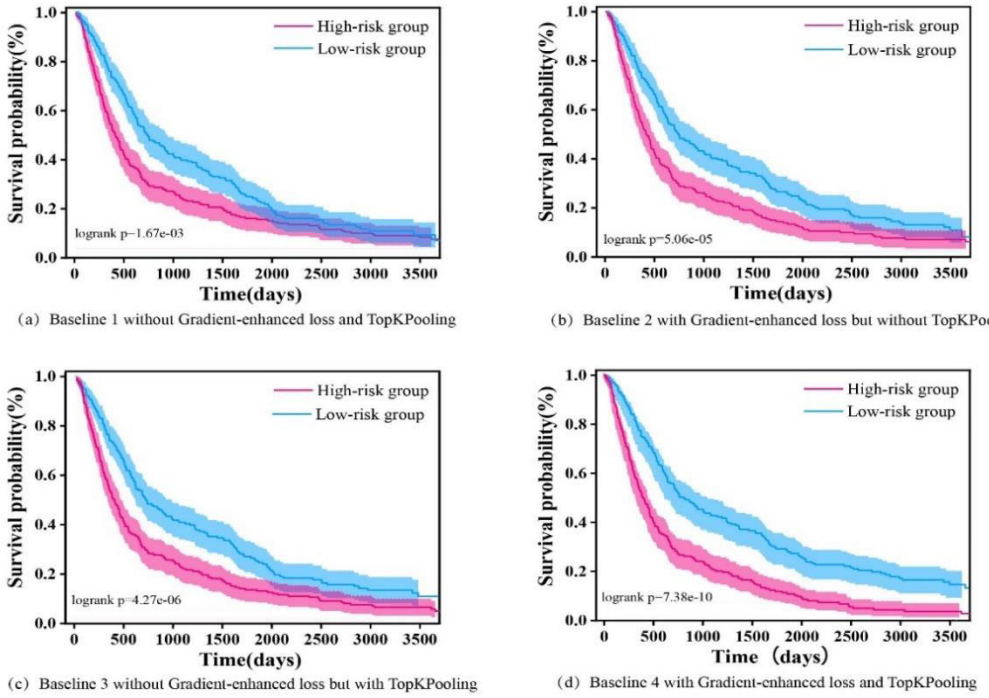

FIGURE 4. Kaplan-Meier Survival Curve and Model Performance Analysis Plot. (a) Baseline 1. (b) Baseline 2. (c) Baseline 3. (d) Baseline 4, which is the methodological model presented in this paper. The predictions of the patients used are combined here for a unified analysis. Patients were categorized into high and low-risk groups based on the median value of all risk coefficients obtained by this study model. In each subplot, the more the two curves overlap, the lower the model performance. Shaded areas indicate 95% confidence intervals.

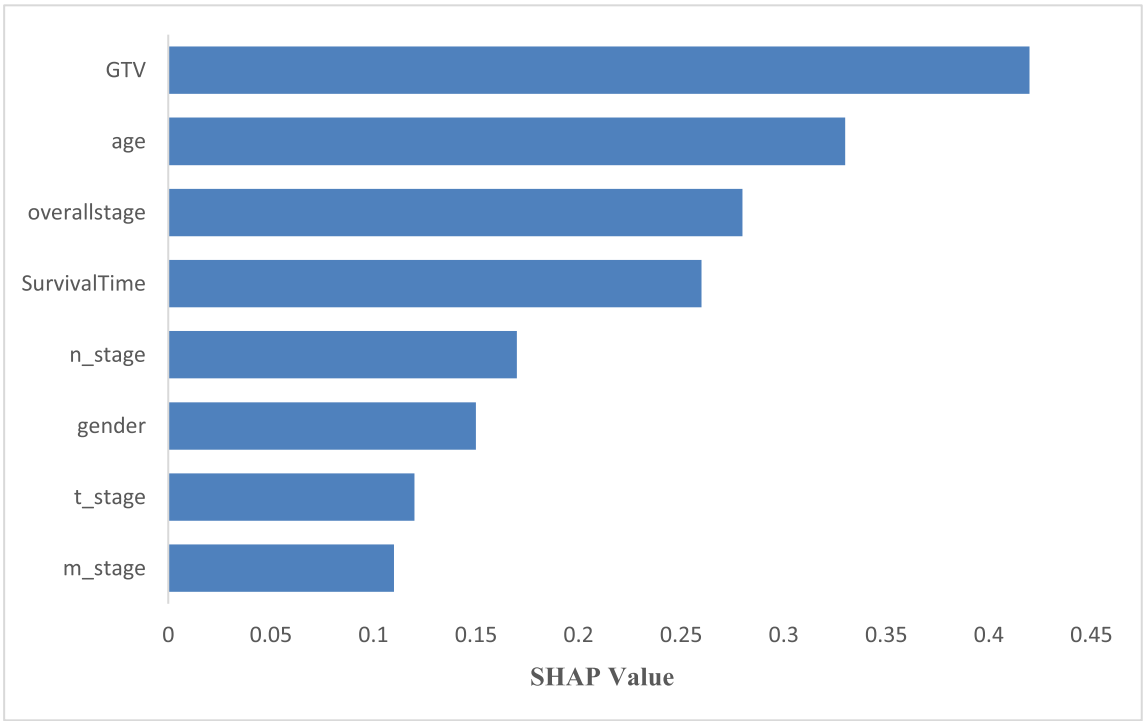

FIGURE 5. SHAP visual analytics. The graph presents the contribution of each feature to the predictive model, with larger SHAP values indicating that the feature contributes more to the results.

obtained a very low p-value of $\mathbf{7.38e^{-10}}$. Therefore, this study conclude that this methodological model plays an active role in survival prediction tasks. It has the potential to be applied to cancer clinical recommendation systems to help physicians make decisions regarding early intervention and personalized treatment strategies for NSCLC patients.

2) SHAP VISUALIZATION AND ANALYSIS

Cancer survival prediction requires not only high performance but also strong justifications to make proper judgments. Benefiting from graph-based modeling of multimodal data, this study framework can provide rich interpretability through backpropagation and SHAP analysis on multimodal graphs. **Figure 5** shows the contribution of each feature to the model prediction, the data shows that GTV and overall stage have a high value of concern in the model prediction process, in addition, patient age is also shown to be a key factor, which is also consistent with the existing knowledge, the above observations can provide a reference for clinicians to formulate a treatment plan according to the patient's needs.

V. CONCLUSION

In this study, this study propose and evaluate a novel survival prediction model (FGCN) for non-small cell lung cancer (NSCLC) based on deep learning and multimodal data. First, this study introduce the TopKPooling strategy to the survival prediction task of GCN multimodal fusion for the first time, which allows this study model to focus on salient features, thus avoiding processing too many redundant and noisy features and improving the efficiency and accuracy of feature map processing. Second, this study improved gradient-enhanced loss function accelerates the convergence of the model and ensures that the risk predicted by the model is as close as possible to the actual observed risk. Through ablation experiments and cross-validation, this study demonstrate the effectiveness of two proposed improved methods that together contribute to the improvement of the survival prediction model. The model shows excellent ability in extracting key feature information from multimodal data and using it for survival prediction, with a C-index value of 0.76, which is better than existing state-of-the-art models.

Additionally, since the model is essentially a predictive regression network based on graph convolution, by replacing the corresponding dataset, changing the structure of the graph data by controlling the informational modality of the inputs, as well as modifying the model labels and the type of outputs, the model can be equally applicable to other tasks, such as cancer diagnostics, classification of the cancer stages or predictive supplementation of missing clinical data. However, in practice the model still has many limitations, first of all, the model requires a high level of dataset, both the patient's clinical data, but also its supporting image data, which is very little existing, and at this stage can not achieve a wider dissemination. Second, the model only uses two types of modal information, clinical data and image data, which may not be able to fully explore the intrinsic interaction between the two modal information during feature extraction and thus affect the model performance. Finally the TopKPooling strategy is only suitable for dealing with feature matrices of the same density, and cannot have a positive effect on the model if the density of features known to each patient is different. The shortcomings and limitations described above are the very same challenges that this work are going to overcome in the work next research.

In summary, this study demonstrates the superior performance of the FGCN model in the field of NSCLC survival prediction, which provides new perspectives and methods for the treatment and management of NSCLC patients, and opens up a brand new way of thinking about some other problems in the clinic. In the future, this study plan to collect relevant datasets on our own and experiment with different data enhancements to solve the problem of the small sample size of the dataset of this model and select other evaluation metrics [55] to test the performance of the model in model training, e.g., AIC, SBIC, and at the same time, validate and improve the model on more datasets, and try to add more modal information, such as genomic information and therapeutic information, to the model, to further explore its adaptability and enhance the accuracy of model predictions. In addition, this study will be adding the tasks of diagnosing cancer and recommending treatments to this work and apply it to a wider range of clinical settings.

REFERENCES

[1] C. Su, "Emerging insights to lung cancer drug resistance," *Cancer Drug Resistance*, vol. 5, no. 3, p. 534, 2022.

[2] J. T. Yang and W. Dai, "Survival prognostic modeling of end-stage cancer patients and its application," *Chin. J. Evidence-Based Med.*, vol. 21, no. 1, pp. 8–14, 2021.

[3] X. Huang, S. Soong, W. H. McCarthy, M. M. Urist, and C. M. Balch, "Classification of localized melanoma by the exponential survival trees method," *Cancer*, vol. 79, no. 6, pp. 1122–1128, Mar. 1997.

[4] S. Solberg, Y. Nilssen, O. T. Brustugun, P. M. Haram, Å. Helland, B. Møller, T.-E. Strand, S. G. F. Wahl, and L. Fjellbirkeland, "Concordance between clinical and pathology TNM-staging in lung cancer," *Lung Cancer*, vol. 171, pp. 65–69, Sep. 2022.

[5] W. Cai, Y. Li, B. Huang, and C. Hu, "Esophageal cancer lymph node metastasis–associated gene signature optimizes overall survival prediction of esophageal cancer," *J. Cellular Biochem.*, vol. 120, no. 1, pp. 592–600, Jan. 2019.

[6] Z. Guo, J. Lan, J. Wang, Z. Hu, Z. Wu, J. Quan, Z. Han, T. Wang, M. Du, Q. Gao, Y. Xue, T. Tong, and G. Chen, "Prediction of lymph node metastasis in primary gastric cancer from pathological images and clinical data by multimodal multiscale deep learning," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105336.

[7] F. A. C. le Noble, J.-J. Mourad, B. I. Levy, and H. A. J. Struijker-Boudier, "VEGF (vascular endothelial growth factor) inhibition and hypertension: Does microvascular rarefaction play a role?" *Hypertension*, vol. 80, no. 5, pp. 901–911, May 2023.

[8] L. Záveský, E. Jandáková, V. Weinberger, L. Minář, V. Hanzíková, D. Dušková, L. Z. Drábková, I. Svobodová, and A. Hořínek, "Ascites-derived extracellular microRNAs as potential biomarkers for ovarian cancer," *Reproductive Sci.*, vol. 26, no. 4, pp. 510–522, Apr. 2019.

[9] K. Qu, Z. Wang, H. Fan, J. Li, J. Liu, P. Li, Z. Liang, H. An, Y. Jiang, Q. Lin, X. Dong, P. Liu, and C. Liu, "Correction to: MCM7 promotes cancer progression through cyclin D1-dependent signaling and serves as a prognostic marker for patients with hepatocellular carcinoma," *Cell Death Disease*, vol. 13, no. 11, p. 950, Nov. 2022.

[10] B. S. Finkelman, H. Zhang, D. G. Hicks, and B. M. Turner, "The evolution of ki-67 and breast carcinoma: Past observations, present directions, and future considerations," *Cancers*, vol. 15, no. 3, p. 808, Jan. 2023.

[11] H. Shimada, Y. Nabeya, S.-I. Okazumi, H. Matsubara, T. Shiratori, Y. Gunji, S. Kobayashi, H. Hayashi, and T. Ochiai, "Prediction of survival with squamous cell carcinoma antigen in patients with resectable esophageal squamous cell carcinoma," *Surgery*, vol. 133, no. 5, pp. 486–494, May 2003.

[12] C. Liu, L. Wang, W. Lu, J. Liu, C. Yang, C. Fan, Q. Li, and Y. Tang, "Computer vision-aided bioprinting for bone research," *Bone Res.*, vol. 10, no. 1, p. 21, Feb. 2022.

[13] Z. Liu, S. Wang, D. Dong, J. Wei, C. Fang, X. Zhou, K. Sun, L. Li, B. Li, M. Wang, and J. Tian, "The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges," *Theranostics*, vol. 9, no. 5, pp. 1303–1322, 2019.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[15] X. Zhu, J. Yao, F. Zhu, and J. Huang, "WSISA: Making survival prediction from whole slide histopathological images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6855–6863.

[16] M. Treppner, H. Binder, and M. Hess, "Interpretable generative deep learning: An illustration with single cell gene expression data," *Human Genet.*, vol. 141, no. 9, pp. 1481–1498, Sep. 2022.

[17] Y.-H. Lai, W.-N. Chen, T.-C. Hsu, C. Lin, Y. Tsao, and S. Wu, "Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning," *Sci. Rep.*, vol. 10, no. 1, p. 4679, Mar. 2020.

[18] S. Wang et al., "Artificial intelligence in lung cancer pathology image analysis," *Cancers*, vol. 11, no. 11, p. 1673, 2019.

[19] X. Pan, Y. Lu, R. Lan, Z. Liu, Z. Qin, H. Wang, and Z. Liu, "Mitosis detection techniques in H&E stained breast cancer pathological images: A comprehensive review," *Comput. Electr. Eng.*, vol. 91, May 2021, Art. no. 107038.

[20] X. Zhang and T. Li, "Breast cancer pathological image classification based on cycle-GAN and improved DPN network," *Zhejiang da Xue Xue Bao. Gong Xue Ban*, no. 4, p. 727, 2022.

[21] W. Han, L. Qin, C. Bay, X. Chen, K.-H. Yu, N. Miskin, A. Li, X. Xu, and G. Young, "Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas," *Amer. J. Neuroradiol.*, vol. 41, no. 1, pp. 40–48, Jan. 2020.

[22] F. M. Howard, S. Kochanny, M. Koshy, M. Spiotto, and A. T. Pearson, "Machine learning–guided adjuvant treatment of head and neck cancer," *JAMA Netw. Open*, vol. 3, no. 11, Nov. 2020, Art. no. e2025881.

[23] L. Zhong, D. Dong, X. Fang, F. Zhang, N. Zhang, L. Zhang, M. Fang, W. Jiang, S. Liang, C. Li, Y. Liu, X. Zhao, R. Cao, H. Shan, Z. Hu, J. Ma, L. Tang, and J. Tian, "A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: A multicentre study," *EBioMedicine*, vol. 70, Aug. 2021, Art. no. 103522.

[24] K. Morani and D. Unay, "Deep learning-based automated COVID-19 classification from computed tomography images," *Comput. Methods biomech. Biomed. Eng., Imag. Visualizat.*, vol. 11, no. 6, pp. 2145–2160, Nov. 2023.

[25] C. Xie, P. Yang, X. Zhang, L. Xu, X. Wang, X. Li, L. Zhang, R. Xie, L. Yang, Z. Jing, H. Zhang, L. Ding, Y. Kuang, T. Niu, and S. Wu, "Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy," *eBioMedicine*, vol. 44, pp. 289–297, Jun. 2019.

[26] S. Akbar, M. Hayat, M. Tahir, S. Khan, and F. K. Alarfaj, "CACP-DeepGram: Classification of anticancer peptides via deep neural network and skip-gram-based word embedding model," *Artif. Intell. Med.*, vol. 131, Sep. 2022, Art. no. 102349.

[27] S. Akbar, A. Raza, T. A. Shloul, A. Ahmad, A. Saeed, Y. Y. Ghadi, O. Mamyrbayev, and E. Tag-Eldin, "pAtbP-EnC: Identifying anti-tubercular peptides using multi-feature representation and genetic algorithm-based deep ensemble model," *IEEE Access*, vol. 11, pp. 137099–137114, 2023.

[28] A. Raza, J. Uddin, A. Almuhaimeed, S. Akbar, Q. Zou, and A. Ahmad, "AIPs-SnTCN: Predicting anti-inflammatory peptides using fasttext and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks," *J. Chem. Inf. Model.*, vol. 63, no. 21, pp. 6537–6554, Nov. 2023.

[29] S. Akbar, A. Raza, and Q. Zou, "Deepstacked-AVPs: Predicting antiviral peptides using tri-segment evolutionary profile and word embedding based multi-perspective features with deep stacking model," *BMC Bioinf.*, vol. 25, no. 1, p. 102, Mar. 2024, doi: 10.1186/s12859-024-05726-5.

[30] W. Liao, C. A. C. Coupland, J. Burchardt, D. R. Baldwin, F. V. Gleeson, and J. Hippisley-Cox, "Predicting the future risk of lung cancer: Development, and internal and external validation of the CanPredict (lung) model in 19·67 million people and evaluation of model performance against seven other risk prediction models," *Lancet Respiratory Med.*, vol. 11, no. 8, pp. 685–697, Aug. 2023.

[31] C.-C. Pu, L. Yin, and J.-M. Yan, "Risk factors and survival prediction of young breast cancer patients with liver metastases: A population-based study," *Frontiers Endocrinol.*, vol. 14, Jun. 2023, Art. no. 1158759.

[32] S. R. Syed and M. A. S. Durai, "A diagnosis model for brain atrophy using deep learning and MRI of type 2 diabetes mellitus," *Frontiers Neurosci.*, vol. 17, Oct. 2023, Art. no. 1291753.

[33] Z. Lin, W. Cai, W. Hou, Y. Chen, B. Gao, R. Mao, L. Wang, and Z. Li, "CT-guided survival prediction of esophageal cancer," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2660–2669, Jun. 2022.

[34] W. Hou, C. Lin, L. Yu, J. Qin, R. Yu, and L. Wang, "Hybrid graph convolutional network with online masked autoencoder for robust multi-modal cancer survival prediction," *IEEE Trans. Med. Imag.*, vol. 42, no. 8, pp. 2462–2473, Aug. 2023.

[35] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[36] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande, "MoleculeNet: A benchmark for molecular machine learning," *Chem. Sci.*, vol. 9, no. 2, pp. 513–530, 2018.

[37] C. Liu, H. Ding, Y. Zhang, and X. Jiang, "Multi-modal mutual attention and iterative interaction for referring image segmentation," *IEEE Trans. Image Process.*, vol. 32, pp. 3054–3065, 2023.

[38] S. Li, Y. Xie, G. Wang, L. Zhang, and W. Zhou, "Adaptive multimodal fusion with attention guided deep supervision net for grading hepatocellular carcinoma," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 4123–4131, Aug. 2022.

[39] A. Vaswani, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[40] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[41] H. Gao and S. Ji, "Graph U-Nets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4948–4960, Sep. 2022.

[42] C. Cangea, P. Veličković, N. Jovanović, T. Kipf, and P. Liò, "Towards sparse hierarchical graph classifiers," 2018, *arXiv:1811.01287*.

[43] B. Knyazev, G. W. Taylor, and M. Amer, "Understanding attention and generalization in graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019, pp. 1–11.

[44] Y. Yang, J. Wu, S. Huang, Y. Fang, P. Lin, and Y. Que, "Multimodal medical image fusion based on fuzzy discrimination with structural patch decomposition," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1647–1660, Jul. 2019.

[45] A. Al-Mamgani, R. Kessels, T. Janssen, A. Navran, S. van Beek, C. Carbaat, W. H. Schreuder, J.-J. Sonke, and C. A. M. Marijnen, "The dosimetric and clinical advantages of the GTV-CTV-PTV margins reduction by 6 mm in head and neck squamous cell carcinoma: Significant acute and late toxicity reduction," *Radiotherapy Oncol.*, vol. 168, pp. 16–22, Mar. 2022.

[46] H. V. Pham, D. H. Thanh, and P. Moore, "Hierarchical pooling in graph neural networks to enhance classification performance in large datasets," *Sensors*, vol. 21, no. 18, p. 6070, Sep. 2021.

[47] R. Chen, G. Vivone, G. Li, C. Dai, and J. Chanussot, "An offset graph U-Net for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520615.

[48] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, no. 1, p. 4006, Jun. 2014.

[49] S. H. Patel, L. M. Poisson, D. J. Brat, Y. Zhou, L. Cooper, M. Snuderl, C. Thomas, A. M. Franceschi, B. Griffith, A. E. Flanders, J. G. Golfinos, A. S. Chi, and R. Jain, "T2–FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: A TCGA/TCIA project," *Clin. Cancer Res.*, vol. 23, no. 20, pp. 6078–6085, Oct. 2017, doi: 10.1158/1078-0432.ccr-17-0560.

[50] D. Y. Lin and L. J. Wei, "The robust inference for cox proportional hazards model," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 1074–1078, 1989, doi: 10.1080/01621459.1989.10478874.

[51] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 24, Feb. 2018.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[53] P. Mobadersany and S. Yousefi, *Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press, 2017, doi: 10.1101/198010.
[54] J. T. Rich, J. G. Neely, R. C. Paniello, C. C. J. Voelker, B. Nussenbaum, and E. W. Wang, "A practical guide to understanding Kaplan-Meier curves," *Otolaryngol.-Head Neck Surg.*, vol. 143, no. 3, pp. 331–336, Sep. 2010.
[55] J. Fan and R. Li, "Variable selection for cox's proportional hazards model and frailty model," *Ann. Statist.*, vol. 30, no. 1, pp. 74–99, Feb. 2002.

XIAOPU MA received the M.S. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, in 2004, and the Ph.D. degree from the School of Computer Science and Technology, Huazhong University of Science and Technology, in 2011. He is currently working as a Professor with the School of Computer Science and Technology, Nanyang Normal University, Henan, China. His research interests include medical image processing and computer network security.

FEI NING received the Academic degree from the School of Computer Science and Technology, Nanyang Normal University, Henan, China, in 2020, where he is currently pursuing the master's degree with the School of Life Science and Agricultural Engineering. His research interests include include AI and medical image processing.

XIAOFENG XU received the Ph.D. degree from Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, in 2020. He is currently working as a Lecturer with the School of Computer Science and Technology, Nanyang Normal University, Henan, China. His research interests include medical image processing and computer artificial intelligence.

JIANGDAN SHAN received the master's degree from the School of Life Science and Agricultural Engineering, Nanyang Normal University. Her research interest includes medical image processing.

HE LI received the M.S. degree in computer science and technology from Henan Polytechnic University, Jiaozuo, China, in 2013, and the Ph.D. degree in communication and information systems from the Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2018. He is currently working as a Lecturer with Nanyang Normal University. His research interests include ad hoc and sensor network management.

XIAO TIAN received the M.S. degree in computer science and technology from Henan Polytechnic University, Jiaozuo, China, in 2014. She is currently working as a Lecturer with Nanyang Medical College. Her research interests include cloud computing and sensor networks.

SHUAI LI is currently pursuing the master's degree with Nanyang Normal University, Henan, China. His research interests include artificial intelligence and medical image processing.