DeepMMSA: A Novel Multimodal Deep Learning Framework for Non-small Cell Lung Cancer Survival Analysis

Yujiao Wu[1] , **Jie Ma**[1] , **Xiaoshui Huang**[2] , **Sai Ho Ling**[1] and **Steven Weidong Su**[1*]
[1]University of Technology Sydney
[2]University of Sydney
yujiao111@gmail.com, jie.ma-5@student.uts.edu.au, xiaoshui.huang@sydney.edu.au, {Steve.Ling, Steven.Su}@uts.edu.au

Abstract

Lung cancer is the leading cause of cancer death worldwide. The critical reason for these deaths is the delayed diagnosis and poor prognosis. With the accelerated development of deep learning techniques, it has been successfully applied extensively in many real-world applications, including health sectors such as medical image interpretation and disease diagnosis. By combining more modalities that being engaged in the processing of information, multimodal learning can extract better features and improve the predictive ability. The conventional method for lung cancer survival analysis normally utilize clinical data and only provide a statistical probability. To improve the survival prediction accuracy, and help prognostic decision-making in clinical practice for medical experts. We for the first time propose a multimodal deep learning framework for non-small cell lung cancer (NSCLC) survival analysis, named DeepMMSA, which leverage CT images in combination with clinical data, enabling the abundant information hold within medical images to be associate with lung cancer survival information. We validate our model on the data of 422 NSCLC patients from The Cancer Imaging Archive (TCIA). Experimental results verify our hypothesis that there is a underlying relation between prognostic information and radiomic images. Besides, quantitative results showing that our method could surpass the state-of-the-art methods by 4% on concordance.

1 Introduction

Lung cancer is the leading cancer killer in both men and women in the world representing 19.4%–27% of all deaths from cancer [Siegel *et al.*, 2016]. It can be broadly classified into non-small cell lung cancer (NSCLC) counting for 85% and small cell lung cancer (SCLC) counting for the remaining 15%. Lung cancer has a poor prognosis. The lung cancer five-year survival rate is lower than many other leading cancer sites, such as colorectal (64.5 %), breast (89.6 %), and prostate (98.2%) [Feuer *et al.*, 2015]. According to the TNM system, 5-years survival for IA, IB, IIA, IIB, IIIA, IIIB and IV stage disease, is about 73%, 58%, 46%, 36%, 24%, 9%, and 13% respectively [Dziedzic *et al.*, 2016]. More than half of people with lung cancer die within one year of being diagnosed [Feuer *et al.*, 2015]. Accurate assessment of disease stage and survival time of lung cancer is essential in deciding the optimal plan and timing treatment for the clinicians.

Nowadays, a large fraction interpretation of medical information is performed by medical expertise. In terms of image interpretation by human experts, a lot of diagnostic errors appear in radiology. Approximately 20 million radiology reports contain clinically significant errors each year [Brady, 2017]. This limitation is mainly due to the subjectivity, the complexity of the image, extensive variations exist across different interpreters, and fatigue [Razzak *et al.*, 2018]. Moreover, 2/3 of the world population lacks adequate access to radiology specialists, this would translate to 4.7 billion people.

Artificial intelligence is a promising tool that has shown its efficacy for diagnostic purposes[Liu *et al.*, 2019] [Zhu *et al.*, 2018] [Xing *et al.*, 2017]. With the rapid development in deep learning-based computer vision, its ability to recognize images or diagnose pictures even exceeds human ability [Russakovsky *et al.*, 2015] [Deng *et al.*, 2009] [Russakovsky *et al.*, 2013]. Convolutional neural networks (CNNs) is the most frequently used deep learning technique. In the last few years, image analysis by deep learning or CNNs has been utilized in low-dose lung CT for early diagnosis, which dramatically decreases the lung cancer mortality rate. To be specific, it enabled computer vision models to assist the doctors to detect suspicious pulmonary nodules or identify the location of the nodule, evaluate whole-lung/pulmonary malignancy, classify candidate nodules into benign or malignant, and predict the risk of lung cancer [Liao *et al.*, 2019] [Gruetzemacher *et al.*, 2018] [Trajanovski *et al.*, 2018] [Ardila *et al.*, 2019] [Zhu *et al.*, 2018] [Ding *et al.*, 2017] [Li and Fan, 2020] [Riquelme and Akhloufi, 2020]; in some cases, the models have reached competitive performance to doctors, and the accuracy even exceed the doctors [Shin *et al.*, 2012] [Esteva *et al.*, 2017] [Gulshan *et al.*, 2016] [Litjens *et al.*, 2017] [Xing *et al.*, 2017].

Although artificial intelligence in combination with CT scans is a promising tool that has shown its utility for diag-

*Contact Author

nostic purposes, it is rarely been used in predicting death, and possibly even determining death, which is a unique and challenging area that could be fraught with the same biases that affect analog physician-patient interactions. Current research for NSCLC survival analysis is largely based on the statistical analysis of clinical data. Traditional approaches generally utilize clinical information such as age, clinical TNM stage, gender information, etc. In the work of [Wanget al., 2019], which use CT images in survival analysis. But the model only learns 2D features from each tumor image slices separately, afterward, averaged features from all image slices for every single patient and dismissed the 3D properties of the tumor. Moreover, the overall performance of currently available works for survival analysis is considerably low. While deep learning To our best knowledge, this is the first time to develop a system based on 3D features extracted from lung CT images fusion with features extracted from clinical data for death prediction using deep learning.

By contrast, deep learning has the potential to reduce diagnostic errors and overcome the human limitations. Thus, it is worth developing a fully automated deep learning system for NSCLC survival prediction based on CT images and clinical data to explore whether the deep learning techniques have the ability to extract useful information from CT images and clinical to predict death. In our work, considering the 3D nature of CT images, in order to reveal the underlying relation between prognostic information and CT images, to fully utilize the potential of the prognostic power existing in the radiomic data, we design a 3D multimodal Resnet framework for NSCLC survival analysis. Since combining CT image and clinical data could provide comprehensive and supplementary information to describe the cancer status, this framework could indicate more accuracy overall trend of survival. To our best knowledge, this is the first time to develop a system based on 3D features extracted from lung CT images for death prediction using deep learning. Quantitative results on the NSCLC- Radiomics data show that the proposed method could surpass the state-of-the-art methods by 4% on concordance, revealing that our method could provide more accurate diagnosis and prognostic decision-making in the future clinical practice. The results of ablation experiments show that using multiple modalities improves the biggest marginal performance compared with using single modality.

To conclude, we mainly have the following contributions:

- The proposed method overcome the weakness of traditional non-parametric methods (KM etc.) which cannot incorporate multiple variables.
- Experiment results shows that our method Could surpass the SOTA methods (Cox-Time etc.) by 4% on concordance.
- To our best knowledge, this is the first attempt to reconstruct a deep 3D convolutional-based model and using images for survival analysis.
- Experiment results verify our hypothesis and reveal the underlying relation between prognostic information and radiomic images.
- Our multimodal framework can provide more accurate survival analysis with sufficient granularity for personalized prognosis and decision-making compared to SOTA methods.

2 Preliminary Knowledge

In this section, we provide a brief survey on multimodal deep learning, 3D ResNets, survival analysis, and survival analysis methods developed in recent years.

2.1 Notations

For notational clarity, we hereby define the key symbols and their meanings in Table 1.

Symbol	Definition
r_i	i-th 3D radiology image
c_i	i-th clinical information
y_i	i-th actucal survival time
e_i	i-th event, 1 for uncensored, 0 for censored
$\hat{y_i}$	predicted i-th survival time
$I(x)$	1 if x=True else 0

Table 1: Notations

2.2 Multimodal deep learning

Multimodal deep learning is a novel framework of deep neural networks to learn features over multiple modalities (e.g., text, images, or audio) [Ngiam *et al.*, 2011]. In medical applications, multiple types of data are related to each patient, including clinical information, radiology images, physician note, medication, to name a few. Thus, when data comes from different sources the approach of multimodal deep learning can help to understand and extract more useful information.

2.3 3D-ResNet

The residual neural network (ResNet) [He *et al.*, 2016] is to handle gradient vanishing or exploding problems in deeper neural networks training, especially in computer vision. The core concept of ResNet is to construct a basic network block in which the output is add up with input. To handle 3D image input, we can simply increase kernel shape from 2 dimensions to 3 dimensions (height, length, and depth) in convolution layers.

2.4 Survival analysis

Survival analysis is the widely used technique to model time-to-event data (e.g., failure, death, admission to hospital, the emergence of disease, etc.) [Lee and Wang, 2003]. Traditional statistical methods for survival analysis, normally contains three options for modeling the survival function: non-parametric methods with no distribution of survival curve predefined (e.g. Kaplan-Meier [Goel *et al.*, 2010], Nelson-Aalen [Nelson, 1972] [Aalen, 1978]), semi-parametric methods such as the Cox proportional hazards model [Cox, 1972] which is most commonly used, and parametric methods with distribution predefined (e.g. Linear regression, Weibull distribution). Besides, machine learning methods such as survival trees, Neural network, Cox-time [Kvamme *et al.*,

2019]), DeepHit [Lee *et al.*, 2018], CoxCC [Kvamme *et al.*, 2019], PC-Hazard [Kvamme and Borgan, 2019] and Ensemble method, to name just a few, are also applied in survival analysis.

Due to the existence of censored survival data (usually right censored), the standard evaluation indexes for regression, such as mean square error (MSE) and R^2, do not fit for quantifying the performance of survival analysis. The most important evaluation index is the concordance index (C-index) which can evaluate uncensored instances and censored instances together.

$$\text{C-index} = \frac{\sum_{i,j} I(\hat{y_i} < \hat{y_j} | e_i = 1, y_i < y_j)}{\sum_{i,j} I(y_i < y_j | e_i = 1)} \quad (1)$$

Besides, as a regression problem, we also use the mean absolute error (MAE) over uncensored instances to evaluate our experiments.

$$\text{MAE} = \frac{1}{\sum_i I(e_i = 1)} \sum_{i=1}^{N} (e_i | y_i - \hat{y_i}|) \quad (2)$$

2.5 Related work

Conventional survival analysis for NSCLC is set of modelling procedures which only harness clinical data and it measures time to an event. For example, in the specific area of NSCLC survival analysis, [Janssen-Heijnen *et al.*, 1998] studied the variation in the prognosis for adult patients with lung cancer within Europe, by age, histology, and country from 1985–1989 with simple statistical methods, such as life-table. The work of [Port *et al.*, 2003] shows that tumor size within stage IA is an important predictor of survival with Kaplan-Meier survival analysis and Cox proportional hazards regression model. [Morita *et al.*, 2009] examined the survival impact of a specific feature, gefitinib, in patients with EGFR mutation-positive NSCLC. [Gyorffy *et al.*, 2013] developed an online survival analysis tool capable of uni- and multivariate analysis (e.g. Kaplan-Meier survival plot, Cox regression analysis) with 1,715 samples of ten independent datasets. Recently, several work has been proposed to use deep learning in survival analysis. DeepConvSurv [Zhu *et al.*, 2016], which for the first time developed a 2D deep convolutional neural network (CNN) for survival analysis with pathological images. The study of [Chaddad *et al.*, 2017] use the random forest model to analyze the manually extracted features from radiology images combined with age to predict a binary classification of survival time. [Wang *et al.*, 2019] propose CNN to extract 2D features from CT images in survival analysis. In their work, features are learned from each 2D tumor image slice, then features from all slices from one patient are directly added together and calculate the average value. This method may lost the spatial and temporal information exists in tumor. [Cui *et al.*, 2020] used a deep neural network to learn cellular features from biomarkers and Cox proportional hazards model to do survival analysis. DeepLung [Zhu *et al.*, 2018] use 3D CT images and 3D CNN for nodule detection and classification, the success of which inspired us to extract 3D features from CT images for survival analysis.

In conclusion, the early stage research for NSCLC survival analysis tends to find specific features to predict the survival curve, most of which using Kaplan-Meier survival analysis and Cox proportional hazards regression model. In this paper, we proposed more complex models, such as machine learning and deep learning models. Besides, to fully utilize the information gathered from all types of data sources, such as CT images and clinical information, we design a fully automated multimodal deep learning framework for individualized NSCLC survival prediction.

3 Methodology

In this section, we describe details about the proposed DeepMMSA. As far as we know, this is the first work to use a multimodal deep learning framework to process CT images together with clinical information for NSCLC survival analysis.

3.1 The structure of DeepMMSA

Inspired by the recent successful applications of CNNs in NSCLC diagnosis and other image recognition tasks, we proposed a DeepMMSA that harness the strategies of analyzing the combined information from multiple modalities. As shown in Figure 1, we use the tumor region of interests(ROIs) which is the lesions on CT scans as low-level image input. Motivated by the work of [Zhu *et al.*, 2018], 3D CNNs was used to extract features from all three-dimensional directions within the tumor volume. Meanwhile, we integrate the clinical information through a 27D high-level clinical layer, the input of which including screening test results such as clinical TNM stage, overall stage, histology, gender, age of the patient. Clinical layer is embedded in the hidden layer directly.

DeepMMSA framework consists of three modules:

- Multimodal feature (CT image features and Clinical record features) extraction.
- Multimodal feature fusion.
- Survival analysis.

3.2 Multimodal feature extraction

CT images feature extraction with 3D-ResNet

As is shown in Figure.1 our framework requires two multimodal inputs, CT images, and clinical data from the according patient. To be specific, CT images are introduced to the radiomics embedding layer and clinical data are introduced to the clinical embedding layer. We propose 3D-ResNets as our network structure for the low-level image feature learning. As shown in Figure 1 and Figure 2, 3D-ResNets can be built by basic blocks or "bottleneck" building blocks, the number of which may vary from 18 to 152 in the whole network. For instance, 3D-ResNet-18 contains 4 basic blocks, each block contains 4 convolutional layers (conv1-conv4). Different from 2D ResNet, features extracted from 3D ResNet are calculated in all three-dimensional directions within the tumour volume, thereby taking the spatial location of each voxel compared with the surrounding voxels into account[Aerts *et al.*, 2014]. It worth noting that, in order to eliminate the vanishing and exploding gradients types of problems in very deep neural network, we add extra shortcut

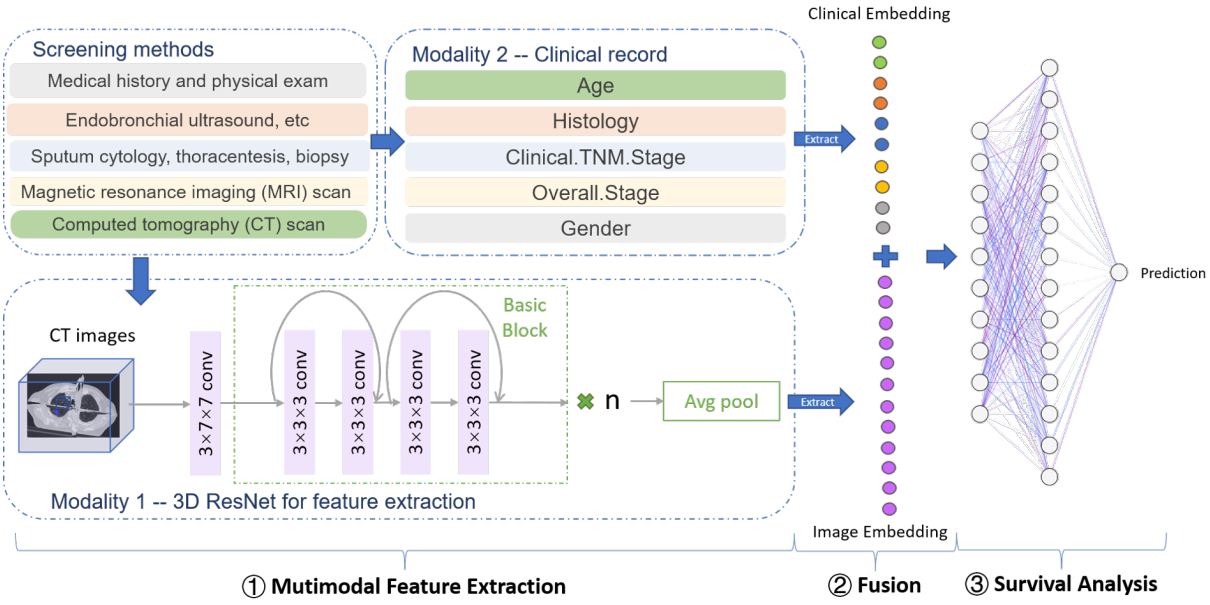

Figure 1: The framework of deepMMSA. DeepMMSA mainly has three module: (1)First it employs the 3D-ResNet in combination with plain networks for multimodal feature extraction; (2) Then it uses simple feature fusion method (early fusion) for multimodal fusion; (3) Lastly, during desicion making stage, plain neural network is designed for the survival prediction.

connections in our model. The advantage of such residual learning is that it enables the reuse of features. Since our dataset is relatively small compared to other general datasets for image recognition, we use data argumentation techniques before importing the data into the model.

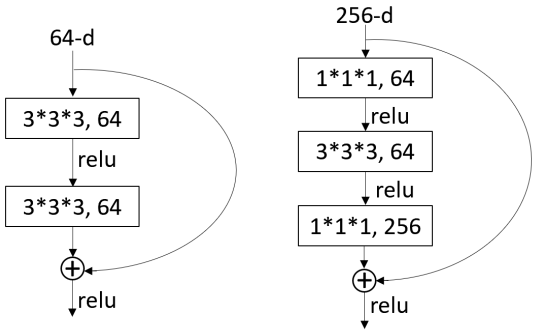

Figure 2: A deeper reseidual function **F** for 3D-ResNet. Left: a building block for 3D-ResNet-18/34. Right: a "bottleneck" building block for 3D-ResNet-50/101/152.

The residual building block can be formulated as:

$$y = \mathbf{F}(x) + x \tag{3}$$

where **F** is the deeper residual function, x is the input, and y is the output.

Clinical record feature extraction network /model

The clinical embedding layers introduce clinical data to the network separately aims to capture the survival information indicated within clinical data. For clinical data, a neural network with two hidden layers was proposed to extract features. As is shown in Figure.1, the 27D non-image features were extracted by a network, which in combination with the image features is processed in later fusion stage.

3.3 Multimodal feature fusion

During feature fusion stage, multimodal features from CT images and clinical records are difficult to be fused together directly. This is due to that the features from different modalities have different scales or statistical properties. Thus, to solve this problem, we applied the Batch Normalization (BN) technique to adjust the mean and variance of extracted features in each modality before fusion procedure. Given the features $z_1, z_2, ..., z_m$ over a batch, $\hat{z}_i$ is calculated as:

$$\hat{z}_i = \gamma_i \frac{z_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} + \beta_i \tag{4}$$

where γ_i, β_i are the parameters to be learned, μ_i is the mean value of z_i over the batch, σ_i is the standard deviation of z_i over the batch, ϵ is set to a very small number such as 10^{-8}.

3.4 Survival analysis network/ model

Survival analysis network/model is the last module in the framework and is intended as a overall feature analyze to get the final survival time prediction. In this module, to meet the requirement for specific problem, any survival analysis model can be used.It can be models from traditional statistical methods such as Kaplan-Meier survival analysis, Cox proportional hazards regression model. Either from novel machine learning methods, such as survival trees, and deep neural network. All the mentioned models can be used to analyze the input from multimodal features in our proposed fusion layer. In this work, we define the survival time as the label, and use

the one hidden layer neural network with one dimension output layer for the regression setting and overall optimization convenience. The only difference is that we normalize the model output with Sigmoid function to be on the same scale with the normalized ground truth,

$$\hat{y_i} = \frac{1}{1 + e^{-f(x_i)}} \tag{5}$$

and we use MSE loss function and L2 regularization penalty term as the objective function which is defined as:

$$\text{minimize } L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2 + \lambda \sum_{j=1}^{M} w_j \tag{6}$$

where w_j is the model parameter and the total number is M. λ is the penalty coefficient.

4 Experiments

We conduct extensive experiments based on NSCLC patients from TCIA to validate the performance of our proposed method DeepMMSA with several state-of-the-art methods in terms of the prediction accuracy for the survival time for each patient. Besides, we also evaluate the prediction result by concordance. Afterwards, we perform several ablation experiments regarding different network structures to determine the best structure.

4.1 Dataset

In this work, we considered 422 NSCLC patients from TCIA to assess the proposed framework. For these patients pretreatment CT scans, manual delineation by a radiation oncologist of the 3D volume of the gross tumor volume and clinical outcome data are available [Clark *et al.*, 2013]. The corresponding clinical data are also available in the same collection. The patients who had neither survival time nor event status were excluded from this work.

4.2 Data preprocessing

For CT images, we resize the raw data which is the 3D volume of the primary gross tumor volume into $96 * 96 * 8$. After that, we transform the range linearity into [0,1]. Then, to prevent overfitting problem, we perform data argumentation which includes three methods: rotate, swap, and flip. Then we get $422 * 8 = 3376$ samples, among which there are $373 * 8 = 2984$ uncensored samples and $49 * 8 = 392$ censored samples.

Clinical data contains categorical data and non-categorical data. Firstly, missing value is a common problem in medical data and may pose difficulties for data analysing and modelling. Specifically, in our dataset, the 'age' category contains a few missing values. After observing the data, we find that the age of patients in the dataset is close to each other. Thus, we impute the mean value and fill it into the missing value. Afterwards, In order to fit into our model, we use the one-hot encoder to encode categorical data into numbers, which allows the representation of categorical data to be more expressive.

Then, We use the min-max feature scaling method and standard score method to perform data normalization, such as age and survival time. For input x, the min-max feature scaling method's output is:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{7}$$

and the standard score method's output is:

$$x' = \frac{x - \mu}{\sigma} \tag{8}$$

where μ is the mean of x, and σ is the standard deviation of x.

For single patient with multiple tumors, we select the primary gross tumor volume ("GTV-1") to be processed in our work, while other tumors such as secondary tumor volumes denoted as "GTV2", "GTV3" to name just a few, which were occasionally present, were not considered in our work.

4.3 Experiment setup

We train and evaluate the framework on the NSCLC-Radiomic dataset following 5-fold cross-validation with the patient-level split. We divided the dataset into training, validation and testing data into 6:2:2 respectively. For hyperparameters tuning such as the penalty coefficient, we use the validation dataset to fine-tune and get the optimized hyperparameters. In training process, we use 200 epochs in total with Adam as optimizer. The batch size parameter is set as 64. The initial learning rate is set as 0.001, then the learning rate is decayed by 0.5 in every 40 epochs.

Since we use the survival time as label, not cumulative hazard. In the training and validation process, we only use the uncensored data for precised survival time and objective function calculation, and in the testing process, we use all data for concordance evaluation and uncensored data for MAE evaluation.

Since this is the first work to use multimodal framework for NSCLC survival analysis, we implement several state of the art survival analysis methods as baselines to compared with our work. The baseline methods including Cox-time [Kvamme *et al.*, 2019]), DeepHit [Lee *et al.*, 2018], CoxCC [Kvamme *et al.*, 2019], PC-Hazard [Kvamme and Borgan, 2019] and the regular cox regression.

4.4 Ablation study

To find a optimal network for our problem, we consider to perform ablation experiments based on the following four aspects of network architecture:

- How the depth of Resnet effect the performance? Which 3D structure is the best?
- Whether multiple modalities outperform single modality?
- What the best ratio set between image data and clinical data in in fusion stage?
- Whether hidden layer should add in survival analysis stage?

Firstly, the evaluation of different depth of Resnet and whether multiple modalities is better than single modality is

	CT images		Multi-modality	
	Loss	C-index	Loss	C-index
r3d18	0.1023	0.5782	0.0847	0.6287
r3d34	0.0975	0.5942	**0.0757**	**0.6490**
r3d50	0.1026	0.5804	0.0760	0.6375
r3d101	0.1071	0.5660	0.0795	0.6142

Table 2: To evaluate the performance of different ResNet structure and effect of whether using multiple modalities.

	Hidden		No hidden	
	Loss	C-index	Loss	C-index
512:27	0.0757	0.6490	0.0760	0.6376
100:27	0.0745	0.6512	0.0755	0.6421
25:27	**0.0739**	**0.6580**	0.0761	0.6450
5:27	0.0765	0.6403	0.0793	0.6215

Table 3: To evaluate the effects of different ratio between modalities features in fusion procedure and the performance of survival analysis neural network with or without hidden layer.

conducted. By fixing the ratio of multiples modalities features in fusion stage equals 512:27 with one hidden layer, four different structures was tested. Table. 2 shows that, our model of r3d34 structure with multiple modalities achieves the best performance. Through experiments, the results validate our assumption and show the effectiveness of using multiple modalities in survival analysis.

Moreover, using the best structure r3d34, we step further to observe the effects of the changing the ratio between modalities with and without hidden layer. It can be done by changing the number of perceptrons for image and non-image features. The comparison results list in Table 3 shows that setting the ratio between images and non-image modalities as 25:27 achieve the best performance. Besides, adding hidden layers in the survival analysis module can further improve the performance. After ablation experiments, the best framework architecture of using r3d34 structure, setting the ratio between multiple modalities as 25:27 with one hidden layer outperform the other structures in this work.

The results indicate that NSCLC prognosis information can be learned from CT images. Using multimodal structure, more information can be jointly extracted and learned from various information sources and non-linearly transformed in to a deep network therefore improve the overall prediction accuracy. Furthermore, using multiple modalities improves the biggest marginal performance compared with using single modalities.

4.5 Results

The loss in the training and testing process is shown in Figure 3.

For a fair comparison, we perform five state-of-the-art methods as baseline methods from previous work for survival analysis. As shown in Table 4, we compare our framework and with other five baseline models. The results show that our framework achieves the best performance. Besides, the results verify that compared to the methods(Cox-time, DeepHit, etc.) which simply use clinical data, DeepMMSA shows

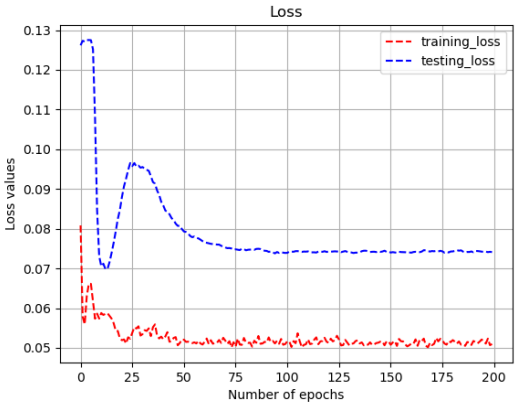

Figure 3: Training and testing process.

Model	MAE	C-index
Cox-time	0.183	0.6152
Cox regression	0.204	0.6009
CoxCC	0.183	0.6120
PC-Hazard	0.191	0.6094
DeepHit	0.183	0.6133
DeepMMSA	**0.162**	**0.6580**

Table 4: Result vs baselines

its superior of effectively extracting the supplementary information from multiple modalities, and can significantly improve the prediction result.

5 Conclusion and Future Work

In this paper, we proposed an fully automated end-to-end multimodal deep network framework for NSCLC survival analysis. Our framework can learn complementary representations from the CT image and non-image clinical data modalities. Extensive experimental result shows that DeepMMSA outperform conventional methods using single source of information alone. But there is still some future work to do. There are some potential ways to improve the performance of the proposed framework. Since there are three basic modules for multimodal deep learning survival analysis framework, we consider to made improvements based on following three aspects:

- Provide more complementary information by adding more modalities to improve the performance, such as e-nose diagnosis time series data, etc, and try to fully exploit the inherent correlations across multiple modalities.
- Perform different multimodal fusion approaches, such as decision fusion and hybrid fusion method, ect.
- In the survival analysis module, theoretically, we can use any survival analysis model, such as cox-time, deepsurv, to improve performance of our framework.

References

[Aalen, 1978] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[Aerts *et al.*, 2014] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5(1):1–9, 2014.

[Ardila *et al.*, 2019] Diego Ardila, Atilla P Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961, 2019.

[Brady, 2017] Adrian P Brady. Error and discrepancy in radiology: inevitable or avoidable? *Insights into imaging*, 8(1):171–182, 2017.

[Chaddad *et al.*, 2017] Ahmad Chaddad, Christian Desrosiers, Matthew Toews, and Bassam Abdulkarim. Predicting survival time of lung cancer patients using radiomic analysis. *Oncotarget*, 8(61):104393, 2017.

[Clark *et al.*, 2013] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

[Cox, 1972] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[Cui *et al.*, 2020] Lei Cui, Hansheng Li, Wenli Hui, Sitong Chen, Lin Yang, Yuxin Kang, Qirong Bo, and Jun Feng. A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. *BMC bioinformatics*, 21(1):1–14, 2020.

[Deng *et al.*, 2009] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[Ding *et al.*, 2017] Jia Ding, Aoxue Li, Zhiqiang Hu, and Liwei Wang. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 559–567. Springer, 2017.

[Dziedzic *et al.*, 2016] Dariusz Adam Dziedzic, Piotr Rudzinski, Renata Langfort, Tadeusz Orlowski, Polish Lung Cancer Study Group, et al. Risk factors for local and distant recurrence after surgical treatment in patients with non–small-cell lung cancer. *Clinical lung cancer*, 17(5):e157–e167, 2016.

[Esteva *et al.*, 2017] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[Feuer *et al.*, 2015] EJ Feuer, N Howlader, AM Noone, M Krapcho, J Garshell, D Miller, et al. National cancer institute seer cancer statistics review. *National Cancer Institute*, 103(7):1975–2012, 2015.

[Goel *et al.*, 2010] Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore. Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, 1(4):274, 2010.

[Gruetzemacher *et al.*, 2018] Ross Gruetzemacher, Ashish Gupta, and David Paradice. 3d deep learning for detecting pulmonary nodules in ct scans. *Journal of the American Medical Informatics Association*, 25(10):1301–1310, 2018.

[Gulshan *et al.*, 2016] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[Gyorffy *et al.*, 2013] Balazs Gyorffy, Pawel Surowiak, Jan Budczies, and Andras Lanczky. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PloS one*, 8(12):e82241–e82241, 2013.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Janssen-Heijnen *et al.*, 1998] MLG Janssen-Heijnen, G Gatta, D Forman, R Capocaccia, JWW Coebergh, EUROCARE Working Group, et al. Variation in survival of patients with lung cancer in europe, 1985–1989. *European Journal of Cancer*, 34(14):2191–2196, 1998.

[Kvamme and Borgan, 2019] Håvard Kvamme and Ørnulf Borgan. Continuous and discrete-time survival prediction with neural networks. *arXiv preprint arXiv:1910.06724*, 2019.

[Kvamme *et al.*, 2019] Håvard Kvamme, Ørnulf Borgan, and Ida Scheel. Time-to-event prediction with neural networks and cox regression. *Journal of machine learning research*, 20(129):1–30, 2019.

[Lee and Wang, 2003] Elisa T Lee and John Wang. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.

[Lee *et al.*, 2018] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Li and Fan, 2020] Yuemeng Li and Yang Fan. Deepseed: 3d squeeze-and-excitation encoder-decoder convolutional

neural networks for pulmonary nodule detection. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1866–1869. IEEE, 2020.

[Liao *et al.*, 2019] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song. Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. *IEEE transactions on neural networks and learning systems*, 30(11):3484–3495, 2019.

[Litjens *et al.*, 2017] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[Liu *et al.*, 2019] Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health*, 1(6):e271–e297, 2019.

[Morita *et al.*, 2009] Satoshi Morita, Isamu Okamoto, Kunihiko Kobayashi, Koichi Yamazaki, Hajime Asahina, Akira Inoue, Koichi Hagiwara, Noriaki Sunaga, Noriko Yanagitani, Toyoaki Hida, et al. Combined survival analysis of prospective clinical trials of gefitinib for non–small cell lung cancer with egfr mutations. *Clinical Cancer Research*, 15(13):4493–4498, 2009.

[Nelson, 1972] Wayne Nelson. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966, 1972.

[Ngiam *et al.*, 2011] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *ICML*, 2011.

[Port *et al.*, 2003] Jeffrey L Port, Michael S Kent, Robert J Korst, Daniel Libby, Mark Pasmantier, and Nasser K Altorki. Tumor size predicts survival within stage ia non-small cell lung cancer. *Chest*, 124(5):1828–1833, 2003.

[Razzak *et al.*, 2018] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[Riquelme and Akhloufi, 2020] Diego Riquelme and Moulay A Akhloufi. Deep learning for lung cancer nodules detection and classification in ct scans. *AI*, 1(1):28–67, 2020.

[Russakovsky *et al.*, 2013] Olga Russakovsky, Jia Deng, Zhiheng Huang, Alexander C. Berg, and Li Fei-Fei. Detecting avocados to zucchinis: what have we done, and where are we going? In *International Conference on Computer Vision (ICCV)*, 2013.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[Shin *et al.*, 2012] Hoo-Chang Shin, Matthew R Orton, David J Collins, Simon J Doran, and Martin O Leach. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1930–1943, 2012.

[Siegel *et al.*, 2016] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2016. *CA: a cancer journal for clinicians*, 66(1):7–30, 2016.

[Trajanovski *et al.*, 2018] Stojan Trajanovski, Dimitrios Mavroeidis, Christine Leon Swisher, Binyam Gebrekidan Gebre, Bastiaan S Veeling, Rafael Wiemker, Tobias Klinder, Amir Tahmasebi, Shawn M Regis, Christoph Wald, et al. Towards radiologist-level cancer risk assessment in ct lung screening using deep learning. *arXiv preprint arXiv:1804.01901*, 2018.

[Wang *et al.*, 2019] Shuo Wang, Zhenyu Liu, Yu Rong, Bin Zhou, Yan Bai, Wei Wei, Meiyun Wang, Yingkun Guo, and Jie Tian. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiotherapy and Oncology*, 132:171–177, 2019.

[Xing *et al.*, 2017] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):4550–4568, 2017.

[Zhu *et al.*, 2016] Xinliang Zhu, Jiawen Yao, and Junzhou Huang. Deep convolutional neural network for survival analysis with pathological images. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 544–547. IEEE, 2016.

[Zhu *et al.*, 2018] Wentao Zhu, Chaochun Liu, Wei Fan, and Xiaohui Xie. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 673–681. IEEE, 2018.