JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 23, No. 1, September 2026, pp. 20-37
DOI: 10.20956/48n17666

Significant Wave Height Forecasting in the Singapore Strait: Deep Learning Ensemble Stacking with ERA5-IDW Data

Sandy Salomo Saruan[1*], Gusti Ayu Dwi Yanti[2]
[1]*Department of Mathematics, Institut Teknologi Batam*
[2]*Department of Mathematics, Universitas Gadjah Mada*
***Email**: [1]sandy@iteba.ac.id, [2]gustiayu9d@gmail.com*
**Corresponding Author*
Received: 13 May 2026, revised: 14 July 2026, accepted: 16 July 2026

Abstract

Accurate significant wave height (SWH) forecasting in the Singapore Strait is critical for maritime safety and marine risk assessment along one of the world's busiest shipping corridors. This study develops a deep learning ensemble stacking framework for one-step-ahead SWH forecasting using hourly ERA5 reanalysis data spanning January 2021 to December 2025, spatially interpolated to a single target point via Inverse Distance Weighting (IDW) from five ERA5 grid points. Four architectures: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Patch Time Series Transformer (PatchTST), and inverted Transformer (iTransformer) are systematically evaluated against a persistence baseline (RMSE = 0.0188 m), with GRU achieving the best individual performance (RMSE = 0.0153 m, MAPE = 4.58%, R^2 = 0.9814), while ridge regression stacking of GRU with Transformer-based models achieves the overall best performance (RMSE = 0.0151 m, MAPE = 4.54%, R^2 = 0.9820) classified as highly accurate. Transformer-based architectures individually fail to capture SWH variability in this fetch-limited regime yet contribute complementary residual information as stacking base learners. The proposed pipeline establishes a replicable methodology for wave climate forecasting in narrow straits under resolved by global reanalysis products, with potential utility as an environment risk covariate for actuarial modeling in marine cargo insurance underwriting.

Keywords: Significant wave height forecasting, ensemble stacking, Singapore strait, ERA5 reanalysis, marine risk.

1. INTRODUCTION

The sinking of the Malaysia-registered tanker Silver Sincere off Pedra Branca on 12 January 2025, and the Vietnam-registered freighter Dolphin 18 a day earlier, exemplify how adverse sea states remain a leading cause of vessel foundering worldwide [5,20,21], accounting for nearly half of all total-loss cases globally [17]. The Singapore Strait's narrow geometry and monsoon-driven wave climate compound this risk, translating directly into hull, cargo, liability, and salvage costs that must be priced into marine insurance premiums before a voyage begins. Within marine cargo

insurance, technical premiums are fundamentally a function of route-specific loss probability and severity, underscoring the need for scientifically rigorous, forward-looking wave forecasting.

The Singapore Strait is a 113-km semi-enclosed waterway connecting the Indian Ocean (via the Strait of Malacca) to the South China Sea, regulated under the IMO Traffic Separation Scheme (TSS) (see Figure 1.1) and narrowing to 1.96 nautical miles at Phillips Channel [15,16]. With over 100,000 vessel transits in 2025, it is classified as a fetch-limited, semi-enclosed sea whose wave climate is governed by local wind forcing and a bimodal monsoon cycle: the energetic Northeast Monsoon (November-March) and the moderate Southwest Monsoon (June-September).

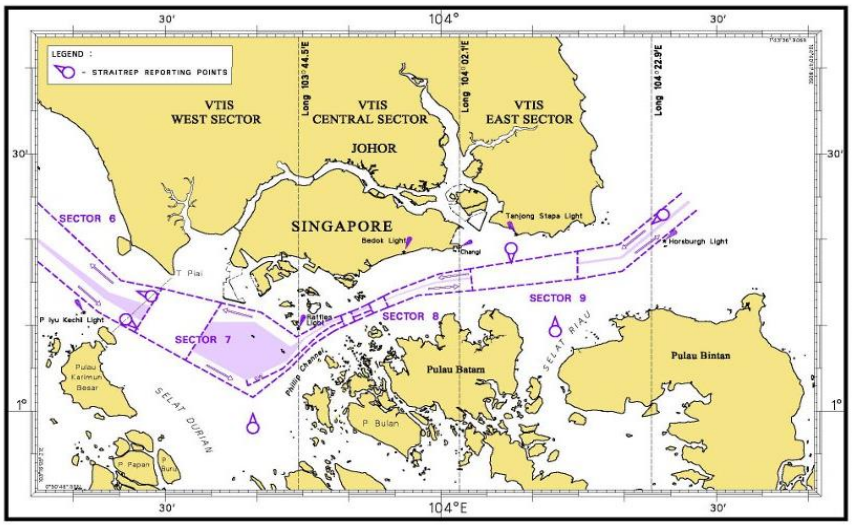

Figure 1.1. STRAITREP operational area of the Singapore Strait (Sectors 7–9) showing the IMO TSS [16]

LSTM and GRU networks are established tools for SWH forecasting using ERA5 data across multiple basins [3,10,14,18], though a persistent time-lag limitation motivates exploration of Transformer-based alternatives such as PatchTST [15] and iTransformer [13], which have shown competitive performance in open-ocean settings [23] but remain untested in narrow, semi-enclosed straits. Three gaps remain unaddressed: no systematic comparison of these architectures in a fetch-limited Southeast Asian strait; no application of IDW-interpolated ERA5 data for the Singapore Strait from an Indonesian research perspective; and no work linking SWH forecasting output to actuarial quantification of marine cargo insurance risk. These gaps motivate three research questions (RQ) addressed in this study: (**RQ1**) Which deep learning architecture achieves the highest accuracy for one-step-ahead SWH forecasting in the Singapore Strait relative to a persistence baseline? (**RQ2**) Do Transformer-based architectures outperform RNN-based models in a fetch-limited, low-variability oceanographic regime? (**RQ3**) Does ensemble stacking improve forecasting accuracy beyond the best individual model?

This study proposes an ensemble stacking framework that combines four models as base learners, with ridge regression as the meta-learner, for one-step-ahead significant wave height forecasting at the center of the Singapore Strait. Wave climate data are derived from ERA5 hourly reanalysis [1,7], spatially interpolated to a single target using IDW from valid ERA5 grid points identified within the strait's bounding region, a methodological adaptation necessitated by the coarse 0.25° spatial resolution of ERA5, which fails to directly resolve the narrow waterway. To provide a comprehensive empirical benchmark, two state-of-the-art Transformer-based architectures, PatchTST and iTransformer are additionally trained and evaluated against the RNN-based models. Three ensemble strategies are compared: simple averaging, inverse-RMSE weighted averaging, and ridge regression

stacking with the stacking approach emerging as the superior configuration. More broadly, this work establishes a replicable pipeline from ERA5 acquisition and IDW interpolation through deep learning ensemble forecasting that can be directly extended to other narrow straits and coastal waterways in the Indonesian archipelago, and whose output can serve as a probabilistic input to actuarial models for maritime cargo insurance risk quantification along the Singapore Strait corridor.

The remainder of this paper is organized as follows. Section 2 details the study area, data sources, preprocessing pipeline, model architectures, ensemble strategies, and evaluation metrics. Section 3 presents the experimental results and discussion, covering wave climate analysis, individual model performance, and ensemble comparison. Section 4 concludes the paper with a summary of key findings, implications for maritime risk assessment, and directions for future research.

2. MATERIAL AND METHODS

2.1. Data ERA5 with IDW Interpolation

This study utilizes hourly reanalysis data from the ERA5 global reanalysis product developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), accessed through the Copernicus Climate Data Store [1]. ERA5 provides a comprehensive set of atmospheric and oceanic variables at a native spatial resolution of 0.25° × 0.25° (~28 km) from 1940 to present, making it one of the most widely adopted reanalysis datasets for oceanographic and climatological research. For this study, six variables were extracted over the bounding region of the Singapore Strait (1.0°N-1.25°N, 103.50°E-104.00°E) for the period January 1, 2021 to December 31, 2025: significant height of combined wind waves and swell (swh), mean wave period (mwp), mean wave direction (mwd), 10-meter eastward wind component (u10), 10-meter northward wind component (v10), and surface pressure (sp), yielding a total of 43,824 hourly observations.

Given that the native spatial resolution of ERA5 does not directly resolve the narrow geometry of the Singapore Strait, IDW interpolation was applied to aggregate data from five valid ERA5 grid points to a single representative target point at (1.20°N, 103.83°E), the target coordinate was selected as it corresponds to the navigational centroid of the Singapore Strait main channel, lying within the TSS corridor where vessel traffic density is highest and wave-induced maritime risk is most operationally relevant [19]. IDW is a deterministic spatial interpolation method that assigns greater weight to observations at nearer grid points, based on the assumption that spatial autocorrelation decreases monotonically with distance. The interpolated value $\hat{z}$ at the target point is defined as

$$\hat{z} = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i}, \tag{2.1}$$

where z_i denotes the observed value at ERA5 grid point i, $w_i = 1/d_i^p$, d_i is the Euclidean distance from grid point i to the target point, $n = 5$ is the number of neighboring grid points, and $p = 2$ is the power parameter controlling the rate of distance-based weight decay. The five valid ERA5 grid points used for interpolation are located at: (1.00°N, 103.50°E), (1.00°N, 103.75°E), (1.00°N, 104.00°E), (1.25°N, 103.75°E), and (1.25°N, 104.00°E), with Euclidean distances to the target point ranging from approximately 0.25° to 0.43°, and corresponding IDW weights ($p = 2$) of 0.154, 0.269, 0.154, 0.269, and 0.154 respectively after normalization.

Following interpolation, three additional features were derived through feature engineering to enrich the physical and temporal information available to the forecasting models. Wind speed magnitude was computed as $U_{10} = \sqrt{u_{10}^2 + v_{10}^2}$, providing a scalar measure of the primary wave-forcing variable. The cyclical nature of the monsoon seasonal cycle and diurnal wind patterns was captured through sine-cosine encoding of month m and hour h: $\text{month\_sin} = \sin(2\pi m/12)$, $\text{month\_cos} = \cos(2\pi m/12)$, $\text{hour\_sin} = \sin(2\pi h/24)$, and $\text{hour\_cos} = \cos(2\pi h/24)$. This encoding preserves the circular continuity of temporal variables ensuring, for example, that December and January are represented as adjacent rather than maximally distant. The final input

feature set comprises 11 variables, as summarized in Table 2.1. All 11 input features are provided to each model simultaneously, including swh as an autoregressive lag feature capturing the persistence structure of the time series. To ensure that model performance reflects genuine predictive skill beyond simple persistence, a naive persistence baseline, wherein the predicted SWH at time $t+1$ equals the observed SWH at time t is included in the evaluation.

Prior to model training, all features were normalized to the range $[0,1]$ using Min-Max scaling fitted exclusively on the training partition, with the learned transformation subsequently applied to the validation and test sets to prevent data leakage. Input sequences were constructed via a sliding window of length $T = 336$ hours (14 days), producing input tensors of shape $(N, 336, 11)$, where N denotes the number of samples. The forecasting target is defined as the SWH value at the immediately subsequent time step, consistent with a one-step-ahead forecasting formulation. The dataset was partitioned chronologically into training 70%, validation 15%, and test 15% sets.

Table 2.1. Summary of input features used in the forecasting models, including source variables, derived features, and encoding methods

Feature	Description	Source
swh	Significant wave height (lag)	ERA5 direct
wind_speed	Wind speed magnitude	Derived: $\sqrt{u_{10}^2 + v_{10}^2}$
u10	Eastward wind component	ERA5 direct
v10	Northward wind component	ERA5 direct
pressure	Surface pressure	ERA5 direct
wave_period	Mean wave period	ERA5 direct
wave_dir	Mean wave direction	ERA5 direct
month_sin	Monthly seasonality (sine)	Cyclical Encoding
month_cos	Monthly seasonality (cosine)	Cyclical Encoding
hour_sin	Diurnal pattern (sine)	Cyclical Encoding
hour_cos	Diurnal pattern (cosine)	Cyclical Encoding

All experiments were implemented in Python 3.11.15 version using TensorFlow 2.21.0. Fixed random seeds were set for NumPy (seed=275), and TensorFlow global random state prior to each model instantiation, ensuring fully deterministic and reproducible results without requiring multiple training runs.

2.2. Recurrent Neural Network Architecture

This subsection presents the model of two recurrent neural network-based architectures, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU).

LSTM is a recurrent neural network architecture introduced by Hochreiter and Schmidhuber in 1997 to address the vanishing gradient problem inherent in standard RNNs, which prevents the learning of long-range temporal dependencies [8]. LSTM achieves this through a gated memory cell mechanism that selectively retains or discards information across time steps. As illustrated in Figure 2.1, each LSTM cell comprises three gates: the forget gate f_t, the input gate i_t, and the output gate o_t, which collectively regulate the flow of information into and out of the cell state C_t. The forget gate determines what proportion of the previous cell state is preserved, the input gate controls what new information is written to the cell state, and the output gate filters what portion of the cell state is exposed as the hidden state h_t to the next time step. The cell state update, which represents the core computational step of LSTM, is defined as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \,, \tag{2.2}$$

where $f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big)$ is the forget gate activation, $i_t = \sigma(W_i\,[h_{t-1}, x_t] + b_i)$ is the input gate activation, $\tilde{C}_t = \tanh(W_C\,[h_{t-1}, x_t] + b_C)$ is the candidate cell state, $\odot$ denotes element-wise

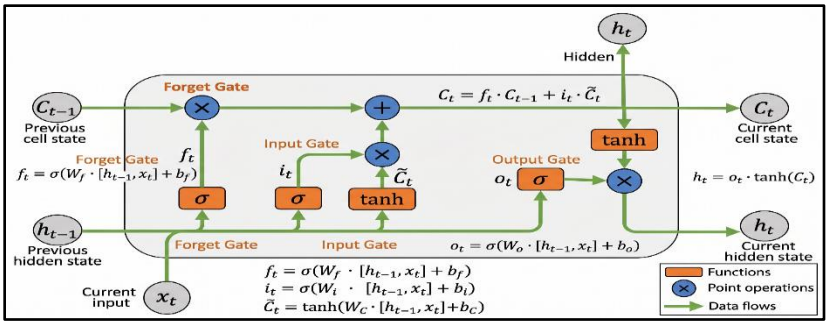

Figure 2.1. Architecture of the LSTM cell

multiplication, σ is the sigmoid activation function, x_t is the input vector at time step t, and W and b are learnable weight matrices and bias vectors respectively. In this study, a stacked two-layer LSTM architecture is employed with 64 hidden units in the first layer and 32 in the second, followed by a fully connected layer with 16 units and ReLU activation, and a linear output layer producing the one-step-ahead SWH forecast.

The GRU, proposed by Cho et al. in 2014, is a streamlined variant of LSTM that consolidates the cell state and hidden state into a single hidden state vector, and replaces the three-gate mechanism with two gates: the reset gate r_t and the update gate z_t [2]. This architectural simplification reduces the number of trainable parameters while preserving the capacity to model temporal dependencies, making GRU particularly well-suited for datasets where model complexity must be balanced against overfitting risk. As shown in Figure 2.2, the reset gate controls how much of the previous hidden state contributes to the candidate hidden state, while the update gate interpolates between the previous hidden state and the candidate, determining the degree to which the hidden state is updated at each time step. The hidden state update equation, the core computational step of GRU is expressed as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t\,, \tag{2.3}$$

where $z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$ is the update gate activation, $r_t = \sigma(W_r\,[h_{t-1}, x_t] + b_r)$ is the reset gate activation, $\tilde{h}_t = \tanh(W_h\,[r_t \odot h_{t-1}, x_t] + b_h)$ is the candidate hidden state, and all other symbols follow the same conventions as defined for LSTM.

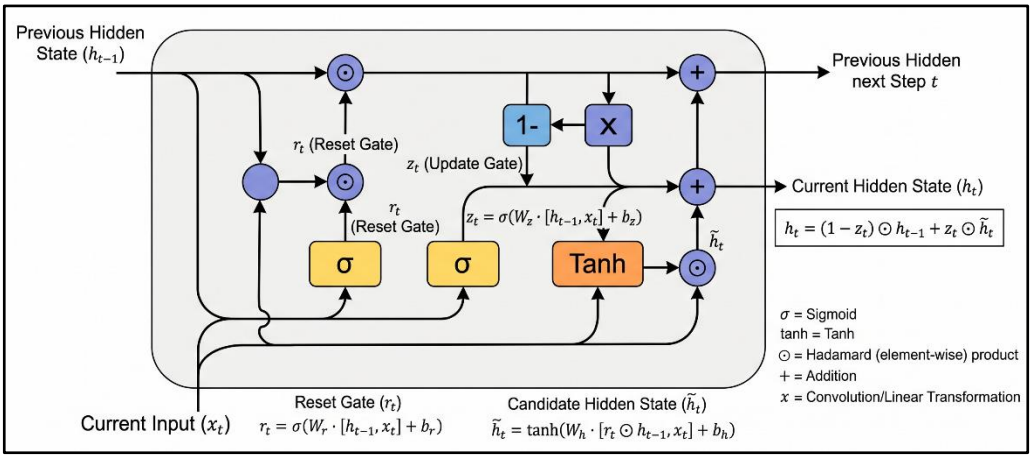

Figure 2.2. Architecture of the GRU cell

2.3. Transformer Architecture

This subsection presents the model of two state-of-the-art Transformer-based architectures, Patch Time Series Transformer (PatchTST) and Inverted Transformer (iTransformer). Both architectures represent recent advances in attention-based sequence modeling and have demonstrated competitive performance in long-horizon forecasting tasks across various domains.

The PatchTST, introduced by Nie et al. in 2022, adapts the Transformer architecture for multivariate time series forecasting by decomposing the input sequence into non-overlapping local patches prior to self-attention computation [15]. This patching mechanism addresses a fundamental limitation of vanilla Transformer applied to time series: when individual time steps are treated as tokens, the quadratic complexity of self-attention scales prohibitively with sequence length, and each token carries insufficient local context for meaningful attention. By grouping consecutive time steps into patches of length P, PatchTST reduces the number of tokens from L to $\lfloor L/P \rfloor$, simultaneously decreasing computational complexity and enriching each token with local temporal semantics. A second distinguishing design principle of PatchTST is channel-independence, whereby each input variate is processed through an independent Transformer encoder sharing the same weights, preventing cross-variate information leakage and improving generalization on datasets where inter-variable relationships are weak or noisy. As illustrated in Figure 2.3, the input sequence of length L and C variates is first segmented into $N = \lfloor L/P \rfloor$ patches per variate, linearly projected to a d_{model}-dimensional embedding space, and passed through N_{layers} Transformer encoder blocks each comprising Multi-Head Self-Attention and a position-wise Feed-Forward Network with residual connections and Layer Normalization. The self-attention mechanism, the core computational step of PatchTST is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2.4}$$

where $Q = ZW^Q$, $K = ZW^K$, $V = ZW^V$ are the query, key, and the value of the patch embedding matrix $Z \in \mathbb{R}^{N \times d_{\text{model}}}$, $d_k = d_{\text{model}}/h$, is the perhead dimensionality, and h is the number of attention heads.

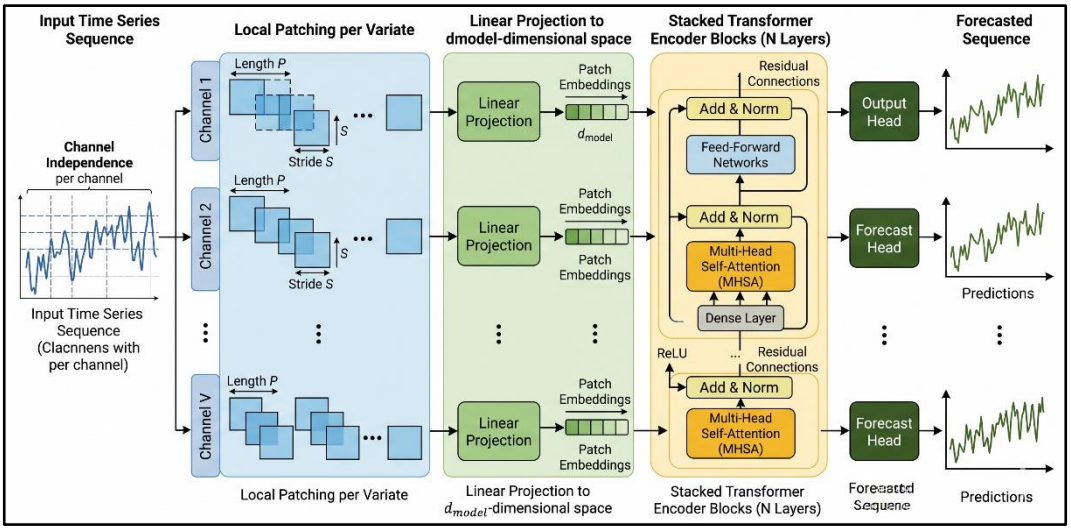

Figure 2.3. Architecture of PatchTST for time series forecasting

The iTransformer, proposed by Liu et al. in 2024, revisits the conventional tokenization paradigm of time series Transformers by inverting the roles of the time and variate dimensions [13]. In standard Transformer-based forecasting models, each time step across all variates constitutes a token, an arrangement that Liu et al. argue is suboptimal because time-point tokens carry limited

semantic meaning and self-attention across time steps fails to capture multivariate correlations effectively. iTransformer instead treats each input variate as a single token, with its entire temporal sequence of length L serving as the token embedding. This inversion, illustrated in Figure 2.4, repositions the self-attention mechanism to model inter-variate correlations rather than temporal dependencies, a design particularly advantageous when the dataset contains many variates with meaningful physical relationships, such as simultaneous wind components, wave period, and wave direction. The layer normalization is applied along the variate dimension rather than the time dimension, and the Feed-Forward Network processes each variate's temporal embedding independently. Following the inverted embedding, the attention output is defined identically to Equation (2.4), with the distinction that $Z \in \mathbb{R}^{M\times L}$ now represents M variate tokens each of length L, such that:

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

where $Q,K,V \in \mathbb{R}^{M\times d_{\text{model}}}$, where M is the number of input variates serving as tokens.

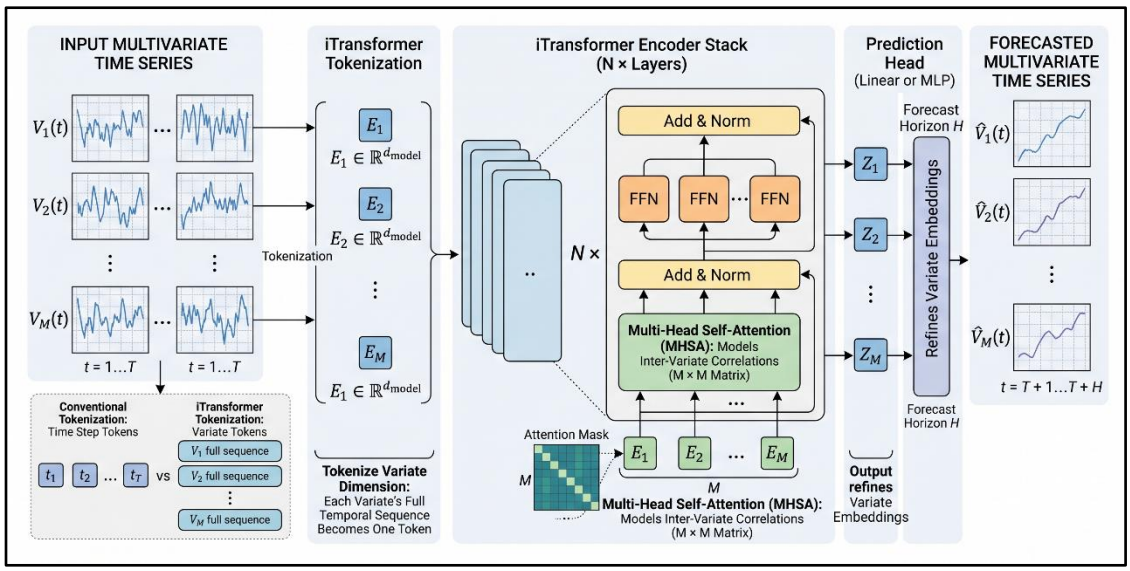

Figure 2.4. Architecture of iTransformer for time series forecasting

All four models share the following training configuration: Adam optimizer with learning rate 10^{-3}, MSE loss function, batch size of 128, maximum 100 epochs with EarlyStopping (patience = 10) monitoring validation loss, ReduceLROnPlateau (factor=0.5, patience=5) and dropout rate 0.2. The architecture-specific hyperparameters that differ across models are summarized in Table 2.2.

Table 2.2. Architecture-specific hyperparameters

Hyperparameter	LSTM	GRU	PatchTST	iTransformer
Layer 1 units	64	64	–	–
Layer 2 units	32	32	–	–
d_model	–	–	64	64
Attention heads	–	–	4	4
Encoder layers	–	–	2	2
Patch size	–	–	24 hours	–

Number of patch	–	–	14	–
Variate tokens	–	–	–	11
FFN dimension	–	–	256(4×)	256(4×)
Dense layer	16 (ReLU)	16 (ReLU)	32 (ReLU)	32 (ReLU)
Recurrent gate	4	2	–	–
Recurrent activation	tanh, sigmoid	tanh, sigmoid	–	–

2.4. Ensemble Methods

Ensemble stacking, or stacked generalization, is a two-level learning framework in which predictions of multiple base learners are treated as input features to a meta-learner that learns the optimal combination [22]. The stacking procedure comprises two stages: in the first stage, all four deep learning models generate predictions on the validation set, forming the meta-feature matrix $\boldsymbol{Z} \in \mathbb{R}^{N_{\text{val}} \times M}$ where M is the number of base learners in a given combination. In the second stage, ridge regression is fitted on $(\boldsymbol{Z}, y_{\text{val}})$ to learn the optimal linear combination:

$$\hat{y}_{\text{stack}} = \boldsymbol{Z}\beta + \beta_0, \quad \hat{\beta} = \arg\min_{\beta} \|y_{\text{val}} - \boldsymbol{Z}\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{2.6}$$

where $\beta \in \mathbb{R}^M$ are the meta-learner coefficients and λ is the Ridge regularization parameter [9]. Ridge regression is selected as the meta-learner for its closed-form solution, interpretable coefficients, and resistance to overfitting in low-dimensional settings, properties appropriate given the small meta-feature dimensionality ($M \leq 4$). At inference time, test set base learner predictions are passed through the fitted meta-learner to produce the final forecast. All three strategies: stacking, simple averaging, and inverse-RMSE weighted averaging are evaluated exhaustively across all possible base learner combinations ($\binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 11$ combinations), yielding 33 ensemble experiments in total. It is acknowledged that the validation set serves a dual role in this framework: as the early stopping monitor during base learner training and as the meta-feature source for ridge regression fitting. This design is a deliberate trade-off given the chronological data constraint, and its potential to introduce mild optimism in meta-learner training is noted as a limitation.

2.5. Evaluation Metrics

Model performance is assessed using four metrics computed in the original physical unit (meters) following inverse Min-Max transformation of scaled outputs. RMSE penalizes large errors quadratically, making it sensitive to peak wave events critical for maritime risk assessment. MAE provides an interpretable mean absolute deviation without disproportionate penalization of extremes. MAPE expresses error as a percentage of observed values, enabling scale-independent comparison across models, with $\epsilon = 10^{-8}$ added to the denominator to prevent numerical instability from near-zero observed values; this value is sufficiently small relative to the minimum observed SWH (0.049 m) to have negligible effect on computed MAPE values. R^2 quantifies the proportion of observed SWH variance explained by the model. The four metrics are formally defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \quad \text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|,$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{|y_i| + \epsilon} \times 100\%, \quad R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}$$

where y_i is the observed SWH, $\hat{y}_i$ is the model prediction, $\bar{y}$ is the observed mean, and N is the number of test samples [11]. Forecast accuracy is further classified following [12]: MAPE < 10% as **highly accurate**, 10-20% as **good**, 20-50% as **reasonable**, and > 50% as **inaccurate**.

2.6. Research Framework

The framework comprises five sequential stages. First, hourly ERA5 reanalysis data are extracted and spatially interpolated to the target point using IDW, followed by feature engineering and cyclical temporal encoding to produce an 11-variable input dataset. Second, the dataset is partitioned chronologically into training, validation, and test sets, with Min-Max normalization fitted exclusively on the training partition to prevent data leakage. Third, sliding window sequences of length $T = 336$ hours are constructed and used to train four deep learning architectures: LSTM, GRU, PatchTST, and iTransformer, independently under identical training configurations, alongside a persistence model as the naive baseline. Fourth, ensemble strategies (simple averaging, weighted averaging, and ridge regression stacking) are constructed from all possible combinations of base learners, with the meta-learner fitted on validation set predictions. Fifth, all models are evaluated on the held-out test set using RMSE, MAE, MAPE, and R^2; statistical significance of performance differences is assessed via the Diebold-Mariano test [4], error complementarity among base learners is quantified through Pearson correlation of prediction residuals, and model robustness under elevated sea states is characterized through a sensitivity analysis restricted to the top 5th percentile of observed SWH.

3. RESULTS AND DISCUSSION

3.1. Wave Climate and Seasonal Pattern Analysis

The IDW-interpolated ERA5 dataset comprises 43,824 hourly SWH observations spanning January 2021 to December 2025. Following sliding window sequence construction with lookback length $T = 336$ hours, the first 336 observations are consumed as the initial input window, yielding a total of 43,488 usable samples that are partitioned chronologically into training (30,441 samples, January 2021-June 2024), validation (6,523 samples, July 2024-March 2025), and test (6,524 samples, April 2025-December 2025) sets, as illustrated in Figure 3.1. The temporal distribution of SWH across all three partitions exhibits consistent seasonal fluctuation, confirming that each partition captures representative wave climate conditions without systematic bias toward any particular monsoon phase. Descriptive statistics of the full dataset are summarized in Table 3.1.

Table 3.1. Descriptive statistics of the SWH

Statistic	SWH
Total observations	43824
Mean	0.275 m
Std. Deviation	0.135 m
Minimum	0.049 m
25th percentile	0.173 m
Median	0.237 m
75th percentile	0.345 m
Maximum	1.165 m

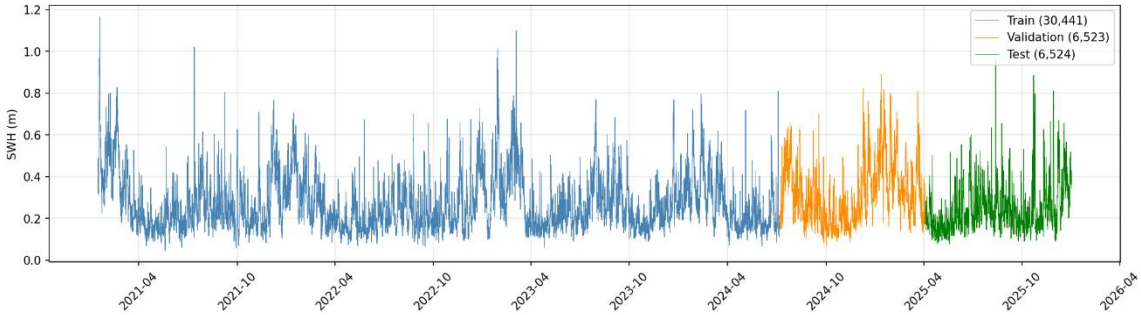

Figure 3.1. Chronological partitioning of the SWH dataset into training, validation, and test sets

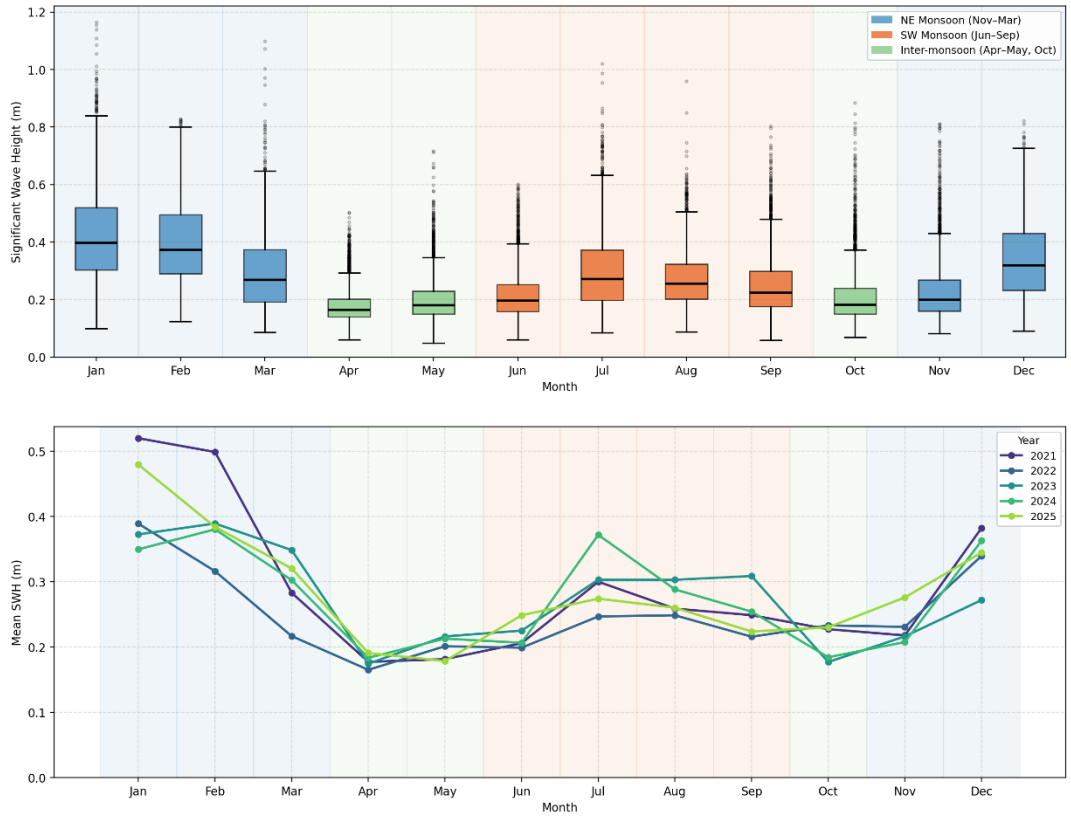

Figure 3.2. Seasonal SWH patterns in the Singapore Strait (2021–2025)

The seasonal wave climate of the Singapore Strait is characterized by a pronounced bimodal pattern driven by two alternating monsoon systems, as revealed in Figure 3.2. Panel (a) shows that the Northeast Monsoon months of January and February exhibit the highest median SWH (~0.40 m) and widest interquartile range, with outliers exceeding 1.0 m, episodic monsoon surge events consistent with the wave conditions that triggered the sinking of Silver Sincere in January 2025. A sharp transition to calmer conditions occurs from April onward, with inter-monsoon months (April-May and October) recording the lowest median SWH (~0.18 m) and minimal spread. The Southwest Monsoon period (June–September) shows moderately elevated values (~0.20-0.28 m), with July exhibiting a secondary energy peak attributable to the mature phase of southwesterly forcing.

December marks the onset of the returning Northeast Monsoon, with median SWH rising sharply toward values comparable to January.

Panel (b) confirms that this seasonal cycle is highly consistent across all five years of the study period (2021–2025). All annual curves follow a near-identical trajectory, declining from January to April, remaining relatively flat through the inter-monsoon and Southwest Monsoon periods, and rising again in November–December. This inter-annual consistency, combined with the overall low SWH variability (σ = 0.135 m), defines the dominant oceanographic characteristic of the Singapore Strait: a fetch-limited, seasonally regular wave regime whose temporal structure is sufficiently stationary to be reliably learned by data-driven forecasting models.

3.2. Individual Model Performance

Table 3.2 summarizes the evaluation metrics for all individual models, and Figure 3.3 illustrates the predicted versus observed SWH curves over the first 180 days of the test period. As a naive baseline, the persistence model wherein $\hat{y}_{t+1} = y_t$, achieves RMSE = 0.0188 m, MAPE = 4.95%, and R^2 = 0.9720, reflecting the high temporal autocorrelation inherent in hourly SWH data. All deep learning architectures except the Transformer-based models exceed this baseline, with GRU reducing persistence RMSE by 18.6% (0.0153 m) and the best stacking configuration by 19.7% (0.0151 m), confirming that the RNN-based models and ensemble stacking provide genuine predictive skill beyond simple temporal persistence.

Table 3.2. Individual model performance on test set (bold indicates best performance per metric) and naive baseline on test set

Model	RMSE (m)	MAE (m)	MAPE (%)	R^2	Category
LSTM	0.0166	0.0112	4.7915	0.9782	Highly Accurate
GRU	**0.0153**	**0.0105**	**4.5795**	**0.9814**	Highly Accurate
PatchTST	0.1072	0.0804	35.3428	0.0897	Reasonable
iTransformer	0.1088	0.0882	42.7863	0.0637	Reasonable
Persistence model	0.0188	0.0125	4.9500	0.9720	Highly Accurate

Both RNN-based models achieve highly accurate classification, with GRU emerging as the superior individual model across all four metrics. The GRU prediction curve closely tracks observed SWH throughout the test period, successfully capturing both the high-frequency oscillations during calm inter-monsoon phases and the episodic spike events associated with monsoon surges, including the extreme event of approximately 0.95 m recorded in August 2025. LSTM similarly demonstrates strong tracking fidelity, though with slightly larger deviations at peak values. The performance advantage of GRU over LSTM despite the latter's more complex four-gate architecture is consistent with the theoretical expectation that GRU's simplified two-gate mechanism reduces overfitting risk on datasets with low variability and smooth temporal dynamics, where the additional capacity of LSTM's separate cell state offers diminishing returns [2].

Both Transformer-based architectures demonstrate substantially degraded performance relative to the RNN-based models, classified as reasonable under [12]. As visible in Figure 3.3, prediction curves of both models converge to a near-constant value approximating the dataset mean (~0.25 m), entirely failing to capture observed SWH variability, the visual manifestation of near-zero R^2 scores. Two complementary explanations account for this outcome. First, PatchTST's self-attention operates on patches derived from a fetch-limited sea with inherently low SWH variability, the signal-to-noise ratio within each 24-hour patch is insufficient for attention weights to discriminate meaningful temporal patterns, causing predictions to collapse toward the mean. Second, iTransformer's inter-variate attention operates on only 11 variate tokens, far below the hundreds for which the architecture was originally designed [13], resulting in an attention map too sparse to extract meaningful cross-variable relationships. These findings constitute an important empirical contribution to the discourse on Transformer applicability in geophysical forecasting: attention-based architectures optimized for high-dimensional, high-variability settings do not generalize to narrow, fetch-limited oceanographic regimes without architectural adaptation.

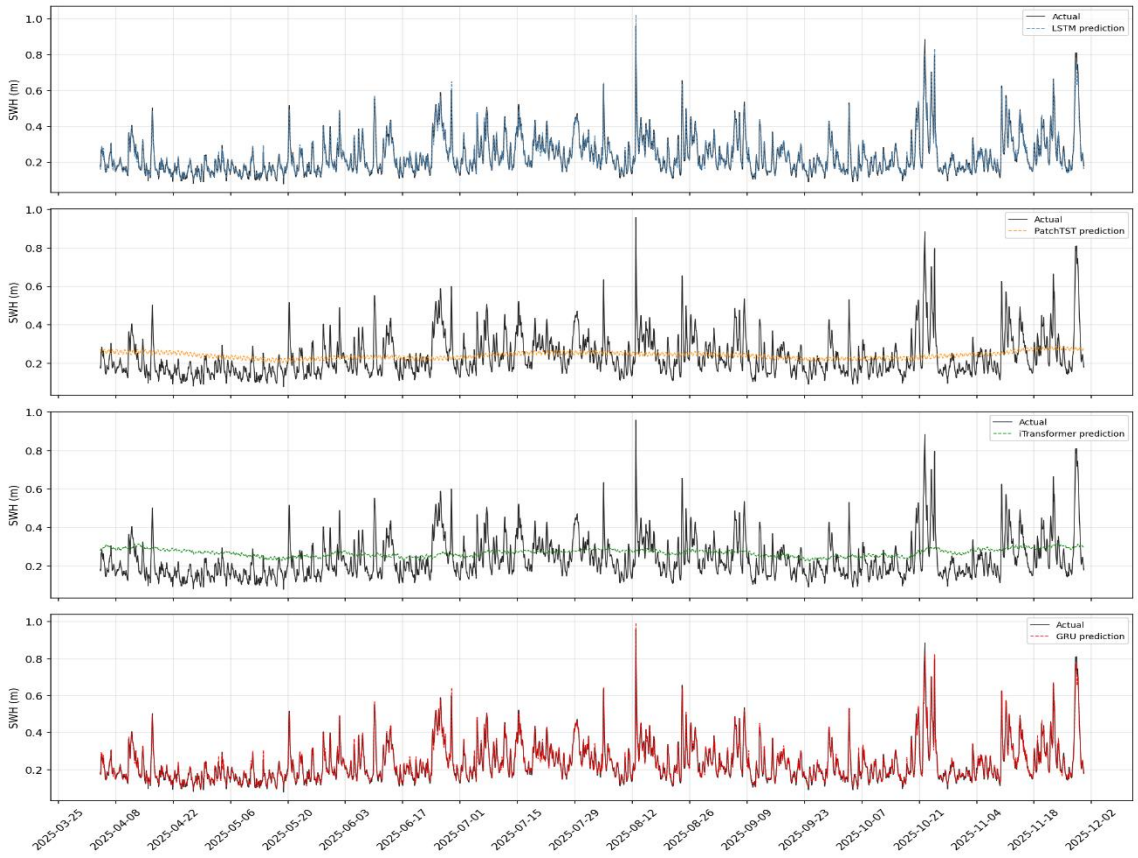

Figure 3.3. Observed versus predicted SWH curves over the first 180 days of the test period of four individual model

3.3. Ensemble Methods Performance

It is noted that reporting the best-performing configuration from 33 experiments introduces a degree of selection optimism inherent to exhaustive search; the complete ranking of all 37 configurations: 4 individuals model and 33 ensembles combinations is provided in Appendix A to assess the full performance distribution independently. Figure 3.4 illustrates the prediction curve of

the best overall configuration, stacking GRU+iTransformer, against observed SWH over the full test period.

Compared to the individual model benchmark where GRU achieved the best individual performance with RMSE = 0.0153 m, MAPE = 4.58%, and R^2 = 0.9814, the exhaustive ensemble evaluation confirms that ridge regression stacking consistently surpasses all individual models. The top three configurations: stacking GRU+iTransformer (RMSE = 0.0151 m, MAPE = 4.55%); stacking GRU+PatchTST (RMSE = 0.0151 m, MAPE = 4.55%); and stacking GRU+PatchTST+ iTransformer (RMSE = 0.0151 m, MAPE = 4.54%), each reduce GRU's individual RMSE by approximately 1.3%, confirming a consistent improvement. Stacking LSTM+GRU (RMSE = 0.0153 m) matches GRU's individual performance exactly, while improving upon LSTM individually by 7.8% in RMSE, demonstrating that ridge regression effectively compensates for a weaker base learner by down-weighting its contribution adaptively.

Most counter-intuitively, the best-performing configurations are not those combining the two strongest individual models. Despite PatchTST (RMSE = 0.1072 m, MAPE = 35.34%) and iTransformer (RMSE = 0.1088 m, MAPE = 42.79%) performing far below GRU individually, their inclusion as base learners in a stacking framework produces the overall best results. This reveals that the Transformer models carry residual information complementary to GRU's error structure. A Diebold-Mariano (DM) test confirms that the improvement of Stacking GRU+iTransformer over GRU individual is statistically significant ($\text{DM} = 5.019, p < 0.001$), demonstrating that the 0.0002 m reduction in RMSE, while numerically small, represents a genuine and reproducible forecasting gain rather than sampling variation. The ridge regression meta-learner exploits the Transformer's mean-reverting bias as a corrective signal in regions where GRU systematically over- or under-predicts. This finding underscores a critical principle of ensemble learning: a base learner's value in stacking cannot be inferred from its individual performance alone; complementarity of error patterns is the decisive criterion [6].

The performance stratification across ensemble strategies is pronounced when contrasted against individual benchmarks. Weighted averaging with RNN-only base learners (LSTM+GRU: RMSE = 0.0157 m) outperforms LSTM individually but remains marginally inferior to GRU. Any inclusion of Transformer models in weighted averaging degrades performance severely, weighted average GRU+iTransformer reaches MAPE = 7.46% and weighted average GRU+PatchTST+iTransformer reaches MAPE = 10.09%, crossing from highly accurate to good and falling below both individual RNN models. Simple averaging is the most sensitive to base learner quality: Simple Avg LSTM+GRU (RMSE = 0.0158 m) performs comparably to GRU individually, yet any Transformer inclusion collapses MAPE to between 12.8% and 43.7% far below even the weakest individual model LSTM at MAPE = 4.79%. The worst overall configuration, Stacking PatchTST+iTransformer (RMSE = 0.1125 m, R^2 = -0.0021), confirms that ridge regression requires at least one competent base learner to produce meaningful predictions.

3.4. Discussion

The results of this study yield three interconnected findings that collectively advance the empirical understanding of deep learning-based SWH forecasting in narrow, fetch-limited straits and its implications for maritime risk assessment.

GRU's advantage over the architecturally more complex LSTM is consistent with the principle that simpler recurrent architectures generalize better on smooth, low-amplitude signals [2,6], as supported by GRU's near-zero bias (-0.0001 m), indicating no systematic directional error. The near-complete failure of PatchTST and iTransformer (MAPE > 35%, $R^2 \approx 0$) reflects a mismatch between architectural assumptions and dataset properties.

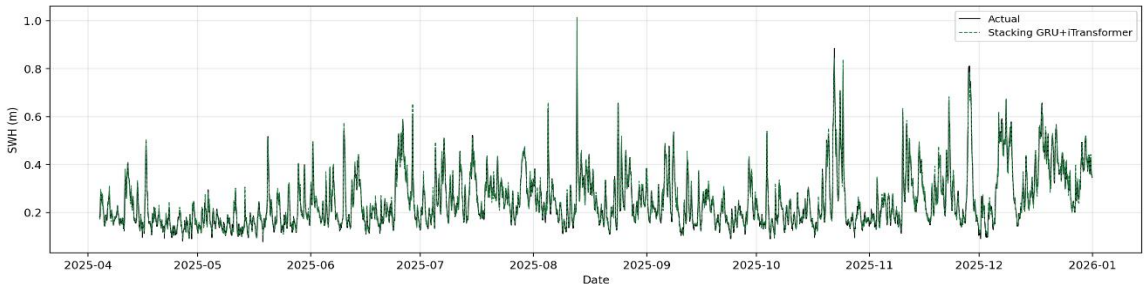

Figure 3.4. Observed versus predicted SWH by the best overall model: stacking GRU+iTransformer over the full test period (April–December 2025)

PatchTST's 24-hour patches carry insufficient discriminative information in a regime dominated by sub-daily local wind forcing, causing predictions to collapse toward the mean [15]. iTransformer's inter-variate attention, designed for high-dimensional settings, is similarly ineffective with only 11 variate tokens, yielding an overly sparse 11×11 attention matrix [13]. Both failure modes confirm that self-attention requires sufficient signal diversity, temporal or cross-variate, to produce non-trivial attention distributions [22], suggesting architectural adaptation (e.g., multi-scale patching) is needed before Transformers can exploit low-variability oceanographic signals.

The superior performance of stacking configurations combining GRU with individually weak Transformer models is consistent with ensemble theory: improvement is driven by error decorrelation across base learners rather than individual accuracy [6]. The Pearson correlation between GRU prediction errors and iTransformer prediction errors on the test set is $r = 0.278$, and between GRU and PatchTST errors is $r = 0.275$, confirming that despite their poor individual performance, Transformer predictions are not redundant with GRU predictions. This low error correlation provides the empirical basis for the complementary signal exploited by the Ridge Regression meta-learner, consistent with the ensemble diversity principle [22]. The failure of averaging-based strategies to replicate this gain confirms that exploiting such complementarity requires an adaptive meta-learner rather than a fixed combination rule.

Several limitations of this study warrant acknowledgment. First, the proposed framework operates in a reanalysis-driven nowcasting mode: ERA5 data, available with an approximate latency of five days from ECMWF, serve as input rather than real-time observational or numerical water prediction (NWP) forecast data. Consequently, the framework is more precisely characterized as a data-driven nowcasting pipeline than an operational real-time forecasting system. Deployment in fully operational contexts would require integration with real-time ERA5 analysis or NWP wind-wave forcing products, which represents a direction for future development. The maximum SWH recorded in the ERA5-derived dataset (1.165 m) likely underestimates actual extreme wave heights due to the spatial averaging inherent in ERA5's 0.25° resolution, which smooths localized wave energy peaks in narrow straits, the sinking of Silver Sincere in January 2025 amid three-meter waves, substantially exceeding the dataset maximum, illustrates this limitation directly. Higher-resolution wave models such as SWAN or WAVEWATCH III, or in-situ buoy measurements, would provide more accurate characterization of extreme sea states in this corridor and are recommended for future work. Second, the one-step-ahead forecasting framework implemented here reflects an operational nowcasting scenario where near-real-time ERA5 reanalysis is available, a configuration that provides an upper-bound estimate of achievable forecasting accuracy but does not directly address multi-step forecasting horizons of 24-168 hours that are more practically relevant for voyage planning and insurance risk quantification. Additionally, IDW interpolation assumes spatial isotropy and does not account for the bathymetric complexity, fetch geometry, and island-induced wave shadowing that characterize the Singapore Strait, potentially limiting the fidelity of the interpolated wave climate relative to high-resolution coastal wave models. The proposed pipeline assumes stationarity of the seasonal wave climate, an assumption that may be violated under projected changes in monsoon

intensity and frequency under climate change scenarios. An analysis of model performance restricted to the top 5% of observed SWH on the test set (threshold = 0.4765 m, n = 327 samples) reveals a markedly different performance hierarchy from that observed across the full test set. Under extreme sea states, LSTM achieves the lowest RMSE (0.0332 m, MAPE = 3.95%) among all evaluated models, substantially outperforming GRU (RMSE = 0.0871 m, MAPE = 10.78%) and the best stacking configuration (RMSE = 0.0824 m, MAPE = 9.97%). The persistence baseline (RMSE = 0.0448 m) also outperforms GRU and stacking under extreme conditions, suggesting that GRU's superior overall performance is driven primarily by its accuracy on the predominant low-to-moderate SWH regime (σ = 0.135 m) rather than its ability to capture episodic wave energy peaks. This finding has important practical implications: for maritime applications where extreme sea state prediction is the primary objective, such as voyage risk assessment and insurance underwriting for high-severity events. LSTM may represent a more appropriate architecture than GRU, despite its inferior aggregate performance metrics. The degradation of GRU and stacking performance under extremes is consistent with the ERA5 spatial averaging bias discussed above, which compresses the dynamic range of training targets and may systematically bias recurrent models toward accurate prediction of moderate conditions at the expense of extreme event fidelity.

4. CONCLUSION

This study developed and evaluated a deep learning ensemble stacking framework for one-step-ahead SWH forecasting in the Singapore Strait, using hourly IDW-interpolated ERA5 data (2021-2025) across four architectures (LSTM, GRU, PatchTST, iTransformer) and an exhaustive experiment spanning 4 individual models and 33 ensemble configurations (37 total).

Three principal conclusions emerge from this study. First, GRU outperforms LSTM as the best individual model (RMSE = 0.0153 m, MAPE = 4.58%, R^2 = 0.9814), demonstrating that architectural simplicity is advantageous in low-variability, fetch-limited oceanographic regimes where the additional complexity of LSTM's four-gate mechanism offers diminishing returns. Second, PatchTST and iTransformer fail to capture SWH variability in this setting, both converging to near-constant mean predictions with MAPE exceeding 35%, confirming that attention-based architectures designed for high-dimensional, high-variability datasets do not generalize to narrow semi-enclosed straits without architectural adaptation. Third, and most significantly, ridge regression stacking of GRU with Transformer-based models achieves the overall best performance (RMSE = 0.0151 m, MAPE = 4.54%, R^2 = 0.9820), surpassing all individual models including GRU, a counter-intuitive result that demonstrates the capacity of stacking to exploit complementary error structures between architecturally heterogeneous base learners, even when Transformer models are individually poor predictors. All RNN-based and ensemble configurations achieve highly accurate classification, confirming the practical viability of the proposed framework for operational sea state forecasting in this corridor.

From an applied perspective, the forecasting pipeline established in this study from ERA5 acquisition and IDW interpolation through deep learning ensemble stacking, provides a replicable and computationally tractable methodology for wave climate characterization in narrow straits that are under resolved by global reanalysis products. The achieved accuracy of MAPE = 4.54% and R^2 = 0.9820 on a held-out test set spanning April to December 2025 demonstrates sufficient predictive precision to serve as a probabilistic input for route-specific marine cargo insurance premium calculation along the Singapore Strait corridor, establishing a replicable oceanographic input pipeline whose predicted SWH can serve as an environmental risk covariate for future actuarial modelling of marine cargo insurance along this strategically critical corridor.

Future research directions include: (1) extending the forecasting framework to multi-step horizons of 24-168 hours, which are more directly relevant to voyage planning and underwriting decisions; (2) incorporating higher-resolution wave model outputs from SWAN or WAVEWATCH

III, or in-situ buoyont measurements, to address the ERA5 spatial averaging bias that likely underestimates extreme SWH events; (3) evaluating the generalizability of the stacking framework to other narrow straits and coastal waterways in the Indonesian archipelago, where similar fetch-limited, low-variability wave regimes are prevalent; (4) investigating architectural adaptations, such as multi-scale patching or domain-specific positional encoding that may improve Transformer performance in low-variability oceanographic settings, potentially yielding stronger base learners for future ensemble configurations; and (5) adopting rolling-origin time-series cross-validation to provide a more robust and less optimistic estimate of generalization performance across multiple forecast origins.

ACKNOWLEDGEMENT

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] C3S, 2018. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS),

[2] Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar

[3] D'Costa, B. & Mandal, S., 2026. Mitigation of the time-lag issue in LSTM-based significant wave height prediction using long-term (1984–2023) ERA5 dataset. *Ocean Engineering*. Vol. 346, 123921.

[4] Diebold, F.X., 2015. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. *Journal of Business & Economic Statistics*. Vol. 33, No. 1, 1–1.

[5] Eliopoulou, E., Alissafaki, A. & Papanikolaou, A., 2023. Statistical Analysis of Accidents and Review of Safety Level of Passenger Ships. *Journal of Marine Science and Engineering*. Vol. 11, No. 2, 410.

[6] Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning*. Springer New York, New York, NY.

[7] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. Vol. 146, No. 730, 1999–2049.

[8] Hochreiter, S. & Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation*. Vol. 9, No. 8, 1735–1780.

[9] Hoerl, A.E. & Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. Vol. 12, No. 1, 55–67.

[10] Hou, B., Fu, H., Li, X., Song, T. & Zhang, Z., 2024. Predicting significant wave height in the South China Sea using the SAC-ConvLSTM model. *Frontiers in Marine Science*. Vol. 11, 1424714.
[11] Hyndman, R.J. & Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*. Vol. 22, No. 4, 679–688.
[12] Lewis, C.D. 1982. *Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting*. Butterworth Scientific, London.
[13] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L. & Long, M., 2024. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. arXiv,
[14] Minuzzi, F.C. & Farina, L., 2023. A deep learning approach to predict significant wave height using long short-term memory. *Ocean Modelling*. Vol. 181, 102151.
[15] Nie, Y., Nguyen, N.H., Sinthong, P. & Kalagnanam, J., 2022. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. arXiv,
[16] Operational Areas. 2026. https://www.mpa.gov.sg/port-marine-ops/operations/vessel-traffic-information-system/operational-areas. [9 Mei 2026]
[17] Safety and Shipping Review 2025, 2025. Allianz Commercial,
[18] Shen, W., Ying, Z., Zhao, Y. & Wang, X., 2024. Significant wave height prediction in monsoon regions based on the VMD-CNN-BiLSTM model. *Frontiers in Marine Science*. Vol. 11, 1503552.
[19] Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM national conference*
[20] Singapore Manages Response to Two Sinkings in One Weekend. 2026. https://maritime-executive.com/article/singapore-manages-response-to-two-sinkings-in-one-weekend. [9 Mei 2026]
[21] Sinking of Tanker Silver Sincere in Singapore Territorial Waters On 12 January 2025, 2025. Transport Safety Investigation Bureau of Singapore, Singapore.
[22] Wolpert, D.H., 1992. Stacked generalization. *Neural Networks*. Vol. 5, No. 2, 241–259.
[23] Zhai, Y., Shi, H., Zhan, C., Wang, Q., You, Z. & Wang, N., 2025. Improving Significant Wave Height Prediction Using Chronos Models. *Ocean Engineering*. Vol. 341, 122502.

APPENDIX

Appendix A. Complete Ensemble Experiment Results

Table A.1. Unified model ranking: individual and ensemble

Type-Method	Models	RMSE (m)	MAE (m)	MAPE (%)	R^2
Ensemble Stacking	GRU+iTransformer	0.0151	0.0103	4.55	0.9820
Ensemble Stacking	GRU+PatchTST	0.0151	0.0103	4.55	0.9820
Ensemble Stacking	GRU+PatchTST+ iTransformer	0.0151	0.0103	4.54	0.9820
Individual	GRU	0.0153	0.0105	4.58	0.9814
Ensemble Stacking	LSTM+GRU	0.0153	0.0104	4.54	0.9816
Ensemble Stacking	LSTM+GRU+ iTransformer	0.0153	0.0104	4.53	0.9816
Ensemble Stacking	LSTM+GRU+PatchTST	0.0153	0.0104	4.53	0.9816
Ensemble Stacking	LSTM+GRU+PatchTST +iTransformer	0.0153	0.0104	4.53	0.9816
Ensemble Weighted Avg	LSTM+GRU	0.0157	0.0106	4.60	0.9804

Type-Method	Models	RMSE (m)	MAE (m)	MAPE (%)	R^2
Ensemble Stacking	LSTM+PatchTST+ iTransformer	0.0158	0.0108	4.74	0.9802
Ensemble Stacking	LSTM+iTransformer	0.0158	0.0109	4.75	0.9802
Ensemble Stacking	LSTM+PatchTST	0.0158	0.0108	4.74	0.9802
Ensemble Simple Avg	LSTM+GRU	0.0158	0.0106	4.60	0.9803
Individual	LSTM	0.0166	0.0112	4.79	0.9782
Ensemble Weighted Avg	LSTM+GRU+ iTransformer	0.0182	0.0129	5.80	0.9737
Ensemble Weighted Avg	LSTM+GRU+PatchTST	0.0185	0.0128	5.54	0.9730
Ensemble Weighted Avg	GRU+iTransformer	0.0214	0.0159	7.46	0.9638
Ensemble Weighted Avg	GRU+PatchTST	0.0216	0.0155	6.82	0.9630
Ensemble Weighted Avg	LSTM+GRU+PatchTST +iTransformer	0.0222	0.0160	7.17	0.9609
Ensemble Weighted Avg	LSTM+iTransformer	0.0232	0.0168	7.65	0.9576
Ensemble Weighted Avg	LSTM+PatchTST	0.0236	0.0164	7.02	0.9558
Ensemble Weighted Avg	GRU+PatchTST+ iTransformer	0.0294	0.0219	10.09	0.9318
Ensemble Weighted Avg	LSTM+PatchTST+ iTransformer	0.0316	0.0232	10.42	0.9209
Ensemble Simple Avg	LSTM+GRU+ iTransformer	0.0405	0.0316	15.10	0.8701
Ensemble Simple Avg	LSTM+GRU+PatchTST	0.0405	0.0294	12.83	0.8705
Ensemble Simple Avg	GRU+PatchTST	0.0564	0.0418	18.36	0.7485
Ensemble Simple Avg	LSTM+GRU+PatchTST +iTransformer	0.0564	0.0432	19.95	0.7483
Ensemble Simple Avg	GRU+iTransformer	0.0569	0.0454	21.96	0.7437
Ensemble Simple Avg	LSTM+PatchTST	0.0571	0.0419	18.25	0.7422
Ensemble Simple Avg	LSTM+iTransformer	0.0574	0.0454	21.83	0.7394
Ensemble Simple Avg	GRU+PatchTST+ iTransformer	0.0729	0.0565	26.22	0.5796
Ensemble Simple Avg	LSTM+PatchTST+ iTransformer	0.0733	0.0566	26.14	0.5752
Ensemble Weighted Avg	PatchTST+iTransformer	0.1069	0.0836	38.76	0.0957
Ensemble Simple Avg	PatchTST+iTransformer	0.1069	0.0836	38.79	0.0956
Individual	PatchTST	0.1072	0.0804	35.34	0.0897
Individual	iTransformer	0.1088	0.0882	42.79	0.0637
Ensemble Stacking	PatchTST+iTransformer	0.1125	0.0896	43.69	-0.0021