**ORIGINAL ARTICLE**

# DCCAFN: deep convolution cascade attention fusion network based on imaging genomics for prediction survival analysis of lung cancer

Liye Jia[1] · Xueting Ren[1] · Wei Wu[3] · Juanjuan Zhao[1,2] · Yan Qiang[1] · Qianqian Yang[2]

## Abstract

Recently, lung cancer prediction based on imaging genomics has attracted great attention. However, such studies often have many challenges, such as small sample size, high-dimensional information redundancy, and the inefficiency of multimodal fusion. Therefore, in this paper, a deep convolution cascade attention fusion network (DCCAFN) based on imaging genomics is proposed for the prediction of lung cancer patients' survival. The network consists of three modules: an image feature extraction module (IFEM), a gene feature extraction module (GFEM), and an attention fusion network (AFN). In the IFEM, a pretrained residual network based on transfer learning is used to extract deep image features to fully capture the computed tomography (CT) image information conducive to prognosis prediction. In the GFEM, the $F$-test is first used for gene screening to eliminate redundant information, and then, a cascade network with the convolution cascade module (CCM) that contains a convolution operation, a pooling operation, and an ensemble forest classifier is designed to better extract the gene features. In the AFN, a bimodal attention fusion mechanism is proposed to fuse deep image features and gene features to improve the performance of predicting lung cancer survival. The experimental results show that the DCCAFN model achieves good performance, and its accuracy and AUC are 0.831 and 0.816, respectively. It indicates that the model is an effective multimodal data fusion method for predicting the survival prognosis of lung cancer, which can greatly help physicians stratify patients' risks, and achieve personalized treatment for improving the quality of patients' lives.

**Keywords** Imaging genomics · Deep features · Convolution cascade · Attention fusion · Survival prediction

## Introduction

Lung cancer is one of the most common malignant tumors in the world, with the highest morbidity and mortality, which seriously threatens human health and life [1, 2]. It is reported that 70% of lung cancer diagnoses are made after the symptoms of advanced local or metastatic diseases appear, and the 5-year survival rate after diagnosis is about 16% [3, 4]. Only when lung cancer patients are diagnosed at an early stage can their survival rate reach more than 50% [5, 6]. Hence, an accurate diagnosis is crucial for the treatment choice and prognosis of each lung cancer patient [7]. Unfortunately, the heterogeneity of lung cancer in many aspects, such as histology, molecular characteristics, and driving genes, makes accurate diagnosis difficult and also makes the prognostic survival time of patients vary from several months to 7 years [8, 9]. Therefore, there is an urgent need for an effective survival prediction model to assist in the selection of treatment plans so as to improve the treatment effect of patients and increase their cure rate and survival rate.

With the rapid development of computer-aided technology, many machine learning and deep learning methods have been developed for the analysis of lung cancer survival prognosis [10–12]. These methods primarily utilize clinical information and image information, such as computed tomography (CT) images and positron emission tomography (PET) images, of lung cancer patients for predicting their survival. For example, Katzman et al. [13] proposed a Cox proportional hazards deep neural network and state-of-the-art survival method, referred to as the DeepSurv model, for establishing the interaction between patient covariates and

✉ Juanjuan Zhao
zhaojuanjuan@tyut.edu.cn

1 College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

2 Jinzhong College of Information, Jinzhong 030600, China

3 Department of Physiology, Shanxi Medical University, Taiyuan 030051, China

treatment effects so as to provide personalized treatment suggestions. She et al. [14] applied the DeepSurv model to the survival analysis of non-small cell lung cancer (NSCLC) and demonstrated that DeepSurv could be used to provide treatment recommendations for better survival outcomes. Astarak et al. [15] designed a novel feature set from the CT and PET images for capturing intra-tumor heterogeneity and used a support vector machine based on the novel feature set and the classic radiomic features for the task of overall survival prediction. Amini et al. [16] proposed a multi-level multi-modal radiomics model based on feature level fusion and image level fusion of PET images and CT images to improve the overall survival prediction accuracy of non-small cell lung cancer (NSCLC) patients. The results show that the concordance index ($C$-index) of the 3D wavelet transform fusion strategy in predicting survival risk is the highest ($C$-index = 0.708). Mukherjee et al. [17] developed a shallow convolution neural network to analyze CT images across four medical centers for predicting the overall survival rate of NSCLC patients. The $C$-index of the overall survival rate of each independent cohort was 0.62, 0.62, 0.62, and 0.58, respectively. Wu et al. [18] proposed a multimodal deep learning method for NSCLC survival analysis. This method uses CT images and clinical data to achieve fully automatic end-to-end lung cancer survival analysis based on 3D ResNets, which allows the rich information associated with survival information in CT images to be preserved and provides personalized prognosis and decision-making with sufficient granularity. However, it is limited to further improve the performance of the above methods using only images and clinical information.

In addition to the above-mentioned clinical and image data related to cancer, a large amount of gene data is also available with the development of high-throughput sequencing technologies. These gene data are involved in the biological processes in many cancers and are associated with the prognostic survival time of patients, which makes them of great interest in the survival prognostic analysis of cancer. Hence, many studies have attempted to use a clinically acceptable combination of gene expression information and images to maximize the prediction performance of cancer survival [19–21]. For example, Wang et al. [22] proposed a deep bilinear network (GPDBN) that effectively integrates gene data and pathological images for improving the performance of breast cancer prognosis prediction. In this model, an inter-modal and two intra-modal bilinear feature coding modules are designed to build complex inter-modal and intra-modal relationships, respectively. Then, a multilayer deep neural network is used to obtain complementary information between inter-modal and intra-modal relationships for the final prognosis prediction. Li et al. [23] proposed a new hierarchical multimodal fusion method, HFBSurv, which is mainly designed with modality-specific and cross-modality

attention factor decomposition bilinear modules and uses multiple fusion strategies to gradually and hierarchically fuse gene and image features. A large number of experiments show that HFBSurv can effectively perform multimodal data hierarchical fusion and achieve good performance in survival prediction. Chen et al. [24] proposed an interpretable strategy for end-to-end multimodal fusion of histology image and genomic features for survival prediction, which can improve prognostic determinations from ground truth grading and molecular subtyping.

Therefore, the above-mentioned deep learning model based on imaging genomics has great potential and application in predicting survival prognosis of cancer using gene data and image data. However, there are still some challenges to achieving high prediction accuracy of cancer survival based on imaging genomics: (1) The multimodal medical image data available for most studies tends to have a small sample size and is prone to overfitting when trained directly using deep learning models. (2) The high-dimensional gene information contains a lot of redundancy and noise, which will affect the model's performance. (3) The existing studies on the fusion of multimodal information such as medical images and genomics still have low fusion efficiency, which is insufficient to establish the complex relationship between multimodalities.

To this end, we propose a deep convolution cascade attention fusion network (DCCAFN) based on imaging genomics to predict the survival prognosis of lung cancer. For CT images, the pretrained residual network is used to extract deep features, effectively preserving the image features related to prognosis prediction. For gene data, feature selection is carried out for high-dimensional gene information, and then features are further extracted through a cascade network with the convolution cascade module (CCM), effectively eliminating redundant data in gene information and retaining the most relevant gene features. In addition, an attention fusion mechanism is proposed, in which deep image features and gene features are used for information fusion and interaction so that the model can not only obtain important features of each modality but also effectively perform deep fusion.

The main contributions of this paper are as follows:

(1) A DCCAFN model based on imaging genomics is proposed to predict the survival risk of lung cancer. The network effectively solves the low-efficiency problem of multimodal feature fusion and provides a new paradigm for multimodal data fusion.
(2) A cascade network with the CCM is proposed to extract gene features, effectively eliminate redundant data in gene information, and retain the most relevant gene features.

(3) An attention fusion mechanism is proposed to fuse deep image features and gene features, which can accurately complete the task of lung cancer survival prediction.

## Related work

### Transfer learning

Deep learning has played an important role in the study of medical imaging. However, the excellent classification ability of deep learning algorithms relies on large-scale datasets, and medical image datasets are often limited and small scale. For this reason, many studies have been proposed to solve the limitations of small medical datasets. Transfer learning techniques improve model performance by transferring image representation capabilities learned from larger-scale natural images to medical small-sample images [25–28]. Esteva et al. [29] used a pretrained GoogleNet Inception v3 Convolutional Neural Network (CNN) on a large-scale natural image set for skin cancer classification. Hassan et al. [30] proposed an efficient and accurate classification method for medical images, which used transfer learning and pretrained ResNet50 model to optimize feature extraction, and then performed linear discriminant analysis classification. The method can be used to retrieve clinical cases from large medical repositories. Therefore, the paper uses transfer learning and pretrained ResNet50 model to extract deep features conducive to survival prognosis, so as to avoid the problem of overfitting caused by the small sample size. The schematic diagram of the transfer learning process is shown in Fig. 1.

### Deep cascade

In machine learning and deep learning, the quality of data features directly affects the performance of the model. High-dimensional data usually contains a large number of redundant features, which will interfere with the subsequent data analysis process, or even cause overfitting, thus affecting the final classification result. Practice has proved that the cascade model has made a new breakthrough, which not only borrows from the advantages of deep learning, but also effectively avoids the phenomenon of overfitting [31, 32]. For example, Zhou et al. [33] proposed a multi-grained cascade forest (gcForest) algorithm based on the idea of constructing deep models with non-differentiable modules such as decision trees. The gcForest can perform representation learning through forests. It works well with small-scale data, and its performance is robust to hyperparameter settings. Ni et al. [34] proposed a Cascade-Gate Forest with gating mechanism. The base classifier of Cascade-Gate Forest only uses random forest, and each layer does not use the previous layer

to extract all features, but uses out-of-bag error estimation (OOB) to filter the base classifier with better classification effect to extract features. Wang et al. [35] improved gcForest based on the idea of DenseNet, and proposed a dense adaptive cascade forest (daForest). In the dense connection, the features extracted by each layer are spliced with all subsequent layers. In the determination of the number of base classifiers in each layer, a linear search method is used to determine the optimal parameters. Mossa et al. [36] proposed a cascade approach based on deep learning for the overall survival (OS) classification of brain tumor patients using multimodal magnetic resonance images (MRI) to improve the performance of CNN model on small volume datasets. Shaaban et al. [37] proposed a dynamic deep cascade model for deep convolution forest (DCF). The model uses the convolution and pooling layers to automatically extract features, and dynamically detects spam based on basic classifiers such as random forest and extreme random tree. The results show that the model achieves remarkable accuracy in text classification.

For small datasets, the performance of individual classifiers may be unsatisfactory. However, the cascade model based on multiple classifiers can achieve satisfactory results. Practice has proved that the cascade method can not only improve the generalization performance and classification accuracy of the model, but also reduce the complexity of the algorithm, and it can be widely used in text, image, gene and other fields [31, 32, 36, 37]. Therefore, based on the idea of deep cascade, this paper proposes convolution cascade modules (CCMs) to extract gene features, and the CCMs improve the model performance in predicting the survival prognosis of lung cancer on small datasets.

### Visualizing CNNs

Deep neural networks are known for their excellent handling of a variety of machine learning and artificial intelligence tasks. However, due to their over-parameterized black-box nature, they are often accused of lacking interpretability, and it is often difficult to understand the prediction results of deep models. In recent years, many interpretative tools have been proposed to help systematically investigate the learned weights and further examine the results of neural networks. Some recent work visualizes the internal representations learned by CNN in an attempt to better understand the features they extract [38]. Zhou et al. [39] elucidated how the global average pooling layer explicitly enables CNN to have significant localization capability, and the proposed network is able to localize the identified image regions, exposing the implicit attention of CNN on images. Selvaraju et al. [40] proposed a technique for generating "visual explanations" for decisions from a large number of CNN-based models, Gradient Weighted Class Activation Mapping (Grad-CAM), which uses the gradient of any target concept that flows into the final
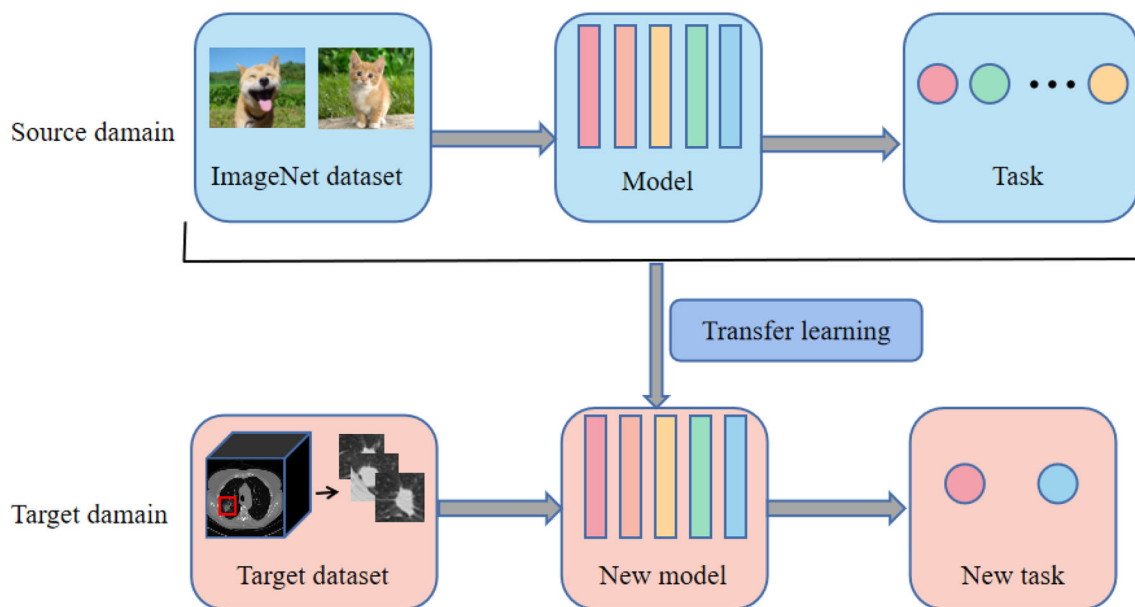
**Fig. 1** The schematic diagram of the transfer learning process

convolutional layer to generate a coarse localization mapping that highlights important regions of the predicted concept in the image, making CNN understanding more transparent. Katafuchi et al. [41] proposed a Layer-wise External Attention Network (LEA-Net) that converts anomaly maps into an attention map and then incorporates the attention map into an intermediate layer network to effectively detect image anomalies. Experiments demonstrate that the proposed layered visual attention mechanism can consistently improve the anomaly detection performance of existing CNN models. Rosso et al. [42] proposed the use of CNN ResNet-50 and a transfer learning approach to accomplish the classification of defects in raw images provided by the GPR instrument and employed a state-of-the-art neuro-vision converter (ViT) architecture to generate attention maps to enhance the interpretability of the model. These works have demonstrated that the introduction of attention maps can indeed improve the prediction results of the models and that the interpretability of neural networks can be effectively enhanced by visualizing the attention maps.

## Method

In this section, DCCAFN based on imaging genomics is proposed to predict the survival of lung cancer. The overall architecture mainly includes three modules: image feature extraction, gene feature extraction, and the attention fusion network. In the image feature extraction module (IFEM), the CT image is transformed into a 2D image block containing nodules by clipping, and the deep features are extracted

using the pretrained ResNet50 network. In the gene feature extraction module (GFEM), the gene expression data of RNA sequencing (RNA-seq) is screened by $F$-test, and then further extracted by a cascade network with the CCM. In the attention fusion network (AFN), the extracted deep image features and gene features are processed by Hadamard product and sumpooling. Then, they are assigned different importance levels and input into the constructed AFN to predict the survival of lung cancer. The final output is obtained from the predicted survival probability value. The overall architecture is shown in Fig. 2. In the following, the image feature extraction, the gene feature extraction, and the attention fusion network will be introduced, respectively.

## Image feature extraction

For image information, we design an IFEM to extract deep image features, as shown in Fig. 2a. In this module, a pretrained ResNet50 network is used to extract more effective deep features from the CT images. This paper mainly regards ResNet50 as a baseline for deep feature extraction, which is mainly composed of the residual block and the jump connection between the block and the block. The residual block consists of a series of convolutional layers, batch normalization, and Relu activation layers. The jump connection makes the gradient of back propagation better by shortening the distance between non-adjacent layers, and it also enables the network to automatically learn the path of feature motions without affecting the performance of the network, thereby enhancing the generalization ability of the network.
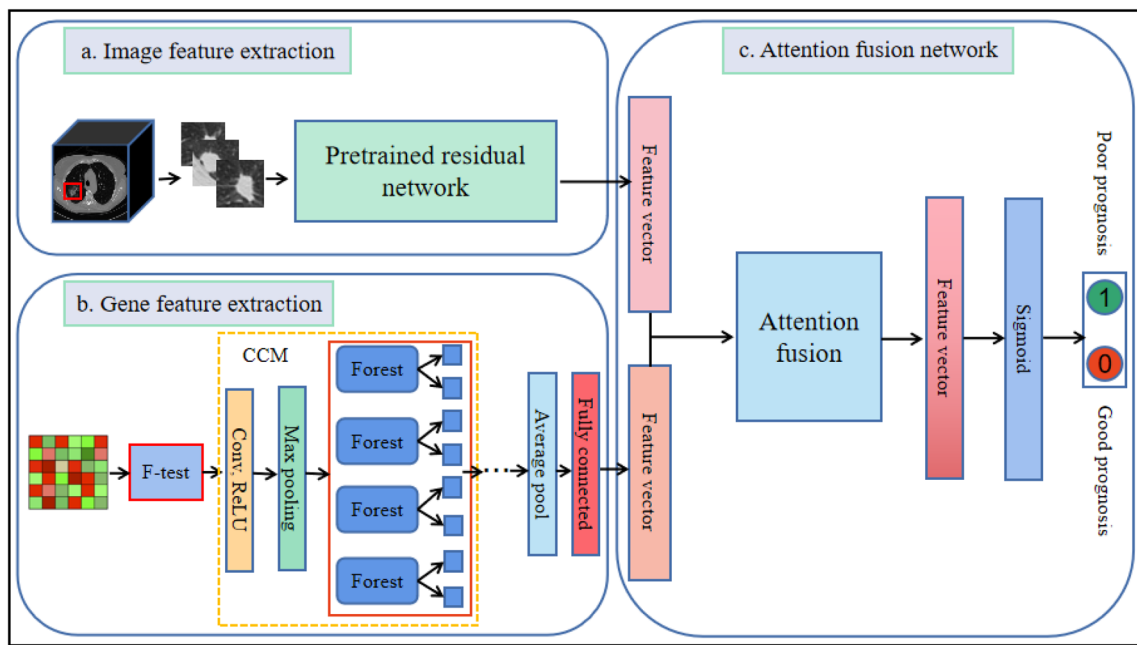
**Fig. 2** The overall architecture of DCCAFN. There are three main modules in DCCAFN: **a** image feature extraction module: The pretraining ResNet50 is used to extract deep features; **b** gene feature extraction module: $F$-test is used for gene feature screening, and then the gene features are further extracted based on deep convolution cascade algorithm; **c** attention fusion network: based on the extracted image features and gene features, the attention fusion network is designed to predict survival



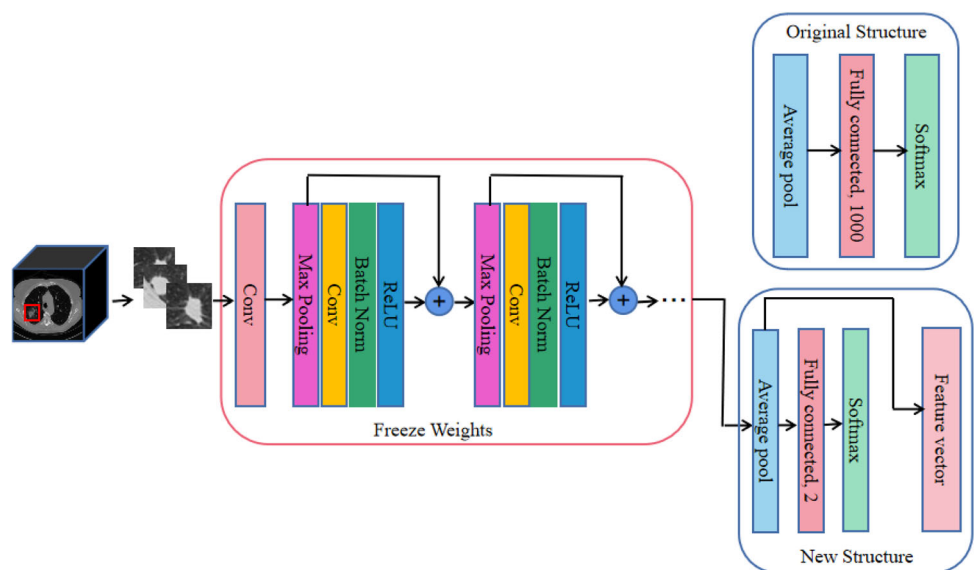**Fig. 3** The details of image feature extraction

Figure 3 shows the details of image feature extraction. In the original network, the fully connected layer of ResNet50 provides an optimal feature, followed by the softmax classification. In this study, we use pretrained ResNet50 based on transfer learning to extract deep image features, which can make better use of the robustness and discrimination learning ability of ResNet50. Firstly, we perform preprocessing steps to adjust the size of images and align them with the ResNet50 network input size. Then, we utilize the pretrained ResNet50 model trained on the ImageNet dataset. Using the transfer learning, the weights except the last fully connected layer are frozen in the setting of the pretrained ResNet50 model, and the original 1000 classes are replaced with the two classes in the last fully connected layer. Lastly, the ResNet50 model is trained on a relatively small dataset to learn and fine-tune the weights of the fully connected layer for CT image classification, and modules before the last fully connected layer in the well-trained model are used as the deep feature extraction

module. Eventually, in the deep feature extraction module, the optimal deep image feature vector $f_{\text{image}}$ of size 2048 is obtained at the "Average pool" layer, and input into the AFN for lung cancer survival prediction.

## Gene feature extraction

The gene expression data used in this study is RNA-seq data. In the RNA-seq data, each patient contains more than 20,000 gene with a large amount of redundant and noisy information, which will significantly increase the computational cost and reduce the prediction accuracy. Therefore, the relevant genes need to be screened from the RNA-seq data before training the model. In the screened genes, some ambiguous gene expressions (N/A) will be deleted. In the section, a GFEM is proposed to extract the gene feature from the RNA-seq data, as shown in Fig. 4. First, genes most relevant to lung cancer survival are selected through the $F$-test algorithm [43]. Then, the deep gene features are further extracted through a cascade network with the CCM.

In the GFEM, the CCM at each level consists of the convolution, pooling, and classification layers, among which the classification layer contains some random forest (RF) base classifiers. The convolution layer is responsible for feature extraction, while the pooling layer helps reduce overfitting in the proposed model. In addition, the classification layer predicts the probability of survival. Each level receives processed feature information from its previous level and outputs its processing results to the next level. The output at each level is the probability of base classifiers, which are then connected to the feature maps of the pooling level output to form the input of the next level. Specifically, the convolution layer extracts the most relevant hidden features from the gene data by performing the convolution operation on the screened gene features and applying the ReLU activation function to the output end. Then, the global maximum pooling operation is used to obtain the maximum values of the output feature maps of the convolution layer, and these maximum values are input to the RF classification layer. At the same time, the new features extracted from the convolution and the pooling layers are concatenated with the output probability features of the classification layer as the input vector of the next convolution layer. Until the last CCM, its output probability features are concatenated with the output features of the pooling layer as the input of the next layer, and then the output feature vector $f_{\text{gene}}$ of GFEM is obtained through average pooling and full connection layer.

The GFEM combines the advantages of ideas in the bagging and boosting methods. Multiple RFs are used in each CCM for classification, which reflects the advantage of bagging methods in reducing variance. Multiple CCMs are cascaded, and the next level continuously corrects the errors of the previous level through the output of the previous level,

reflecting the advantage of boosting in reducing deviation. In this paper, the adaptive method is used to determine the number of levels (i.e. CCM). When the three consecutive levels of the model cannot extract new features based on the previous level to obtain better classification accuracy, the model stops training. Unlike deep neural networks in which the number of hidden layers is a predefined parameter, the GFEM can adjust the complexity during training by itself. Therefore, the proposed model is very suitable for high-dimensional small sample gene data.

## Attention fusion network

In order to effectively fuse the extracted image and gene information, we design the AFN based on the extracted deep features and gene features to fuse information of different modalities. The network can clearly explore the complex relationship between different modalities and give them different importance, as shown in Fig. 5. In our work, using feature representations $f_{\text{image}}$ and $f_{\text{gene}}$ of different modalities, the cross-modality representation $f_{\text{fusion}}$ can be obtained by:

$$f_{\text{fusion}} = \text{Sumpooling}\left(U^T f_{\text{image}} \odot V^T f_{\text{gene}}, k\right), \quad (1)$$

where $\text{Sumpooling}(f, k)$ represents performing sumpooling operation over $f$ by using non-overlapping windows of size $k$; $U^T$ and $V^T$ represent learnable weight matrices; $\odot$ is the Hadamard product of two feature vectors.

Furthermore, a bimodal attention is introduced to determine the importance of the cross modality representation. In our work, the importance of deep image features and deep gene features is first measured by $\alpha_1$ and $\alpha_2$, as follows,

$$\alpha_1 = \text{Sigmoid}\left(w_1 f_{\text{image}} + b_1\right),$$
$$\alpha_2 = \text{Sigmoid}\left(w_2 f_{\text{gene}} + b_2\right), \quad (2)$$

where $w_m$ and $b_m (m = 1,2)$ are the parameter matrix and bias terms (from image feature modality and gene feature modality) of the fully connection layer from image feature modality and gene feature modality, respectively. Then, we consider the similarity $S_{\text{fusion}}$ between $f_{\text{image}}$ and $f_{\text{gene}}$, and it is estimated as follows:

$$S_{\text{fusion}} = \sum \left(\frac{e^{\alpha_1 f_{\text{image}}}}{\sum e^{\alpha_1 f_{\text{image}}}}\right)\left(\frac{e^{\alpha_2 f_{\text{gene}}}}{\sum e^{\alpha_2 f_{\text{gene}}}}\right). \quad (3)$$

The calculated similarity is in the range of 0 to 1. The importance $\alpha$ of the final cross modality is obtained as follows:

$$\alpha = \frac{e^{\hat{\alpha}}}{\sum e^{\hat{\alpha}}}, \quad \hat{\alpha} = \frac{\alpha_1 + \alpha_2}{S_{\text{fusion}} + S_0}, \quad (4)$$
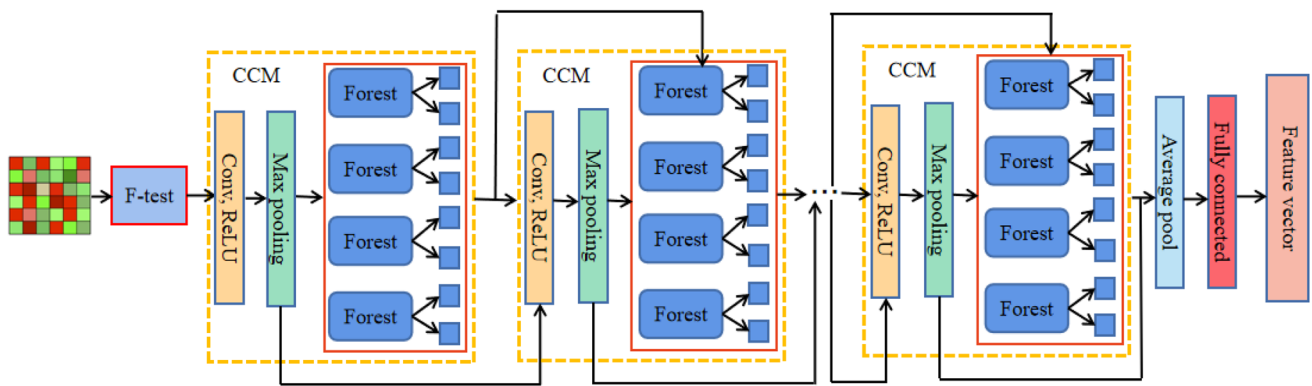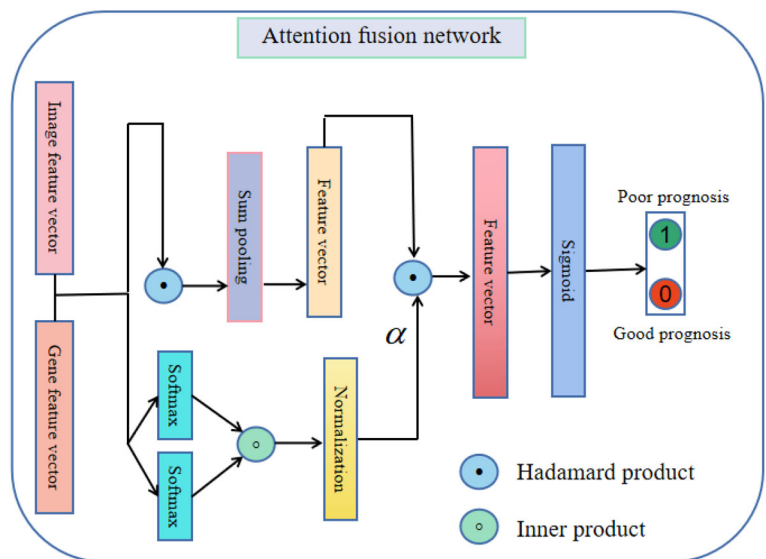
**Fig. 4** The details of gene feature extraction

**Fig. 5** The details of attention fusion network



where $S_0$ represents a predefined term that controls the relative contribution between the similarity and the importance of a specific modality, which is set to 0.5 here. Therefore, the output feature $\hat{f}_{\text{fusion}}$ of the AFN can be denoted as the following weighted cross modality representation,

$$\hat{f}_{\text{fusion}} = \alpha f_{\text{fusion}}. \tag{5}$$

The output features of the last fully connected layer are input into the output layer (i. e. sigmoid layer) to generate the final prediction scores for shorter term and longer term survivors.

## Experimental results

In this section, we first introduce specific data preprocessing, experiment details, and evaluation metrics. Then, some ablation experiments are performed to verify the validity of

each module in the DCCAFN model. Finally, the proposed model is compared with other studies to demonstrate that the proposed DCCFAN model can make full use of the CT image information and the gene information and effectively improve the performance of predicting survival.

## Datasets

Some experiments are conducted on the public datasets NSCLC Radiogenomics, TCGA-LUSC, and TCGA-LUAD downloaded from the TCIA website (https://wiki.cancerim agingarchive.net). The patients involved in the dataset have received ethical approval. Meanwhile, patients from the public dataset need to meet the following inclusion criteria:

(1) Primary lung cancer is confirmed by histology;
(2) All selected patients included follow-up data for 5-year survival time;
(3) Cases contain both CT data and RNA-seq data.

**Fig. 6** Lung cancer CT images including the shorter term survivors and longer term survivors image samples
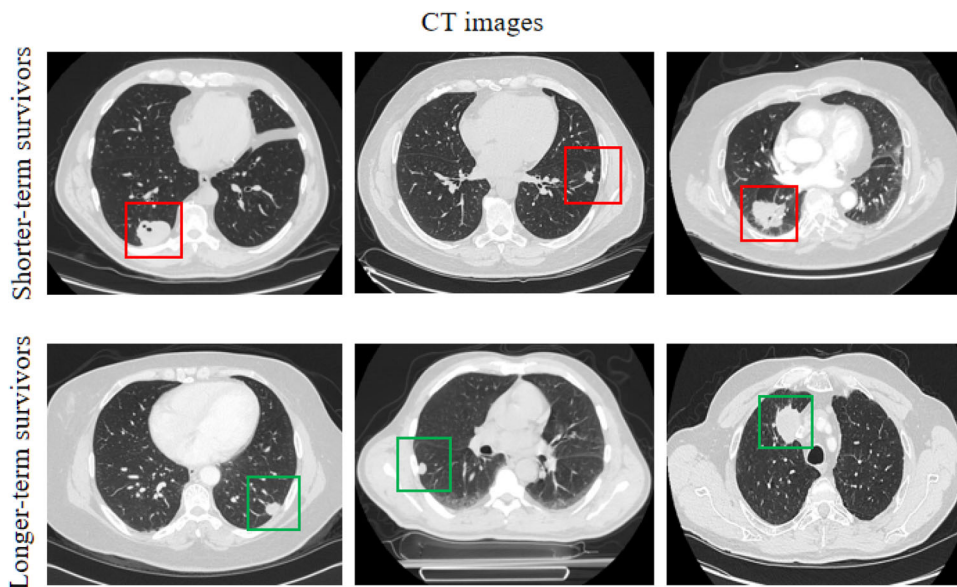


Besides, in the training and testing datasets, patients will be excluded with the following situation, such as (1) the lack of clinical data; (2) the lack of RNA-seq data; (3) the lack of CT data; (4) the lack of follow-up data.

The lesion areas in all CT images from 168 patients in the public dataset (NSCLC Radiogenomics 117 cases, TCGA-LUSC 30 cases, and TCGA-LUAD 21 cases) are marked by these experienced radiologists (lung imaging practice for 5 years) in the partner hospital. Based on these marked lesion areas, the CT images are cropped to a region of interest (ROI) with a size of $64 \times 64$, and a total of 6467 ROI images are obtained. Among the 168 cases, there are 5268 genes after deleting the missing values in the RNA-seq data. After the z-score standardization of the remaining genes, the F-Test algorithm is used to screen the genes, and the screened genes are used as the input of the cascade network with the CCM. These patients are further classified into longer term and shorter term survivors using the 5-year survival criterion based on their clinical information. Accordingly, shorter term survivors are labeled as 1 (i.e., a poor prognosis), while longer-term survivors are labeled as 0 (i.e., a good prognosis). Figure 6 shows some CT images, including the shorter term survivors and longer term survivors image samples. To comprehensively evaluate the proposed survival prediction methods as well as ensure the robustness of the results, we employ fivefold cross-validation. In particular, the dataset is randomly divided into the training set and the testing set at a ratio of 4:1.

Table 1 presents the clinical characteristics of patients, including the number of patients, average age, sex, smoking status, histology, and survival status in the training set and testing set, and the corresponding $p$ value between the two datasets. It is evident from the table that the $p$ values of

**Table 1** The clinical characteristics of patients

| Attributes | Training set | Testing set | $p$ value |
|---|---|---|---|
| The number of Patients | 134 | 34 | |
| Average age | 68.7 | 67.2 | 0.102 |
| Sex | | | |
| Female | 38 (22.6%) | 10 (6.0%) | 0.336 |
| Male | 96 (57.1%) | 24 (14.3%) | |
| Smoking status | | | |
| Non smoker | 22 (13.1%) | 5 (3.0%) | 0.145 |
| Former | 82 (48.8%) | 21 (12.5%) | |
| Current | 30 (17.9%) | 8 (4.8%) | |
| Histology | | | |
| Adenocarcinoma | 87 (51.8%) | 22 (13.1%) | 0.763 |
| Squamous cell carcinoma | 45 (26.8%) | 11 (6.5%) | |
| NSCLC NOS (not otherwise specified) | 2 (1.2%) | 1 (0.6%) | |
| Five-year survival time | | | |
| Shorter-term survivors | 88 (52.4%) | 22 (13.1%) | |
| Longer-term survivors | 46 (27.4%) | 12 (7.1%) | |

age, sex, smoking status, and histology are greater than 0.05, which implies that there are no significant differences in age, sex, smoking status, and histology between the training set and the testing set. Note that when the $p$ value is less than 0.05, there is statistical significance for the corresponding characteristic between the training set and the testing set.

**Table 2** Optimum parameter setting in the GFEM of the DCCAFN model

| Parameters | Best value |
|---|---|
| The number of convolution layers in each CCM | 1 |
| The number of trees in each CCM | 100 |
| The number of cascade levels | 4 |

## Implementation details

In the experiment, CT images and RNA-seq data are used as inputs to IFEM and GFEM, respectively. Deep image features of CT images are extracted by pre-trained ResNet50, which ensures that in the case of insufficient samples, low-level features can be learned quickly and high-level features can also be obtained by fine-tuning the pre-trained ResNet50 only, thus improving the convergence speed and prediction accuracy of the model. Then, after removing the missing values from the RNA-seq data, the genes are screened by the $F$-Test algorithm, and the deep gene features are extracted by the cascade network with the CCM. Finally, the obtained deep image features and deep gene features are fused by the AFN to output the prediction results.

To reduce the influence of imbalanced datasets, a mini-batch training strategy is used. When a mini-batch is created, the overlapping selections of a few samples are allowed to balance the numbers of the two classes. In the experiment, the Adam gradient optimization algorithm is used to optimize the parameters of the model; the learning rate is set to 1e–4 and the number of training epochs to 30. The batch size is set to 24, and the cross-entropy loss is used. Moreover, the performance of the model is evaluated on the testing set at each epoch.

Furthermore, the other best parameters of the DCCAFN model are shown in Table 2. In the cascade network of the DCCAFN model, four CCMs are used, and each CCM contains a convolution layer and 100 random trees.

The experiments in this work are carried out on a workstation with NVIDIA RTX A5000 GPU. Besides, all the deep learning frameworks are realized using Python 3.7.9 with Keras 2.3.1 and TensorFlow 1.15.0.

## Evaluation metrics

In order to comprehensively evaluate the prediction performance, we take the accuracy (ACC), recall, precision rate (precision), $F1$ score ($F1$), and receiver operating characteristic area under the curve (AUC) as the evaluation indicators, which are widely used in classification and prediction tasks. They are defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$Precision = \frac{TP}{TP + FP}, \tag{8}$$

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision}, \tag{9}$$

where TP is true positive, TN is true negative, FP is false positive, FN is false negative, and AUC is the area under the ROC curve.

## Ablation study

To validate the effectiveness of the DCCAFN model in predicting lung cancer survival, we conducted a series of ablation experiments, including the effect of IFEM, the effect of GFEM, and the effect of AFN.

### Effect of IFEM

To demonstrate the effectiveness of the IFEM for the DCCAFN model, the DCCAFN with IFEM and the DCCAFN without IFEM are compared. In the DCCAFN without IFEM, the deep image features obtained by the IFEM are replaced with handcrafted features extracted using 3D slicer software, including shape, first-order statistics, the Gray Level Cooccurence Matrix (glcm), the Gray Level Run Length Matrix (glrlm), and so on. In these two experiments, the relevant parameters are set according to Section "Implementation details".

Table 3 lists the prediction performance of the DCCAFN model with IFEM and without IFEM on the survival prognosis of lung cancer, respectively. In the DCCAFN without
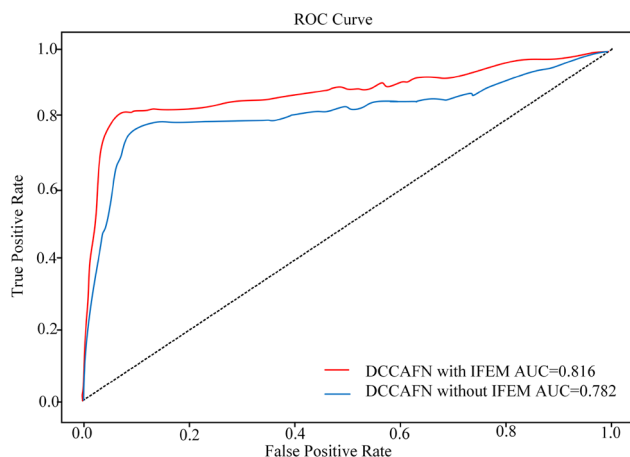
**Table 3** Prediction performance of the DCCAFN model with IFEM and without IFEM

| Method | Precision | Recall | $F1$ | ACC | AUC |
|---|---|---|---|---|---|
| DCCAFN without IFEM | 0.792 | 0.803 | 0.792 | 0.807 | 0.782 |
| DCCAFN with IFEM | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |

The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting

**Fig. 7** ROC curve for predicting lung cancer survival using the DCCAFN models with and without IFEM
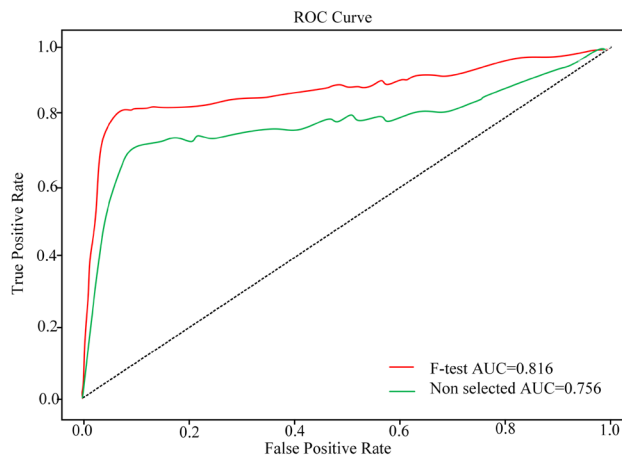


**Fig. 8** The ROC curves of models using the screened genes and the non-selected genes for the prediction of lung cancer survival

IFEM, the AUC value of the model on the testing set reaches 0.782, and the ACC is 0.807. In the DCCAFN with IFEM, the AUC value of the model on the testing set reaches 0.816, and the ACC is 0.831. From the table, the AUC value and the ACC of the DCCAFN with IFEM are improved by 0.034 and 0.024, respectively. The results show that the IFEM provides more information, and these deep image features are effective in improving the accuracy of the model.

Figure 7 shows the ROC curve for predicting lung cancer survival using the DCCAFN models with and without IFEM. The performance of the model with IFEM is better than that without IFEM in ACC, AUC, recall, precision, and $F$1 score. The results show that the deep image features obtained by the IFEM are helpful for the prediction of the model.

### Effect of GFEM

In the subsection, we consider the effect of gene feature screening, the number of convolution layers in CCM, classification algorithms in CCM, and the number of cascade levels in GFEM on the performance of the DCCAFN model in predicting lung cancer survival.

(1)   Effect of gene feature screening

In the DCCAFN model, the original genes (non-selected genes) and the screened genes using the $F$-test algorithm are used to investigate the effect of gene feature screening,

respectively. In the experiments, the deep image features are still obtained through the IFEM, and the relevant parameters are set according to Section "Implementation details".

Table 4 lists the performance of the DCCAFN models using the non-selected genes and the screened genes. In the experiments, the $F$-test algorithm selects 735 features with a $p$-value less than 0.05. The results show that the model using the screened genes by the $F$-test algorithm has improved in ACC, AUC, recall, precision, and $F$1 score compared with the non-selected genes, among which the ACC and AUC values increased by 0.067 and 0.06, respectively. This implies that $F$-test feature screening is effective, which can reduce the redundancy of gene data, reduce the interference of genes with weak correlation to the model, and improve the performance of the model in predicting lung cancer survival.

Figure 8 shows the ROC curves of models using the screened genes and the non-selected genes for the prediction of lung cancer survival. It can be seen from Fig. 8 that the performance of the model using the $F$-test algorithm outperforms the non-selected method. The results show that the $F$-test feature screening can effectively remove a large amount of irrelevant and redundant information and improve the prediction accuracy of the model.

(2)   Effect of the number of convolution layers in CCM

We consider the effect of the number of convolution layers in the CCM of the GFEM on the performance of models. Table

**Table 4** Performance of the DCCAFN models using the non-selected genes and the screened genes

| Selection method | Selected genes | Precision | Recall | $F$1 | ACC | AUC |
|---|---|---|---|---|---|---|
| Non-selected | 5268 | 0.684 | 0.669 | 0.632 | 0.764 | 0.756 |
| $F$-test | **735** | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |

The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting

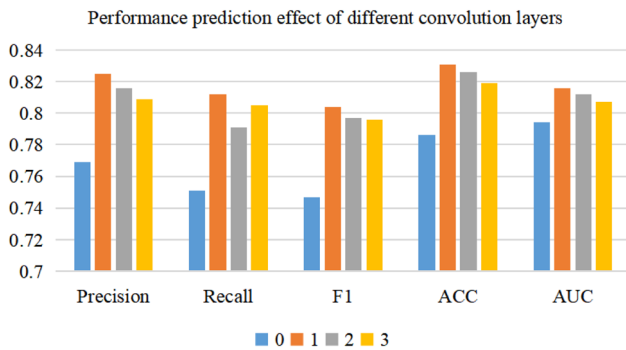**Table 5** Performance of DCCAFN models with different numbers of convolutional layers in CCM

| Convolution layers | Precision | Recall | $F1$ | ACC | AUC |
|---|---|---|---|---|---|
| 0 | 0.769 | 0.751 | 0.747 | 0.786 | 0.794 |
| 1 | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |
| 2 | 0.816 | 0.791 | 0.797 | 0.826 | 0.812 |
| 3 | 0.809 | 0.805 | 0.796 | 0.819 | 0.807 |

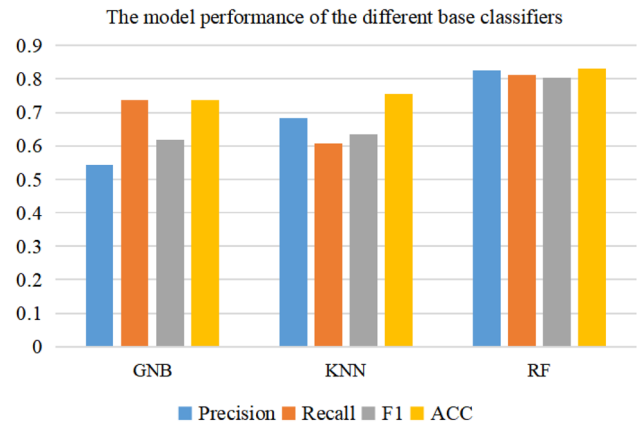The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting



**Fig. 9** Performance of DCCAFN models with different numbers of convolutional layers in CCM



**Fig. 10** Comparison of DCCAFN model performance for the different classification layer algorithms

5 and Fig. 9 show the effect of different numbers of convolutional layers in CCM on network performance. From these results, we find that the DCCAFN model with one convolution layer in each CCM has the best performance (precision = 0.825, recall = 0.812, $F1$ = 0.804, ACC = 0.831, AUC = 0.816). The results also show that the convolution layer is effective for extracting gene information to a certain extent. But the performance of the DCCAFN model does not significantly improve as the number of convolutional layers increases. Therefore, we select one convolution layer in the CCM for all experiments.

(3) Effect of classification algorithms in CCM

The effect of classification algorithms in the CCM is considered. In the experiments, three types of classification algorithms are compared, including random forest (RF), gaussian naive bayes (GNB), and K-nearest neighbor (KNN). Table 6 and Fig. 10 show the performance of the DCCAFN

models with different classification layer algorithms in CCM in predicting lung cancer survival. It can be seen from Table 6 that the performance of the model using the RF algorithm in the classification layer of the CCM to predict lung cancer survival is higher than that using other algorithms (GNB, KNN), improving by 0.093 and 0.075 in accuracy and by 0.091 and 0.037 in AUC. The results show that CCM using RF algorithm can effectively extract gene features, which is positive for improving the performance of lung cancer survival prediction.

(4) Effect of the number of cascade levels in GFEM

We study the impact of the number of cascade levels in the GFEM to obtain the optimum cascade levels. In the experiments, we use the cascade networks with 1 level, 2 levels,…7

**Table 6** Comparison of DCCAFN model performance for different classification layer algorithms

| Classification layer algorithms | Precision | Recall | F1 | ACC | AUC |
|---|---|---|---|---|---|
| GNB | 0.543 | 0.738 | 0.617 | 0.738 | 0.725 |
| KNN | 0.682 | 0.608 | 0.635 | 0.756 | 0.779 |
| RF | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |

The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting
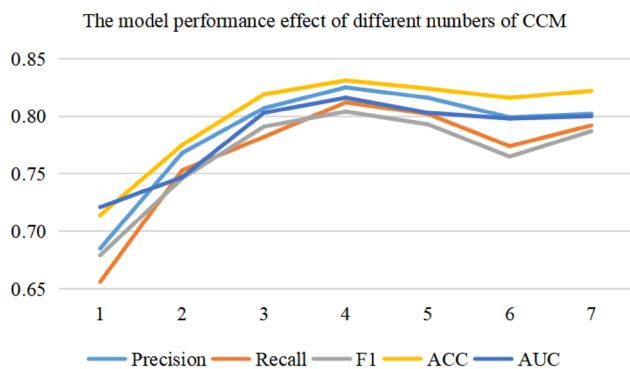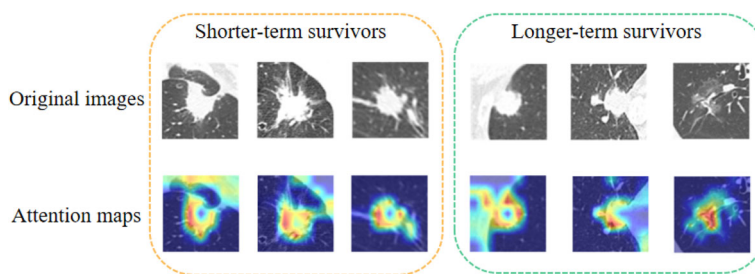
**Fig. 11** The model performance for different numbers of cascade levels in the GFEM

levels in the GFEM. Figure 11 shows the effect of different numbers of cascade levels in the GFEM on network performance. From the figure, we find that when the number of cascade levels is set to 4, the prediction effect is the best (precision = 0.825, recall = 0.812, $F1$ = 0.804, ACC = 0.831, and AUC = 0.816). The number of cascade levels is not positively correlated with the prediction performance of models. When the number of cascade levels is greater than 4, the prediction effect may become stable and will not increase or decrease significantly. Therefore, the proper number of cascade levels can improve the final prediction efficiency of the network to some extent, and the number of cascade levels is set to 4 in the other experiments.

### Effect of AFN

To verify the impact of AFN on the performance of the DCCAFN model, the DCCAFN models without AFN and with AFN are compared. In the DCCAFN model without AFN, deep image features and deep gene features are directly concatenated for lung cancer survival prediction. Table 7 lists the performance of the DCCAFN model without AFN and with AFN. From the table, we find that the use of AFN can improve the performance of the DCCAFN model in predicting lung cancer survival, increasing by 1.6% and 2.4% in the ACC and AUC values, respectively.

To demonstrate the advantages of the AFN module, we visualized the attention maps, where the attention maps are the feature maps after the fusion of deep image features and deep gene features. In Fig. 12, the first row is the original CT images, the second row is the corresponding attention maps, the left side is the images corresponding to poor survival prognosis, and the right side is the images corresponding to good survival prognosis. We observe that the attention maps after the fusion not only focus on the morphological information of the nodules themselves (such as contour and location), but also pay attention to some features of the edges. The attention maps place a greater weight on the tumor edges and their vicinity than other irrelevant information in the whole lung parenchyma, which allows the model to better learn the rich semantic information in and around the tumor features. In addition, it is also observed that when there is pleural invasion or vascular attachment (e.g., the third column of Shorter term survivors and the third column of Longer term survivors in Fig. 12), the attention maps focus well on the tumor and its surrounding features. The experimental results fully demonstrate the effectiveness of the proposed AFN module, which can greatly improve the accuracy of lung cancer survival prediction.

### Comparison with other advanced methods

In the experiments, DCCAFN is further compared with the recent deep learning-based survival prediction methods DeepSurv [13], DeepMMSA [18], GPDBN [22], and HFBSurv [23] to evaluate their performance in the survival

**Fig. 12** Visualizations of attention maps



**Table 7** Performance of the DCCAFN model without AFN and with AFN

| Fusion results | Precision | Recall | $F1$ | ACC | AUC |
|---|---|---|---|---|---|
| DCCAFN without AFN | 0.806 | 0.785 | 0.779 | 0.815 | 0.792 |
| DCCAFN with AFN | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |

The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting

**Table 8** Performance of different deep learning-based survival prediction methods

| Author | Method | Precision | Recall | F1 | ACC | AUC |
|---|---|---|---|---|---|---|
| Katzman et al. [13] | DeepSurv | 0.695 | 0.674 | 0.657 | 0.682 | 0.719 |
| Wu et al. [18] | DeepMMSA | 0.709 | 0.687 | 0.662 | 0.734 | 0.740 |
| Wang et al. [22] | GPDBN | 0.738 | 0.692 | 0.721 | 0.781 | 0.767 |
| Li et al. [23] | HFBSurv | 0.796 | 0.745 | 0.774 | 0.814 | 0.803 |
| Our research | DCCAFN | **0.825** | **0.812** | **0.804** | **0.831** | **0.816** |

The bold values indicate that the DCCAFN model achieves the best performance under the optimal parameter setting
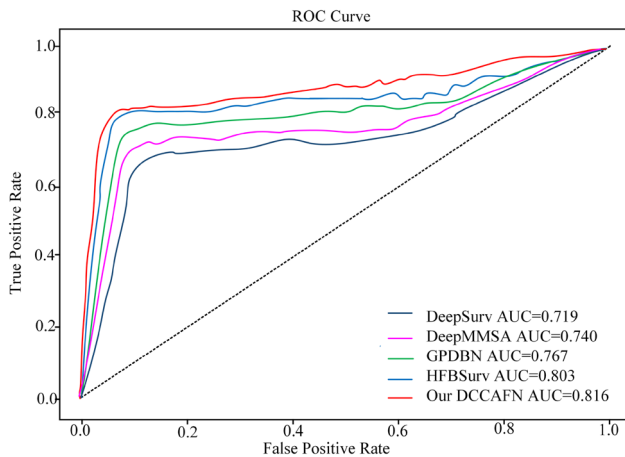


**Fig. 13** The ROC of survival prediction using state-of-the-art methods and the proposed DCCAFN method

prognosis prediction task for lung cancer. The results are listed in Table 8. In order to make a fair comparison, all the above prediction methods use the same multimodal data for performance evaluation throughout the experiment. It is clear from Table 8 that all these methods have satisfactory performance when combining multimodal information. Compared with DeepSurv, DeepMMSA, GPDBN, and HFBSurv, the DCCAFN model has the highest performance, and its AUC values are 9.7%, 7.6%, 4.9%, and 1.3% higher, respectively. These results show that our method performs an effective and specialized multimodal data fusion for survival prediction. It further shows that the proposed architecture has certain improvement abilities and that the DCCAFN is effective in predicting the survival prognosis of lung cancer.

Furthermore, Fig. 13 shows the comparison of the ROC curves for DCCAFN with other survival prediction methods. As shown in Table 8 and Fig. 13, our DCCAFN model achieves better performance than other deep learning methods such as DeepSurv, DeepMMSA, GPDBN, and HFBSurv. We observe that the proposed method significantly improves performance, and the ACC reaches 83.1%. The results show that the DCCAFN based on image genomics is of great significance for learning more useful features and improving

the accuracy of survival prognosis prediction based on CT images.

To further evaluate the performance of DCCAFN, we plot the Kaplan–Meier curves of the above methods, as shown in Fig. 14. In Fig. 14, we can observe that deep learning methods such as DeepSurv, DeepMMSA, GPDBN, HFBSurv, and the proposed DCCAFN all achieve good performance. But compared with DeepSurv ($p = 0.012$), DeepMMSA ($p = 0.0087$), GPDBN ($p = 0.0055$) and HFBSurv ($p = 0.0015$), the DCCAFN can more easily classify patients into low-risk and high-risk groups, and its stratification is significantly better ($p < 0.0001$). Moreover, it is worth noting that, in contrast, the p value of DCCAFN is the most significant, less than 0.0001, providing a more favorable prognosis prediction, which again confirms the effectiveness of this method in predicting survival. All these results clearly demonstrate the superiority of our multimodal fusion method for survival prediction.

## Discussion

The non-invasive automatic prediction of lung cancer survival is challenging due to the small size and imbalance of most lung cancer image datasets. The small sample size of image data and the redundant information in high-dimensional gene data all affect the prediction performance of the deep learning framework. In the paper, we investigate the prediction of lung cancer survival using CT images and gene data to help physicians and patients prepare for the risks that may occur. To achieve high prediction performance, DCCAFN is proposed for predicting lung cancer survival to overcome the problems of small sample size, high feature dimension, and poor multimodal feature fusion effect.

The DCCAFN model proposes a series of strategies to solve the above problems, including IFEM for deep image features based on transfer learning, GFEM based on deep convolution cascades, and AFN based on image and gene multimodal information. The IFEM uses the deep learning transfer technique to extract the deep image features of CT images that are conducive to improving the accuracy of the

**Fig. 14** Kaplan–Meier curves of the DCCAFN and other state-of-the-art methods
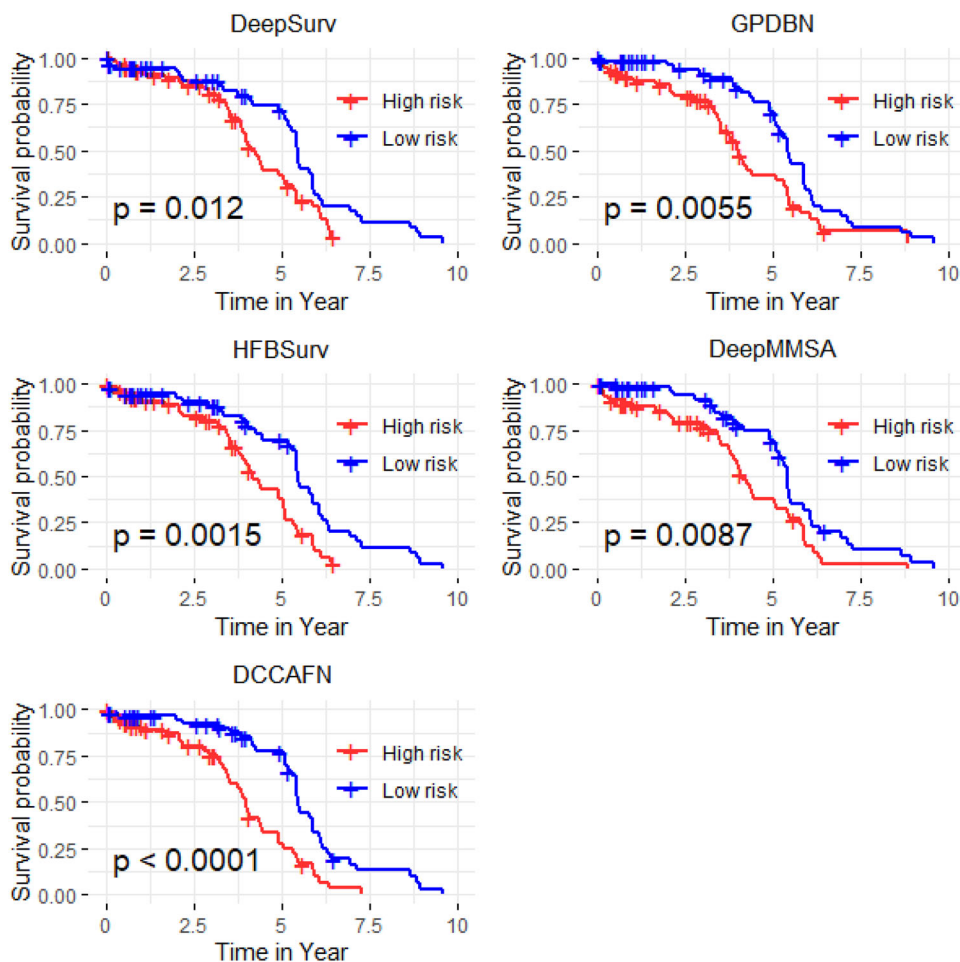


survival prediction. The GFEM, based on a deep convolution cascade, is proposed to solve the high-dimensional redundancy of gene data. In the experiments, we investigate the impact of gene feature screening, the number of convolution layers in CCM, classification algorithms in CCM, and the number of cascade levels in GFEM on the network performance, which implies that the DCCAFN model with GFEM is effective in predicting lung cancer survival prognosis. Firstly, using the F-test algorithm to select genes can eliminate some redundant information in genes and improve the accurate expression of gene features. Then, when one convolution layer and RF classification algorithm in the CCM are designed and four cascade levels are set, the cascade network with the CCM can better extract deep gene features. In addition, the proposed AFN is effective and practical for predicting the survival performance of lung cancer and can improve the prediction performance of the model to a certain extent. Therefore, the DCCAFN model can make full use of information correlation and diversity to fuse multimodal information from CT images and gene data for predicting the survival of lung cancer.

## Conclusion

In this paper, we propose a DCCAFN model based on small sample CT image data and gene data for predicting the survival of lung cancer. In the model, the deep image features and deep gene features can be better extracted through a pretrained ResNet50 model and the proposed cascade network with the CCM, respectively. Then, an attention fusion mechanism is constructed to better fuse the extracted deep image features and deep gene features. Through a series of ablation experiments and comparison experiments, we find the following conclusions:

(1) In the DCCAFN model, the IFEM can better extract deep image features from CT images; the GFEM can better extract deep gene features associated with lung cancer survival; and the AFN can better fuse deep image features and deep gene features. They both have positive effects on improving the performance of lung cancer survival prediction.

(2) Compared with the existing models, the proposed model based on image genomics is of great significance for

learning more useful features and improving the accuracy of survival prediction, with ACC and AUC values of 0.831 and 0.816, respectively.

Therefore, the proposed DCCAFN model can effectively fuse the extracted features from CT images and gene data and improve the prediction accuracy of the survival of lung cancer. Although DCCAFN enhances prediction performance, there is still considerable room for further expansion and improvement:

(1) The sample size of lung cancer in this study is relatively small, which impedes the development of a more powerful and robust survival prediction model. The follow-up studies need to collect more patient samples and actively promote large sample sizes in multicenter studies to reduce differences and improve prognosis performance.

(2) This study conducts a fusion study on CT image features and RNA gene features and does not make full use of other modalities (such as copy number variation, gene methylation data, miRNA gene expression data, etc.). More modalities can be further considered and included in future work.

In the future, the model can be extended to the fusion of different modalities for other prediction tasks and can provide clues for further cancer prognostic studies. Meanwhile, we need to continuously optimize the algorithm to meet clinical needs so as to establish models with high generalization and accuracy and realize the clinical application of artificial intelligence-assisted diagnosis.

**Data availability** The data used to support the findings of this study are included within the paper.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Siegel RL, Miller KD, Jemal A (2022) Cancer statistics, 2022. CA Cancer J Clin 72(1):7–33
2. Barta JA, Powell CA, Wisnivesky JP (2019) Global epidemiology of lung cancer. Ann Glob Health 85(1):1–16
3. Valente IRS, Cortez PC, Neto EC et al (2016) Automatic 3D pulmonary nodule detection in CT images: a survey. Comput Methods Programs Biomed 124:91–107
4. Thawani R, McLane M, Beig N et al (2018) Radiomics and radiogenomics in lung cancer: a review for the clinician. Lung Cancer 115:34–41
5. Palma JF, Das P, Liesenfeld O (2016) Lung cancer screening: utility of molecular applications in conjunction with low-dose computed tomography guidelines. Expert Rev Mol Diagn 16(4):435–447
6. Wang S, Yang DM, Rong R et al (2019) Artificial intelligence in lung cancer pathology image analysis. Cancers 11(11):1673
7. D'Antonoli TA, Farchione A, Lenkowicz J et al (2020) CT radiomics signature of tumor and peritumoral lung parenchyma to predict nonsmall cell lung cancer postsurgical recurrence risk. Acad Radiol 27(4):497–507
8. Brenner DR, Poirier A, Woods RR et al (2022) Projected estimates of cancer in Canada in 2022. CMAJ 194(17):E601–E607
9. Marentakis P, Karaiskos P, Kouloulias V et al (2021) Lung cancer histology classification from CT images based on radiomics and deep learning models. Med Biol Eng Compu 59(1):215–226
10. Gao J, Lyu T, Xiong F et al (2021) Predicting the survival of cancer patients with multimodal graph neural network. IEEE/ACM Trans Comput Biol Bioinform 19(2):699–709
11. Subramanian V, Syeda-Mahmood T, Do MN (2021) Multimodal fusion using sparse CCA for breast cancer survival prediction. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021, pp 1429–1432
12. Xiao Y, Wu J, Lin Z et al (2018) A deep learning-based multimodel ensemble method for cancer prediction. Comput Methods Programs Biomed 153:1–9
13. Katzman JL, Shaham U, Cloninger A et al (2018) DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 18(1):1–12
14. She Y, Jin Z, Wu J et al (2020) Development and validation of a deep learning model for non-small cell lung cancer survival. JAMA Netw Open 3(6):e205842–e205842
15. Astaraki M, Wang C, Buizza G et al (2019) Early survival prediction in non-small cell lung cancer from PET/CT images using an intratumor partitioning method. Phys Med 60:58–65
16. Amini M, Nazari M, Shiri I, et al (2020) Multi-Level PET and CT Fusion Radiomics-based Survival Analysis of NSCLC Patients. In: 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, pp 1–4
17. Mukherjee P, Zhou M, Lee E et al (2020) A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets. Nat Mach Intell 2(5):274–282
18. Wu Y, Ma J, Huang X, et al (2021) DeepMMSA: A novel multimodal deep learning method for non-small cell lung cancer survival analysis. In: 2021 IEEE International Conference on systems, man, and cybernetics (SMC). IEEE, 2021, pp 1468–1472
19. Lafata KJ, Corradetti MN, Gao J et al (2021) Radiogenomic analysis of locally advanced lung cancer based on CT imaging and

intratreatment changes in cell-free DNA. Radiol Imaging Cancer 3(4):e200157

20. Wang H, Subramanian V, Syeda-Mahmood T (2021) Modeling uncertainty in multi-modal fusion for lung cancer survival analysis. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021, pp 1169–1172

21. Singh A, Wang Z, Katz S, et al (2021) Development of a radiogenomic biomarker for tumor characterization and prognosis in non-small cell lung cancer patients. In: Medical Imaging 2021: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2021, 11597: 115972W

22. Wang Z, Li R, Wang M et al (2021) GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. Bioinformatics 37(18):2963–2970

23. Li R, Wu X, Li A et al (2022) HFBSurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. Bioinformatics 38(9):2587–2594

24. Chen RJ, Lu MY, Wang J et al (2022) Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans Med Imaging 41(4):757–770

25. Lu Y, Jha A, Deng R, et al (2022) Contrastive learning meets transfer learning: a case study in medical image analysis. In: Medical Imaging 2022: Computer-Aided Diagnosis. SPIE, 2022, 12033: 715–722

26. Morid MA, Borjali A, Del Fiol G (2021) A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 128:104115

27. Karimi D, Warfield SK, Gholipour A (2021) Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. Artif Intell Med 116:102078

28. Godasu R, Zeng D, Sutrave K (2020) Transfer learning in medical image classification: challenges and opportunities. Transfer 5:28

29. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118

30. Hassan M, Ali S, Alquhayz H et al (2020) Developing intelligent medical image modality classification system using deep transfer learning and LDA. Sci Rep 10(1):1–14

31. Pang M, Ting K M, Zhao P, et al (2020) Improving deep forest by screening. IEEE Trans Knowl Data Eng 34(9):4298–4312

32. Dong Y, Yang W, Wang J et al (2019) MLW-gcForest: a multi-weighted gcForest model towards the staging of lung adenocarcinoma based on multi-modal genetic data. BMC Bioinform 20(1):1–14

33. Zhou ZH, Feng J (2019) Deep forest. Natl Sci Rev 6(1):74–86

34. Ni S, Kao HY (2020) PSForest: Improving deep forest via feature pooling and error screening. In: Asian Conference on machine learning. PMLR, 2020, pp 769–781

35. Wang H, Tang Y, Jia Z et al (2020) Dense adaptive cascade forest: a self-adaptive deep ensemble for classification problems. Soft Comput 24(4):2955–2968

36. Mossa AA, Cevik U (2021) Ensemble learning of multiview CNN models for survival time prediction of braintumor patients using multimodal MRI scans. Turk J Electr Eng Comput Sci 29(2):616–631

37. Shaaban MA, Hassan YF, Guirguis SK (2022) Deep convolutional forest: a dynamic deep ensemble approach for spam detection in text. Complex Intell Syst 8:1–13

38. Li X, Xiong H, Li X et al (2022) Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. Knowl Inf Syst 64(12):3197–3234

39. Zhou B, Khosla A, Lapedriza A, et al (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on computer vision and pattern recognition, 2016, pp 2921–2929

40. Selvaraju R R, Cogswell M, Das A, et al (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on computer vision, 2017, pp 618–626

41. Katafuchi R, Tokunaga T (2021) LEA-Net: layer-wise external attention network for efficient color anomaly detection. arXiv preprint arXiv:2109.05493

42. Rosso MM, Marasco G, Aiello S et al (2023) Convolutional networks and transformers for intelligent road tunnel investigations. Comput Struct 275:106918

43. Aonpong P, Iwamoto Y, Wang W et al (2020) Hand-crafted and deep learning-based radiomics models for recurrence prediction of non-small cells lung cancers. In: Chen Y-W, Tanaka S, Howlett RJ, Jain LC (eds) Innovation in medicine and healthcare. Springer, Singapore, pp 135–144

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