PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

WBMCF: Robust Spatiotemporal Forecasting of Terrorism Fatalities with Multi-Scale Time–Frequency Fusion

Zhenkai Qin[1] , **Baozhong Wei**[1] , **Huan Zeng**[1] , **Caifeng Gao**[1] and **Ziqian Lin**[2*]
[1]School of Information Technology, Guangxi Police College, Nanning 530028, China
[2]Editorial Department, Guangxi Police College, Nanning 530028, China
qinzhenkai@gxjcxy.edu.cn, weibaozhong@gxjcxy.edu.cn, huanzeng705@gmail.com, gaocaifeng22@gmail.com, linziqian@gxjcxy.edu.cn

Abstract

Reliable forecasting of cross-regional terrorism fatalities is essential for early warning and security planning. However, real-world terrorism data are characterized by extreme sparsity, abrupt volatility, and weak, rapidly shifting non-Euclidean spatial relationships, posing substantial challenges for existing spatiotemporal forecasting models in learning stable multi-scale dependencies. To address these challenges, we propose a robust spatiotemporal forecasting framework, termed WBMCF. In the temporal dimension, a single-level Discrete Wavelet Transform (DWT) is introduced to stabilize temporal structures, and a Bidirectional Gated Mamba module is employed to capture stage-dependent long-range temporal dependencies; additionally, in the spatial dimension, a topology-agnostic Convolutional Feed-Forward Network (ConvFFN) is incorporated to model cross-regional dependencies via implicit feature coupling, thereby eliminating reliance on fixed spatial topologies. Extensive experiments on five regional subsets of the Global Terrorism Database (GTD) demonstrate that WBMCF achieves stable and competitive forecasting performance under sparse and noisy conditions. The source code is available at https://github.com/weibaozhong/WBMCF.

1 Introduction

Terrorism remains a critical destabilizing force in the persistently volatile global security landscape [Wei *et al.*, 2024]. Long-term analysis of the Global Terrorism Database (GTD) [LaFree and Dugan, 2007] reveals that incidents in South Asia, the Middle East, and Sub-Saharan Africa exhibit pronounced episodic fluctuations [Xu *et al.*, 2024] and cross-border spillovers [Rosenberg and Avdan, 2025] driven by political conflict, armed confrontations, and organizational restructuring. Crucially, these cross-regional interactions are non-Euclidean and transcend fixed geographical adjacency, instead reflecting shifting geopolitical alignments, emergent militant networks, and the episodic activation of latent conflict zones. Consequently, spatial relationships among regions reorganize continuously and abruptly, often manifesting as "jump diffusion" behavior [Kibris, 2021] where instability transmits between geographically distant but ideologically connected areas. Capturing such dynamic, discontinuous, and topology-agnostic dependencies poses a persistent challenge for terrorism forecasting [Shao *et al.*, 2022].

Terrorism-induced fatality sequences [Hou *et al.*, 2023] diverge fundamentally from conventional socioeconomic time series, characterized by extreme sparsity, abrupt spikes, long dormant intervals, and multi-scale non-stationarity [Cheng *et al.*, 2023]. These features violate the stationarity and linearity assumptions [Petropoulos *et al.*, 2022] inherent in classical techniques like ARIMA and VAR [Zeng *et al.*, 2023]. Furthermore, decomposition-based methods (e.g., STL, Prophet) often conflate high-frequency bursts with trend shifts. While sophisticated modal decomposition approaches (e.g., VMD, CEEMDAN) and fixed-mapping time–frequency representations (e.g., Wavelet Scattering) offer improvements [Deng *et al.*, 2024], they frequently yield unstable components or fail to adapt to evolving diffusion patterns. These limitations underscore the necessity for a representation that is robust to noise and capable of disentangling long-term geopolitical tendencies from sharp local outbreaks [Cao *et al.*, 2025; Dong *et al.*, 2025].

Although deep learning has advanced nonlinear temporal modeling, significant gaps remain in addressing the unique spatial dynamics of terrorism data. Recurrent models (LSTM, GRU) suffer from memory degradation under extreme sparsity as the underlying information flow is frequently disrupted by prolonged dormant intervals [Schirmer *et al.*, 2022]. Similarly, Transformers, despite their expressiveness, are noise-sensitive [Cheng *et al.*, 2023; Qin *et al.*, 2025b] and prone to overfitting anomalies in weakly structured and sparse sequences [Liu *et al.*, 2022c]. More critically, conventional spatiotemporal models (e.g., GCN, GAT, DCRNN) typically rely on explicit graph topologies to guide information propagation. This approach imposes a strong inductive bias favoring smooth diffusion across defined neighbors, which is ill-suited for terrorism's sparse and erratic propagation paths. When spatial relationships undergo sudden reorganization—due to shifting alliances or cross-border interventions [Liang *et al.*, 2024]—forcing an explicit graph structure often intro-

*Corresponding author

duces noise through false connections [Zhang *et al.*, 2023], rendering prediction unstable. While State Space Models like Mamba enhance long-range temporal reasoning [Gu and Dao, 2024], they lack intrinsic mechanisms to incorporate dynamic, implicit cross-regional coupling and robustly filter the sporadic impulse noise [Weng *et al.*, 2023] characteristic of sparse datasets.

To address the multifaceted challenges of extreme sparsity, multi-scale non-stationarity, and explicit graph topology mismatch, we propose a unified forecasting framework: Wavelet–Bidirectional Gated Mamba–Convolution Fusion (WBMCF). The core of the framework lies in a hierarchical information reconstruction mechanism: first, a single-level discrete wavelet transform (DWT) is utilized to map disjointed raw impulses into a time–frequency representation with a more continuous energy distribution, effectively stabilizing the feature structure of sparse sequences at the signal source. Building upon this robust foundation, a Bidirectional Gated Mamba module adaptively modulates information gain through content-aware state transitions, which filters out interference from prolonged silent intervals while enhancing the capacity to capture critical volatility feedback in non-stationary environments. Regarding spatial dependencies in weak-topology settings, we introduce a topology-agnostic enhancement module (ConvFFN) that leverages implicit channel-mixing to facilitate feature transfer and pattern sharing across regions. This allows the model to adaptively learn a global interaction map of shifting geopolitical dynamics without requiring a predefined adjacency matrix, ensuring stability even in information-poor scenarios.

Overall, WBMCF integrates multi-scale temporal structure, bidirectional reasoning, and implicit spatial modeling into a unified framework tailored to the volatility of terrorism data. Experiments across five regional datasets demonstrate that WBMCF consistently outperforms state-of-the-art temporal, frequency-domain, and explicit graph-based models, particularly in long-horizon, sparse, and noisy scenarios. The framework thus provides a principled solution for forecasting fatalities in geopolitically dynamic environments where explicit topology is weak or unavailable.

The main contributions of this paper are summarized as follows:

- We investigate terrorism fatality forecasting under a weak-topology spatiotemporal setting, where sequences are sparse, bursty, non-stationary, and affected by non-Euclidean cross-regional diffusion, differing from conventional tasks with stable physical graphs.
- We propose WBMCF, a unified framework that integrates multi-scale time–frequency representation, bidirectional state-space modeling, and topology-agnostic cross-regional feature coupling to learn robust spatiotemporal dependencies without predefined adjacency structures.
- We introduce a single-level DWT representation and a Bidirectional Gated Mamba encoder to stabilize sparse temporal patterns while preserving burst dynamics, together with ConvFFN to enhance implicit inter-regional modeling under weak spatial relations.
- Experiments on five GTD regional subsets show that WBMCF achieves competitive and stable performance, while ablation, sensitivity, robustness, and efficiency analyses further validate the effectiveness of the proposed design.

2 Problem Setting

We study the problem of cross-regional terrorism fatality forecasting, which aims to model how terrorism-induced death counts evolve jointly over time across multiple regions at a fixed temporal resolution. To keep the formulation concise and consistent, we focus on region-level aggregated fatality sequences rather than event-level or individual-level modeling.

Regional System. Consider a regional system consisting of C regions, denoted by the set $\mathcal{R} = \{r_1, r_2, \ldots, r_C\}$. At each time step t, the system state is represented by a C-dimensional vector, where each entry corresponds to the number of fatalities caused by terrorism in a specific region at time t. The temporal evolution of all regions together thus forms a multivariate time series describing the global dynamics of cross-regional risk.

Weak-Topology Diffusion Scenario. Unlike spatiotemporal systems such as transportation or energy networks that rely on stable physical adjacency, cross-regional terrorism diffusion does not follow fixed or continuous spatial propagation paths. Interactions among regions are driven by geopolitical ties, organizational networks, and stage-wise conflict dynamics, exhibiting weak-topology, non-Euclidean, and potentially discontinuous jump diffusion patterns. Under this scenario, no reliable and static spatial adjacency structure is assumed to be available, i.e., we do not assume the existence of a fixed adjacency matrix $A \in \mathbb{R}^{C \times C}$ to explicitly model inter-regional dependencies. Instead, such dependencies are regarded as latent and must be inferred implicitly from historical observations.

Problem Formulation. Given a continuous time interval, suppose that the historical states of the regional system are observed over the past T time steps. The objective is to predict the evolution of terrorism-induced fatalities for all regions over the subsequent H time steps. Let the historical observation sequence be denoted as $X_{t-T:t}$ and the future sequence as $X_{t:t+H}$. The forecasting task can then be formulated as

$$X_{t:t+H} = f(X_{t-T:t}), \tag{1}$$

where $f(\cdot)$ denotes the spatiotemporal prediction function to be learned, mapping historical dynamics to future risk evolution.

Under a supervised learning framework, the model parameters are optimized by minimizing the discrepancy between predictions and ground-truth observations. Given a set of training samples, the learning objective is defined as

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^{N} \ell\Big(f_\theta\Big(X_{t-T:t}^{(n)}\Big), X_{t:t+H}^{(n)}\Big), \tag{2}$$

where θ denotes the model parameters and $\ell(\cdot)$ is a prediction error metric consistent with the evaluation protocol used in the experiments.

3 Methodology

Despite substantial progress in multi-scale temporal modeling, long-range dependency learning, and region-level interaction mechanisms, existing approaches often become unstable when confronted with the extreme sparsity, abrupt volatility, and weak spatial structure inherent in terrorism-related time series. To address these challenges, we propose WBMCF (Figure 1), a unified architecture designed to robustly encode spatiotemporal dynamics under highly non-stationary conditions.

WBMCF first applies RevIN to mitigate inter-region statistical bias, and then performs a single-level DWT to decouple slow-varying trends from high-frequency disruptions, yielding a stabilized multi-scale representation. Next, a Bidirectional Gated Mamba encoder captures complementary long-range dependencies from forward and backward temporal flows, where a lightweight gate suppresses spike-induced instabilities before the state-space update. In parallel, a topology-agnostic ConvFFN enables implicit cross-region coupling via shared and grouped convolutions, encoding latent inter-regional dependencies without relying on explicit adjacency. Finally, the framework reconstructs the time-domain sequence through Inverse DWT (IDWT) and denormalization to produce multi-horizon forecasts.

3.1 Discrete Wavelet Transform

Terrorism sequences often combine slowly evolving structural trends with abrupt local fluctuations. Direct time-domain modeling may entangle these heterogeneous patterns and reduce stability. We therefore apply a single-level DWT to obtain a disentangled multi-scale representation. Compared with deep recursive decompositions, a single level better preserves temporal localization, which is crucial for sparse and instantaneous spikes in early warning scenarios.

For notational simplicity, we denote the historical input sequence $X_{t-T:t}$ in a batched form as

$$X \in \mathbb{R}^{B\times C\times T}, \tag{3}$$

where B is the batch size, C is the number of regions, and T corresponds to the historical window length defined in the problem formulation.

A single-level DWT produces low-frequency approximation coefficients Y_L and high-frequency detail coefficients Y_H:

$$(Y_L, Y_H) = \mathrm{DWT}(X). \tag{4}$$

Under the Haar basis, the low-pass and high-pass filters are

$$h_\phi = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right], \quad h_\psi = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right], \tag{5}$$

and the decomposition is implemented via convolution and downsampling:

$$Y_L = (X * h_\phi) \downarrow 2, \quad Y_H = (X * h_\psi) \downarrow 2, \tag{6}$$

where $*$ denotes 1D convolution and $\downarrow 2$ indicates temporal downsampling. The resulting subbands are stacked to form the multi-scale embedding:

$$Y_{\text{emb}} = \mathrm{Stack}(Y_L, Y_H) \in \mathbb{R}^{B\times C\times 2\times T/2}. \tag{7}$$

3.2 Bidirectional Gated Mamba

To capture both cumulative long-range effects and short-term corrections, we adopt a bidirectional state-space encoder. However, under extreme sparsity and volatility, bidirectional fusion can amplify spikes and noise. We thus introduce a lightweight gating mechanism to modulate inputs before state updates, acting as a content-aware filter that suppresses impulsive perturbations while preserving informative variations.

Given an input token (or channel-wise vector) X_t, the gate is defined as

$$g_t = \sigma(W_g X_t + b_g), \quad X_t' = g_t \odot X_t, \tag{8}$$

where $\sigma(\cdot)$ denotes the Sigmoid function and $\odot$ is element-wise multiplication. The gated input X_t' is then fed into the Mamba state-space update:

$$h_t = \bar{A}_t h_{t-1} + \bar{B}_t X_t'. \tag{9}$$

Following the standard formulation of Mamba state-space models, A_t and B_t denote learnable state transition parameters. With a learnable discretization step size Δ_t, the discretized parameters are given by

$$\bar{A}_t = \exp(-\Delta_t A_t), \tag{10}$$

$$\bar{B}_t = (\Delta_t A_t)^{-1}\big(\exp(-\Delta_t A_t) - I\big)(\Delta_t B_t). \tag{11}$$

The bidirectional representations are obtained as

$$H_t^{(f)} = \mathrm{Norm}\big(\mathrm{Mamba}_{\text{fwd}}(X_t') + X_t\big), \tag{12}$$

$$H_t^{(b)} = \mathrm{Norm}\big(\mathrm{Flip}(\mathrm{Mamba}_{\text{bwd}}(\mathrm{Flip}(X_t'))) + X_t\big), \tag{13}$$

which are fused as $H_t = H_t^{(f)} + H_t^{(b)}$. A feed-forward layer with residual normalization yields the final output:

$$\tilde{H}_t = \mathrm{FFN}(H_t), \quad H_t^{\text{out}} = \mathrm{Norm}(\tilde{H}_t + H_t). \tag{14}$$

3.3 Implicit Cross-Regional Modeling via ConvFFN

Traditional spatiotemporal models often rely on explicit topologies and local message passing, which can be unreliable under weak and rapidly shifting spatial structures. We therefore model cross-regional dependency through a topology-agnostic ConvFFN, using shared kernels and implicit channel mixing to capture latent interactions (including "jump diffusion") without explicit adjacency.

Given $X \in \mathbb{R}^{B\times C\times T}$, ConvFFN first applies grouped convolution to extract local features:

$$X_1 = W_1 * X + b_1, \tag{15}$$

followed by nonlinearity and dropout:

$$X_2 = \mathrm{GELU}(\mathrm{Dropout}(X_1)). \tag{16}$$

A second convolution aggregates refined patterns to complete implicit spatial mixing:

$$\mathrm{ConvFFN}(X) = \mathrm{Dropout}(W_2 * X_2 + b_2). \tag{17}$$

With global parameter sharing, ConvFFN enables pattern transfer from data-rich to sparse regions and captures synchronized bursts among non-adjacent areas in latent space, complementing the long-range temporal modeling of the gated Mamba.

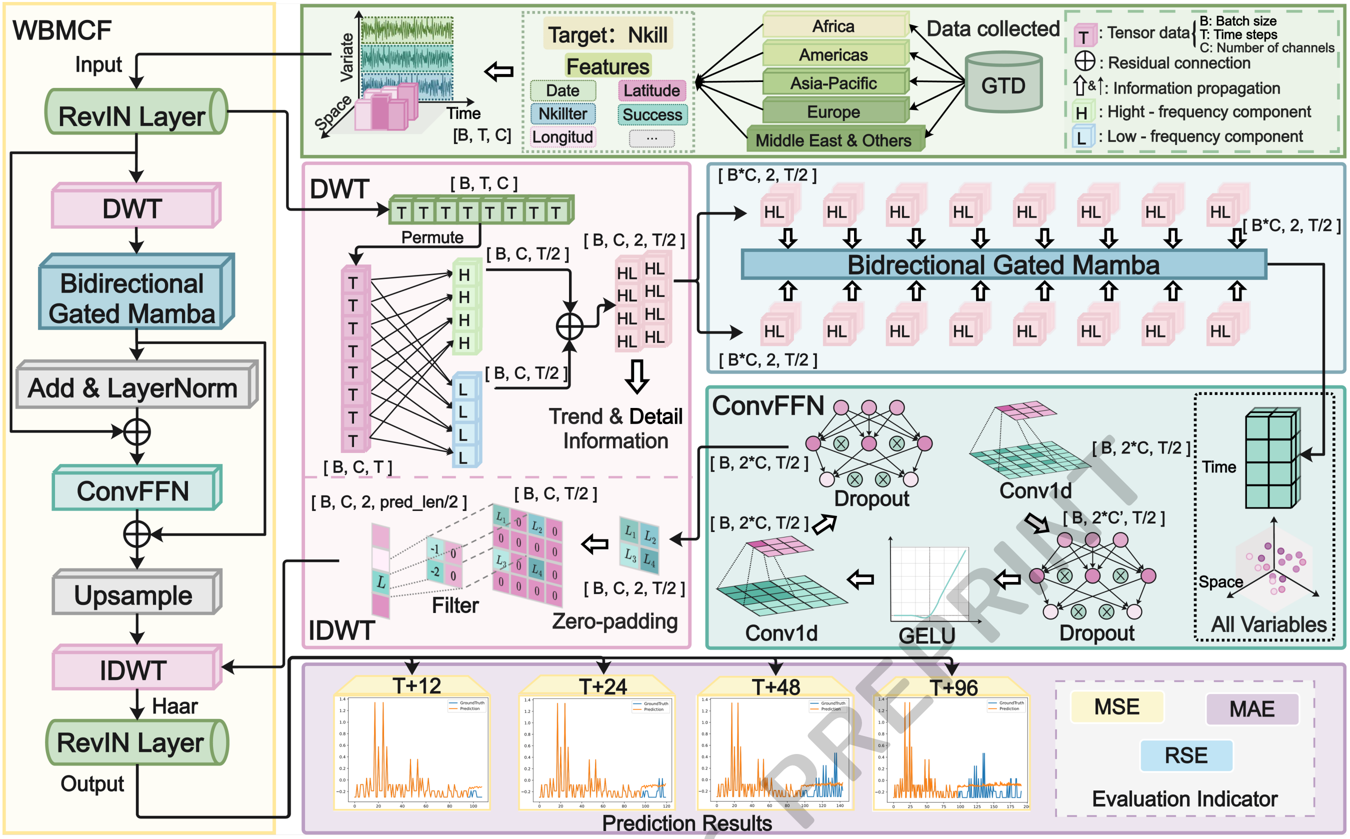


Figure 1: Overall architecture of WBMCF: RevIN + single-level DWT for multi-scale representation, Bidirectional Gated Mamba for long-range temporal modeling, ConvFFN for topology-agnostic cross-regional coupling, and IDWT for time-domain reconstruction.

3.4 Inverse Discrete Wavelet Transform

After multi-scale modeling, the model predicts trend and detail components separately. We apply IDWT to reconstruct a coherent time-domain forecast. Let the predicted representation be $Y' \in \mathbb{R}^{B\times C\times 2\times T'/2}$, where T' denotes the temporal length of the prediction horizon, with subbands (Y'_L, Y'_H). Under the Haar basis, reconstruction can be written as

$$\hat{Y} = (Y'_L \uparrow 2) * h_\phi + (Y'_H \uparrow 2) * h_\psi, \tag{18}$$

where $\uparrow 2$ denotes zero-insertion upsampling. Equivalently,

$$\hat{Y} = \mathrm{IDWT}(Y'_L, Y'_H). \tag{19}$$

Finally, we apply upsampling (if required) and denormalization (RevIN inversion) to output multi-horizon forecasts, preserving both smooth long-term evolution and localized disruptive events.

4 Experiment

We conduct extensive experiments on multiple regional subsets of the Global Terrorism Database to evaluate WBMCF from different perspectives and address the following research questions: **Q1)** How does WBMCF compare with state-of-the-art forecasting models in terms of predictive accuracy under sparse, volatile, and weak-topology terrorism data? **Q2)** How does each core component of WBMCF contribute to the overall performance? **Q3)** How stable is WBMCF under different hyperparameter settings? **Q4)** How robust is WBMCF to noise and partial observation? **Q5)** How does WBMCF perform in terms of computational efficiency?

4.1 Experiment Settings

Dataset Description. We conduct experiments on the Global Terrorism Database (GTD), which contains 186,116 terrorist incident records from 1970 to 2020 with 15 task-relevant attributes. Following the United Nations M49 classification, we construct five regional subsets: Africa (AFR), the Americas (AMER), Asia-Pacific (APAC), Europe (EU), and the Middle East and Others (MEO). Event-level records are aggregated into region-level fatality sequences, which are chronologically split into training, validation, and test sets with a ratio of 7:1:2 and normalized using z-score standardization.
Implementation Details. All experiments are performed under a unified setting on a single NVIDIA RTX3090 GPU using PyTorch with CUDA support. The input sequence length is fixed to 96, and prediction horizons are selected from $\{12, 24, 48, 96\}$. Models are trained with the Adam optimizer (learning rate 1×10^{-4}) using MSE loss, a batch size of 32, and early stopping; each experiment is repeated five times and averaged.
Metrics and Baselines. Following prior studies on spatiotemporal forecasting, we adopt Mean Squared Er-

ror (MSE), Mean Absolute Error (MAE), and Relative Squared Error (RSE) to evaluate forecasting performance. We compare WBMCF with representative baseline methods spanning diverse modeling paradigms, including PFN-based TimePFN [Taga *et al.*, 2025], Mamba-based S_Mamba [Wang *et al.*, 2024], Transformer-based methods such as PatchTST [Nie *et al.*, 2023], Crossformer [Zhang and Yan, 2023], and FEDformer [Zhou *et al.*, 2022], the linear model DLinear [Zeng *et al.*, 2023], and the CNN-based Pyraformer [Liu *et al.*, 2022b].

4.2 Comparison Results (RQ1)

We conduct comparative experiments on five regional terrorism datasets (AFR, AMER, APAC, EU, and MEO). Table 1 reports the averaged performance over four prediction horizons $O \in \{12, 24, 48, 96\}$. For each method, experiments are independently repeated five times with different random seeds, and the variation across runs is consistently within 0.01%, indicating high experimental stability. Overall, WBMCF achieves the best or second-best results across all regions, demonstrating consistent advantages under cross-horizon evaluation.

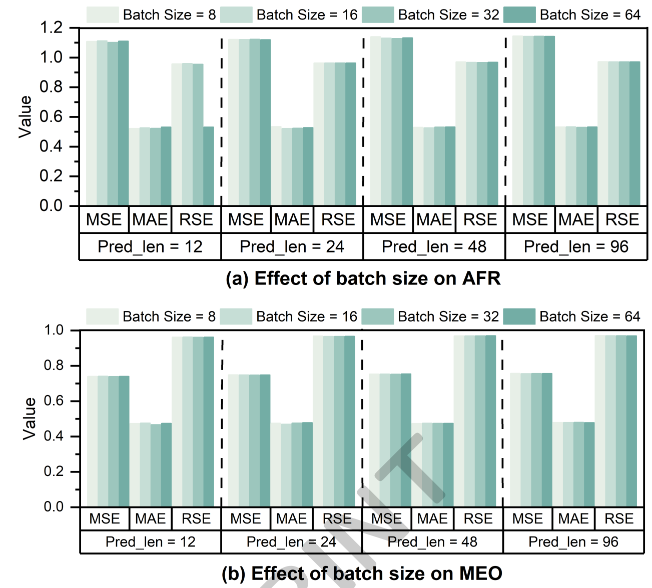
Notably, WBMCF yields clear improvements on challenging regions. In AFR, it improves the averaged MSE, MAE and RSE by 1.5%, 19.2% and 0.8% over the strongest baseline. In MEO, the corresponding improvements reach 1.2%, 19.2% and 0.6%. WBMCF also remains competitive on AMER, APAC, and EU, consistently attaining leading or near-leading performance across metrics. These results indicate that WBMCF maintains robust and reliable forecasting accuracy even when evaluated under an averaged, multi-horizon setting.

4.3 Effectiveness of Components (RQ2)

To examine the contributions of individual components in WBMCF, we conduct ablation experiments on two representative datasets (AFR and MEO), comparing the full model with several variants, including w/o DWT, w/o Mamba, w/o ConvFFN, and the minimal w/o Mamba + ConvFFN. Table 2 reports the averaged results over four prediction horizons ($O = 12, 24, 48, 96$). The full WBMCF consistently achieves the best performance across both datasets and all evaluation metrics, indicating the complementary roles of the proposed components. In particular, removing DWT or simultaneously removing Mamba and ConvFFN leads to the most pronounced performance degradation, highlighting the foundational role of time–frequency decoupling in stabilizing highly non-stationary sequences. By contrast, removing ConvFFN results in a milder yet consistent drop in performance, suggesting that implicit cross-regional coupling contributes steadily to overall accuracy. These results validate the hierarchical design of WBMCF, in which time–frequency stabilization provides a reliable basis that is further refined by temporal modeling and topology-agnostic spatial interactions.

4.4 Hyperparameter Sensitivity Analysis (RQ3)

Effect of Batch Size. Experiments on the AFR and MEO datasets show that varying the batch size among {8, 16, 32, 64} results in only minor fluctuations in MSE, MAE, and RSE across all forecasting horizons, as illustrated in Figure 2. This stability persists despite substantial differences in volatility and non-stationarity between the two regions, indicating that WBMCF is insensitive to gradient noise induced by small-batch training. Moreover, increasing the batch size does not yield consistent performance gains, suggesting that training stability is primarily attributed to the model architecture rather than large-batch smoothing effects.



(a) Effect of batch size on AFR

(b) Effect of batch size on MEO

Figure 2: Performance of WBMCF with different batch sizes.

Effect of State Expansion Factor. The State Expansion Factor controls the state-space dimensionality of the Bidirectional Gated Mamba. Under different SEF settings ({8, 16, 32, 64}), the MSE curves for both AFR and the highly volatile MEO region remain smooth and confined within narrow ranges across all horizons, as shown in Figure 3. While larger state dimensions provide marginal gains in certain configurations, no structural degradation is observed. This indicates that WBMCF does not rely on brute-force state expansion to achieve expressive power, but instead derives its robustness from the synergy between time–frequency modeling and bidirectional temporal dynamics, enabling stable performance even under constrained computational settings.

Effect of Decomposition Depth. We evaluate the Effect of the wavelet decomposition depth by comparing $J = 1, 2, 3$ across multiple horizons on AFR and MEO. As shown in Figure 4, the single-level setting ($J = 1$) consistently achieves the best or near-best performance and maintains stable behavior across prediction lengths. Increasing the decomposition depth does not lead to further improvements and instead introduces mild but consistent degradation in several cases, implying that overly deep multi-scale decomposition may reduce representation coherence in highly noisy and bursty terrorism sequences.

Method	AFR			AMER			APAC			EU			MEO		
	MSE	MAE	RSE	MSE	MAE	RSE	MSE	MAE	RSE	MSE	MAE	RSE	MSE	MAE	RSE
Pyraformer	0.781	0.510	0.987	0.939	0.585	1.040	0.910	0.556	0.994	1.163	0.650	1.008	1.192	0.544	0.993
FEDformer	0.772	0.477	0.982	0.903	0.584	1.020	0.891	0.545	0.983	1.094	0.612	0.978	1.165	0.554	0.982
Crossformer	0.760	0.486	0.974	0.870	0.534	1.001	0.881	0.540	0.978	1.085	0.583	0.974	1.144	0.538	0.972
DLinear	0.766	0.488	0.978	0.873	0.530	1.003	0.898	0.544	0.987	1.061	0.576	0.963	1.151	0.536	0.975
PatchTST	1.077	0.607	1.155	1.171	0.664	1.159	1.330	0.720	1.197	1.306	0.632	1.069	1.425	0.613	1.084
S_Mamba	0.760	0.489	0.976	0.883	0.613	1.009	0.880	0.524	0.977	1.067	0.626	0.966	1.137	0.544	0.969
TimePFN	0.763	0.454	0.976	0.889	0.517	1.012	0.880	0.523	0.977	1.083	0.555	0.973	1.137	0.526	0.970
WBMCF(ours)	**0.749**	**0.367**	**0.967**	**0.857**	**0.438**	**0.994**	**0.876**	**0.425**	**0.975**	**1.049**	**0.512**	**0.958**	**1.124**	**0.425**	**0.964**
Improvements	+1.5%	+19.2%	+0.8%	+1.5%	+15.3%	+0.7%	+0.4%	+18.7%	+0.2%	+1.1%	+7.8%	+0.6%	+1.2%	+19.2%	+0.6%

Table 1: Multivariate forecasting performance of all methods on five datasets, where results are averaged over four prediction horizons ($O = 12, 24, 48, 96$). The best results are highlighted in **bold**, and the suboptimal results are underlined.

Method	AFR			MEO		
	MSE	MAE	RSE	MSE	MAE	RSE
WBMCF(ours)	**0.749**	**0.367**	**0.967**	**1.124**	**0.424**	**0.964**
w/o Mamba	0.764	0.402	0.977	1.140	0.463	0.971
w/o ConvFFN	0.790	0.464	0.993	1.154	0.529	0.977
w/o DWT	0.759	0.368	0.974	1.133	0.428	0.968
w/o Mamba + ConvFFN	0.761	0.369	0.974	1.128	0.425	0.967

Table 2: Ablation results of WBMCF on two representative datasets (AFR and MEO), where results are averaged over four prediction horizons ($O = 12, 24, 48, 96$). The best results are highlighted in bold, and the suboptimal results are underlined.

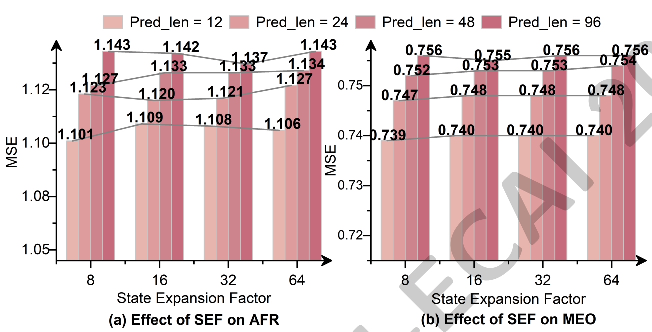


Figure 3: Performance of WBMCF with different State Expansion Factors (SEF).

4.5 Noise Robustness Evaluation (RQ4)

To assess robustness under disturbed conditions, we inject Gaussian noise into continuous variables of the AFR and MEO datasets with two perturbation levels (50% and 100%), and compare the resulting prediction errors, as shown in Figure 5. WBMCF consistently demonstrates superior robustness across both noise settings, exhibiting markedly smaller increases in MSE, MAE, and RSE than baseline methods. Notably, even under the 100% noise level, WBMCF maintains stable predictive behavior, whereas baselines such as PatchTST and DLinear suffer from pronounced error inflation. The consistent performance on AFR and MEO—two regions with contrasting event characteristics—indicates that WBMCF adapts well to diverse noise patterns, further supporting its reliability in partially observed and highly unstable forecasting scenarios.

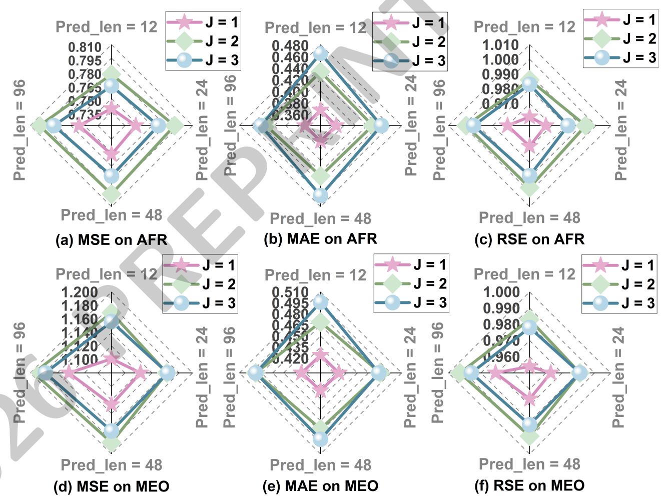


Figure 4: Performance of WBMCF with different decomposition depths J.

4.6 Model Efficiency Analysis (RQ5)

To examine the trade-off between predictive accuracy and computational cost, we conduct an efficiency analysis (Figure 6) on the AFR and MEO regions under a unified experimental configuration. We compare representative models in terms of prediction error, per-iteration runtime, and memory consumption, and visualize their relative positions in the efficiency–performance space.

The results indicate that WBMCF consistently occupies a favorable region in both datasets, achieving low prediction error with fast computation and moderate memory usage. This compact positioning reflects a well-balanced efficiency profile. In contrast, S_Mamba attains competitive runtime performance but at the expense of higher prediction error, while Transformer-based models generally incur increased computational or memory overhead and exhibit less stable efficiency–performance trade-offs across regions. Overall, WBMCF maintains strong predictive accuracy while effectively controlling computational and memory costs, demon-

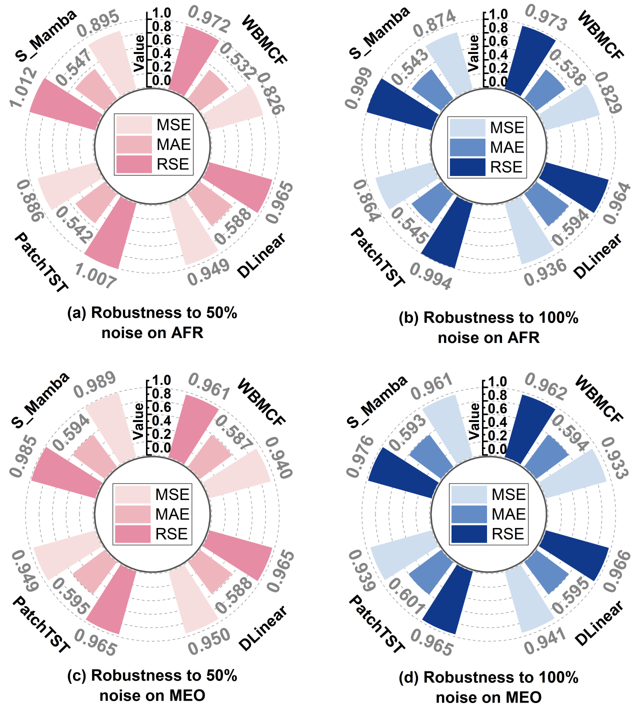


Figure 5: Performance of all methods with different noise perturbation levels.

strating its practicality for large-scale cross-regional terrorism forecasting.

5 Related Work

Spatiotemporal forecasting. Forecasting terrorism fatalities requires modeling complex spatiotemporal dependencies driven by geopolitical shifts and organizational dynamics. Early approaches relied on convolution-based architectures, such as ST-ResNet and 3D CNNs, to capture temporal evolution and local spatial neighborhoods through fixed receptive fields [Jiang and Luo, 2022], but they struggle with non-Euclidean or dynamically changing regional interactions. Graph Neural Networks (GNNs), including STGCN, Graph WaveNet, and diffusion-based models like DCRNN, propagate information over predefined non-Euclidean structures [Zhang *et al.*, 2023]. To relax static-topology assumptions, adaptive and dynamic graph models such as AGCRN [Kim *et al.*, 2023] and DGCRN [Li *et al.*, 2023] introduce latent relation learning or time-varying adjacency mechanisms [Lan *et al.*, 2022; Qin *et al.*, 2025a]. However, these methods generally assume smooth temporal and spatial evolution, which is often violated in sparse and bursty terrorism data [Liu *et al.*, 2022c; Weng *et al.*, 2023].

Multi-scale temporal modeling. Terrorism-related time series exhibit pronounced multi-scale non-stationarity due to the coexistence of long-term geopolitical trends and short-term attacks [Hou *et al.*, 2023]. Decomposition- and frequency-based models, including Autoformer [Wu *et al.*, 2021], FEDformer [Zhou *et al.*, 2022], SCINet [Liu *et al.*, 2022a], TimesNet [Wu *et al.*, 2023], and Wavelet-

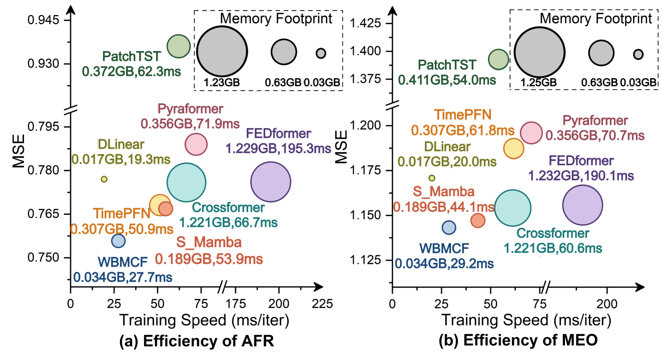


Figure 6: Training efficiency and memory costs of different approaches.

former [Masserano *et al.*, 2025], aim to separate trends from local fluctuations. Lightweight methods such as DLinear [Zeng *et al.*, 2023], PatchTST [Nie *et al.*, 2023], and TiDE [Das *et al.*, 2023] reduce model complexity while maintaining stable performance. State-space models, notably Mamba [Gu and Dao, 2024], further enhance long-range dependency modeling, yet existing approaches still struggle to jointly suppress noise and respond to abrupt anomalies under frequent structural shifts.

Spatial dependency modeling. Compared to temporal dynamics, terrorism events exhibit greater irregularity in spatial diffusion, often characterized by weak topology and jump diffusion [Liang *et al.*, 2024; Zhan *et al.*, 2024], rendering static adjacency matrices inadequate [Capone *et al.*, 2025]. Explicit graph-based models perform well under stable spatial structures but degrade when spatial relations are unstable or missing [Han *et al.*, 2024]. Adaptive and dynamic graph approaches partially alleviate static assumptions yet remain limited in capturing abrupt structural reorganization [Kong *et al.*, 2024; Kieu *et al.*, 2024]. In contrast, topology-agnostic methods model spatial dependencies through implicit feature interactions without explicit graphs [Qin *et al.*, 2023; Mao *et al.*, 2024], offering greater flexibility under weak-topology conditions.

6 Conclusion

This paper proposes WBMCF, a wavelet–bidirectional state-space–convolution fusion framework for cross-regional terrorism fatality forecasting under highly non-stationary and weak-topology conditions. By decoupling long-term trends and short-term bursts in the time–frequency domain, and jointly modeling long-range temporal dependencies with implicit spatial interactions, WBMCF achieves robust and stable performance across diverse regions, prediction horizons, and disturbed environments. Extensive experiments demonstrate that the proposed framework consistently balances predictive accuracy, robustness, and computational efficiency, highlighting its practical potential for reliable medium-term risk assessment in complex and volatile terrorism scenarios. Future work may explore the incorporation of explicit dynamic spatial signals and uncertainty-aware extensions to further enhance its applicability.

Acknowledgments

The authors would like to thank Guangxi Police College for providing academic support and a supportive research environment.

References

[Cao *et al.*, 2025] Lingxiao Cao, Bin Wang, Guiyuan Jiang, Yanwei Yu, and Junyu Dong. Spatiotemporal-aware trend-seasonality decomposition network for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11463–11471, 2025.

[Capone *et al.*, 2025] Vincenzo Capone, Angelo Casolaro, and Francesco Camastra. Spatio-temporal prediction using graph neural networks: A survey. *Neurocomputing*, page 130400, 2025.

[Cheng *et al.*, 2023] Yunyao Cheng, Peng Chen, Chenjuan Guo, Kai Zhao, Qingsong Wen, Bin Yang, and Christian S Jensen. Weakly guided adaptation for robust time series forecasting. *Proceedings of the VLDB Endowment*, 17(4):766–779, 2023.

[Das *et al.*, 2023] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan K Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tiDE: Time-series dense encoder. *Transactions on Machine Learning Research*, 2023.

[Deng *et al.*, 2024] Jinliang Deng, Feiyang Ye, Du Yin, Xuan Song, Ivor Tsang, and Hui Xiong. Parsimony or capability? decomposition delivers both in long-term time series forecasting. *Advances in Neural Information Processing Systems*, 37:66687–66712, 2024.

[Dong *et al.*, 2025] Hao Dong, Ziyue Qiao, Zhiyuan Ning, Qi Hao, Yi Du, Pengyang Wang, and Yuanchun Zhou. Disentangled multi-span evolutionary network against temporal knowledge graph reasoning. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Findings of the Association for Computational Linguistics: ACL 2025*, pages 12697–12707, Vienna, Austria, July 2025. Association for Computational Linguistics.

[Gu and Dao, 2024] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*, 2024.

[Han *et al.*, 2024] Jindong Han, Weijia Zhang, Hao Liu, Tao Tao, Naiqiang Tan, and Hui Xiong. Bigst: Linear complexity spatio-temporal graph neural network for traffic forecasting on large-scale road networks. *Proceedings of the VLDB Endowment*, 17(5):1081–1090, 2024.

[Hou *et al.*, 2023] Zhiwen Hou, Yuchen Zhou, Xiaowei Wu, and Fanliang Bu. Attention-based spatial–temporal multi-graph convolutional networks for casualty prediction of terrorist attacks. *Complex & Intelligent Systems*, 9(6):6307–6328, 2023.

[Jiang and Luo, 2022] Weiwei Jiang and Jiayun Luo. Graph neural network for traffic forecasting: A survey. *Expert systems with applications*, 207:117921, 2022.

[Kibris, 2021] Arzu Kibris. The geo-temporal evolution of violence in civil conflicts: A micro analysis of conflict diffusion on a new event data set. *Journal of Peace Research*, 58(5):885–899, 2021.

[Kieu *et al.*, 2024] Duc Kieu, Tung Kieu, Peng Han, Bin Yang, Christian S Jensen, and Bac Le. Team: Topological evolution-aware framework for traffic forecasting. *Proceedings of the VLDB Endowment*, 18(2):265–278, 2024.

[Kim *et al.*, 2023] SangYoun Kim, SungKu Heo, KiJeon Nam, TaeYong Woo, and ChangKyoo Yoo. Flexible renewable energy planning based on multi-step forecasting of interregional electricity supply and demand: Graph-enhanced ai approach. *Energy*, 282:128858, 2023.

[Kong *et al.*, 2024] Weiyang Kong, Ziyu Guo, and Yubao Liu. Spatio-temporal pivotal graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 8627–8635, 2024.

[LaFree and Dugan, 2007] Gary LaFree and Laura Dugan. Introducing the global terrorism database. *Terrorism and political violence*, 19(2):181–204, 2007.

[Lan *et al.*, 2022] Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International conference on machine learning*, pages 11906–11917. PMLR, 2022.

[Li *et al.*, 2023] Fuxian Li, Jie Feng, Huan Yan, Guangyin Jin, Fan Yang, Funing Sun, Depeng Jin, and Yong Li. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Trans. Knowl. Discov. Data*, 17(1), February 2023.

[Liang *et al.*, 2024] Ke Liang, Sihang Zhou, Meng Liu, Yue Liu, Wenxuan Tu, Yi Zhang, Liming Fang, Zhe Liu, and Xinwang Liu. Hawkes-enhanced spatial-temporal hypergraph contrastive learning based on criminal correlations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8733–8741, 2024.

[Liu *et al.*, 2022a] Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022.

[Liu *et al.*, 2022b] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*, 2022.

[Liu *et al.*, 2022c] Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in neural information processing systems*, 35:9881–9893, 2022.

[Mao *et al.*, 2024] Yuntao Mao, Jinwei Zhu, Ziwei Chen, You Li, and Ziyang Deng. Predicting urban carbon emissions through spatio-temporal encoding. In *Proceedings* *of the International Conference on Computer Vision and Deep Learning*, CVDL '24, New York, NY, USA, 2024. Association for Computing Machinery.

[Masserano *et al.*, 2025] Luca Masserano, Abdul Fatir Ansari, Boran Han, Xiyuan Zhang, Christos Faloutsos, Michael W Mahoney, Andrew Gordon Wilson, Youngsuk Park, Syama Sundar Rangapuram, Danielle C Maddix, et al. Enhancing foundation models for time series forecasting via wavelet-based tokenization. In *International Conference on Machine Learning*, pages 43248–43275. PMLR, 2025.

[Nie *et al.*, 2023] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

[Petropoulos *et al.*, 2022] Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J Bessa, Jakub Bijak, John E Boylan, et al. Forecasting: theory and practice. *International Journal of forecasting*, 38(3):705–871, 2022.

[Qin *et al.*, 2023] Yanjun Qin, Haiyong Luo, Fang Zhao, Yuchen Fang, Xiaoming Tao, and Chenxing Wang. Spatio-temporal hierarchical mlp network for traffic forecasting. *Information Sciences*, 632:543–554, 2023.

[Qin *et al.*, 2025a] Zhenkai Qin, Baozhong Wei, Caifeng Gao, Feng Zhu, Weiqi Qin, and Qian Zhang. Acsaformer: A crime forecasting model based on sparse attention and adaptive graph convolution. *Frontiers in Physics*, 13:1596987, 2025.

[Qin *et al.*, 2025b] Zhenkai Qin, Baozhong Wei, Huan Zeng, and Qian Zhang. Tc-linear: lightweight traffic speed forecasting via trend-seasonal decomposition and temporal convolution. *Systems Science & Control Engineering*, 13(1):2546843, 2025.

[Rosenberg and Avdan, 2025] Andrew S Rosenberg and Nazli Avdan. Assessing border walls' varied impacts on terrorist group diffusion. *Conflict Management and Peace Science*, 42(4):438–461, 2025.

[Schirmer *et al.*, 2022] Mona Schirmer, Mazin Eltayeb, Stefan Lessmann, and Maja Rudolph. Modeling irregular time series with continuous recurrent units. In *International conference on machine learning*, pages 19388–19405. PMLR, 2022.

[Shao *et al.*, 2022] Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 1567–1577, 2022.

[Taga *et al.*, 2025] Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. Timepfn: Effective multivariate time series forecasting with synthetic data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 20761–20769, 2025.

[Wang *et al.*, 2024] Zihan Wang, Fanheng Kong, Shi Feng, Ming Wang, Xiaocui Yang, Han Zhao, Daling Wang, and Yifei Zhang. Is mamba effective for time series forecasting? *Neurocomputing*, page 129178, 2024.

[Wei *et al.*, 2024] Xiao Wei, Qi Xu, Hang Yu, Qian Liu, and Erik Cambria. Through the MUD: A multi-defendant charge prediction benchmark with linked crime elements. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2864–2878, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[Weng *et al.*, 2023] Wenchao Weng, Jin Fan, Huifeng Wu, Yujie Hu, Hao Tian, Fu Zhu, and Jia Wu. A decomposition dynamic graph convolutional recurrent network for traffic forecasting. *Pattern Recognition*, 142:109670, 2023.

[Wu *et al.*, 2021] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

[Wu *et al.*, 2023] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

[Xu *et al.*, 2024] Zonghuang Xu, Yao Lin, Hongyu Cai, Wei Zhang, Jin Shi, and Lingyun Situ. Risk assessment and categorization of terrorist attacks based on the global terrorism database from 1970 to 2020. *Humanities and Social Sciences Communications*, 11(1):1–14, 2024.

[Zeng *et al.*, 2023] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[Zhan *et al.*, 2024] Chuanhong Zhan, Wei Xiang, Liang Chao, and Bang Wang. What would happen next? predicting consequences from an event causality graph. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 820–832, 2024.

[Zhang and Yan, 2023] Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

[Zhang *et al.*, 2023] Qianru Zhang, Chao Huang, Lianghao Xia, Zheng Wang, Siu Ming Yiu, and Ruihua Han. Spatial-temporal graph learning with adversarial contrastive adaptation. In *International Conference on Machine Learning*, pages 41151–41163. PMLR, 2023.

[Zhou *et al.*, 2022] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pages 27268–27286. PMLR, 2022.