廣東工業大學
Guangdong University of Technology

Speechmaker： Yejing Huo

BIBM paper report

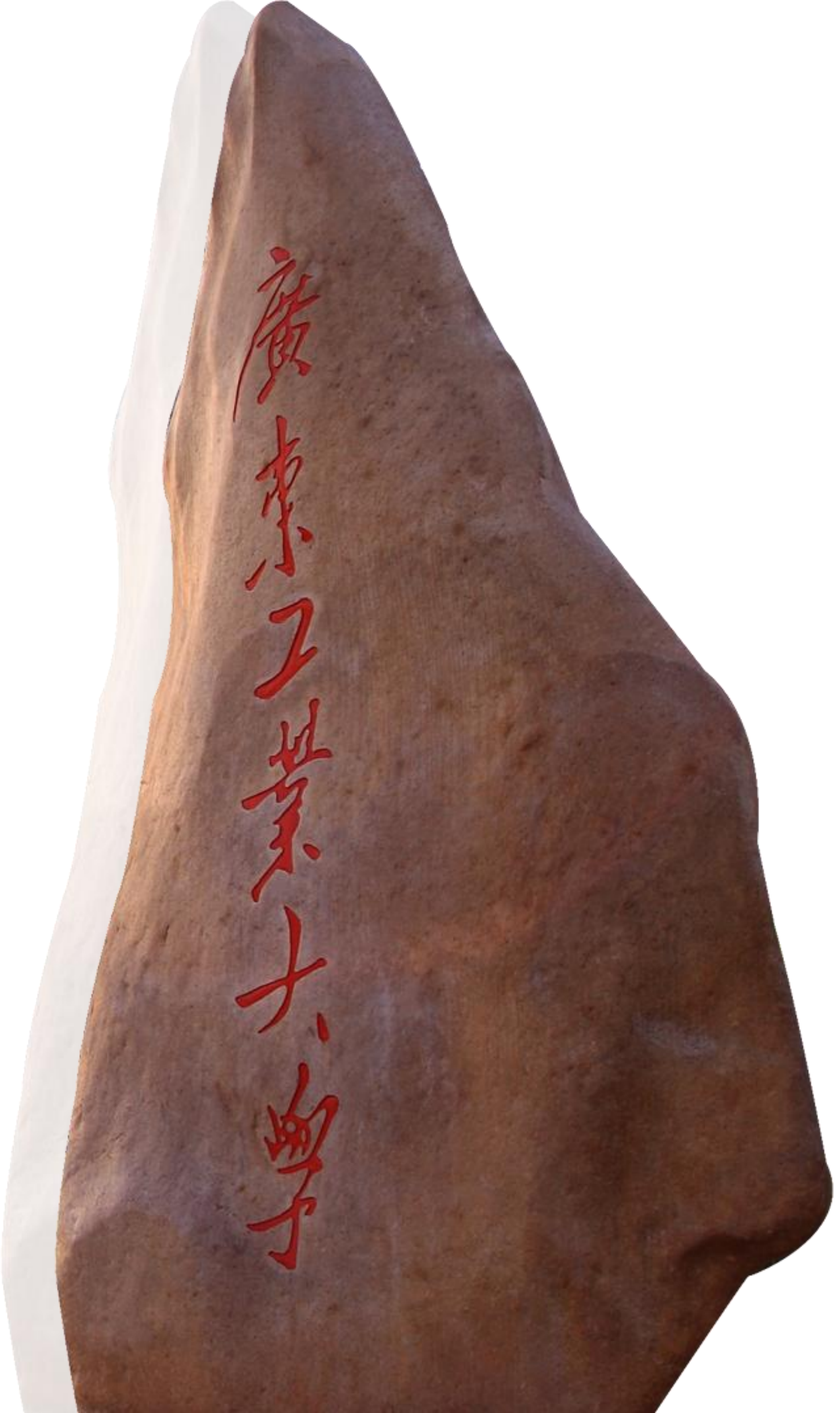
Catalogue

1、Background

2、Our Method

3、Result and Discussion

4、Conclusion and Prospect



Background

Background

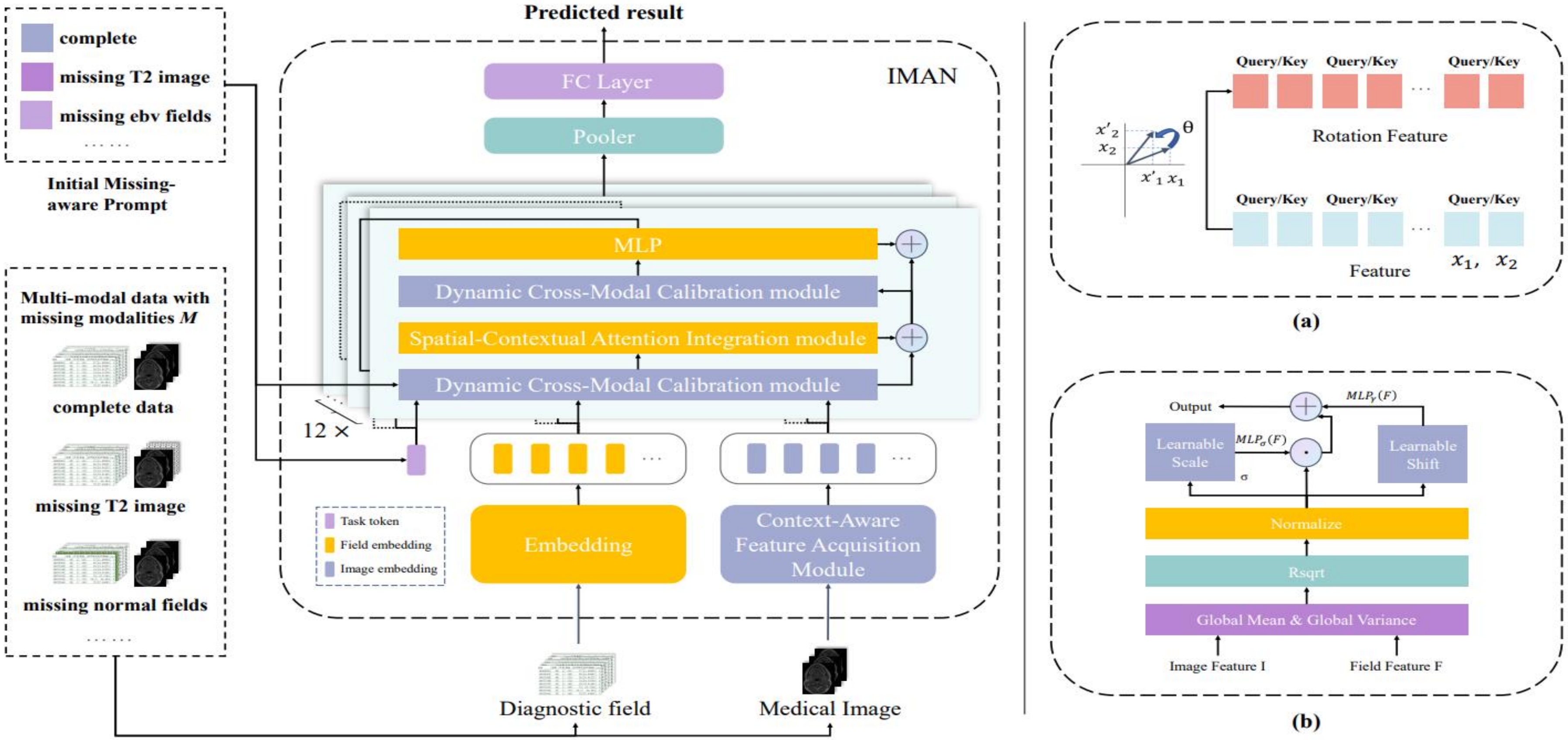
Nasopharyngeal cancer is a highly prevalent malignancy in Southeast Asia, with approximately 80,000 new cases and 50,000 deaths each year. The disease is far more prevalent in southern China and Southeast Asia than in other parts of the globe, with rates 50 to 100 times higher. Early accurate diagnosis and mortality prediction are essential to optimize treatment and improve patient survival. However, the diagnosis and prognosis of nasopharyngeal carcinoma often rely on multimodal data, including imaging data, pathological reports, and gene expression. Due to the high dimensionality, heterogeneity and modal missing of data, it brings great challenges to accurate prediction.

Existing traditional machine learning methods, such as RuleFit and Lasso, exhibit significant performance degradation when dealing with missing data and incomplete modes(a common occurrence in medical procedures). Even the rise of deep learning technologies in recent years has not fully solved this problem. Although the multimodal learning method based on Transformer can better integrate different types of data in theory, it still performs poorly when faced with the absence of high dimensional data. Therefore, faced with the challenges brought by the loss of multimodal data, the existing technical solutions need to be improved.

Our Method

Our Method

Based on this, we propose IMAN model, which is an adaptive network specifically designed to deal with the multi-modal data missing in the prediction of nasopharyngeal cancer mortality. The IMAN(as shown in the following picture) model consists of three core modules, each specifically designed for a specific multimodal data processing problem.



Dynamic Cross-Modal Calibration Module

We know that when dealing with multimodal data, data of different modes often have different scales and feature distributions, especially the difference between medical images and structured text data is very large. Therefore, simply feeding this data directly into the model can result in significant performance degradation. To solve this problem, the DCMC module dynamically adjusts and aligns image data and text data with learnable parameters. Specifically, the DCMC module computes its global mean and variance for each mode of the input, and then normalizes these modes using learnable scaling and offset parameters. This method can not only effectively calibrate the input data of different modes, but also enhance the alignment between different modal features through adaptive adjustment, so as to improve the fusion effect of multi-modal data.

The key formula is as follows:

$$\mathrm{DCMC}(I, \mathbf{F}) = \sigma(\mathbf{F}) \odot \left(\frac{I - \mu(I)}{\sigma(I)} \right) + \gamma(\mathbf{F})$$

$$\sigma(\mathbf{F}) = \mathrm{MLP}_{\sigma}(\mathbf{F}), \quad \gamma(\mathbf{F}) = \mathrm{MLP}_{\gamma}(\mathbf{F})$$

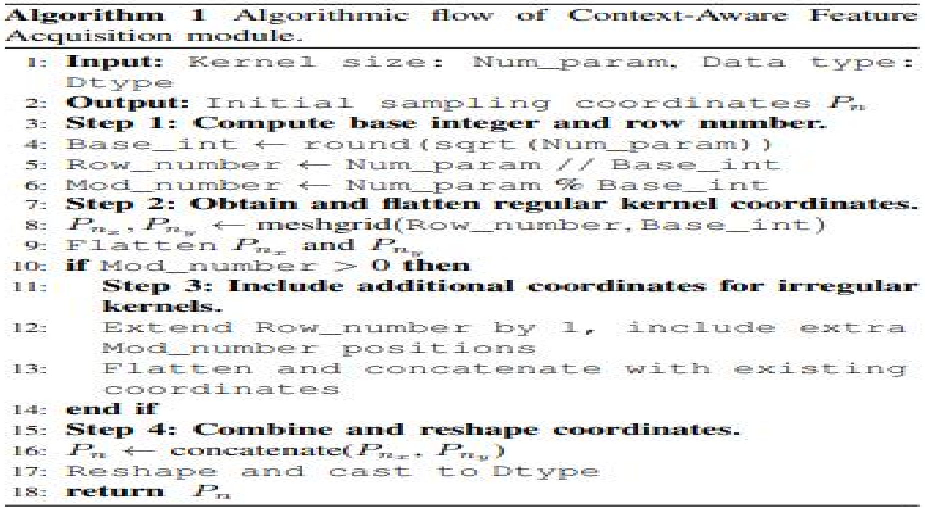
Spatial-Contextual Attetion Integration Module

In multimodal learning, how to effectively integrate the characteristics of each mode is the key. Although the traditional Transformer model has achieved remarkable success in natural language processing and image processing tasks, it cannot adequately capture the complex spatial information in medical image data. Therefore, we have improved Transformer's self-attention mechanism in the SCAI module to include location information. This module imparts positional relationships by rotationally encoding, enabling the model to better capture spatial patterns in the image. Different from fixed absolute position coding, rotation coding can dynamically adjust position representation according to the change of input sequence, so it takes into account the preservation of spatial structure in the image and the information fusion between modes while processing medical images. SCAI module enables IMAN to capture more delicate cross-modal patterns when integrating multi-modal features, effectively improving the fusion effect.

$$
\begin{aligned}
&< P_q(x_s, s), P_k(x_t, t) > \\
&= \left(\begin{pmatrix} \cos(s\theta) & -\sin(s\theta) \\ \sin(s\theta) & \cos(s\theta) \end{pmatrix}^T \begin{pmatrix} q_s^{(1)} \\ q_s^{(2)} \end{pmatrix} \right)^T \\
&\quad \begin{pmatrix} \cos(t\theta) & -\sin(t\theta) \\ \sin(t\theta) & \cos(t\theta) \end{pmatrix} \begin{pmatrix} k_t^{(1)} \\ k_t^{(2)} \end{pmatrix} \\
&= \begin{pmatrix} q_s^{(1)} & q_s^{(2)} \end{pmatrix} \begin{pmatrix} \cos((s-t)\theta) & -\sin((s-t)\theta) \\ \sin((s-t)\theta) & \cos((s-t)\theta) \end{pmatrix} \begin{pmatrix} k_t^{(1)} \\ k_t^{(2)} \end{pmatrix}
\end{aligned}
$$

Context-Aware Feature Acquisition Module

Feature acquisition of medical images usually relies on convolutional neural networks (CNNS), but traditional convolutional operations have a fixed receptive field, which limits its ability to capture features at different scales and orientations. However, in nasopharyngeal carcinoma image data, different lesion areas may exist in different scales and directions, and a single fixed convolution operation is difficult to adapt. In order to deal with this problem, CAFA module dynamically adjusts the position of convolution kernel by introducing learnable convolution kernel offset. This mechanism allows the model to generate more adaptive feature representations for the input image data in the convolution operation, thus improving the ability of the model to extract image features at different scales and directions. This dynamic adjustment not only applies to standard images, but also ensures effective information acquisition when modes are missing.



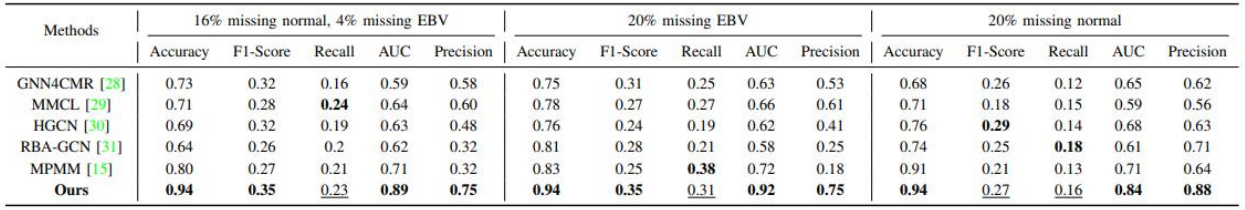
Algorithm 1 Algorithmic flow of Context-Aware Feature Acquisition module.

```
Input: Kernel size: Num_param, Data type: Dtype
Output: Initial sampling coordinates $P_n$
Step 1: Compute base integer and row number.
Base_int ← round(sqrt (Num_param))
Row_number ← Num_param // Base_int
Mod_number ← Num_param % Base_int
Step 2: Obtain and flatten regular kernel coordinates.
$P_{n_x}$, $P_{n_y}$ ← meshgrid(Row_number, Base_int)
Flatten $P_{n_x}$ and $P_{n_y}$
if Mod_number > 0 then
    Step 3: Include additional coordinates for irregular kernels.
    Extend Row_number by 1, include extra Mod_number positions
    Flatten and concatenate with existing coordinates
end if
Step 4: Combine and reshape coordinates.
$P_n$ ← concatenate($P_{n_x}$, $P_{n_y}$)
Reshape and cast to Dtype
return $P_n$
```

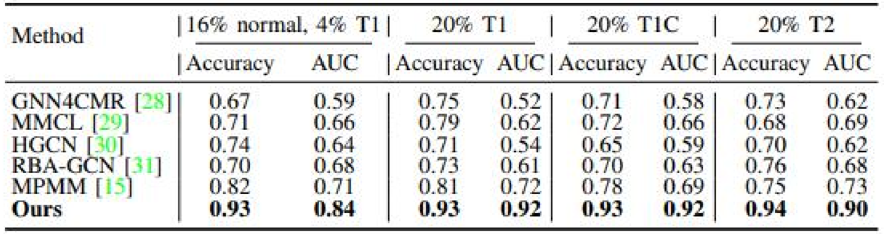
Result and Discussion

Result and Discussion

Our comprehensive experiments comparing IMAN with state-of-the-art methods across various missing data scenarios demonstrate IMAN's superior performance, particularly in Accuracy and AUC. For example, with 20% missing EBV data, IMAN achieved an accuracy of 0.94 and an AUC of 0.92, significantly outperforming MPMM's accuracy of 0.83 and AUC of 0.72. However, the relatively lower F1-Score and Recall metrics may be due to the inherent class imbalance in nasopharyngeal carcinoma datasets, where negative cases typically outnumber positive ones. Additionally, IMAN may reflect a conservative diagnostic approach, leading to fewer false positives but potentially more false negatives. The complexity of cancer progression, influenced by factors like genetic expression and treatment response, may result in some rare predictive features being overlooked. Furthermore, predicting long-term mortality poses challenges in capturing all relevant factors. While IMAN excels in handling missing data, the absence of certain key predictive factors may still impact its performance in identifying all positive cases. The model's design, focusing on optimizing accuracy and AUC, might have led to trade-offs in F1-Score and Recall. Despite these limitations, IMAN's overall performance surpasses existing methods, especially in scenarios with incomplete data, highlighting its potential to enhance the accuracy and reliability of mortality predictions in clinical settings for nasopharyngeal carcinoma.



Methods	16% missing normal, 4% missing EBV					20% missing EBV					20% missing normal				
	Accuracy	F1-Score	Recall	AUC	Precision	Accuracy	F1-Score	Recall	AUC	Precision	Accuracy	F1-Score	Recall	AUC	Precision
GNN4CMR [28]	0.73	0.32	0.16	0.59	0.58	0.75	0.31	0.25	0.63	0.53	0.68	0.26	0.12	0.65	0.62
MMCL [29]	0.71	0.28	**0.24**	0.64	0.60	0.78	0.27	0.27	0.66	0.61	0.71	0.18	0.15	0.59	0.56
HGCN [30]	0.69	0.32	0.19	0.63	0.48	0.76	0.24	0.19	0.62	0.41	0.76	**0.29**	0.14	0.68	0.63
RBA-GCN [31]	0.64	0.26	0.2	0.62	0.32	0.81	0.28	0.21	0.58	0.25	0.74	0.25	**0.18**	0.61	0.71
MPMM [15]	0.80	0.27	0.21	0.71	0.32	0.83	0.25	**0.38**	0.72	0.18	0.91	0.21	0.13	0.71	0.64
Ours	**0.94**	**0.35**	0.23	**0.89**	**0.75**	**0.94**	**0.35**	0.31	**0.92**	**0.75**	**0.94**	0.27	0.16	**0.84**	**0.88**



Method	16% normal, 4% T1		20% T1		20% T1C		20% T2	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
GNN4CMR [28]	0.67	0.59	0.75	0.52	0.71	0.58	0.73	0.62
MMCL [29]	0.71	0.66	0.79	0.62	0.72	0.66	0.68	0.69
HGCN [30]	0.74	0.64	0.71	0.54	0.65	0.59	0.70	0.62
RBA-GCN [31]	0.70	0.68	0.73	0.61	0.70	0.63	0.76	0.68
MPMM [15]	0.82	0.71	0.81	0.72	0.78	0.69	0.75	0.73
Ours	**0.93**	**0.84**	**0.93**	**0.92**	**0.93**	**0.92**	**0.94**	**0.90**

Conclusion and Prospect

Conclusion and Prospect

In this study, we introduce IMAN, an adaptive network for predicting mortality in nasopharyngeal carcinoma cases with missing modalities. IMAN includes several key modules: the DCMC module normalizes heterogeneous inputs by scaling medical images and field data using learnable parameters, the SCAI module enhances multi-modal feature fusion with positional information in a self-attention mechanism, and the CAFA module adjusts convolution kernel positions for adaptive feature capture. These components work together to improve the precision of the system in handling medical data, enabling the detection of subtle patterns and effective feature capture across various scales and orientations. Experimental results in different missing data scenarios show that our method outperforms current alternatives. Future work will explore the performance of our model under varying data missing rates and more complex missing data scenarios.

廣東工業大學
Guangdong University of Technology

Thank you for listening！