CoC: Chain-of-Cancer based on Cross-Modal Autoregressive Traction for Survival Prediction

Haipeng Zhou[1], Sicheng Yang[2], Sihan Yang[2],
Jing Qin[3], Lei Chen[1,4], and Lei Zhu[1,4]

[1] The Hong Kong University of Science and Technology (Guangzhou)
[2] Xi'an Jiaotong University
[3] The Hong Kong Polytechnic University
[4] The Hong Kong University of Science and Technology

Abstract. Survival prediction aims to evaluate the risk level of cancer patients. Existing methods primarily rely on pathology and genomics data, either individually or in combination. From the perspective of cancer pathogenesis, epigenetic changes, such as methylation data, could also be crucial for this task. Furthermore, no previous endeavors have utilized textual descriptions to guide the prediction. To this end, we are the first to explore the use of four modalities, including three clinical modalities and language, for conducting survival prediction. In detail, we are motivated by the Chain-of-Thought (CoT) to propose the Chain-of-Cancer (CoC) framework, focusing on intra-learning and inter-learning. We encode the clinical data as the raw features, which remain domain-specific knowledge for intra-learning. In terms of inter-learning, we use language to prompt the raw features and introduce an Autoregressive Mutual Traction module for synergistic representation. This tailored framework facilitates joint learning among multiple modalities. Our approach is evaluated across five public cancer datasets, and extensive experiments validate the effectiveness of our methods and proposed designs, leading to producing *state-of-the-art* results. Codes will be released [1] .

Keywords: Survival Prediction, Multimodal Learning

1 Introduction

Survival prediction [5,4], particularly in oncology, plays a pivotal role in guiding clinical decision-making and personalized treatment strategies. It utilizes clinical data as biomarkers, aiming to provide risk stratification for patients. Recently, the use of data in survival analysis has primarily focused on genes and pathological images, leading to the development of both single-modality [12,27,9,16,20] and multimodal learning methods [3,30,25,10,24] that integrate the two.

Relying solely on genomic data poses challenges due to the high dimensionality of genomic profiles [17]. Consequently, single-modality approaches often

✉ Corresponding Author: Lei Zhu (leizhu@ust.hk)

[1]:https://github.com/haipengzhou856/CoC

adopt Multiple Instance Learning (MIL) [1] for processing Whole Slide Images (WSIs), *i.e.*, pathology images. By dividing the giga-pixel resolution of WSIs into smaller patches, these models can focus on learning specific patterns from regions of interest. However, recent advances in combining genomic data with WSIs have demonstrated improved performance. These two modalities provide complementary information, enabling a more comprehensive understanding of the disease. For example, SurvPath [10] suggests that tumor morphologies correspond well to the pathways in transomics. Specific genes influence the morphological features observed in pathology, and their synergistic relationship can be effectively uncovered through multimodal learning.

Nevertheless, epigenetics (e.g., methylation data)[22] also plays a vital role in regulating gene expression, impacting various biological processes and resulting in specific pathological imaging features like increased cell density and abnormal cell morphology [29] in WSIs. By integrating genomic, epigenetic, and WSI data, we can achieve a comprehensive understanding of diseases from molecular, epigenetic, and morphological perspectives. Furthermore, with advancements in Large Language Models (LLMs), an increasing number of downstream tasks [18,13,15,19] are exploring textual guidance to enhance their outcomes. Inspired by this, we aim to leverage these multiple modalities to unleash their potential to improve survival prediction.

In this paper, we propose a cross-modal autoregressive model, named Chain-of-Cancer (CoC). Our core idea is to utilize task-specific handcrafted language descriptions as the initial prompt, encouraging mutual learning across different modalities in an autoregressive manner. In particular, **1) we introduce a CoC-Adapter.** By providing clinical-related descriptions, we can embed textual guidance into different modalities to enhance representation. **2) Moreover, we propose an Autoregressive Mutual Traction (AMT) module** to facilitate synergistic learning. This module establishes dependencies among different modalities to enable cross-modal learning. **3) We leverage the inter-learning and intra-learning for survival prediction**. We encode the clinical data to extract raw features for intra-learning, and we deploy the CoC-Adapter and AMT module to conduct inter-learning. Through their combined effects, we achieve promising results on this task. To the best of our knowledge, **we are the first to apply these new modalities (*i.e.*, methylation and language) to survival prediction.** The tailored autoregressive learning manner also empowers our model to achieve promising results. We evaluate our method on five public datasets, and the experimental results demonstrate that our CoC consistently surpasses *state-of-the-art* methods.

2 Method

2.1 Formulation and Overview

Given the clinical data $\mathbb{X}$ and time-to-event label $\mathbb{Y} = \{t, c\}$, where $t \in \mathcal{R}^+$ is the overall survival time and $c \in \{0, 1\}$ is event censorship at t, our target is to estimate the death probability via the hazard function $f(t) = f(T = t|T \geq t, \mathbb{X})$.

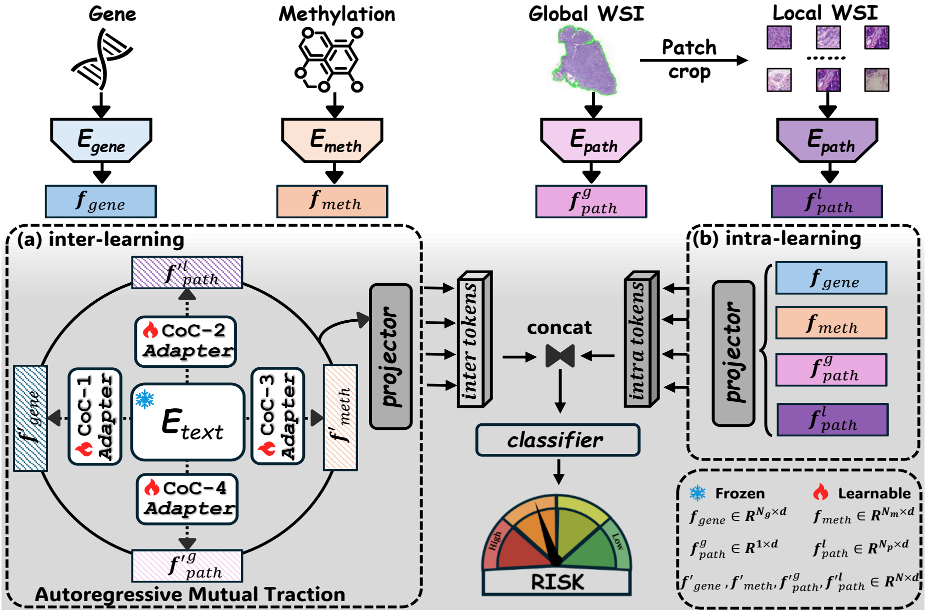
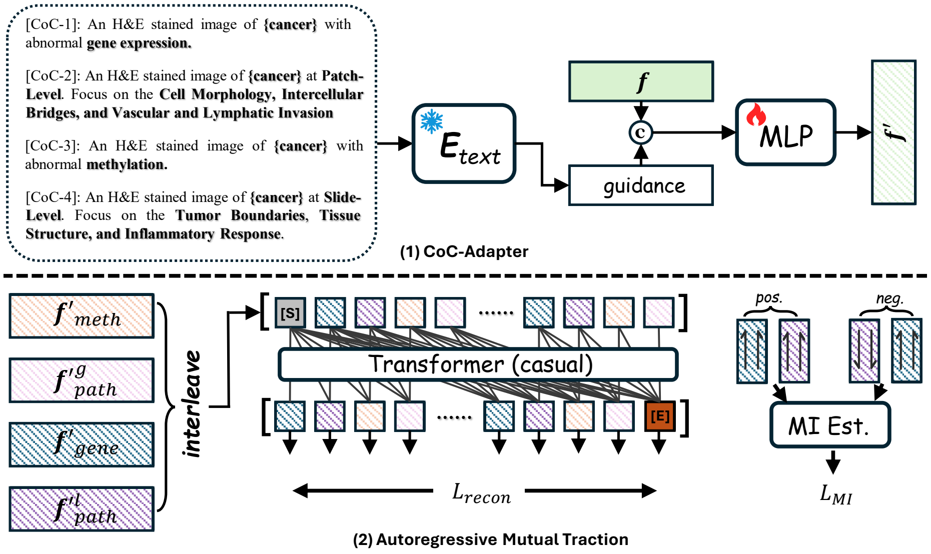


Fig. 1. An overview of our CoC framework. It consists of the basic feature extraction, the (a) inter-learning, and the (b) intra-learning. In (a) we use CoC-Adapter to embed language guidance to produce homogeneity features autoregressively. In (b) we use a simple projector to yield heterogeneity features. Finally, by concatenating these features we can utilize a classifier to conduct survival prediction.

Such that, we can estimate the ordinal risk of death event occurring at time point t via a discrete-time format [6] by $S(t|\mathbb{X}) = \prod_{j}^{t}(1-f(j))$. The hazard-based task can be converted into classification [26] by setting the number of time bins being 4, and we have negative log-likelihood loss with the censorship c termed as:

$$\mathcal{L}_{surv} = -c\,\log(S(t|\mathbb{X})). \tag{1}$$

As shown in Fig. 1, in our study we have $\mathbb{X} = \{\mathcal{X}_{gene}, \mathcal{X}_{meth}, \mathcal{X}^{g}_{path}, \mathcal{X}^{l}_{path}\}$ which denotes the 1-D gene profile, and the 1-D methylation profile, the pathology images at global-level, and the cropped patches, respectively. We first conduct the feature engineering to obtain their features yielding $f_{gene} \in \mathcal{R}^{N_g \times d}$, $f_{meth} \in \mathcal{R}^{N_m \times d}$, $f^{g}_{path} \in \mathcal{R}^{1 \times d}$, and $f^{l}_{path} \in \mathcal{R}^{N_p \times d}$, where N_g and N_m are the number of tokens for gene and methylation features, N_p is the number of local patches, and d is the latent dimension. These raw features are heterogeneous, and we simply retain them and feed them into a projector to conduct intra-learning. In terms of inter-learning, we explicitly use descriptive text to embed these features into synergistic knowledge space via CoC-Adapter. Then, we deploy Autogregressive Mutual Tranction to continue mine the homogeneous features. By leverage intra-learning and inter-learning, our network can effectively explore the synergistic and independent contributions of each modality.


[CoC-1]: An H&E stained image of **{cancer}** with abnormal **gene expression.**

[CoC-2]: An H&E stained image of **{cancer}** at **Patch-Level**. Focus on the **Cell Morphology, Intercellular Bridges, and Vascular and Lymphatic Invasion**

[CoC-3]: An H&E stained image of **{cancer}** with abnormal **methylation.**

[CoC-4]: An H&E stained image of **{cancer}** at **Slide-Level**. Focus on the **Tumor Boundaries, Tissue Structure, and Inflammatory Response**.

Fig. 2. The Inter-Learning branch and it consists of two steps. (1) We first use CoC-Adapter to embed the raw feature with the text guidance. (2) We deploy Autoregressive Mutual Traction (AMT) module to conduct synergistic representation.

2.2 Chain-of-Cancer Adapter

Prompt methods like Chain-of-Thought (CoT) [23,13,28] demonstrate that explicitly using language prompts can enhance the model's reasoning ability for multimodal learning. Motivated by it, we propose the Chain-of-Cancer Adapter, which makes use of clinical-related descriptions as textual guidance to integrate with different modalities.

Fig. 2 (1) shows the workflow of our CoC-Adapter. We first design tailored descriptions. For all the modalities, we give a vanilla prompt `"An H&E stained image of {cancer}"` where 'cancer' denotes the cancer type. Considering the specific knowledge, we detail the prompt according to the modality. For the visual end, we analogize the pathologists' diagnosis in which we provide the priors from global and local perspectives, *e.g.*, the Tumor Boundaries *vs.* Intercellular Bridges. In terms of the 1-D data, we add the suffix class (*e.g.*, 'methylation') to inject priors. Thus, we can deploy a text encoder E_{text} to generate language guidance embedding. Then, we concatenate the raw feature f with the guidance and feed it into a learnable MLP layer, yielding the text-embedded feature f'.

2.3 Autoregressive Mutual Traction (AMT)

We first introduce the preliminary of autoregressive models (ARM). Given a sequence of tokens $x = \{x^1, x^2, ..., x^n\}$, ARM predict the current token x_i based on previous tokens $\{x^j\}_{j\leq i-1}$, and the next token prediction can be termed as:

$$p(x^1, ..., x^n) = \prod_{i=1}^{n} p(x^i|x^1, ..., x^{i-1}). \tag{2}$$

Thus, an ARM parameterized by θ is to optimize the probability $p_\theta(x^i|x^1, ..., x^{i-1})$.

We propose that the autoregressive structure facilitates cross-modal interaction and homogeneity representation by enforcing causal dependencies between different modalities through the next-step prediction paradigm. As shown in Fig. 2, in our approach, we interleave the features to form a sequence x, which helps prevent the model from focusing exclusively on features within a single modality. Following previous ARM [14], we add a start token [S] to serve as the context for predicting the first element in the sequence and an end token [E] to denote the end of sequence. Such that, we can deploy a standard Transformer with 2 layers to decode the reconstruction. To optimize it, we have:

$$\mathcal{L}_{rec} = ||x - \hat{x}||^2. \tag{3}$$

Besides, to prevent over-reconstruction we design a Mutual Information regulation to make the model retain meaningful cross-modal relationships. Given a pair of reconstructed features from different modalities, m_1 and m_2, we have:

$$\mathcal{L}_{MI}^{m1,m2} = \sum_{(m_1,m_2)} -\mathbb{E}_{(x_i^{m_1}, x_i^{m_2}),(\tilde{x}_j^{m_2})} \left[\log \sigma \left(f_\phi(x_i^{m_1}, x_i^{m_2}) - f_\phi(x_i^{m_1}, \tilde{x}_j^{m_2})\right)\right], \tag{4}$$

where $x_i^{m_2}$ is the positive samples predicted by the ARM. For negative samples $\tilde{x}_j^{m_2}$, since in ARM we generate them in an interleaved causal manner, we can use negative samples by randomly shuffling the order of this modality. The ϕ is a mutual information estimator which can be a simple MLP layer. Thus, considering the four traction chains (including the global and local pathology features), we will have a total of 6 (3×2) pairs to compute the mutual information.

2.4 Objective

For our intra-learning, we can concatenate the raw features and use a linear projector layer to out intra-tokens. Similarly, we also deploy a linear projection for the output of AMT yielding the inter-token. Considering the specific and synergistic knowledge jointly, we concatenate them and feed them into the classifier to predict the patient risk. The total term of the loss function is computed as:

$$\mathcal{L} = \mathcal{L}_{surv} + \mathcal{L}_{rec} + \lambda * \mathcal{L}_{MI}, \tag{5}$$

where λ is empirically set as 0.3.

3 Experiments and Results

3.1 Datasets

We use the public available data from TCGA[2] , including WSIs, genomic data, the methylation data, and the ground truth of survival time. A total of 5 cancer datasets are used, including Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC, 270 cases), Liver hepatocellular carcinoma (LIHC, 350 cases), Breast invasive carcinoma (BRCA, 1058 cases), Colon adenocarcinoma (COAD, 444 cases), and Kidney renal clear cell carcinoma (KIRC, 509 cases). All the cases are cleaned with 60,660 genomics features and 80,000 methylation features. The overall data volume is about 2.5 TB.

3.2 Implementation Details

Metrics. Following previous works, we adopt C-Index [7] as the quantitative metric to evaluate the performance. We conduct 5-fold cross-validation to ensure the reproduction and robustness. The Kaplan-Meier analysis [21,11] is also presented, along with an evaluation of the log-rank test.

Configurations. We use AdamW optimizer, a learning rate of 1e-4, and 20 epochs to train the model. In line with previous works[27,16,20,30,25,10,24], we utilize ResNet-50 [8] to encode pathology data and use SNN [12] to encode 1-D data (*i.e.*, gene and methylation). In terms of language-end, we deploy the tokenizer from CONCH [15] which is trained by WSI-Text pairs. This enables us to make use of handcrafted descriptive texts in a medical domain instead of vanilla CLIP [18]. In our implementation, we have $N_g = 6$, $N_m = 8$, $N = 4$, and $d = 512$, respectively. The N_p is determined by the WSI tiling resulting in various values. For other methods, we reproduce them by their official codes.

3.3 Performance Evaluation

Comparisons with *state-of-the-art* Methods. We present the quantitative comparisons in Tab. 1, showcasing a total of 12 methods. For 1-D data, we utilize MLP [2], SNN [12], and SNNTrans [12] to generate single-modal and cross-modal results. For single pathology data, we reproduce Deep-sets [27], AttnMIL [9], CLAM [16], and TransMIL [20]. Additionally, we reconstruct dual-modal methods involving genomics and WSIs, such as MCAT [3], MOTCAT [25], SurvPath [10], and MoME [24]. From the table, it is evident that our method outperforms all others across all five datasets. Specifically, our approach achieves improvements of 0.6%, 1.4%, 1.5%, 2.0%, and 1.3% on CESC, LIHC, BRCA, COAD, and KIRC datasets, resulting in an overall 1.7% boost compared to the runner-up (0.655 *vs.* 0.638). Besides, dual-modal methods show better performance compared to single-modal methods. Our method, driven by descriptive text, effectively leverages triple clinical modal data, leading to the best results.

[2] https://portal.gdc.cancer.gov

Table 1. Quantitative comparisons with *state-of-the-art* methods under the metric of C-index (mean ± std) on 5 cancer datasets with 5-fold cross-validation. The best ones and runner-ups are highlighted with **red** and **blue**, respectively. The 🧬, 🔬, and ⌬ denote the modality of gene, pathology and methylation, respectively.

Methods	Modality	CESC (N=270)	LIHC (N=350)	BRCA (N=1058)	COAD (N=444)	KIRC (N=509)	Overall
MLP [2]	🧬	$0.589_{\pm 0.054}$	$0.586_{\pm 0.044}$	$0.621_{\pm 0.028}$	$0.597_{\pm 0.025}$	$0.655_{\pm 0.059}$	0.610
SNN [12]	🧬	$0.573_{\pm 0.052}$	$0.591_{\pm 0.037}$	$0.622_{\pm 0.027}$	$0.600_{\pm 0.018}$	$0.664_{\pm 0.052}$	0.610
SNNTrans [12]	🧬	$0.581_{\pm 0.046}$	$0.600_{\pm 0.057}$	$0.624_{\pm 0.012}$	$0.602_{\pm 0.037}$	$0.661_{\pm 0.043}$	0.614
SNN [12]	⌬	$0.611_{\pm 0.039}$	$0.594_{\pm 0.036}$	$0.594_{\pm 0.006}$	$0.588_{\pm 0.037}$	$0.651_{\pm 0.045}$	0.608
SNNTrans [12]	⌬	$0.609_{\pm 0.033}$	$0.596_{\pm 0.046}$	$0.611_{\pm 0.034}$	$0.591_{\pm 0.030}$	$0.653_{\pm 0.031}$	0.612
Deep-sets [27]	🔬	$0.541_{\pm 0.023}$	$0.501_{\pm 0.003}$	$0.541_{\pm 0.026}$	$0.518_{\pm 0.031}$	$0.512_{\pm 0.027}$	0.523
AttnMIL [9]	🔬	$0.613_{\pm 0.065}$	$0.505_{\pm 0.053}$	$0.602_{\pm 0.014}$	$0.574_{\pm 0.036}$	$0.546_{\pm 0.035}$	0.568
CLAM-SB [16]	🔬	$0.618_{\pm 0.066}$	$0.526_{\pm 0.071}$	$0.602_{\pm 0.015}$	$0.569_{\pm 0.035}$	$0.555_{\pm 0.036}$	0.574
CLAM-MB [16]	🔬	$0.604_{\pm 0.073}$	$0.508_{\pm 0.062}$	$0.591_{\pm 0.014}$	$0.567_{\pm 0.034}$	$0.554_{\pm 0.035}$	0.565
TransMIL [20]	🔬	$0.612_{\pm 0.044}$	$0.611_{\pm 0.040}$	$0.607_{\pm 0.017}$	$0.595_{\pm 0.026}$	$0.602_{\pm 0.009}$	0.605
SNN [12]	🧬+⌬	$0.613_{\pm 0.044}$	$0.590_{\pm 0.057}$	$0.613_{\pm 0.027}$	$0.604_{\pm 0.020}$	$0.669_{\pm 0.013}$	0.618
SNNTrans [12]	🧬+⌬	$0.610_{\pm 0.038}$	$0.603_{\pm 0.041}$	$0.628_{\pm 0.016}$	$0.606_{\pm 0.029}$	$0.674_{\pm 0.051}$	0.624
MCAT [3]	🧬+🔬	$0.573_{\pm 0.034}$	$0.604_{\pm 0.040}$	$0.588_{\pm 0.007}$	$0.590_{\pm 0.051}$	$0.656_{\pm 0.017}$	0.602
CMTA [30]	🧬+🔬	$0.622_{\pm 0.061}$	$0.595_{\pm 0.020}$	$0.636_{\pm 0.013}$	$0.605_{\pm 0.020}$	$0.675_{\pm 0.029}$	0.627
MOTCAT [25]	🧬+🔬	$0.637_{\pm 0.035}$	$0.613_{\pm 0.028}$	$0.630_{\pm 0.016}$	$0.615_{\pm 0.016}$	$0.696_{\pm 0.054}$	0.638
SurvPath [10]	🧬+🔬	$0.627_{\pm 0.044}$	$0.616_{\pm 0.047}$	$0.639_{\pm 0.032}$	$0.618_{\pm 0.017}$	$0.685_{\pm 0.036}$	0.637
MoME [24]	🧬+🔬	$0.621_{\pm 0.037}$	$0.610_{\pm 0.031}$	$0.625_{\pm 0.049}$	$0.610_{\pm 0.031}$	$0.677_{\pm 0.038}$	0.628
Ours	🧬+🔬+⌬	$\mathbf{0.643}_{\pm 0.028}$	$\mathbf{0.630}_{\pm 0.047}$	$\mathbf{0.654}_{\pm 0.036}$	$\mathbf{0.638}_{\pm 0.046}$	$\mathbf{0.709}_{\pm 0.048}$	**0.655**

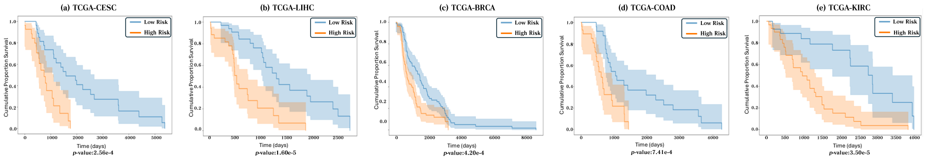


Fig. 3. Kaplan-Meier survival curves. The prognostic separation between high-risk and low-risk cohorts are stratified by median prognostic scores. The shaded areas denote the confidence intervals. Please zoom in for the best view.

Kaplan-Meier Analysis. We further conduct statistical analysis based on Kaplan-Meier analysis, and the visualization can be found in Fig. 3. The patients are divided into high-risk and low-risk groups based on the cut-off of the predicted median prognostic scores. Our method demonstrates a clear distinction between the two groups, as evidenced by the p-values with log-rank test, which are consistently below 0.01 across all five datasets. This indicates a high level of statistical significance and robustness in our method, underscoring the effectiveness of our approach in differentiating between the groups.

Table 2. Ablation study of methods on the five datasets. The † denotes using vanilla text prompt, *i.e.*, `"An H&E stained image of {cancer}"`.

`CONFIG`	CESC	LIHC	BRCA	COAD	KIRC	Overall
Only Intra-Learning	$0.610_{\pm 0.034}$	$0.602_{\pm 0.047}$	$0.619_{\pm 0.033}$	$0.601_{\pm 0.026}$	$0.652_{\pm 0.017}$	0.617
Basic	$0.621_{\pm 0.011}$	$0.611_{\pm 0.024}$	$0.633_{\pm 0.010}$	$0.603_{\pm 0.026}$	$0.656_{\pm 0.034}$	0.625
`w/o CoC-1` (M1)	$0.629_{\pm 0.062}$	$0.616_{\pm 0.043}$	$0.640_{\pm 0.041}$	$0.623_{\pm 0.030}$	$0.690_{\pm 0.046}$	0.640
`w/ CoC-1` (M2)	$0.638_{\pm 0.021}$	$0.626_{\pm 0.014}$	$0.641_{\pm 0.031}$	$0.626_{\pm 0.034}$	$0.694_{\pm 0.024}$	0.645
`w/o CoC-3` (M3)	$0.613_{\pm 0.037}$	$0.622_{\pm 0.025}$	$0.620_{\pm 0.051}$	$0.615_{\pm 0.042}$	$0.686_{\pm 0.034}$	0.631
`w/ CoC-3` (M4)	$0.624_{\pm 0.030}$	$0.618_{\pm 0.024}$	$0.636_{\pm 0.017}$	$0.627_{\pm 0.049}$	$0.701_{\pm 0.034}$	0.641
+`CoC-1` & `CoC-3`(†M5)	$0.630_{\pm 0.028}$	$0.620_{\pm 0.037}$	$0.642_{\pm 0.041}$	$0.630_{\pm 0.041}$	$0.697_{\pm 0.034}$	0.644
+`CoC-1` & `CoC-3`(Ours)	$\mathbf{0.643}_{\pm 0.028}$	$\mathbf{0.630}_{\pm 0.047}$	$\mathbf{0.654}_{\pm 0.036}$	$\mathbf{0.638}_{\pm 0.046}$	$\mathbf{0.709}_{\pm 0.048}$	**0.655**

3.4 Ablation Studies

We present ablation studies in Tab. 2. Note that the 'Only Intra-Learning' means we only conduct the branch (b) in Fig. 1. On top of it, the 'Basic' merely uses WSIs and CoC-2&4 in branch (a). And we gradually introduce other modalities to form M1 to M4, and our final method.

Text Prompt. In the comparison of using versus not using the CoC-Adapter, we denote them by 'w/' and 'w/o', respectively. Note we give a linear projection for 'w/o' to adjust the shape in order to use AMT. After deploying CoC-Adapter, almost every dataset receives an improvement, as demonstrated in the overall comparisons between M2 and M1 (0.645 *vs.* 0.640), and M4 and M3 (0.641 *vs.* 0.631). This indicates the effectiveness of our text prompt and adapter design. We also explore the use of chain-of-thought textual descriptions. When deploying the vanilla text prompt for these adapters (M5), performance decreases by 1.1% compared to our final model. This suggests that the chain-of-thought mechanism is effective for our task.

Clinical Modal Contribution. With the introduced AMT, our 'Basic' model can interact with the local patches (CoC-2) and global slides (CoC-4) for pathology image, yielding an overall 0.8% improvement compared to the 'Only Intra-Learning'. Based on this 'Basic' setting, we can observe that after adding other modalities (M1-M4) in our intra-learning branch (Fig. 1 (a)), our methods receive gains as well. Moreover, without using methylation data(M1 and M2), our methods still overhead the counterparts (*i.e.*, gene+WSI, methods in Tab. 1). Our final model can leverage triple clinical modalities, producing the best practice. These findings illustrate our AMT can encourage mutual learning among different modalities, and each modality contributes to the final prediction.

4 Conclusion

In this paper, we propose a Chain-of-Cancer (CoC) framework for survival prediction, based on the pathogenesis of cancer, utilizing three modalities of clinical data and language guidance for the first time. The core idea of CoC is to innovatively introduce clinical-related descriptions as textual guidance embedded into the clinical features. We propose an Autoregressive Mutual Traction (AMT) module to encourage synergistic learning among different modalities. Experimental results confirm the effectiveness of our approach, demonstrating its superiority over *state-of-the-art* methods. Additionally, ablation studies indicate that the proposed designs are effective. Our research indicates that by introducing language guidance for multimodal learning, the interpretation can offer significant benefits for feature enhancement. We hope our study offers valuable insights for future research in multimodal learning for survival prediction.

Acknowledgment. This work is supported by the Guangdong Science and Technology Department (No. 2024ZDZX2004) and the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007).

Disclosure of Interests The authors declare that they have no competing interests.

References

1. Amores, J.: Multiple instance classification: Review, taxonomy and comparative study. Artificial intelligence **201**, 81–105 (2013)
2. Bengio, Y., Ducharme, R., Vincent, P.: A neural probabilistic language model. NIPS **13** (2000)
3. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: ICCV. pp. 4015–4025 (2021)
4. COX, D.R.: Partial likelihood. Biometrika **62**(2), 269–276 (08 1975). https://doi.org/10.1093/biomet/62.2.269
5. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
6. Haider, H., Hoehn, B., Davis, S., Greiner, R.: Effective ways to build and evaluate individual survival distributions. JMLR **21**(85), 1–63 (2020)
7. Harrell Jr, F.E., Lee, K.L., Mark, D.B.: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine **15**(4), 361–387 (1996)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
9. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: ICML. pp. 2127–2136. PMLR (2018)
10. Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: CVPR. pp. 11579–11590 (2024)
11. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American statistical association **53**(282), 457–481 (1958)

12. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. NIPS **30** (2017)
13. Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., Iwasawa, Y.: Large language models are zero-shot reasoners. NIPS **35**, 22199–22213 (2022)
14. Li, T., Tian, Y., Li, H., Deng, M., He, K.: Autoregressive image generation without vector quantization. NIPS **37**, 56424–56445 (2024)
15. Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. Nature Medicine **30**, 863–874 (2024)
16. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
17. Qiu, Y.L., Zheng, H., Devos, A., Selby, H., Gevaert, O.: A meta-learning approach for genomic survival analysis. Nature communications **11**(1), 6350 (2020)
18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763. PMLR (2021)
19. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al.: Photorealistic text-to-image diffusion models with deep language understanding. NIPS **35**, 36479–36494 (2022)
20. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. NIPS **34**, 2136–2147 (2021)
21. Syriopoulou, E., Wästerlid, T., Lambert, P.C., Andersson, T.M.L.: Standardised survival probabilities: a useful and informative tool for reporting regression models for survival data. British journal of cancer **127**(10), 1808–1815 (2022)
22. Verma, M., Srivastava, S.: Epigenetics in cancer: implications for early detection and prevention. The lancet oncology **3**(12), 755–763 (2002)
23. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. NIPS **35**, 24824–24837 (2022)
24. Xiong, C., Chen, H., Zheng, H., Wei, D., Zheng, Y., Sung, J.J., King, I.: Mome: Mixture of multimodal experts for cancer survival prediction. In: MICCAI. pp. 318–328. Springer (2024)
25. Xu, Y., Chen, H.: Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In: ICCV. pp. 21241–21251 (2023)
26. Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. IEEE TPAMI **43**(9), 3126–3137 (2020)
27. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. NIPS **30** (2017)
28. Zheng, G., Yang, B., Tang, J., Zhou, H.Y., Yang, S.: Ddcot: Duty-distinct chain-of-thought prompting for multimodal reasoning in language models. NIPS **36**, 5168–5191 (2023)
29. Zheng, H., Momeni, A., Cedoz, P.L., Vogel, H., Gevaert, O.: Whole slide images reflect dna methylation patterns of human tumors. NPJ genomic medicine **5**(1), 11 (2020)
30. Zhou, F., Chen, H.: Cross-modal translation and alignment for survival analysis. In: ICCV. pp. 21485–21494 (2023)