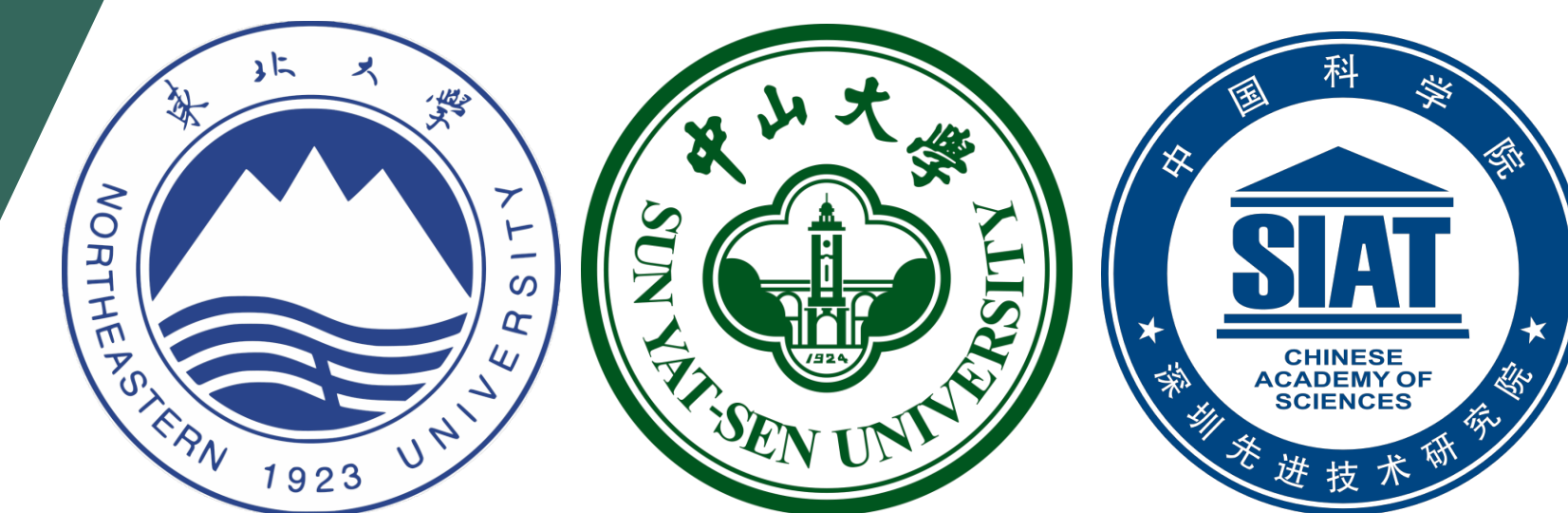


A Multiscale Frequency Domain Causal Framework for Enhanced Pathological Analysis

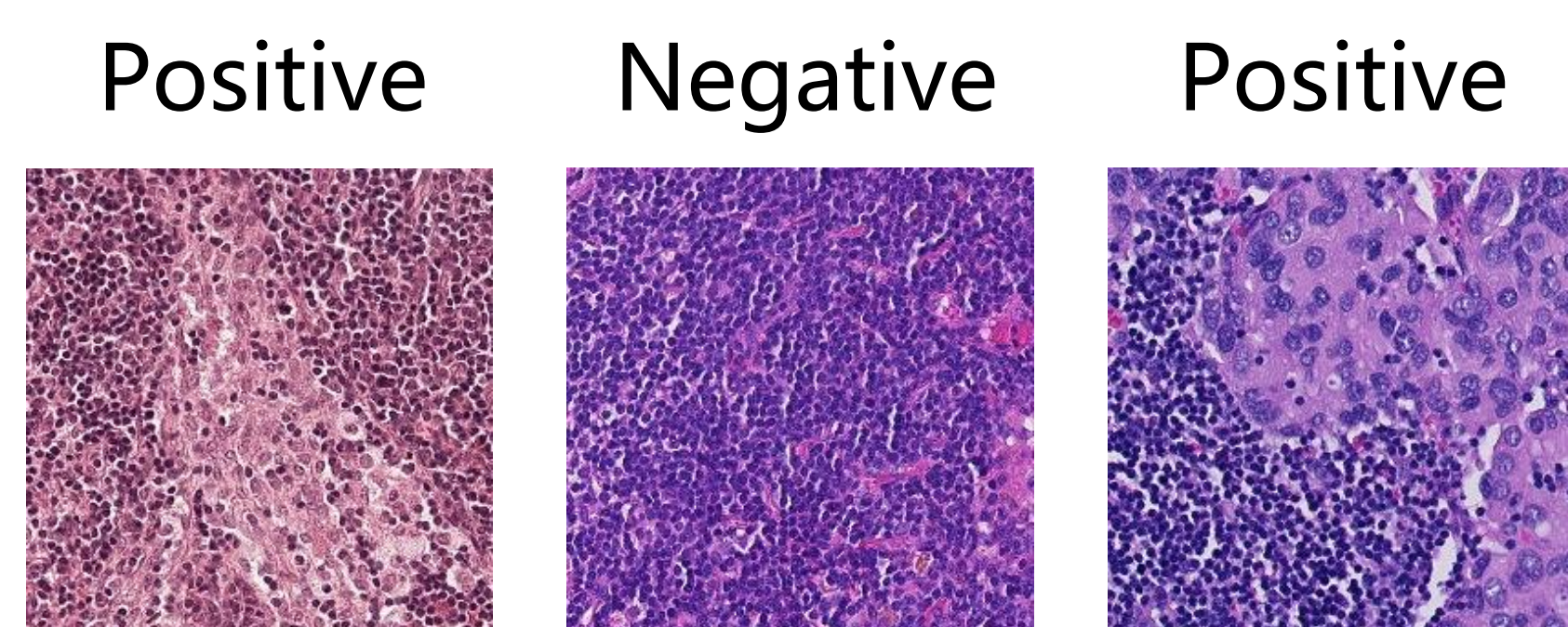
Xiaoyu Cui*, Weixing Chen*, Jiandong Su*



Contributions

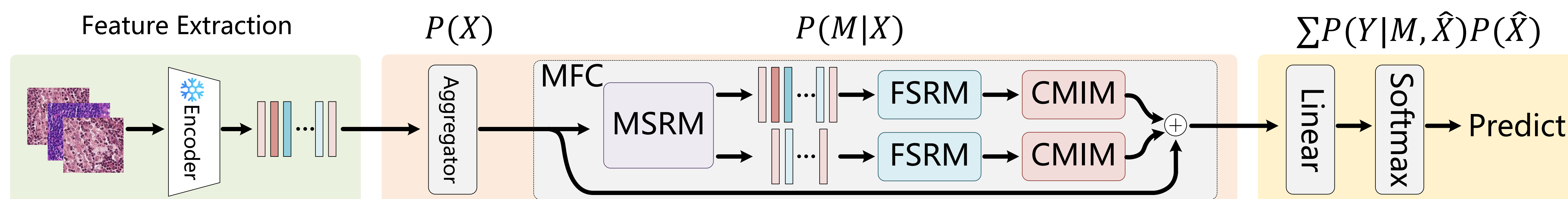
- The MFC framework aims to **eliminate spurious correlations** caused by confounders, ensuring the model relies on true diagnostic features.
- It integrates **multi-scale spatial information** to capture both global tissue and local cellular details comprehensively.
- It enhances robustness by **leveraging frequency-domain analysis**, reducing interference from staining and color biases.

Motivations



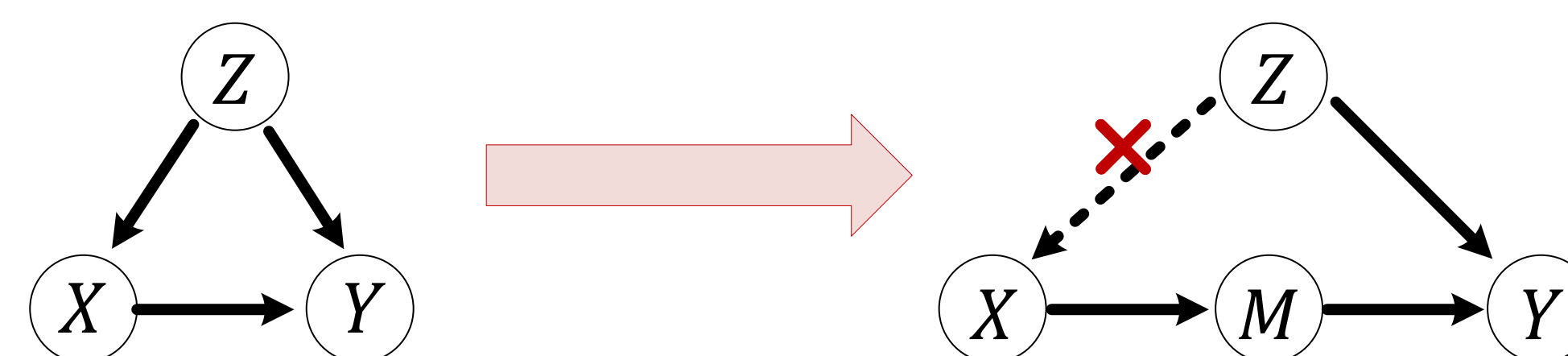
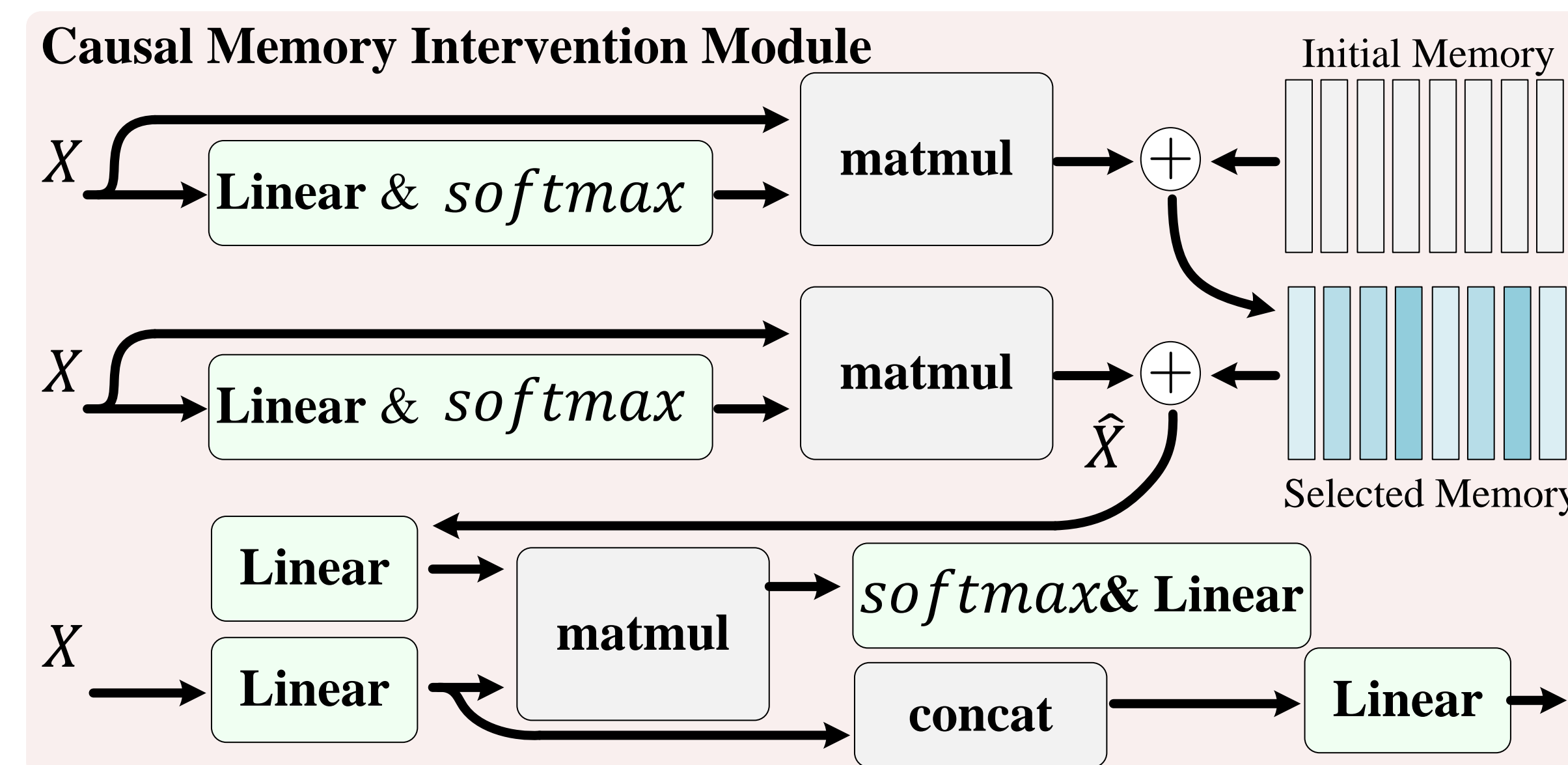
- **Causal intervention** is employed to address confounders like color biases, which can lead to spurious correlations and incorrect classifications.
- **Frequency-domain information** is used to capture robust structural details that are critical for accurate diagnosis but often missed in the spatial domain.
- Medical diagnosis requires integrating **high-magnification cellular structures** with **low-magnification tissue structures** to provide a comprehensive understanding for accurate classification.

Overview



The MFC framework combines a Causal Memory Intervention Module (**CMIM**) to mitigate spurious correlations, a Multi-scale Spatial Representation Module (**MSRM**) to integrate multi-level features, and a Frequency-domain Structural Representation Module (**FSRM**) to extract robust structural information

CMIM



Due to the presence of confounders, the non-causal model does not follow the correct causal path $X \rightarrow Y$ for prediction. Instead, it relies on spurious correlations established by the confounders, following the incorrect path $X \leftarrow Z \rightarrow Y$, which can be formulated as:

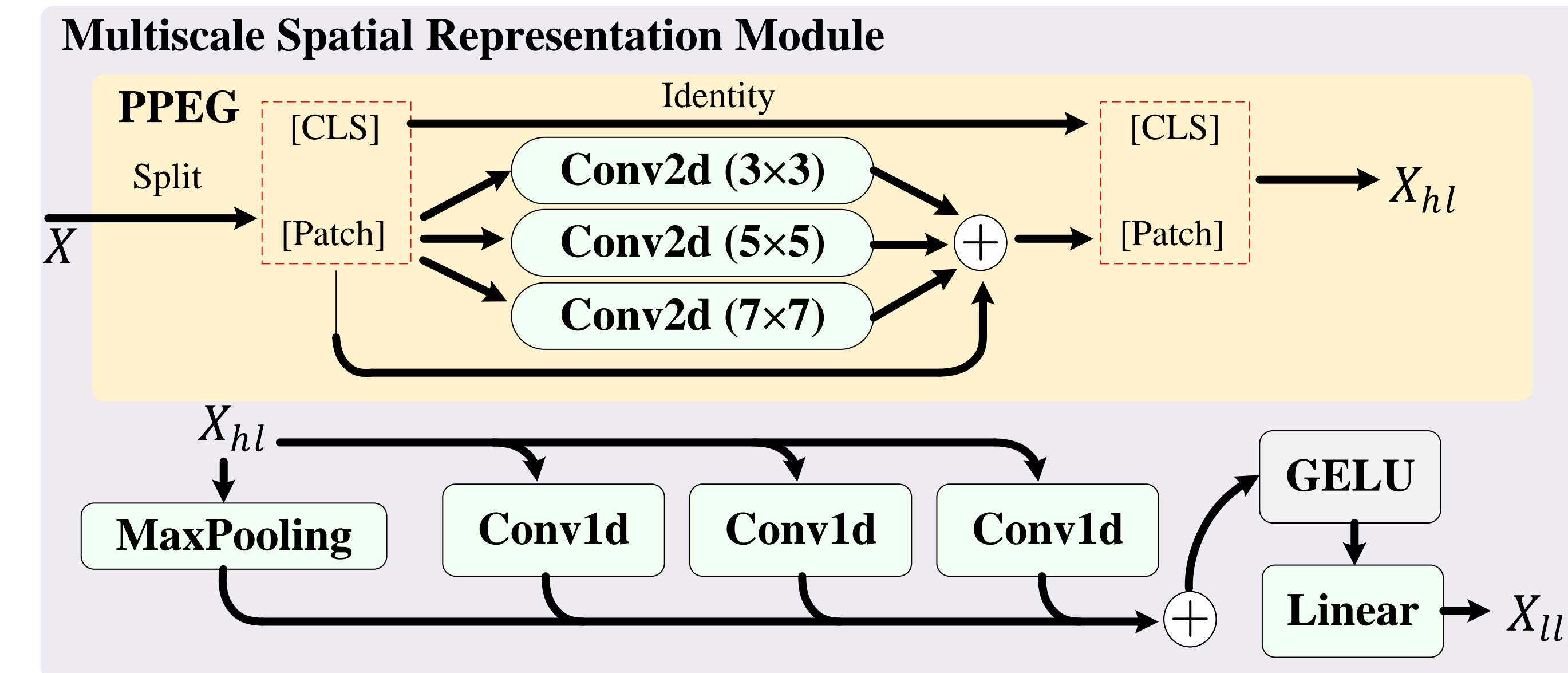
$$P(Y|X) = \sum_z P(Y|X, Z=z)(Z=z|X)$$

We introduce a mediator M derived from the X , and estimates the causal effect $P(Y|do(X))$ through the mediator as:

$$P(Y|do(X)) = \sum_m P(M=m|X=x) \sum_{\hat{x}} P(Y|X=\hat{x}, M=m) P(X=\hat{x})$$

where \hat{X} represents potential values of X .

MSRM



The MSRM applies Position-aware Patch Embedding Generation (PPEG) for multi-scale positional encoding, followed by sampling with multiple large-kernel convolutions to capture spatial relationships across different scales.

FSRM

The FSRM utilizes the Hilbert transform to convert input features into analytic signals, capturing both amplitude and phase information, and then maps these transformed features back to the original feature space using an output projection layer, enhancing the model's ability to capture complex structural and textural variations. The transformation is defined as:

$$x_a(t) = x(t) + j\hat{x}(t)$$

where $x(t)$ is the original signal, $\hat{x}(t)$ is its Hilbert transform, and j is the imaginary unit. The overall transformation function is:

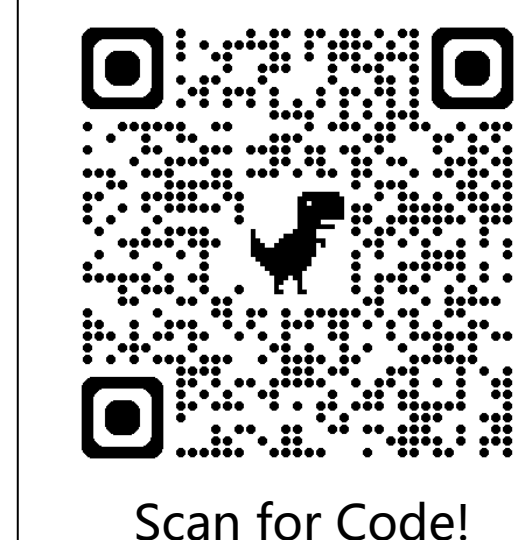
$$F(x) = x + g(H(f(x)))$$

where f and g are linear mappings, and H is the Hilbert transform operator.

Result

Method	ACC	AUC	Pre.	Rec.	F1	Spe.
Baseline	86.98	94.95	96.14	68.98	79.41	98.00
IBMIL	91.78	96.31	94.85	83.67	88.50	96.75
MFC-MIL	92.25	95.41	94.96	84.08	89.13	97.25

Comparison of the results (%) of our MFC-MIL and IBMIL on the Camelyon16 dataset, and the baseline is DSMIL



Scan for Code!