Supplementary Material

Jingwen Xu[1], Ye Zhu[1], Fei Lyu[1], Grace Lai-Hung Wong[2], and Pong C. Yuen[1]

[1] Department of Computer Science, Hong Kong Baptist University
{csjwxu,csyzhu,feilyu, pcyuen}@comp.hkbu.edu.hk
[2] Department of Medicine and Therapeutics, The Chinese University of Hong Kong
wonglaihung@mect.cuhk.edu.hk

Table 1: Data preprocessing and implementation in our framework

Raw Data		**Preprocessing**
CT scans		Multi-Organ Segmentation [3] $S = 3$ Slices Selection (largest liver mask)
Clinical Time Series		-
Modules		**Implementation**
Modal-specific Encoders	CT Encoder	ConvNeXt [1] ($D_{\mathbf{CT}} = 128$)
	TS Encoder	DATA-GRU [2] ($D_{\mathbf{TS}} = 32$)
Fusion Modules (**TNformer-MP**)	Kernel scales (K)	6
	Hidden size ($D_{\mathbf{Fuse}}$)	32
	Prompt Number (P)	4
	Transformer Layers (L)	3

Table 2: Training procedure of our framework

	Uni-Modal Pre-training		**Multi-Modal Fine-tuning**		
	CT Encoder	TS Encoder	CT Encoder	TS Encoder	Fusion Module (**TNformer-MP**)
Initial Lr	0.001	0.0005	0.0001	0.0001	0.001
Optimizer	SGD	Adam	SGD	Adam	Adam
Batch Size	32	32	32		
Iteration	50	50	50		
Prediction Supervision	✓	✓	✓		

References

1. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)

2. Tan, Q., Ye, M., Yang, B., Liu, S., Ma, A.J., Yip, T.C.F., Wong, G.L.H., Yuen, P.: Data-gru: Dual-attention time-aware gated recurrent unit for irregular multivariate time series. In: AAAI. vol. 34, pp. 930–937 (2020)
3. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: CVPR. pp. 20730–20740 (2022)