Algorithm S1 The FRCM-MIL training algorithm

Input:

- X : Bag of instance from WSIs, $X = \{x_1, x_2, ..., x_m\}$

Output:

- $\hat{Y}$: Bag-level predicted label

1. Preprocess 20× WSIs into 1024 × 1024 image patches;

2. Extract feature vectors from image patches using a pretrained feature extractor:

$F = \{f_1, f_2, ..., f_m\} \leftarrow X$, where $f_i \in \mathbb{R}^{1 \times d}, F \in \mathbb{R}^{m \times d}$

3. Select discriminative features:

$F' = \{f_1, f_2, ..., f_n\} \leftarrow F$, where $F' \in \mathbb{R}^{n \times d}$

4. Apply the SFRM-WT to the reconstructed features F_r :

$F_r = SFRM - WT(F')$, where $F_r \in \mathbb{R}^{1 \times n \times d}$

5. Use the CQAM module to aggregate confidence features F_q :

$F_q = CQAM(F')$, where $F_q \in \mathbb{R}^{1 \times k \times d}$

6. Feed F_r and F_q into the feature cross-attention module (CAM):

$\hat{F} = CAM(F_r, F_q)$, where $\hat{F} \in \mathbb{R}^{1 \times k \times d}$

7. Transform $\hat{F}$ from to obtain the final bag-level prediction $\hat{Y}$:

$logits = MLP(LN(\hat{F}))$

$\hat{Y} = argmax(logits)$

8. Return $\hat{Y}$

Algorithm S2 Pseudocode of the SFRM-WT Module

Input:

- X: Token features of shape (c,n,d)

- D_M: Feature modulation matrix of shape $(c,h,w,2)$, where $d = h*w$

Output:

- X: Processed token features of shape (c,n,d)

1. Initialize D_M weight with shape $(c,h,w,2)$;

2. Reshape X from (c,n,d) to (n,c,h,w):
$X = reshape(X,(n,c,h,w))$

3. Apply 2D discrete wavelet transform (DWT2) on X:
$X\_ll, X\_lh, X\_hl, X\_hh = DWT2(X)$

4. Concatenate the DWT subbands along the channel dimension:
$X = cat([X\_ll, X\_lh, X\_hl, X\_hh], dim = 1)$

5. Apply feature modulation with the matrix D_M:
$X\_tilde = X*D_M$

6. Apply 2D inverse discrete wavelet transform (IDWT2) on X_tilde:
$X = IDWT2(X\_tilde)$

7. Reshape X back to (c,n,d):
$X = reshape(X,(c,n,d))$

8. Return X