

Survival Prediction in Lung Cancer through Multi-Modal Representation Learning

Aiman Farooq

Indian Institute of Technology Jodhpur
NH 62, Nagaur Road, Jodhpur
farooq.1@iitj.ac.in

Deepak Mishra

Indian Institute of Technology Jodhpur
NH 62, Nagaur Road, Jodhpur
dmishra@iitj.ac.in

Santanu Chaudhury

Indian Institute of Technology Delhi
Hauz Khas, New Delhi
schaudhury@gmail.com

Abstract

Survival prediction is a crucial task associated with cancer diagnosis and treatment planning. This paper presents a novel approach to survival prediction by harnessing comprehensive information from CT and PET scans, along with associated Genomic data. Current methods rely on either a single modality or the integration of multiple modalities for prediction without adequately addressing associations across patients or modalities. We aim to develop a robust predictive model for survival outcomes by integrating multi-modal imaging data with genetic information while accounting for associations across patients and modalities. We learn representations for each modality via a self-supervised module and harness the semantic similarities across the patients to ensure the embeddings are aligned closely. However, optimizing solely for global relevance is inadequate, as many pairs sharing similar high-level semantics, such as tumor type, are inadvertently pushed apart in the embedding space. To address this issue, we use a cross-patient module (CPM) designed to harness inter-subject correspondences. The CPM module aims to bring together embeddings from patients with similar disease characteristics. Our experimental evaluation of the dataset of Non-Small Cell Lung Cancer (NSCLC) patients demonstrates the effectiveness of our approach in predicting survival outcomes, outperforming state-of-the-art methods.

1. Introduction

Survival prediction is critical to medical care for various diseases, including cardiovascular, chronic respiratory,

neurological conditions, and multiple cancers. Non-small cell lung cancer (NSCLC), which constitutes 85% of lung cancer cases [36], exemplifies the importance of this task. About 25% of lung cancer patients experience relapse post-surgery, making alternative treatment strategies essential when surgical options are not viable. Therefore, accurate survival prediction is vital for patient classification and assisting doctors in making informed treatment decisions. Recently, the prediction of survival risk has become increasingly significant. It supports treatment planning, patient staging, and monitoring, improving cancer patients' care. In the era of precision medicine, data-driven approaches and deep learning models empower healthcare professionals to make precise future outcome predictions through survival models. However, survival prediction is complex and influenced by various factors such as disease physiology, clinical data, genomics, and treatment courses. Enhanced predictions require the availability of multiple data modalities, yet datasets often lack one or more of these modalities, posing a challenge. Despite this, the available modalities still contain valuable information, making it crucial to address missing data to maximize the utility of limited data in survival prediction models.

Survival prediction for NSCLC involves CT, PET, and molecular data. Features derived from CT scans [7, 8, 53] have shown high relevance with the survival prognosis of lung cancer. While PET scans [26, 38] and genomics data [46, 54] have also been linked to survival prediction studies. Traditionally, survival prediction models like random survival forest [15], XGBoost [18], and Cox proportional hazards model [31] rely on deriving features from imaging modalities or clinical data. With deep learning gaining popularity over the years, several Convolutional Neu-

ral Network (CNN) based models [12] have been developed for prediction, surpassing the handcrafted approaches. For instance, Huang et al. [17] extracted features using CNN from FDDG-PET scans and constructed random forest (RF) for the prediction. For survival prediction, Lian et al. [28] used a graph convolutional network. In particular, models such as DeepSurv [22, 42] and MVAESA [49] have shown promising results. However, using only a single modality doesn't allow us to capture all the biomarkers associated with the tumor. While image biomarkers provide the advantages of characterizing heterogeneous cancer in its entirety, genomic data offers insights into the underlying molecular mechanisms driving cancer progression.

Several multi-modal models are also proposed to capture and leverage complementary contextual cues across diverse modalities. Works such as [16, 24, 39] proposed incorporating imaging and clinical data for improved survival prediction in lung cancer patients. MultiSurv [47] was the first model to combine information from multiple domains, including clinical, genomic, and whole-slide images (WSI), for a pan-cancer dataset survival prediction. To circumvent the reliance on labels, [6, 11] used similarity metrics to guide the fusion of multi-modal representations and showed improved performance.

Despite their respective advantages, most of the methods mentioned above encounter limitations due to their dependence on labeled data and lack of feature generalizability, especially in tasks involving multi-modal data. Most multi-modal methods discussed above target WSI images for prediction, with little attention to common imaging modalities such as PET and CT. Methods handling limited data labels don't consider tumor characteristics and across-patient correspondences. We propose to circumvent the label requirement and use self-supervised learning to acquire embeddings for the patient modalities, along with feature-based correspondence via contrastive learning, aiming to bolster the generalizability of visual representations. Our framework exploits inter-subject correspondence across CT, PET, and RNA-Seq data, enforced through cross-modal cluster alignment without reliance on labels, thus excelling in the specified downstream task of survival prediction. It also exhibits robust generalization and maintains predictive accuracy when RNA-Seq data is unavailable for some patients. In summary, our contribution includes :

- We propose to utilize CT, PET, and molecular (RNA-Seq) information to improve survival prediction in NSCLC patients
- We rely not only on cross-modality relationships for a patient but also on relationships across patients based on disease similarity.
- We also deal with cases where certain modalities of a patient are missing and still provide adequate results.

- We present experimental results showing that our Multi-Modal patient embedding module and cross-patient embedding module bring significant performance gains on the NSCLC dataset.

2. Related Work

2.1. Survival Prediction

Several studies have proposed models to improve survival prediction for cancer patients. Traditionally, Cox-based survival models [43], such as the linear proportional hazards model, were used across cancers; however, they need substantial feature engineering or prior medical expertise to model treatment interactions at an individual level effectively. On the other hand, nonlinear survival methods like neural networks and survival forests can naturally capture these complex interaction terms. Still, they have not yet demonstrated effectiveness as treatment recommender systems. DeepConvSurv [55] proposed using CNNs for survival analysis; they used radiological scans alongside clinical data to improve predicted outcomes. Lee et al. [25] proposed DeepHit [25] learns the distribution of survival times without assuming an underlying stochastic process, allowing for time-varying relationships between covariates and risks. It outperformed all baselines for breast cancer survival prediction. Katzman et al. proposed a deep learning-based survival model DeepSurv [22] initially proposed for breast cancer. It outperformed state-of-the-art for nasopharyngeal [29], lung cancer [23] as well as for glioblastoma [34]. It extends the Cox proportional hazards model and utilizes neural networks to model complex, nonlinear relationships in survival data. [19] saw an improvement in survival prediction for cancer patients using RNA-seq data. However, all the methods use a single modality, while multi-modal data is generally present for cancer patients.

Kalakoti [20] used multi-omics data for lung adenocarcinoma patients to stratify patients into high and low-risk categories and improve the survival prediction. While dealing with multi-modal data, [52] proposed integrating both genomic data and pathological images using a bilinear network structure, which captures the interactions between the two data modalities. The network can model complex relationships and provide more accurate prognostic predictions by fusing the high-dimensional genomic features with detailed visual features from pathology images.

Multisurv [47] also proposed fusing learned representations from clinical, imaging, and genomic data for improved predictions across cancer datasets. Their fusion is achieved by taking the row-wise maximum, which performs better than the keyless multi-modal attention mechanism. Saeed et al. [40] proposed using an ensemble model that averages out results from a CNN-based prediction model for images and a Cox proportional hazard model for clinical

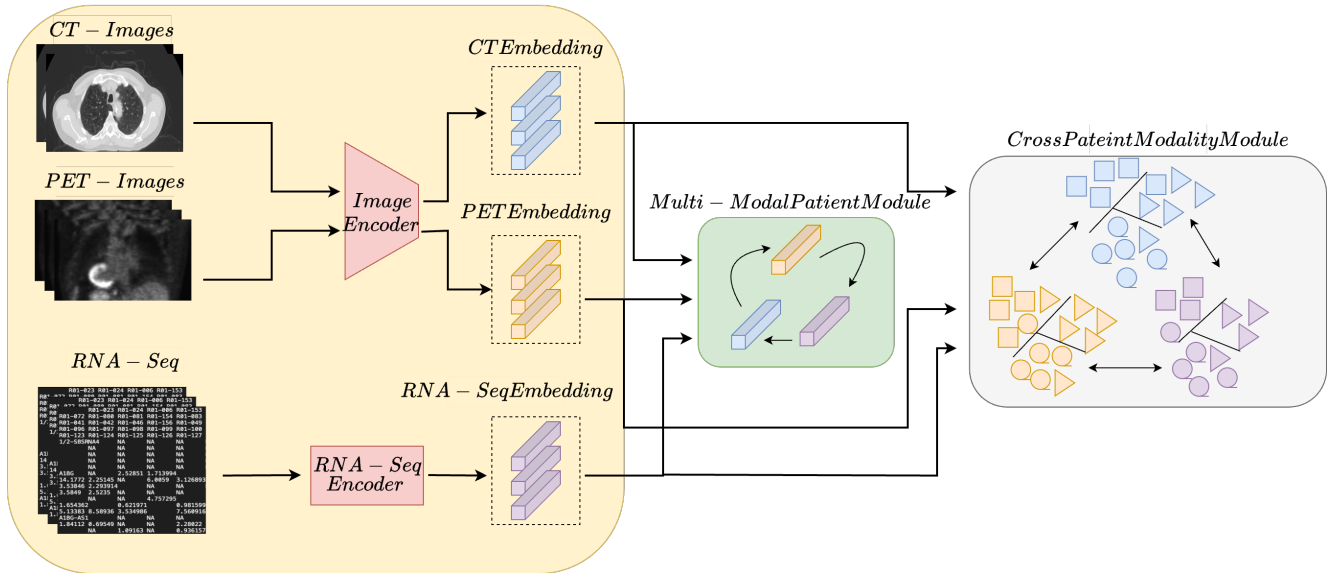


Figure 1. Overview of the multi-modal pretraining model. We show the pipeline of extracting embeddings for each modality (CT, PET, and RNA-Seq). The embeddings are refined through the MPE module, which takes care of associations between the different modalities of a patient, and the CPM module, which associates modalities across patients.

data. With the rise of the use of transformers for imaging tasks, [41] proposed encoding joint CT-PET embeddings using transformers and using the learned features for survival prediction, this model outperforms all state of the art for head and neck tumor datasets. Using both CT and PET features [32] proposed a merging-diverging learning framework. The model features a merging encoder with a Hybrid Parallel Cross-Attention (HPCA) block to fuse multi-modality information through parallel convolutional layers and cross-attention transformers. The diverging decoder incorporates a Region-specific Attention Gate (RAG) block to extract region-specific prognostic information. Ding et al. [11] proposed integrating genomic and pathology information for survival prediction in colon cancer patients. To circumvent the reliance on labels, [6, 11] used similarity metrics to guide the fusion of multi-modal representations and showed improved performance.

2.2. Survival prediction for NSCLC

Specifically for NSCLC, much research has been done over the years to improve the survival prediction. Radiomic features [48] were shown to improve the survival prediction using CT images. Sun et al. [44] studied survival prediction in NSCLC patients using Cox’s partial likelihood model. Haarbuerger [14] was one of the first studies using CNNs to extract quantitative features from 3D medical images to predict lung cancer patient survival. They simplified survival analysis to median survival classification and trained the model with small batch sizes, overcoming challenges related to large batch processing and survival loss

functions. [35]. [10] employed a parallel transformer mechanism to capture the global context of multi-scale encoder feature maps, integrating external attention to learn the positional properties of each small patch within the dataset for segmentation, the output of which is fed to a Multimodality-based Survival Network for survival prediction in NSCLC patients. Kar et al. [21] studied using clinical features for survival prediction using deep learning models. Li et al. [27] used features extracted from WSI using a transformer encoder to show impressive c-index values.

Several multi-modal models are also proposed to capture and leverage complementary contextual cues across diverse modalities. Oh et al. [39] proposed incorporating FDG-PET data and clinical data for improved survival prediction in lung cancer patients. Hou et al. proposed using the DeepSurv [16] model, which leverages radiomic features extracted from PET scans in conjunction with clinical features for the survival prediction time of NSCLC patients. Lai et al. [24] developed a Deep Neural Network (DNN) that combines gene expression and clinical data for prediction. Wang et al. [51] proposed a multi-modal deep learning model using biological knowledge that incorporated gene expression, images, and clinical data.

3. Methodology

Our framework consists of two main stages: a self-supervised multi-modal pretraining as shown in Fig. 1 and a supervised prediction model as shown in Fig. 2. The model’s primary objective is to learn embeddings from imaging and genomic data to facilitate overall survival pre-

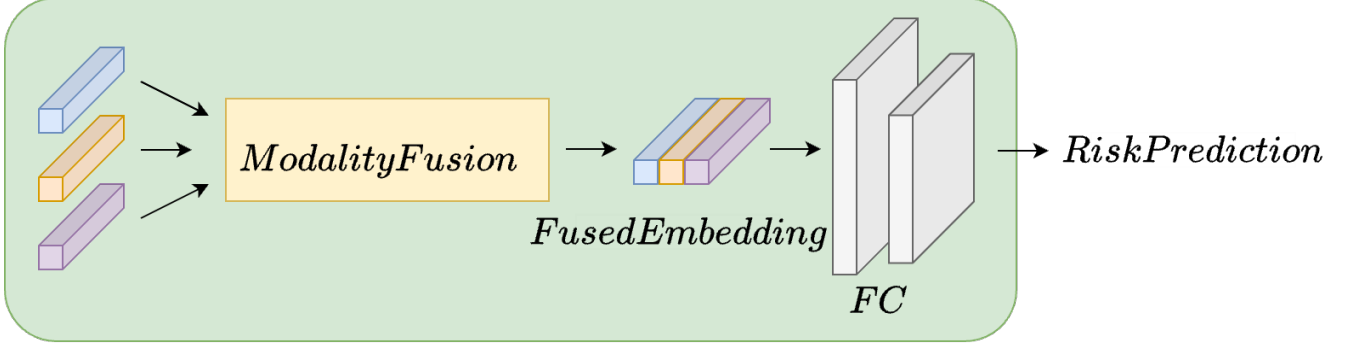


Figure 2. Our supervised prediction model utilizes the fused feature representation and generates a conditional survival probability for each predefined follow-up time interval.

diction. The pretraining phase captures interaction patterns among CT, PET, and genomic features, aiming to align the embeddings closely. Additionally, we leverage similarities across patients to enhance the learned embeddings. This approach allows us to extract valuable information from the dataset, even when data for specific modalities may be limited or missing.

For learning the embeddings, consider the dataset D comprising of N triplets of CT(X_{tN}), PET(X_{pN}) and RNA-Seq(X_{rN}) data of each patient. We plan on learning each embedding individually. We utilize Vision Transformer (ViT) [13] as the encoders f_c and f_p for CT and PET images, respectively. Additionally, for RNA-Seq embeddings, an FC net with up to 6 hidden layers f_r is employed to generate the embeddings [47]. Following [1] each input triplet $(x_{t_i}, x_{p_i}, x_{r_i})$ is mapped to (t_i, p_i, r_i) in the latent space. Following the extraction of embeddings, our approach delves into exploiting the inherent correspondences across modalities. Our framework recognizes that each modality about a single patient inherently exhibits a natural correspondence. Building upon this, subsection 3.1 presents a multi-modal patient embedding module to minimize the distances among the learned embeddings within each patient. This facilitates a closer alignment of the representations derived from CT, PET, and RNA-Seq data. Moreover, we extend our exploration to identify correspondences among the embeddings of patients with similar semantic characteristics, such as the pathological state (cancer type) in subsection 3.2. By leveraging these shared semantic features, our model can discern commonalities across patients and modalities, enriching the overall representation learning process and enhancing the predictive capabilities of our framework for survival prediction in NSCLC patients.

3.1. Multi-Modal Patient Embedding (MPE) Module

The learned embeddings are normalized to ensure consistency across modalities. For multi-modal approaches

following [2, 50], we project the i -th pair, (t_i, p_i, r_i) into a lower dimensional embedding $(\tilde{t}_i, \tilde{p}_i, \tilde{r}_i)$ and proceed to calculate the contrastive loss across each pair as:

$$\ell_i^{ab} = -\log \frac{\exp(\text{sim}(\tilde{a}_i, \tilde{b}_i)/\tau_1)}{\sum_{j=1}^B \exp(\text{sim}(\tilde{a}_i, \tilde{b}_j)/\tau_1)}, \quad (1)$$

Here, $a, b \in \{t, p, r | a \neq b\}$, $\text{sim}(\tilde{a}, \tilde{b})$ calculates the cosine similarity between a pair, τ_1 is the temperature parameter that scales the similarity values and B is the batch size. The overall loss for the MPE module is :

$$L_{\text{MPE}} = \frac{1}{3N} \sum_{i=1}^N (\ell_i^{tp} + \ell_i^{pr} + \ell_i^{rt}). \quad (2)$$

3.2. Cross Patient Modality (CPM) Module

The MPE module allows us to align the information coming from the different modalities of a patient. However, across the patients, there may exist semantic similarity [5, 50] based on the cancer type, which gets pushed apart due to the MPE module. We propose a Cross Patient-Modality (CPM) module to account for patient similarity. From [50], we define $C = \{c_1, c_2, \dots, c_K\}$, a set of K trainable cross-modal prototypes and obtain the cosine similarity between each of the low dimension embedding and the cross-modal prototypes. The softmax probability (p_{ai}) for each modality is calculate as :

$$p_{a,i}^{(k)} = \frac{\exp(\text{sim}(\tilde{a}_i, c_k)/\tau_2)}{\sum_{j=1}^K \exp(\text{sim}(\tilde{a}_i, c_j)/\tau_2)}, \quad (3)$$

Here, $a \in \{t, p, r\}$ represents each of the modalities, k indicates the k -th element of the vector, and τ_2 is the temperature hyperparameter. To leverage the cross-modal cross-patient similarity, we perform K-means clustering [30] over the triplet $(\tilde{c}_i, \tilde{p}_i, \tilde{r}_i)$ by individually assigning each of the modality to a cluster to obtain cluster assignments z_{ti} , z_{pi} and z_{ri} for the CT, PET and RNA-Seq embedding. The

Table 1. Evaluation of the prediction results (C-Index) using the NSCLC-Radiogenomics Dataset

Model	Modality	C-index
XGBoost [18]	CT	0.638 ± 0.025
	RNA-Seq	0.610 ± 0.018
RF [15]	CT	0.641 ± 0.034
	RNA-Seq	0.648 ± 0.019
DeepSurv [22]	CT	0.663 ± 0.016
MCSP [37]	CT & PET	0.670 ± 0.010
Multimodal dropout [6]	CT, PET & RNA-Seq	0.695 ± 0.013
DeepMTS [33]	CT, PET	0.719 ± 0.040
MultiSurv [47]	CT, PET & RNA-Seq	0.713 ± 0.077
XSurv [32]	CT, PET	0.721 ± 0.063
TMSS [41]	CT, PET & RNA-Seq	0.724 ± 0.053
Ours(MPE+CPM)	CT, PET & RNA-Seq	0.756 ± 0.020

alignment across the clusters is done by the cross entropy loss as follows:

$$\ell(\tilde{a}_i, z_{b,i}) = \sum_{k=1}^K z_{b,i}^{(k)} \log p_{a,i}^{(k)}, \quad (4)$$

Here $a, b \in \{t, p, r | a \neq b\}$. Thus, the loss across the three modalities for the cross-modality patient module becomes :

$$L_{CPM} = \frac{1}{3N} \sum_{i=1}^N (\ell(\tilde{t}_i, z_{pi}) + \ell(\tilde{p}_i, z_{ri}) + \ell(\tilde{r}_i, z_{ti})). \quad (5)$$

The overall loss for the pretraining is calculated as the weighted sum of the above two losses:

$$L_T = \alpha_1 \cdot L_{MPE} + \alpha_2 \cdot L_{CPM}. \quad (6)$$

Here, α_1 and α_2 are hyperparameters that determine the relative importance of each loss term in equation 6.

3.3. Modality Fusion and Prediction

Following the pretraining, the embeddings can be utilized for several downstream tasks. Here, specifically, we are targeting survival prediction. We obtain a multi-modal fused representation by concatenating the embeddings obtained from the previous task, as shown in Fig. 2. This fused output serves as input to the FC layer for the final prediction using the Cox Proportional Hazard [9] model

$$L_{\text{risk}} = - \sum_{i=1}^N \left[\delta_i \left(\beta^T f_i - \log \left(\sum_{j \in R_i} \exp(\beta^T x_j) \right) \right) \right] \quad (7)$$

Where f_i is the fused embedding for individual i , δ_i is an indicator variable taking the value 1 if individual i experienced an event, and 0 if it was censored, R_i is the set of

individuals who are at risk at the same time as individual i and β is the vector of coefficients learned by the model.

In most lung cancer datasets, while the imaging modalities may be easily accessible, the genomic data may not be available due to cost constraints. Our model allows us to effectively leverage the relationships between the available modalities to impute missing values. To mimic said behavior, we replace the RNA-Seq modality with an average of the other two modalities. As a result, the reliance on complete genomic information for survival prediction is alleviated.

4. Experiments

4.1. Dataset

In this work, we use a public radiogenomics dataset of NSCLC Radiogenomics [3, 4] available in The Cancer Imaging Archive (TCIA). The imaging and clinical data are already anonymized, along with all the ethical clearances. The data includes CT images with tumor segmentation on the CT image, tumor characteristics, PET scans, and RNA-Sequence data. It contains data from about 211 patients, out of which genomic data is available for a subset of 130 subjects. For the pre-processing, CT & PET scan slices were resampled with a slice thickness of 1 mm³ and a set number of 256 slices. Standard normalization operation was performed for both modalities.

4.2. Implementation details & Evaluation metric

For the NSCLC dataset, we propose a 4-fold cross-validation setup for training and evaluation purposes. In all runs, we computed the average performance on the test

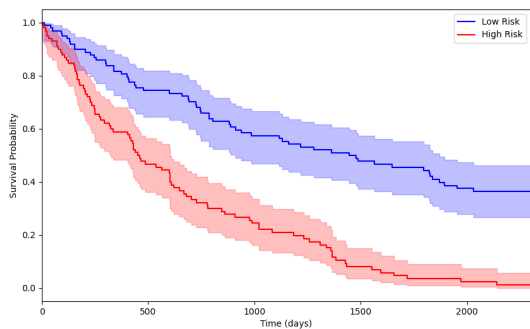


Figure 3. Kaplan-Meier curves of identified low-risk and high-risk patients for survival prediction.

set based on the results obtained from the best-performing models. We use the cross-validated concordance index (C-Index) to assess the predictive performance. The modified C-Index [45] evaluates the ability of the model to rank accurately predicted patient risk scores in relation to overall survival as :

$$C\text{-index}(D) = \frac{\sum_{i,j \in D} 1_{T_i < T_j} \cdot 1_{R_i < R_j} \cdot \delta_i}{\sum_{i,j \in D} 1_{T_i < T_j} \cdot \delta_i} \quad (8)$$

Here, T_i and T_j represent the survival time of individual i and j , R_i and R_j represent the predicted risk score of individual i and j , δ_i is the censoring indicator for the individual and $1_{T_i < T_j}$ is 1 if $T_i < T_j$, else 0. Kaplan-Meier survival curves analysis is done to evaluate the differences in event-free survival distribution. The number of epochs is kept at 100, and the batch size is 4. The learning rate is 0.0001 for pretraining and 0.00005 for the prediction task. For the pretraining setup, the dimensionality is set to 128, $\tau_1 = 0.1$, $\tau_2 = 0.2$, and $k = 500$. We have trained our model using NVIDIA A100 Tensor Core and RTX 3090 GPU using the Pytorch framework.

For comparison purposes we consider state-of-the-art survival models, including TMSS [41], XSurv [32], DeepSurv [22] and MCSP [37] as baselines. For TMSS [41], we replace the EHR embeddings with RNA-seq data to study the impact of all three modalities. Furthermore, due to ML-based models’ success in the survival prediction task, we include RF [15] and XGBoost [18] also in the baselines. To compare against the existing multi-modal methodologies, we include MultiSurv [47] and Multi-modal dropout [6]. We have used the same split, hyperparameters, and pre-processing strategies to compare all models fairly.

5. Results and Discussion

Our method demonstrates improved survival prediction on the NSCLC Radiogenomics dataset as shown in

Tab.1. Our self-supervised data fusion methodology outperforms all state-of-the-art methods, including all uni-modal and multi-modal baselines. Our best-performing model is achieved by a combined multi-modal approach using CT, PET, and RNA-Seq data; the details of other combinations are presented in the ablation study subsection. We also study the impact of missing modality, particularly RNA-Seq, on the overall results. The performance sees a slight dip by replacing the missing modality with the average of the other two modalities. Specifically, even when RNA-Seq data for 10% of patients is unavailable, our model attains a C-index value of 0.725, an improvement over existing multi-modal approaches. This shows that the model can generate meaningful embedding even with limited RNA-Seq data available. The Kaplan-Meier analysis is provided in Fig. 3

There are some limitations to our study. Mainly due to the limited nature of datasets publicly available with all three modalities available, we are unable to test our approach across multiple cancer variants, which hinders the model’s generalizability. Also, a significant drop in the C-index is observed when we replace around 40% of RNA-seq. This suggests that for larger amounts of missing data, our model is not able to achieve a good embedding.

Table 2. Ablation study on the impact of individual modalities

Data Modalities			C-index
CT	PET	RNA-Seq	
✓	✓	✓	0.756 ± 0.016
✓	✓	-	0.673 ± 0.021
✓	-	✓	0.728 ± 0.018
-	✓	✓	0.613 ± 0.025

5.1. Ablation Study

To evaluate the impact of individual modalities during pretraining, we conducted ablation studies where we trained the model using varying combinations of the modalities. Tab. 2 shows that the combination of CT, PET, and RNA-Seq outperforms all the rest. In a two-pair scenario, where the third modality is omitted altogether, the combination of CT and RNA-Seq is the best performer. To study the effect of the missing modality, we train models with varying percentages of missing RNA-Seq values in Tab. 3; we observe that even with increasing percentages of missing modality, our model continues to generate embeddings strong enough to predict accurate survival status. We also tried imputing the missing modality with zero value (dropout) and a value predicted from the available two modalities. The model yields a C-index of 0.692 with the predicted value, while

zero-embedding gives a 0.673 score for 20% missing values.

Table 3. Ablation study on the impact of missing RNA-Seq modality

% of RNA-Seq replaced	C-index
10	0.725 ± 0.013
20	0.715 ± 0.021
30	0.697 ± 0.027
40	0.681 ± 0.022

6. Conclusion

This study presents a novel approach for predicting survival in non-small cell lung cancer (NSCLC) patients by leveraging comprehensive information from CT and PET scans and genomic data. Using the Cross Patient Modality module, our proposed model is designed to capture associations between multiple modalities for individual patients while aligning embeddings closely based on semantic similarities across patients. This work presents the first instance of utilizing multi-modal representation learning to predict overall survival in lung cancer patients. Our experimental evaluation of the NSCLC dataset patients demonstrated a remarkable 9% increase in predictive performance compared to uni-modal approaches. Furthermore, our method exhibited a significant 3% improvement compared to existing multi-modal approaches. Our model achieved competitive C-index scores even in scenarios with RNA-Seq modality missing, highlighting its robustness. This capability is particularly crucial when genomic data for specific subjects is unavailable. Overall, our proposed approach holds promise for enhancing patient outcomes and guiding treatment planning in the battle against lung cancer, the leading cause of cancer-related deaths worldwide.

Acknowledgement

This research is funded through the grant No. TIH/iHub Drishti/Project/2021-22/43 from the iHub Drishti Foundation, IIT Jodhpur.

References

- [1] Hassan Akbari, Liangzhe Yuan, Rui Qian, Wei-Hong Chuang, Shih-Fu Chang, Yin Cui, and Boqing Gong. Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text. *Advances in Neural Information Processing Systems*, 34:24206–24221, 2021. 4
- [2] Jean-Baptiste Alayrac, Adria Recasens, Rosalia Schneider, Relja Arandjelović, Jason Ramapuram, Jeffrey De Fauw, Lucas Smaira, Sander Dieleman, and Andrew Zisserman. Self-supervised multimodal versatile networks. *Advances in Neural Information Processing Systems*, 33:25–37, 2020. 4
- [3] Shaimaa Bakr, Olivier Gevaert, Sebastian Echeagaray, Kelsey Ayers, Mu Zhou, Majid Shafiq, Hong Zheng, Jalen Anthony Benson, Weiruo Zhang, Ann NC Leung, et al. A radiogenomic dataset of non-small cell lung cancer. *Scientific data*, 5(1):1–9, 2018. 5
- [4] S. Bakr, O. Gevaert, S. Echeagaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, et al. Data for nscl radiogenomics collection. The Cancer Imaging Archive, 2017. Available online: <https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics>. 5
- [5] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in neural information processing systems*, 33:12546–12558, 2020. 4
- [6] Anika Cheerla and Olivier Gevaert. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*, 35(14):i446–i454, 2019. 2, 3, 5, 6
- [7] Nai-Bin Chen, Mai Xiong, Rui Zhou, Yin Zhou, Bo Qiu, Yi-Feng Luo, Su Zhou, Chu Chu, Qi-Wen Li, Bin Wang, et al. Ct radiomics-based long-term survival prediction for locally advanced non-small cell lung cancer patients treated with concurrent chemoradiotherapy using features from tumor and tumor organismal environment. *Radiation Oncology*, 17(1):1–12, 2022. 1
- [8] Y Choi, J Aum, SH Lee, HK Kim, J Kim, S Shin, JY Jeong, CY Ock, and HY Lee. Deep learning analysis of ct images reveals high-grade pathological features to predict survival in lung adenocarcinoma. *cancers* 2021, 13, 4077, 2021. 1
- [9] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–220, 1972. 5
- [10] Duy-Phuong Dao, Hyung-Jeong Yang, Ngoc-Huynh Ho, Sudarshan Pant, Soo-Hyung Kim, Guee-Sang Lee, In-Jae Oh, and Sae-Ryung Kang. Survival analysis based on lung tumor segmentation using global context-aware transformer in multimodality. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 5162–5169. IEEE, 2022. 3
- [11] Kexin Ding, Mu Zhou, Dimitris N Metaxas, and Shaoting Zhang. Pathology-and-genomics multimodal transformer for survival outcome prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 622–631. Springer, 2023. 2, 3
- [12] Shreyesh Doppalapudi, Robin G Qiu, and Youakim Badr. Lung cancer survival period prediction and understanding: Deep learning approaches. *International Journal of Medical Informatics*, 148:104371, 2021. 2
- [13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Trans-

- formers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 4
- [14] Christoph Haarbuerger, Philippe Weitz, Oliver Rippel, and Dorit Merhof. Image-based survival prediction for lung cancer patients using cnns. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 1197–1201. IEEE, 2019. 3
- [15] Tingshan He, Jing Li, Peng Wang, and Zhiqiao Zhang. Artificial intelligence predictive system of individual survival rate for lung adenocarcinoma. *Computational and Structural Biotechnology Journal*, 20:2352–2359, 2022. 1, 5, 6
- [16] Kuei-Yuan Hou, Jyun-Ru Chen, Yung-Chen Wang, Ming-Huang Chiu, Sen-Ping Lin, Yuan-Heng Mo, Shih-Chieh Peng, and Chia-Feng Lu. Radiomics-based deep learning prediction of overall survival in non-small-cell lung cancer using contrast-enhanced computed tomography. *Cancers*, 14(15):3798, 2022. 2, 3
- [17] Brian Huang, John Sollee, Yong-Heng Luo, Ashwin Reddy, Zhushi Zhong, Jing Wu, Joseph Mammarrappallil, Terrance Healey, Gang Cheng, Christopher Azzoli, et al. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment fdg-pet/ct. *EBioMedicine*, 82, 2022. 2
- [18] Zhangheng Huang, Chuan Hu, Changxing Chi, Zhe Jiang, Yuexin Tong, and Chengliang Zhao. An artificial intelligence model for predicting 1-year survival of bone metastases in non-small-cell lung cancer patients based on xgboost algorithm. *BioMed research international*, 2020, 2020. 1, 5, 6
- [19] Zhi Huang, Travis S Johnson, Zhi Han, Bryan Helm, Sha Cao, Chi Zhang, Paul Salama, Maher Rizkalla, Christina Y Yu, Jun Cheng, et al. Deep learning-based cancer survival prognosis from rna-seq data: approaches and evaluations. *BMC medical genomics*, 13:1–12, 2020. 2
- [20] Yogesh Kalakoti, Shashank Yadav, and Durai Sundar. Survcn: a discrete time-to-event cancer survival estimation framework using image representations of omics data. *Cancers*, 13(13):3106, 2021. 2
- [21] İrem Kar, Gökhan Kocaman, Farrukh İbrahimov, Serkan Enön, Erdal Coşgun, and Atilla Halil Elhan. Comparison of deep learning-based recurrence-free survival with random survival forest and cox proportional hazard models in stage- i nslc patients. *Cancer Medicine*, 12(18):19272–19278, 2023. 3
- [22] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deep-surv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18:1–12, 2018. 2, 5, 6
- [23] Young Jae Kim, Hyun-Ju Lee, Kwang Gi Kim, and Seung Hyun Lee. The effect of ct scan parameters on the measurement of ct radiomic features: a lung nodule phantom study. *Computational and Mathematical Methods in Medicine*, 2019(1):8790694, 2019. 2
- [24] Yu-Heng Lai, Wei-Ning Chen, Te-Cheng Hsu, Che Lin, Yu Tsao, and Semon Wu. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Scientific reports*, 10(1):4679, 2020. 3
- [25] 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. 2
- [26] Natalie Si-Yi Lee, Jesmin Shafiq, Matthew Field, Caroline Fiddler, Suganthy Varadarajan, Senthilkumar Gandhidasan, Eric Hau, and Shalini Kavita Vinod. Predicting 2-year survival in stage i-iii non-small cell lung cancer: the development and validation of a scoring system from an australian cohort. *Radiation Oncology*, 17(1):1–12, 2022. 1
- [27] Butuo Li, Linlin Yang, Huan Zhang, Haoqian Li, Chao Jiang, Yueyuan Yao, Shuping Cheng, Bing Zou, Bingjie Fan, Taotao Dong, et al. Outcome-supervised deep learning on pathologic whole slide images for survival prediction of immunotherapy in patients with non-small cell lung cancer. *Modern Pathology*, 36(8):100208, 2023. 3
- [28] Jie Lian, Yonghao Long, Fan Huang, Kei Shing Ng, Faith MY Lee, David CL Lam, Benjamin XL Fang, Qi Dou, and Varut Vardhanabhuti. Imaging-based deep graph neural networks for survival analysis in early stage lung cancer using ct: A multicenter study. *Frontiers in Oncology*, 12:868186, 2022. 2
- [29] Kuiyuan Liu, Weixiong Xia, Mengyun Qiang, Xi Chen, Jia Liu, Xiang Guo, and Xing Lv. Deep learning pathological microscopic features in endemic nasopharyngeal cancer: Prognostic value and potential role for individual induction chemotherapy. *Cancer medicine*, 9(4):1298–1306, 2020. 2
- [30] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982. 4
- [31] Chip M Lynch and Joshua D BehnazAbdollahi. Fuqua, alexandra r. de carlo, james a. bartholomai, rayeanne n. balgemann, victor h. van berkel, hermann b. frieboes, prediction of lung cancer patient survival via supervised machine learning classification techniques. *International Journal of Medical Informatics*, 108:1–8, 2017. 1
- [32] Mingyuan Meng, Lei Bi, Michael Fulham, Dagan Feng, and Jinman Kim. Merging-diverging hybrid transformer networks for survival prediction in head and neck cancer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 400–410. Springer, 2023. 3, 5, 6
- [33] Mingyuan Meng, Bingxin Gu, Lei Bi, Shaoli Song, David Dagan Feng, and Jinman Kim. Deepmts: deep multi-task learning for survival prediction in patients with advanced nasopharyngeal carcinoma using pretreatment pet/ct. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4497–4507, 2022. 5
- [34] Hajar Moradmand, Seyed Mahmoud Reza Aghamiri, Reza Ghaderi, and Hamid Emami. The role of deep learning-based survival model in improving survival prediction of patients with glioblastoma. *Cancer Medicine*, 10(20):7048–7059, 2021. 2
- [35] Ju Gang Nam, Samina Park, Chang Min Park, Yoon Kyung Jeon, Doo Hyun Chung, Jin Mo Goo, Young Tae Kim, and Hyungjin Kim. Histopathologic basis for a chest ct deep

- learning survival prediction model in patients with lung adenocarcinoma. *Radiology*, 305(2):441–451, 2022. 3
- [36] S Navada, P Lai, AG Schwartz, and GP Kalemkerian. Temporal trends in small cell lung cancer: analysis of the national surveillance, epidemiology, and end-results (seer) database. *Journal of Clinical Oncology*, 24(18_suppl):7082–7082, 2006. 1
- [37] Dong Nie, Junfeng Lu, Han Zhang, Ehsan Adeli, Jun Wang, Zhengda Yu, Lu Yan Liu, Qian Wang, Jinsong Wu, and Dinggang Shen. Multi-channel 3d deep feature learning for survival time prediction of brain tumor patients using multimodal neuroimages. *Scientific reports*, 9(1):1103, 2019. 5, 6
- [38] Seungwon Oh, Jaena Im, Sae-Ryung Kang, In-Jae Oh, and Min-Soo Kim. Pet-based deep-learning model for predicting prognosis of patients with non-small cell lung cancer. *IEEE Access*, 9:138753–138761, 2021. 1
- [39] Seungwon Oh, Sae-Ryung Kang, In-Jae Oh, and Min-Soo Kim. Deep learning model integrating positron emission tomography and clinical data for prognosis prediction in non-small cell lung cancer patients. *BMC bioinformatics*, 24(1):1–13, 2023. 2, 3
- [40] Numan Saeed, Roba Al Majzoub, Ikboljon Sobirov, and Mohammad Yaqub. An ensemble approach for patient prognosis of head and neck tumor using multimodal data. In *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, pages 278–286. Springer, 2021. 2
- [41] Numan Saeed, Ikboljon Sobirov, Roba Al Majzoub, and Mohammad Yaqub. Tmss: an end-to-end transformer-based multimodal network for segmentation and survival prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 319–329. Springer, 2022. 3, 5, 6
- [42] Yunlang She, Zhuochen Jin, Junqi Wu, Jiajun Deng, Lei Zhang, Hang Su, Gening Jiang, Haipeng Liu, Dong Xie, Nan Cao, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA network open*, 3(6):e205842–e205842, 2020. 2
- [43] Michael J Shulman and Elie A Benaim. Prognostic model of event-free survival for patients with androgen-independent prostate carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 103(11):2280–2286, 2005. 2
- [44] Wenzheng Sun, Mingyan Jiang, Jun Dang, Panchun Chang, and Fang-Fang Yin. Effect of machine learning methods on predicting nsccl overall survival time based on radiomics analysis. *Radiation oncology*, 13:1–8, 2018. 3
- [45] Zhenyu Tang, Zhenyu Zhang, Huabing Liu, Dong Nie, and Jing Yan. Pre-operative survival prediction of diffuse glioma patients with joint tumor subtyping. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 786–795. Springer, 2023. 6
- [46] Suyan Tian. Classification and survival prediction for early-stage lung adenocarcinoma and squamous cell carcinoma patients. *Oncology letters*, 14(5):5464–5470, 2017. 1
- [47] Luís A Vale-Silva and Karl Rohr. Long-term cancer survival prediction using multimodal deep learning. *Scientific Reports*, 11(1):13505, 2021. 2, 4, 5, 6
- [48] Janna E van Timmeren, Ralph TH Leijenaar, Wouter van Elmpt, Bart Reymen, Cary Oberije, René Monshouwer, Johan Bussink, Carsten Brink, Olfred Hansen, and Philippe Lambin. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam ct images. *Radiotherapy and Oncology*, 123(3):363–369, 2017. 3
- [49] Thanh-Hung Vo, Guee-Sang Lee, Hyung-Jeong Yang, In-Jae Oh, Soo-Hyung Kim, and Sae-Ryung Kang. Survival prediction of lung cancer using small-size clinical data with a multiple task variational autoencoder. *Electronics*, 10(12):1396, 2021. 2
- [50] Fuying Wang, Yuyin Zhou, Shujun Wang, Varut Vardhanabhuti, and Lequan Yu. Multi-granularity cross-modal alignment for generalized medical visual representation learning. *Advances in Neural Information Processing Systems*, 35:33536–33549, 2022. 4
- [51] Shuo Wang, Hao Zhang, Zhen Liu, and Yuaning Liu. A novel deep learning method to predict lung cancer long-term survival with biological knowledge incorporated gene expression images and clinical data. *Frontiers in Genetics*, 13:800853, 2022. 3
- [52] Zhiqin Wang, Ruiqing Li, Minghui Wang, and Ao Li. Gpdbn: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction. *Bioinformatics*, 37(18):2963–2970, 2021. 2
- [53] Sheng Xu, Jing Qi, Bin Li, and Xiao-Guang Li. Survival prediction for non-small cell lung cancer patients treated with ct-guided microwave ablation: development of a prognostic nomogram. *International Journal of Hyperthermia*, 38(1):640–649, 2021. 1
- [54] Bin Zhang, Lianmin Zhang, Dongsheng Yue, Chenguang Li, Hua Zhang, Junyi Ye, Liuwei Gao, Xiaoliang Zhao, Chen Chen, Yansong Huo, et al. Genomic characteristics in chinese non-small cell lung cancer patients and its value in prediction of postoperative prognosis. *Translational Lung Cancer Research*, 9(4):1187, 2020. 1
- [55] 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. 2