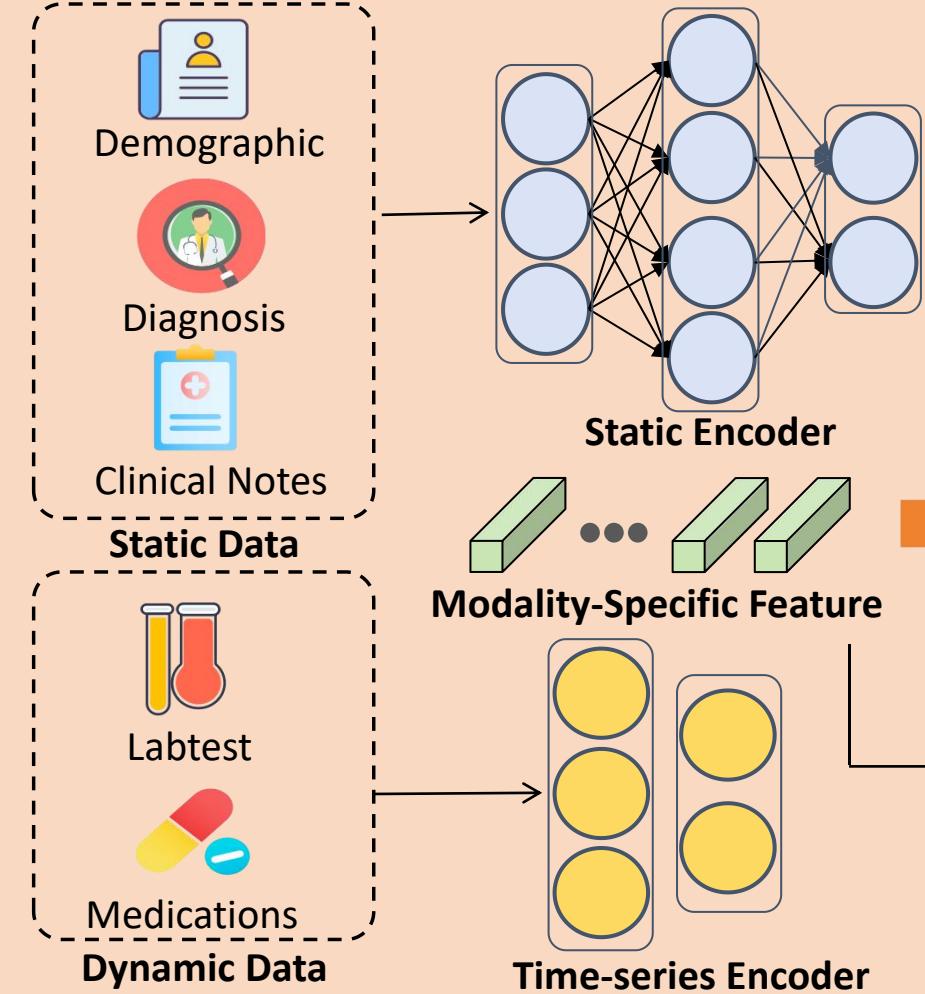


Modality-Specific Feature Encoding



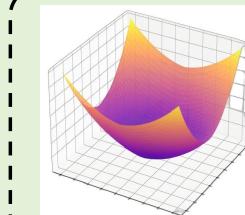
Task-aware Variational Information Bottleneck Modulation

Modality-Specific Feature Task Embedding

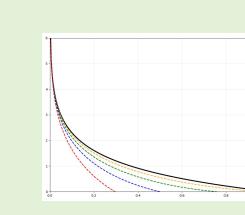
Modality-Specific Feature + Task Embedding



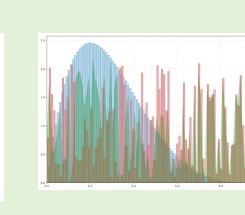
Dimension-wise Gating Network



L1 Regularization



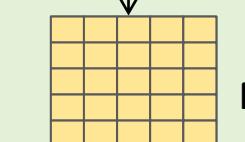
Entropy Regularization



Sparsity Matching

Sparsity Regularization

Element-wise Multiplication



Weight Matrix

Modulated Feature

↓

↓

Convert to Matrix

Sigmoid



Weight Logits

$$W_i^m = \text{MultiHeadAttention}(\text{Input}_i^m)$$

EHR Downstream Task

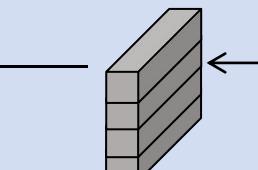
Classification Loss

Regression Loss

Prediction Layer

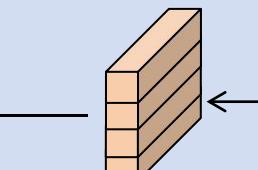
Fusion Layer

Enhanced Static Feature



$$[r_{in}] = E_n([r_{in-1}, l_i])$$

Enhanced Dynamic Feature



Feed Forward

Feed Forward

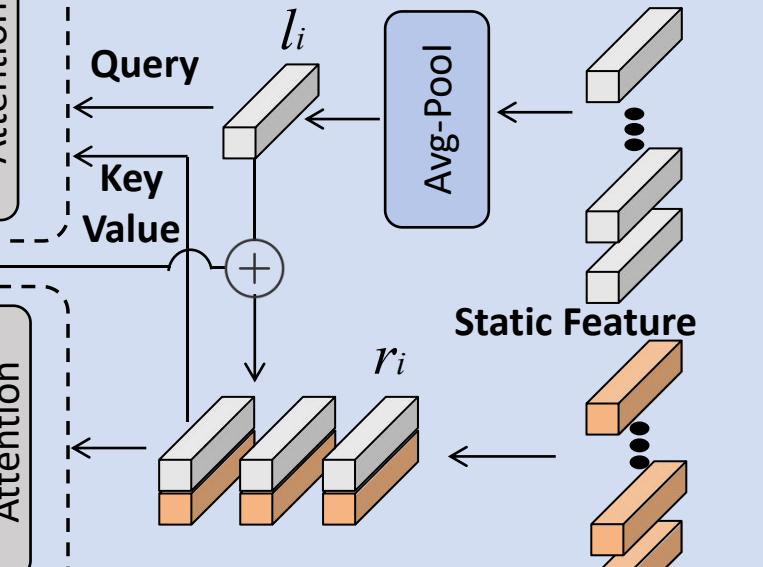
Cross Attention

Cross Attention

Self Attention

Self Attention

Asymmetric Feature Fusion



Asymmetric Feature Fusion

Asymmetric Feature Fusion