Transformer-Based Models for Short-to-Medium Range Tropical Cyclone Track and Intensity Forecasting

Wisnu Syahid
Faculty of Meteorology, State College of Meteorology, Climatology and Geophysics, Tangerang City, Indonesia
E-mail : wisnoesjahid@stmkg.ac.id
Received 3 July 2026; Revised 24 July 2026; Accepted 25 July 2026

Abstract - This study presents a systematic benchmark of six Transformer-based architectures, Vanilla Transformer, Informer, Autoformer, FEDformer, PatchTST, and iTransformer, for short-to-medium range tropical cyclone (TC) track and intensity forecasting over the Southern Indian Ocean basins surrounding Indonesia. Using a 3-hourly resampled IBTrACS dataset during (1965-2025), multi-lead-time supervised sequences were constructed with a 72-hour (24-step) input window across six forecast horizons (12–72 hours). All models were implemented under a unified dual-output framework with identical hyperparameters and training conditions to ensure fair comparison. Performance was evaluated using MAE, RMSE, MAPE, R^2, and categorical skill scores at positional thresholds of 50, 100, and 200 km. A Weighted Sum Model (WSM) composite ranking identified iTransformer as the top performer (WSM = 0.075), followed by PatchTST (0.092) and Autoformer (0.105). Horizon-stratified analysis shows Autoformer excels at short lead times (≤24 h), while iTransformer dominates from 36 to 72 hours. FEDformer ranked last, highlighting limitations of frequency-domain mixing on sparse kinematic data.

Keywords - Tropical Cyclone Forecasting; Transformer Architecture; IBTrACS; Weighted Sum Model; Deep Learning.

1. INTRODUCTION

Tropical cyclones (TCs) represent one of the most destructive classes of natural hazards in the tropical and subtropical Indo-Pacific, generating extreme winds, storm surges, and catastrophic precipitation that annually affect hundreds of millions of people [1,2]. Although the Indonesian archipelago lies outside the primary TC genesis zones, the Southern Indian Ocean basins that flank it sustain some of the highest TC activity on Earth, and the indirect effects of these systems, including anomalous rainfall, destructive swells, and enhanced monsoonal convection, regularly impact coastal and maritime communities across the region [3,4]. As sustained ocean warming progressively expands the Indo-Pacific Warm Pool and modulates the Madden–Julian Oscillation, the thermodynamic environment supporting TC intensification in these basins is expected to become increasingly favourable, with consequent increases in both TC lifetime maximum intensity and the frequency of rapid-intensification events [5]. Accurate, timely TC track forecasting is therefore a critical component of Indonesia's early-warning and disaster-risk-reduction infrastructure.

Among the dimensions of TC prediction, track (trajectory) forecasting holds a privileged operational position: errors in the predicted storm centre position propagate directly into errors in landfall timing, evacuation-zone delineation, and the spatial targeting of severe-weather warnings [6]. Large-scale steering-flow variability further amplifies position uncertainty at extended lead times [7]. Traditional Numerical Weather Prediction (NWP) systems, which explicitly integrate the primitive equations of atmospheric motion, remain the operational backbone of meteorological agencies worldwide; however, their skill is constrained by imperfect initial conditions, the high computational cost of ensemble integration, and persistent limitations in resolving sub-grid-scale inner-core dynamics that govern rapid intensification [8,9]. These structural limitations have motivated sustained interest in complementary, data-driven approaches

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

capable of learning complex, nonlinear spatiotemporal patterns directly from historical best-track records.

Deep learning methods have emerged as a credible complement to NWP-based TC forecasting. Recurrent neural network (RNN) architectures, particularly Long Short-Term Memory (LSTM; [10]) and Gated Recurrent Unit (GRU; [11]) networks, exploit explicit gating mechanisms to model multi-step temporal dependencies in TC track sequences and have demonstrated competitive accuracy across multiple ocean basins [12,13,14]. A companion study benchmarking six gated-recurrent architectures for the Indonesian region found that Bidirectional GRU and Attention Bidirectional LSTM achieved the most balanced joint performance on position and intensity forecasting across the 12–72-hour range, while also observing that positional error amplifies substantially faster than intensity error as the forecast horizon extends, a finding consistent with the chaotic sensitivity of TC steering to large-scale atmospheric flow. Despite these successes, the sequential processing constraint of recurrent architectures limits their ability to capture long-range dependencies efficiently and restricts training parallelism, motivating investigation of alternative sequence modeling paradigms.

The Transformer architecture introduced by Vaswani et al. [15] replaces recurrent state propagation entirely with multi-head self-attention, which computes pairwise dependencies across all sequence positions simultaneously. This design offers two structural advantages directly relevant to TC forecasting: the capacity to attend to distant historical observations without information dilution through sequential hidden states, and full training parallelism that substantially reduces wall-clock training time. Building on this foundation, a family of specialized Transformer variants has been developed for long-sequence time series forecasting. The Informer [16] introduces ProbSparse self-attention and encoder distillation to reduce the quadratic complexity of full attention. The Autoformer [17] embeds moving-average series decomposition and an FFT-based Auto-Correlation mechanism to exploit periodicity in temporal data. The FEDformer [18] performs attention in the frequency domain via a learnable Fourier mixing block. The PatchTST [19] divides each channel into overlapping patches, enabling the model to attend over longer effective time horizons per token. The iTransformer [20] inverts the conventional time-token paradigm, treating each input variable's full history as a single token and computing attention across variables rather than time steps. Despite rapid adoption in general-purpose forecasting benchmarks, these architectures have not been systematically compared for TC track and intensity prediction in the Indonesian maritime region, where the operational stakes and climatological context are distinct.

This paper addresses that gap with four specific contributions. First, it presents, to the authors' knowledge, the first systematic comparison of all six major Transformer architecture families, canonical, distilling, decomposition-based, frequency-domain, patch-based, and variate-inverted, applied to a unified TC forecasting task. Second, it introduces a shared dual-output modeling framework that predicts storm position and intensity simultaneously under identical training conditions, enabling architecturally fair comparison. Third, it provides a multi-horizon (12–72 h) evaluation that integrates deterministic error metrics and operational skill scores, quantifying how different architectural inductive biases perform as forecast uncertainty grows with lead time. Fourth, it applies a Weighted Sum Model (WSM) multi-criteria decision framework to produce a composite, reproducible architecture ranking that directly supports operational model selection. Through this approach, the research aims to provide a scientific basis for determining the most effective deep learning model for tropical cyclone early warning systems in Indonesia.

2. RESEARCH METHOD

2.1. Data Source and Preprocessing

The dataset comprises historical TC best-track records for the North and Southern Indian Ocean basins, sourced from the IBTrACS archive (1965-2025) [21] and stored in a quality-

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

controlled spreadsheet. Each record contains storm name, observation timestamp, geodetic center position (latitude φ, longitude λ), maximum sustained wind speed (knots), and storm bearing. A column-alias resolution procedure reconciled heterogeneous field-name conventions across data sources (e.g., LAT/LATITUDE, INTENSITY/VMAX) into a standardized internal schema prior to any further processing.

Because raw best-track observations may be reported at irregular intervals, each storm's position sequence was resampled to a uniform 3-hour interval by constructing the union of original and target timestamps, applying time-based linear interpolation over each numeric field, and reindexing to the regular grid. This procedure preserves temporal continuity for storms with non-standard reporting intervals while avoiding artificial discontinuities. The great-circle (haversine) distance between consecutive resampled positions is the kinematic foundation for two derived features:

$$d = 2R \cdot si\,n^{-1}\sqrt{\left[sin^2\left(\frac{\Delta\varphi}{2}\right) + cos\varphi_1 cos\varphi_2 sin^2\left(\frac{\Delta\lambda}{2}\right)\right]} \quad (1)$$

where φ_1 and φ_2 are the latitudes of two consecutive positions, $\Delta\varphi$ and $\Delta\lambda$ are their latitude and longitude differences, and R = 6371 km is the mean Earth radius. Translational speed is the distance per elapsed time interval Δt (hours):

$$v_{\{trans\}} = \frac{d}{\Delta t} \quad (2)$$

and the initial bearing from one position to the next is the forward azimuth:

$$\theta = atan * 2[sin(\Delta\lambda)cos\varphi_2\,, cos\varphi_1 sin\varphi_2 - sin\varphi_1 cos\varphi_2 cos(\Delta\lambda)] \quad (3)$$

These two derived quantities, combined with the resampled latitude and longitude, constitute the four-dimensional input feature vector $F = \{lat, lon, v_{\{trans\}}, \theta\}$ at each time step. To prevent scale differences among features from distorting gradient-based optimization, each input feature was standardized using z-score normalization with parameters estimated exclusively from the training partition:

$$z = \frac{(x - \mu_{\{train\}})}{\sigma_{\{train\}}} \quad (4)$$

where $\mu_{\{train\}}$ and $\sigma_{\{train\}}$ denote the training-set mean and standard deviation, respectively. Wind intensity targets were normalized using an independently fitted scaler applied to the subset of training samples for which measured intensity is available. For samples without a valid intensity observation, we assign a zero-value target and a training sample weight of zero on the intensity loss term, ensuring that intensity-free sequences do not corrupt the intensity gradient while still contributing to positional loss.

2.2. Problem Formulation

We cast the forecasting task as a multi-horizon, dual-output sequence-to-vector regression problem. Let T = 24 be the input sequence length (72 hours at 3-hour resolution), and let F = 4 be the number of input features. For a given storm observation at time t, the input tensor is:

$$X_t = \left[x_{\{t-T+1\}}, x_{\{t-T+2\}}, \ldots, x_t\right] \in R^{\{T\times F\}} \quad (5)$$

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

The model learns a parameter set θ defining a mapping from this history window to two simultaneous regression targets at a future lead time τ ∈ {12, 24, 36, 48, 60, 72} hours:

$$f_\theta: R^{\{T\times F\}} \rightarrow (R^2, R) - \left(\hat{y}_{\{pos\}}, \hat{y}_{\{int\}}\right) = f_\theta(X_t) \quad (6)$$

where $\hat{y}_{\{pos\}} = \left(\hat{y}_{\{lat\}}, \hat{y}_{\{lon\}}\right)$ is the predicted storm center position and $\hat{y}_{\{int\}}$ is the predicted maximum sustained wind speed at time t + τ. Supervised (X, y) pairs are constructed independently for each of the six lead times via a sliding window over each storm's resampled trajectory; windows in which any input feature value is missing, or for which the target position is undefined, are discarded. This procedure yields six lead-time-specific datasets that share a common 24-step input representation but differ in their target offset, enabling the trained model to be evaluated across all horizons without retraining.

2.3 Transformer Model Architectures

We implemented six Transformer-based architectures within a unified Keras functional-API model factory. All six share an identical input shape (batch, 24, 4), identical dual output heads, and an identical loss and optimizer configuration, differing only in the design of their encoder modules. This unified framework ensures architecturally fair comparisons by eliminating confounding factors related to hyperparameters, training procedures, and output structures. This section first presents the components shared across all architectures, then details the encoder-specific mechanisms that distinguish each model.

2.3.1 Shared Architectural Components

Every model begins with a linear input projection that maps the raw F-dimensional feature at each time step into a D-dimensional latent space:

$$Z^{\{(0)\}} = X \cdot W_{\{proj\}} b_{\{proj\}} \in R^{\{T\times D\}}, W_{\{proj\}} \in R^{\{F\times D\}} \quad (7)$$

For architectures that operate over the temporal axis, a fixed sinusoidal positional encoding is added element-wise to convey sequence-order information without learnable parameters [15]. For position $t \in \{0, \dots, T-1\}$ and channel index $i \in \left\{0, \dots, \frac{D}{2} - 1\right\}$:

$$PE_{\{t,2i\}} = sin(t \cdot \omega_i), PE_{\{t,2i+1\}} = cos(t \cdot \omega_i), \omega_i = 10000^{\left\{-\frac{2i}{D}\right\}} \quad (8)$$

so that the encoder input becomes $Z = Z^{\{(0)\}} + PE$. The standard scaled dot-product multi-head self-attention mechanism ($H = 4\ heads, head\ dimension\ d_k = \frac{D}{H} = 16$) used verbatim in the Vanilla Transformer, Informer, PatchTST, and iTransformer is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

$$MHA(Z) = Concat(head_1, \dots, head_H) \cdot W^O; head_h = Attention\left(ZW_h^Q, ZW_h^K, ZW_h^V\right) \quad (10)$$

Each encoder layer also contains a position-wise feed-forward sublayer with GELU activation and inner dimension $D_{\{ff\}} = 256$:

$$FFN(Z) = GELU(ZW_1 + b_1)W_2 + b_2; W_1 \in R^{\{D\times D_{\{ff\}}\}}, W_2 \in R^{\{D_{\{ff\}}\times D\}} \quad (11)$$

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

Both sub-layers employ a post-norm residual connection (Layer Normalization with ε = 10^{-6}):

$$Z' = LayerNorm\big(Z + SubLayer(Z)\big); LayerNorm(z) = \gamma \odot (z - \mu) \cdot (\sigma^2 + \varepsilon)^{\{-½\}} + \beta \quad (12)$$

After L = 2 encoder layers, the sequence representation $Z^{\{(L)\}} \in R^{\{T \times D\}}$ is collapsed by global average pooling to a single context vector, which then feeds two independent regression heads:

$$c = \left(\frac{1}{T}\right) \Sigma_{\{t=1\}}^{\{T\}} Z_{\{(L)\}}^{t} \in R^{D} \quad (13)$$

$$\hat{y}_{\{int\}} = W_{\{int,2\}} ReLU\big(W_{\{int,1\}} c + b_{\{int,1\}}\big) + b_{\{int,2\}} \in R \quad (14)$$

$$\hat{y}_{\{pos\}} = W_{\{pos,2\}} \cdot ReLU\big(W_{\{pos,1\}c} + b_{\{pos,1\}}\big) + b_{\{pos,2\}} \in R^2 \quad (15)$$

The composite training loss combines mean absolute error (MAE) on intensity with mean squared error (MSE) on position, each with unit weight, optimised by Adam with learning rate $\eta = 0.001$:

$$L = MAE\big(y_{\{int\}}, \hat{y}_{\{int\}}\big) + MSE\big(y_{\{pos\}}, \hat{y}_{\{pos\}}\big) \quad (16)$$

2.3.2 Vanilla Transformer

The baseline architecture follows the encoder-only design of Vaswani et al. [15] without modification. For each of the L encoder layers indexed i:

$$Z_i = LayerNorm\left(Z_{\{i-1\}} + MHA\big(Z_{\{i-1\}}\big)\right) \quad (17)$$

$$Z_i = LayerNorm\big(Z_i + FFN(Z_i)\big) \quad (18)$$

The context vector is obtained directly as $c = GAP(Z_L)$, followed by the output heads in Eqs. 14–15.

2.3.3 Informer

The Informer [16] augments the standard encoder with a self-attention distillation operation that progressively compresses the sequence after each layer. Following the standard attention/FFN sub-layers in Eqs. 17–18, a 1-D convolution with ELU activation and subsequent max-pooling halve the sequence length at every intermediate layer ($i < L - 1$):

$$Z_i^{\{conv\}} = ELU\big(Conv1D_{\{k=3\}(Z_i)}\big) \quad (19)$$

$$Z_i^{\{distill\}} = MaxPool_{\{k=2,s=2\}\left(Z_i^{\{conv\}}\right)} \quad (20)$$

Each distillation step reduces the effective sequence length by approximately half, so that after L = 2 encoder layers, the final sequence has length ≈ 6 before global average pooling. In the present TensorFlow/Keras implementation, the ProbSparse attention of the original paper is approximated with standard multi-head attention (Eq. 10), as the dynamic index-sampling required by the exact mechanism is incompatible with static TensorFlow graphs; the distillation convolution in Eqs. 19–20, which constitutes the architecture's primary computational contribution, is retained in full.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

2.3.4 Autoformer

The Autoformer [17] introduces two interdependent innovations: a series decomposition block that separates each hidden representation into a seasonal and a trend component, and an Auto-Correlation mechanism that replaces dot-product attention with FFT-based temporal cross-correlation. The decomposition, applied after both the attention and feed-forward sub-layers, uses a same-padded moving average with kernel k = 25:

$$trend = AvgPool_{k(Z)};\ season = Z - trend \tag{21}$$

The Auto-Correlation mechanism computes cross-correlation between projected queries and keys in the frequency domain via the Wiener–Khinchin identity, returning time-domain attention weights $A = softmax(Corr)$ applied elementwise to the value projection, AutoCorr(Z) = (A⊙V)W^O, following the formulation of Wu et al. [17]. At each encoder layer, the Auto-Correlation block replaces MHA in Eq. 17, and the series decomposition (Eq. 21) is applied after both sub-layers, with only the seasonal component propagated forward.

2.3.5 FEDformer

The FEDformer [18] replaces the attention sub-layer entirely with a Frequency-Enhanced Block that performs learnable low-rank mixing in the frequency domain. Given the hidden representation Z, the block retains only the M = 64 lowest-frequency components of the real-valued FFT, applies a complex-valued linear transformation parameterized by learnable real and imaginary weight matrices $W\{real\}, W\{imag\} \in R^{\{D\times D\}}$, before reconstructing the time-domain representation via an inverse FFT following the complex-valued transform and zero-padding procedure of Zhou et al. [18].

$$X_f = RFFT(Z^{\mathrm{T}}) \in C^{\left\{D\times\left(\left\lfloor\frac{T}{2}\right\rfloor+1\right)\right\}};\ X_f^{\{low\}} = X_{f[:,:M]} \tag{22}$$

Series decomposition (Eq. 21) is applied after both the FourierBlock and FFN sub-layers. Positional encoding is omitted, as the frequency-domain representation implicitly encodes global temporal structure.

2.3.6 PatchTST

The PatchTST [19] adopts a channel-independent tokenization strategy in which each of the F input features is processed by a dedicated Transformer encoder operating over subsequence patches rather than individual time steps. For each channel $c \in \{1, \dots, F\}$, overlapping patches of length P = 4 with stride S = 2 are embedded by a strided 1-D convolution:

$$N_{\{patch\}} = \left\lfloor\frac{T-P}{S}\right\rfloor + 1 = 11 \tag{23}$$

$$Z^{\{(c)\}\{patch\}} = Conv1D_{\{k=P,s=S\}}\left(X^{\{(c)\}}\right) + PE^{\{(c)\}} \in R^{\{N_{\{patch\}}\times D\}} \tag{24}$$

Each channel's patch sequence is then processed by an independent L-layer Transformer encoder with the standard attention/FFN sub-layers (Eqs. 17–18). The resulting per-channel context representations are pooled, concatenated across all channels, and linearly projected back to dimension D via global average pooling, channel concatenation, and a fusion layer. This channel-independent design prevents spurious cross-variable interference during patch-level attention, while the 4-step patch window effectively quadruples the temporal receptive field relative to a single-step token.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

2.3.7 iTransformer

The iTransformer [20] inverts the conventional tokenization axis by transposing the input tensor, so that each of the F input features, rather than each time step, becomes an individual attention token. The entire 24-step time series of each feature is linearly projected into a single D-dimensional embedding:

$$Z^{\{(0)\}} = X^{\mathrm{T}} \cdot W_{\{tok\}} + b_{\{tok\}} \in R^{\{F \times D\}}(F \ = \ 4 \ tokens) \tag{25}$$

With positional encoding unnecessary under this formulation, multi-head self-attention is computed across the F = 4 variable tokens (Eqs. 9–10), modeling inter-variable correlations, for instance, the dependency between translational speed and forward bearing, while delegating intra-variable temporal pattern extraction to the FFN sub-layers (Eq. 11). After L encoder layers, the context vector is obtained by pooling over the feature tokens:

$$c = \left(\frac{1}{F}\right) \Sigma_{\{f=1\}}^{\{F\}} Z_{\{(L)\}}^{f} \in R^{D} \tag{26}$$

2.4 Experimental Setup

Table 1 shows the hyperparameter configuration applied uniformly to all six architectures.

Table 1. Hyperparameter Configuration for All Six Transformer Architectures

Configuration	Values
Dimension (D)	64
Attention Heads (H)	4
Feed-Forward (D_{ff})	256
Encoder Depth (L)	2
Learning Rate (η)	0.001
Epochs	100
Batch Size	32

Data were partitioned temporally to prevent information leakage: training set (storms with target year ≤ 2019), validation set (2020–2024), and test set (≥2025). Table 2 reports the number of individually named tropical cyclones per year across the full 1965–2025 IBTrACS record used in this study; aggregated by split, this corresponds to 465 storms in the training set, 38 storms in the validation set, and 8 storms in the test set (511 storms in total), providing the explicit sample basis for assessing the statistical significance of the test-set results reported in Section 3. Where the temporal split yields fewer than 200 training sequences a safeguard against thin temporal coverage the procedure falls back to a stratified random 70/15/15 split seeded with the global reproducibility seed (SEED = 42). Each architecture was trained independently on the 12-hour lead-time training partition. Three training callbacks were applied uniformly: early stopping (monitoring validation total loss, patience = 15 epochs, restoring the best-weight checkpoint),

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

learning-rate reduction on plateau (reduction factor = 0.5, patience = 7, minimum rate = 10^{-6}), and model checkpointing (retaining the lowest-validation-loss weights). After training, each model was evaluated without parameter update on the test partitions of all six lead-time datasets, thereby directly measuring cross-horizon generalisation from a single set of learned weights.

Table 2. Number of Historical Tropical Cyclones per Year in the IBTrACS Dataset Used in This Study (1965–2025)

Year	N	Year	N	Year	N
1965	9	1966	11	1967	4
1968	12	1969	4	1970	11
1971	12	1972	4	1973	16
1974	12	1975	13	1976	7
1977	7	1978	5	1979	8
1980	13	1981	8	1982	8
1983	9	1984	12	1985	10
1986	12	1987	6	1988	5
1989	9	1990	9	1991	8
1992	6	1993	7	1994	12
1995	8	1996	17	1997	7
1998	11	1999	10	2000	11
2001	10	2002	5	2003	8
2004	11	2005	8	2006	9
2007	8	2008	11	2009	8
2010	7	2011	9	2012	6
2013	7	2014	6	2015	7
2016	3	2017	8	2018	7
2019	7	2020	9	2021	7
2022	11	2023	6	2024	6
2025	8				

2.5 Evaluation Metrics

Forecast accuracy was assessed with four deterministic error metrics, one explained-variance metric, and three categorical skill scores, applied separately to the position and intensity outputs at each lead time. Position error for each prediction was first computed as the great-circle distance between the predicted and observed storm centres using Eq. 1. The four deterministic metrics are then defined as follows. Mean Absolute Error (MAE) measures the average magnitude of error in physical units (km for position, kt for intensity):

$$MAE = \left(\frac{1}{n}\right) \Sigma_{\{i=1\}}^{\{n\}} |y_i - \hat{y}_i| \tag{27}$$

Root Mean Square Error (RMSE) assigns disproportionately greater penalty to large errors and is therefore a sensitive indicator of forecasting skill during rapidly evolving or extreme events:

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right) \Sigma_{\{i=1\}}^{\{n\}} (y_i - \hat{y}_i)^2 \right]} \tag{28}$$

Mean Absolute Percentage Error (MAPE) expresses error as a fraction of the true value, facilitating scale-independent comparison across lead times:

$$MAPE = \frac{\left(\frac{100}{n}\right) \Sigma_{\{i=1\}}^{\{n\}} |y_i - \hat{y}_i|}{|y_i|} \tag{29}$$

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

The coefficient of determination R^2 quantifies the proportion of variance in the observations explained by the model:

$$R^2 = 1 - \frac{[\Sigma(y_i - \hat{y}_i)^2]}{[\Sigma(y_i - \bar{y})^2]} \quad (30)$$

Categorical skill scores were computed at three position-error thresholds $d \in \{50,100,200\}$ km to quantify operational forecast utility. Hit Rate (HR) is the proportion of test cases whose positional error falls within threshold d; False Alarm Rate (FAR) is its complement; and Pierce Skill Score (PSS) measures forecast discrimination on the interval [−1, 1]:

$$HR = \frac{n_{\{hit\}}}{n_{\{total\}}}; FAR = 1 - HR; PSS = HR - FAR \quad (31)$$

To synthesise the multi-metric, multi-horizon evaluation into a single comparative ranking, a Weighted Sum Model (WSM) was applied. For each metric $j \in \{Pos_{MAE}, Pos_{RMSE}, Pos_{MAPE}, Wind_{MAE}, Wind_{RMSE}\}$, the average value across all six lead times was first computed per architecture. Each averaged metric was then min–max normalised across architectures so that 0 corresponds to the best-performing architecture and 1 to the worst:

$$r_{\{ij\}} = \frac{x_{\{ij\}} - min_{j(x)}}{max_{j(x)} - min_{j(x)}} \quad (32)$$

The composite WSM score for architecture (i) is the weighted sum of its normalised metric vector:

$$S_i = \Sigma_{\{j\}w_j} \cdot r_{\{ij\}}, w = \{0.30, 0.20, 0.15, 0.20, 0.15\} \quad (33)$$

where the weights correspond to $\{Pos_{MAE}, Pos_{RMSE}, Pos_{MAPE}, Wind_{MAE}, Wind_{RMSE}\}$ and sum to unity. The positional sub-weights total 0.65, reflecting the operational primacy of track accuracy in TC hazard guidance; the intensity sub-weights total 0.35. Architectures are ranked by ascending S_i, with $S_i = 0$ indicating a hypothetical model that minimises every metric simultaneously. An independent per-lead-time ranking is computed by applying the same normalisation and weighting within each horizon's test results.

3. RESULTS AND DISCUSSION

3.1 Overall Predictive Performance

Figure 1 shows the training loss (blue) and validation loss (orange) curves for six Transformer architectures: Vanilla Transformer, Informer, Autoformer, FEDformer, PatchTST and iTransformer.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

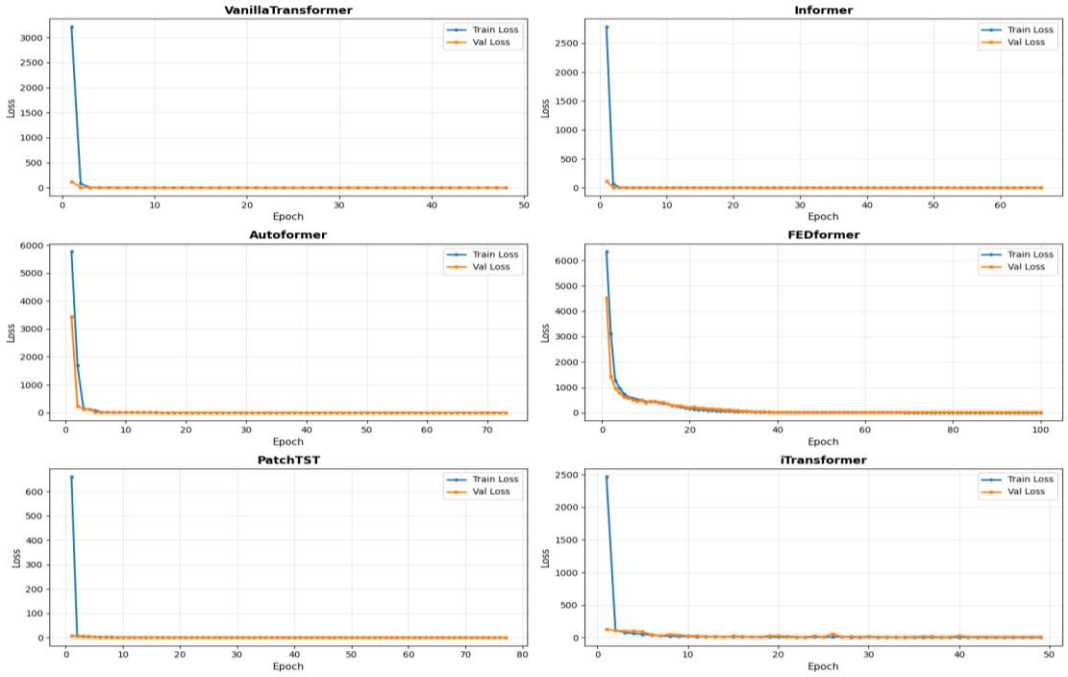

Figure 1. Training and validation loss curves for all six Transformer architectures

All models exhibit a very sharp decline in loss during the first few epochs, followed by rapid stabilization, with training and validation losses being very close together, thus confirming stable convergence with no signs of overfitting.

Variations in the number of epochs are clearly evident, ranging from around 50 epochs (Vanilla Transformer and iTransformer) to as many as 100 epochs (FEDformer), whilst Autoformer (~70 epochs), PatchTST (~80 epochs) and Informer (~60 epochs) also exhibit consistent training behavior. The variation in convergence epoch count is architecturally informative: FEDformer required the full 100 epochs, consistent with the greater parameter count of its complex-valued Fourier mixing block, whereas the Vanilla Transformer and iTransformer converged in approximately 50 epochs, suggesting that simpler or variate-inverted attention mechanisms adapt more rapidly to this low-dimensional dataset. Overall, these training dynamics indicate that all architectures achieved reliable convergence, with differences primarily reflecting their architectural complexity rather than optimization instability.

Table 3 reports mean values of each error metric, averaged over the six forecast lead times, for all architectures evaluated on the held-out test set. A clear stratification is immediately apparent: the top four architectures (PatchTST, Informer, VanillaTransformer, iTransformer) achieve mean positional MAE values clustered between 368 and 382 km, while FEDformer is an outlier at 538 km, some 46–70 km worse than every other model. Wind intensity errors show a different ordering: Autoformer, FEDformer, and iTransformer share the lowest wind MAE (approximately 17 kt), whereas Informer posts the highest at 24.07 kt despite having competitive positional accuracy (371.49 km). This inversion of relative rankings between positional and intensity metrics motivates the multi-criteria WSM aggregation in Section 3.2.

Table 3. Mean Error Metrics Averaged Across All Lead Times (12–72 h)

Model	Pos MAE (km)	Pos RMSE (km)	Pos MAPE (%)	Wind MAE (kt)	Wind RMSE (kt)
VanillaTransformer	381.7	427.6	32.9	22.19	27.04

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

Informer	371.49	414.5	32.03	24.07	29.58
Autoformer	389.01	437.9	33.51	16.98	23.16
FEDformer	537.9	585	46.02	16.94	22.13
PatchTST	367.99	415.1	31.74	19	23.78
iTransformer	**381.21**	**434**	**32.79**	**17.36**	**22.43**

3.2 Weighted Sum Model Composite Ranking

The overall predictive performance of the six Transformer architectures was further synthesised through a Weighted Sum Model (WSM) to provide a comprehensive, multi-criteria ranking that balances positional accuracy and intensity prediction across all forecast horizons.

Table 4. WSM Composite Ranking Aggregated Across All Lead Times

Rank	Model	WSM Score	Pos MAE	Pos RMSE	Wind MAE	Wind RMSE
1	**iTransformer**	**0.075**	**381.21**	**433.95**	**17.36**	**22.43**
2	PatchTST	0.092	367.99	415.11	19	23.78
3	Autoformer	0.105	389.01	437.87	16.98	23.16
4	VanillaTransformer	0.298	381.7	427.62	22.19	27.04
5	Informer	0.359	371.49	414.48	24.07	29.58
6	FEDformer	0.65	537.9	585.02	16.94	22.13

Table 4 presents the WSM composite scores computed using Eqs. 32–33; the composite ranking is visualised in Figure 2. The ranking reveals a pronounced two-tier structure. The top tier iTransformer (S = 0.075), PatchTST (0.092), and Autoformer (0.105) comprises architectures whose WSM scores fall below 0.11, indicating a broadly balanced profile across both positional and intensity metrics. This distinct separation demonstrates that architectural design choices exert a substantial influence on predictive performance in this domain. A substantial gap then separates this tier from VanillaTransformer (0.298), Informer (0.359), and FEDformer (0.650).

The iTransformer's top-overall placement is instructive: it is not individually the best model on any single raw metric, yet it is the only architecture that simultaneously achieves near-minimum wind error (MAE = 17.36 kt, the second-lowest alongside Autoformer) while maintaining competitive positional accuracy (381.21 km). Because the WSM weights positional performance at 0.65 and intensity at 0.35 (Eq. 33), architectures that excel on one dimension at the cost of the other, such as Informer (lowest Pos MAE but worst Wind MAE) or FEDformer (lowest Wind RMSE but worst Pos MAE), receive inflated normalised scores on their weak metric, pulling their composite score upward.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

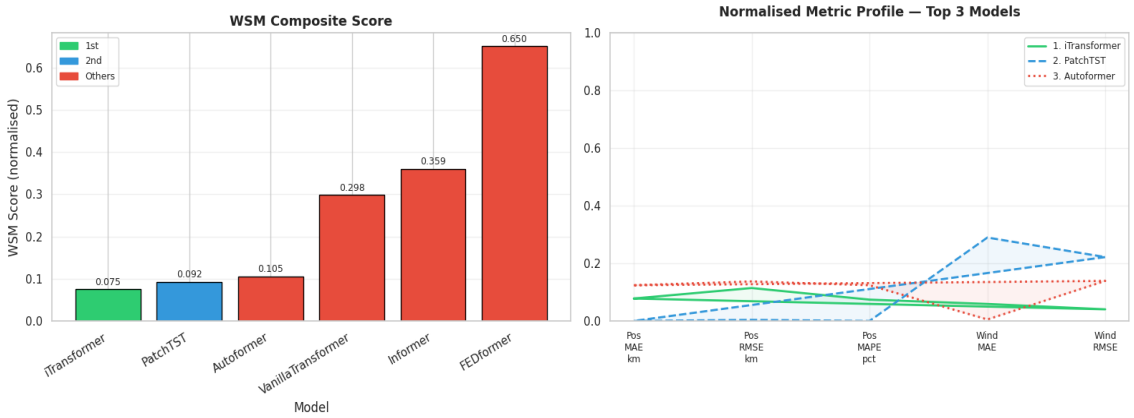

Figure 2. Weighted Sum Model (WSM) composite ranking

FEDformer's last-place ranking (S = 0.650) warrants specific attention, given that its wind-intensity errors are competitive with those of the top three architectures. The architecture's positional error (MAE = 537.90 km) is 38–46% higher than any other model, implying that the learnable low-frequency Fourier mixing in Eq. 22 is poorly adapted to this particular feature space. The TC best-track input, four kinematic variables sampled at 3-hour intervals over 24 steps, does not exhibit the strongly periodic multi-scale structure that motivated FEDformer's design, and the complex weight parameterization of the FourierBlock introduces additional overfitting risk given the comparatively limited volume of training data. These aggregate scores, however, conceal systematic variation in model ranking across individual forecast horizons, which is examined in Section 3.3.

3.3 Horizon-Stratified Analysis

Table 5 presents the WSM rankings recomputed independently at each of the six lead times. This horizon-stratified view reveals a consistent transition in model superiority across the forecast range that is obscured by the aggregate analysis in Table 4.

Table 5. WSM Per-Lead-Time Ranking: Horizon-Stratified Results

Lead	1st	2nd	3rd	4th	5th	6th
12 h	Autoformer (0.069)	PatchTST (0.125)	iTransformer (0.220)	Vanilla (0.235)	Informer (0.351)	FEDformer (0.699)
24 h	Autoformer (0.084)	PatchTST (0.101)	iTransformer (0.103)	Vanilla (0.280)	Informer (0.356)	FEDformer (0.675)
36 h	iTransformer (0.062)	PatchTST (0.089)	Autoformer (0.106)	Vanilla (0.311)	Informer (0.365)	FEDformer (0.655)
48 h	iTransformer (0.040)	PatchTST (0.086)	Autoformer (0.134)	Vanilla (0.324)	Informer (0.367)	FEDformer (0.650)
60 h	iTransformer (0.058)	PatchTST (0.108)	Autoformer (0.181)	Vanilla (0.350)	Informer (0.363)	FEDformer (0.650)
72 h	iTransformer (0.097)	PatchTST (0.127)	Autoformer (0.241)	Vanilla (0.366)	Informer (0.370)	FEDformer (0.650)

The lower-tier rankings remain consistent in ordering across all horizons; only the top-three positions change. At the two shortest lead times (12 and 24 h), Autoformer ranks first by a meaningful margin over PatchTST (WSM gap of 0.056 at 12 h and 0.017 at 24 h). This is

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

consistent with the behavior anticipated from Autoformer's series decomposition mechanism (Eq. 21): at short horizons, TC motion is predominantly governed by a smooth, slowly evolving trend in the steering-layer flow. The explicit separation of seasonal and trend components, and the retention of only the seasonal component through the encoder, appears to provide a useful inductive bias that suppresses irrelevant high-frequency noise at these horizons.

A clear regime transition occurs at 36 h, where iTransformer overtakes Autoformer for the first time and sustains top-ranked performance through to 72 h. iTransformer's best lead-time score occurs at 48 h (S = 0.040), precisely the horizon where forecasting uncertainty is most substantially elevated by the chaotic divergence of atmospheric steering-flow trajectories. The inverted-attention formulation (Eqs. 25–26), which models pairwise cross-variable dependencies among the four input features rather than temporal dependencies, may confer an advantage at these longer horizons by capturing the multivariate coupling between translational speed, bearing, and geodetic position that is most informative for medium-range steering-flow extrapolation. This interpretation is consistent with the broader finding in Liu et al. [20] that inverted attention is particularly beneficial when cross-variate interactions, rather than autocorrelations within individual series, carry the dominant predictive information.

PatchTST occupies second place at every lead time from 24 to 72 h. Within the medium-range window (36–72 h) where iTransformer holds first place, the WSM gap between the two architectures remains narrow, ranging from only 0.027 (at 36 h) to 0.030 (at 72 h). This consistency reflects the patch tokenization strategy (Eqs. 23–24): by aggregating four consecutive 3-hour observations into each patch token, PatchTST effectively operates on 12-step (36-hour) local temporal windows during attention computation, enlarging the receptive field relative to single-step tokenization and providing some degree of smoothing analogous to the series decomposition in Autoformer, albeit in a learnable, position-aware form rather than a moving-average form.

The lower half of the ranking, VanillaTransformer (4th), Informer (5th), FEDformer (6th), is stable across all six lead times. VanillaTransformer's mid-tier placement suggests that standard scaled dot-product attention without any architectural specialization provides a competitive but not superior baseline for this dataset. The Informer's consistent 5th-place finish is somewhat counterintuitive given its theoretical advantages: the distillation mechanism (Eqs. 19–20) was designed to improve long-sequence forecasting by progressively concentrating attention mass, but at T = 24 the computational savings are marginal, and the aggressive max-pooling may discard fine-grained kinematic transitions that are predictive of storm-motion changes. FEDformer's ceiling-like WSM behavior- its score remains near 0.65–0.70 across all lead times rather than degrading progressively- indicates that its errors saturate early relative to the other architectures, converging to a poor-but-stable regime rather than compounding further.

These results carry practical implications for operational deployment. Where a single fixed architecture is required across the full 12–72-hour forecast range, as is the norm in operational meteorological systems, PatchTST offers the most consistently high performance with the lowest inter-horizon variance in WSM rank. Where a horizon-adaptive strategy is feasible, pairing Autoformer for guidance at ≤24 h with iTransformer at ≥36 h is predicted to yield the best combined performance. Such targeted deployment strategies could serve as a valuable complement to traditional NWP systems in Indonesia's early-warning infrastructure. The weak results for FEDformer and, to a lesser extent, Informer indicate that the most computationally sophisticated architectural modifications in this set are not universally beneficial, particularly when applied to a kinematic feature space as sparse and low-dimensional as the four-variable best-track dataset used here.

3.4 Case Study: Tropical Cyclone Taliah

Aggregate metrics such as Tables 3–5 average prediction error across all storms, lead times, and track geometries, which can obscure how architectural differences manifest along a single, physically continuous trajectory. TALIAH is the best-represented storm in the 24-h test

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

partition, contributing n = 50 sequences, the largest single-storm sample of any track evaluated in this study, and is used here as a spatially resolved case study of the two architectures at the extremes of the WSM ranking: iTransformer (rank 1, S = 0.075) and FEDformer (rank 6, S = 0.650). Figure 3 compares the observed track against each architecture's 24-h-ahead forecasts for this storm.

The observed track exhibits three kinematically distinct phases. Between 90°E and 93°E the storm undergoes a rapid equatorward translation, from approximately 19.7°S to 14°S over only three degrees of longitude, corresponding to a brief interval of elevated translational speed and a large, abrupt change in heading. From 93°E to approximately 110°E, the storm enters an extended quasi-stationary phase, oscillating within a narrow 14–16.5°S band with repeated small-scale reversals in heading; this segment is characterized by low temporal persistence in the bearing feature, meaning the direction of motion at time t is a poor predictor of the direction at t + 1. Beyond 110°E, the track straightens, and translational speed increases again toward 120°E, consistent with the storm entering a more strongly steered regime. The physical driver of the mid-track looping behavior is most plausibly a weakened or laterally displaced steering current, a hypothesis consistent with the known sensitivity of TC motion to ridge-axis perturbations, though full attribution requires reanalysis-derived steering-layer wind data beyond the kinematic feature set available here. This quasi-stationary segment is highlighted as the most demanding forecast condition across TALIAH's observed trajectory.

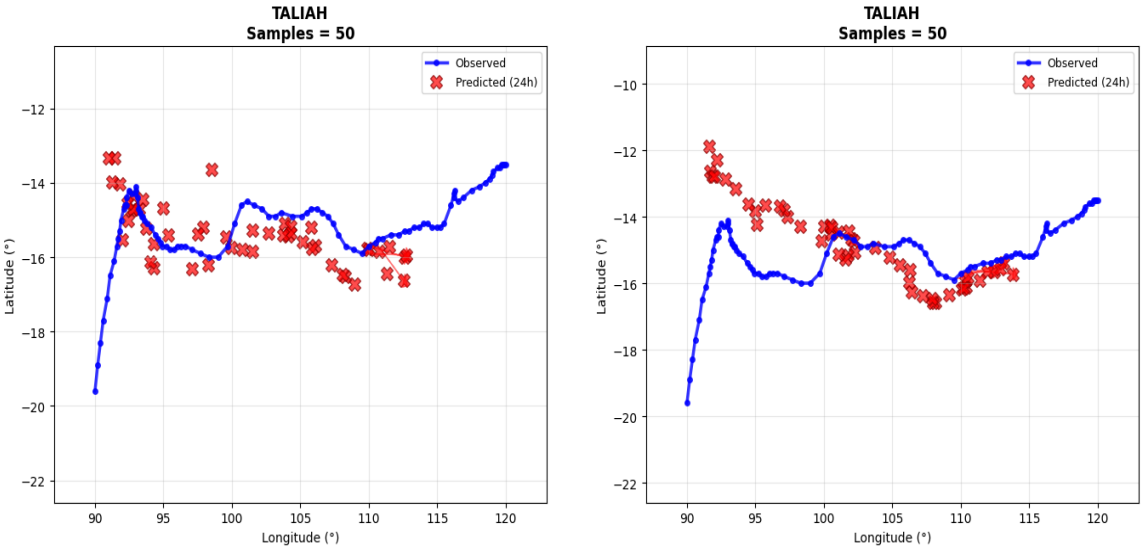

Figure 3. Observed versus predicted 24-h track for storm TALIAH: (a) iTransformer (best-ranked, S = 0.075) versus (b) FEDformer (worst-ranked, S = 0.650)

The two architectures diverge most visibly in the low-persistence segment. iTransformer's predictions (Figure 3a) track the observed curve within approximately 0.5–1° of latitude through each of the small-scale reversals between 95°E and 110°E. FEDformer's predictions (Figure 3b) show markedly larger and more systematic deviations across the same segment: near 92–93°E several forecasts overshoot the initial equatorward jog by 1.5–2° of latitude, predicting continued southward motion after the observed track has already turned north; a similar lag is visible again near 108–110°E, where FEDformer forecasts a continued southward drift roughly 1–1.5° south of the observed position. Both errors are directionally consistent rather than randomly scattered, i.e., FEDformer systematically under-reacts to reversals, which is

consistent with the effect of the FourierBlock's low-frequency truncation (Eq. 22, M = 64 retained modes): components of the input sequence responsible for short-period heading reversals are attenuated before reconstruction, biasing the forecast toward the sequence's smoother, lower-frequency trend. iTransformer's inverted-attention formulation (Eqs. 25–26)

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

instead operates on the raw variable tokens across the full 24-step history, which appears to preserve sensitivity to the abrupt bearing changes that the frequency-domain approach suppresses.

This case-level pattern is consistent with TALIAH's per-sample error distribution at the 24-h lead time: across the storm's 50 test sequences, FEDformer's median positional error is approximately 385–390 km, with an interquartile range extending to roughly 480 km and a small number of sequences exceeding 650–700 km, versus a median near 205–210 km for iTransformer with a comparable interquartile spread but a substantially lower ceiling. The gap between the two distributions is therefore not driven by a small number of outlier sequences; it reflects a systematic tendency that is present throughout the quasi-stationary segment of the track.

These storm-level observations reinforce the aggregate WSM finding that FEDformer's frequency-domain attention is poorly matched to the sparse, non-periodic kinematics of short-to-medium-range TC motion (Section 3.2), while offering a physically interpretable mechanism- delayed response to abrupt heading reversals- for why that mismatch occurs. The TALIAH case also illustrates a practical forecasting implication: architectures that smooth over short-period directional changes are disproportionately penalized during a storm's most operationally consequential phases, near-stationary or looping motion, when even small directional errors translate into large absolute displacement errors relative to the storm's already-slow translational speed.

4. CONCLUSION

This study conducted the first systematic multi-architecture Transformer benchmark for short-to-medium range tropical cyclone track and intensity forecasting in the Western Pacific and Southern Indian Ocean basins surrounding Indonesia. Six architectures spanning the major design paradigms in the Transformer time-series literature- canonical, distilling, decomposition-based, frequency-domain, patch-based, and variate-inverted- were evaluated under a strictly controlled experimental protocol with identical training conditions, a shared dual-output framework predicting both storm position and wind intensity, and a multi-horizon evaluation spanning 12 to 72 hours.

The Weighted Sum Model composite ranking identifies iTransformer as the best overall performer (S = 0.075), followed by PatchTST (0.092) and Autoformer (0.105). These three architectures form a top tier clearly separated from VanillaTransformer (0.298), Informer (0.359), and FEDformer (0.650). Horizon-stratified analysis reveals that no single architecture is uniformly optimal across the full 12–72-hour range: Autoformer's series decomposition inductive bias gives it an advantage at short lead times (≤24 h), whereas iTransformer's cross-variable inverted attention yields superior balanced performance at medium range (36–72 h). PatchTST's consistent second-place rank across all horizons establishes it as the most robust single-model choice for operational deployment. Conversely, FEDformer and Informer underperform relative to the simpler VanillaTransformer baseline, suggesting that frequency-domain mixing and ProbSparse distillation provide limited benefit for the sparse, four-dimensional kinematic feature space used in this study.

The Tropical Cyclone TALIAH case study (Section 3.4) grounds these aggregate rankings in a physically interpretable, storm-level mechanism. Examined along a single continuous track, FEDformer's largest positional errors concentrate specifically in TALIAH's quasi-stationary, low-heading-persistence phase, where its low-frequency Fourier mixing under-reacts to abrupt directional reversals and instead extrapolates the sequence's smoother recent trend, an effect visible as a systematic 1–2° latitude overshoot at the storm's two sharpest track inflections. iTransformer's inverted-attention formulation, which preserves per-variable temporal detail across the full input window rather than compressing it into a truncated frequency basis, tracks these reversals substantially more closely and shows a markedly lower and less dispersed per-sample error distribution across the same 50-sequence test set. This storm-level evidence indicates that the aggregate WSM ranking is not solely an artefact of averaging across heterogeneous tracks: the same frequency-domain smoothing that separates FEDformer from the

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

top-tier architectures in Table 4 manifests as a specific, diagnosable failure mode exactly where TC motion is least persistent, and where forecast errors are operationally most consequential given the storm's already-slow translational speed.

Several limitations circumscribe these conclusions and point toward productive avenues for future research. All architectures were evaluated on a kinematic best-track feature set alone; incorporating atmospheric reanalysis predictors (steering-layer winds, vertical wind shear, sea-surface temperature) would substantially enrich the input representation and may alter the relative ranking of architectures that benefit most from high-dimensional multivariate inputs (notably iTransformer and PatchTST), and could additionally help attribute the physical driver of quasi-stationary or looping motion phases such as the one identified in the TALIAH case study. The WSM weight vector assigns a fixed 0.65:0.35 ratio to positional versus intensity accuracy; sensitivity analysis of this weighting scheme is recommended before any single architecture is selected for operational deployment in an application where intensity accuracy is of primary concern (e.g., wind-hazard guidance). The temporal train/test split confines the test set to storms from 2025 onward, a relatively narrow evaluation window; validation on a larger held-out sample spanning multiple basin-active seasons would strengthen confidence in generalisation. The storm-level case-study.

REFERENCES

[1] Emanuel K. 100 Years of Progress in Tropical Cyclone Research. *Meteorological Monographs. 2018; 59: 15.1–15.68*.
[2] Knutson T, et al. Tropical Cyclones and Climate Change Assessment: Part II. Projected Response to Anthropogenic Warming. *Bulletin of the American Meteorological Society. 2019; 100(9): 1987–2007.*
[3] Mahubessy P, et al. The Influence of Tropical Cyclone on Extreme Rainfall in Eastern Indonesia. *Journal of Physics: Conference Series. 2018; 1008: 012007.*
[4] Mulyana E, et al. Various Characteristics of Tropical Cyclone Affecting Indonesian Territory. *IOP Conference Series: Earth and Environmental Science. 2018; 149: 012017.*
[5] Roxy MK, et al. Twofold Expansion of the Indo-Pacific Warm Pool Warps the MJO Life Cycle. *Nature. 2019; 575(7784): 647–651*.
[6] World Meteorological Organization. Global Guide to Tropical Cyclone Forecasting (WMO-No. 1194). *Geneva: World Meteorological Organization. 2017.*
[7] Navarro RA, Hakim GJ. Ensemble-Based Sensitivity Analysis Applied to Pacific-Storm Track Variability. *Monthly Weather Review. 2014; 142(7): 2527–2545.*
[8] Huang L, et al. A Systematic Review of Deep Learning for Tropical Cyclone Track Forecasting. *Earth-Science Reviews. 2025; 248: 104631.*
[9] Zhang JA, Chen SS. Rapid Intensification of Tropical Storm Fanapi (2010): What Is the Role of Upper-Level Warm Core? *Journal of Geophysical Research: Atmospheres. 2012; 117(D17): D17108.*
[10] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation. 1997; 9(8): 1735–1780.*
[11] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha. 2014: 1724–1734*.
[12] Rahman MS, et al. Comparative Analysis of RNN, LSTM, and GRU for TC Track Prediction over the North Indian Ocean. *Meteorological Applications. 2025; 32(1): e2234.*
[13] Song X, et al. Attention-Based Bidirectional GRU for TC Track Forecasting. *Frontiers in Earth Science. 2022; 10: 854755.*
[14] Wang D, et al. Comparative Study of Deep Learning Models for TC Track Forecasting in the Northwest Pacific. *Atmosphere. 2023; 14(3): 551.*

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122

[15] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention Is All You Need. *Advances in Neural Information Processing Systems. 2017; 30: 5998–6008.*
[16] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence. 2021; 35(12): 11106–11115.*
[17] Wu H, Xu J, Wang J, Long M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems. 2021; 34: 22419–22430.*
[18] Zhou T, Ma Z, Wen Q, Wang X, Sun L, Jin R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. *International Conference on Machine Learning (ICML). Baltimore. 2022: 27268–27286.*
[19] Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. *International Conference on Learning Representations (ICLR). Kigali. 2023.*
[20] Liu Y, Hu T, Zhang H, Wu H, Wang S, Ma L, Long M. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *International Conference on Learning Representations (ICLR). Vienna. 2024.*
[21] Knapp KR, et al. The International Best Track Archive for Climate Stewardship (IBTrACS). *Bulletin of the American Meteorological Society. 2010; 91(3): 363–376.*

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.17122