Supplementary Material A: Detailed Model Architecture and Components

DETAILS OF THE MOTCN ARCHITECTURE

The MoTCN (Modern Temporal Convolutional Network) component is designed to effectively extract long-term, multi-variable, and local features from flight state sequences while strictly enforcing causality. Its architecture is shown in Figure S1.

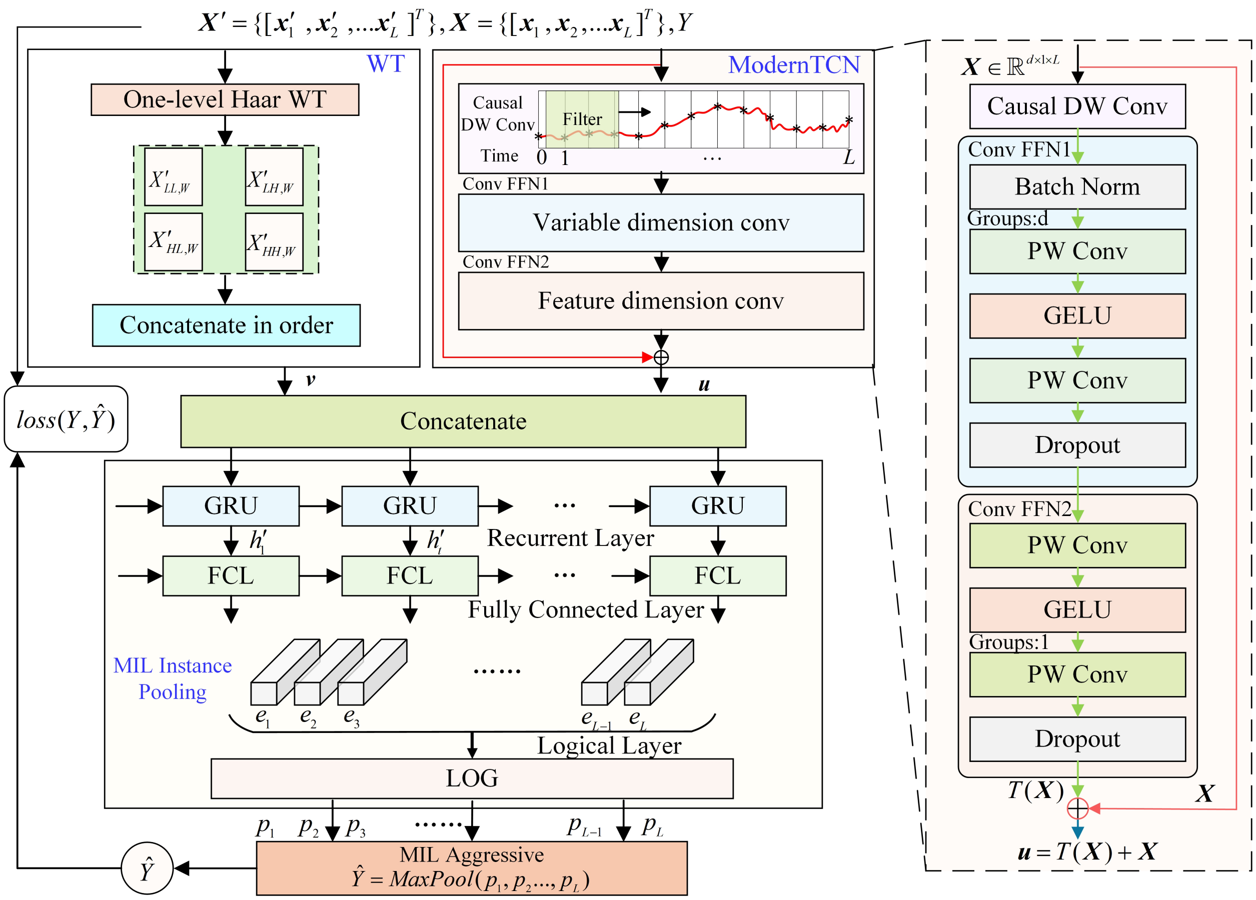

FIG. S1: The detailed structure of the MoTCN component and the data flow within the model. The *Groups* is the group number in group convolution.

The MoTCN block utilizes a residual connection for stable training and features two primary components:

1. **Causal Feature Extraction:** We employ Causal Depth-Wise Convolution (CDW Conv) and causal padding to restrict the model from accessing future data. This design ensures that the feature extraction at time t depends strictly on data from time $\leq t$, which is fundamental to the CaM-MIL's real-time precursor identification capability.
2. **Group Convolution for Dependency Modeling:** The architecture utilizes Point-Wise Convolution (PW Conv) with specific group settings (indicated by $Groups$). The ConvFFN1 and ConvFFN2 modules can separately Temporal dependencies for each variable, and coupling relationships between different variables .

The GELU activation function is used for stable and efficient training, and the residual connection ensures the flow of information across network layers. The input data $\boldsymbol{X}$ passes through the MoTCN component $T(\cdot)$ and is combined via the residual connection to output $\boldsymbol{u}$.

DETAILS OF THE CAUSAL WAVELET TRANSFORM (WT)

To ensure that operational features from control variables (e.g., flap handle positions) are adequately modeled while strictly maintaining the causal relationship required for real-time detection, we employ a custom Causal Wavelet Transform (WT). This stream uses a sliding window approach to extract multi-scale temporal dependencies.

The input operational data $\boldsymbol{X'} \in \mathbb{R}^{L \times fs \times d'}$ is divided into multiple non-overlapping segments. A one-level 2-dimensional Haar wavelet transform is performed within each fixed-length window W using the following four depth-wise convolutional filters:

$$f_{LL} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad f_{LH} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad f_{HL} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad f_{HH} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{S1}$$

Here, f_{LL} is the low-pass filter, capturing approximation information, while f_{LH}, f_{HL}, and f_{HH} are high-pass filters, capturing horizontal, vertical, and diagonal detail components, respectively. The resulting outputs are concatenated in a timely order to form the feature vector $\boldsymbol{v}$, ensuring the temporal characteristics are preserved. The utilization of a fixed-length sliding window is the key mechanism to guarantee strict causality during the WT process, as data from future time steps is never included in the current window's calculation.

Figure S2 illustrates the detailed data flow of our Causal Wavelet Transform implementation. A fixed-length window is applied sequentially to the input data, and the WT results within that window are concatenated to ensure that all generated features maintain the correct temporal order and causal dependency.

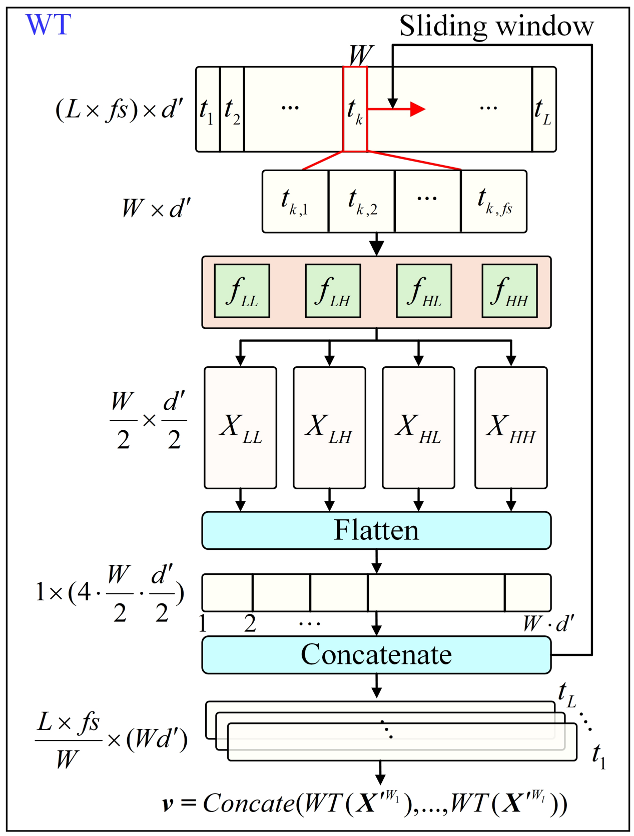

FIG. S2: Detailed process of the Causal Wavelet Transform (Causal WT).