Ocean Engineering 343 (2026) 123191

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

Ship fuel consumption prediction based on ResGCN and iLSTM with multi-scale dynamic attention mechanism

Weibo Zhong [a], Kang Bai [a], Yang Gu [b,a], Nan Ye [c,a,*]

[a] Ocean college, Jiangsu University of Science and Technology, Zhenjiang, 212003, China
[b] Ocean college, Zhejiang University, Zhoushan, 316022, China
[c] CSSC PRIDe (Nanjing) Atmospheric and Oceanic Information System Co, Ltd. Nanjing, 211100, China

ARTICLE INFO

Keywords:
Ship fuel consumption prediction
Residual sparse graph convolutional network
Improved long short-term memory network
Dynamic attention mechanism

ABSTRACT

Accurate prediction of ship fuel consumption is essential for maritime energy optimization and emission reduction. Due to the complex interactions among navigation-related factors such as ship speed, wind, draft, waves, and currents, ship fuel consumption exhibits strong nonlinear, non-stationary, and multi-scale characteristics, posing challenges to conventional prediction methods. To address this, a hybrid prediction framework is proposed, integrating multi-scale wavelet decomposition (MWD), a residual sparse graph convolutional network (ResGCN), and an improved long short-term memory network (iLSTM). Key features are selected using random forest, and a residual sparse graph based on mutual information is constructed to capture nonlinear inter-feature dependencies while enhancing information propagation and mitigating over-smoothing. The iLSTM combines the strengths of structural and memory-augmented LSTM variants, improving spatiotemporal modeling and computational efficiency. MWD decomposes the input into frequency sub-bands to separate trends from fluctuations, with a dynamic attention mechanism enhances cross-scale feature fusion. Experimental results show the proposed model achieves a peak R^2 of 0.9667 and a MAPE of 4.051 %, demonstrating exceptional accuracy, robustness in ship fuel consumption forecasting.

1. Introduction

Amid tightening IMO decarbonization mandates and global carbon neutrality pledges, ship fuel optimization has become a make-or-break factor for achieving Poseidon Principles compliance. Accurate prediction of fuel consumption offers vital decision support for reducing operating costs (Zhou et al., 2023), advancing energy conservation and emission reduction (Xiao et al., 2025), achieving carbon neutrality goals, and optimizing routes (Hu et al., 2022).

In particular, ship fuel consumption is influenced by factors such as speed, wind, waves, currents, and load, which exhibit non-stationarity, nonlinearity, and multi-scale spatiotemporal complexity. Current ship fuel consumption prediction methods are broadly categorized as: white box models (WBMs), black box models (BBMs), and gray box models (GBMs) (Fan et al., 2022).

WBMs build on ship propulsion physics, which ensures strong physical interpretability (Tillig and Ringsberg, 2019; Orihara and Tsujimoto, 2018). However, WBMs require high-precision parameter measurements and accurate modeling in complex conditions, making dynamic adjustments and reliable predictions challenging (Vinayak et al., 2021). BBMs use historical data exclusively, applying statistical or machine learning methods, which makes them ideal for data-rich practical applications (Yan et al., 2021). GBMs merge physical mechanisms with data-driven approaches, integrating domain knowledge while using data to calibrate and supplement parameters (Chen et al., 2019). This hybrid methodology excels in modeling complex navigation scenarios (Odendaal et al., 2023). Moreover, their data-driven components suffer from data constraints, while parameter calibration adds complexity.

Consequently, recent studies increasingly shift toward deep learning-offering powerful nonlinear modeling in high-dimensional spaces-with initial efforts primarily using feedforward neural networks. Moreira (Moreira et al., 2021) applied a backpropagation (BP) network to model fuel consumption relative to ship speed under varying sea conditions, achieving good accuracy. Yan (Yan et al., 2020) employed radial basis function (RBF) networks, known for their faster convergence and superior generalization compared to traditional BP models. However, both BP and RBF networks suffer from limitations, including sensitivity

* Corresponding author.
E-mail addresses: vebo@just.edu.cn (W. Zhong), 495172992@qq.com (K. Bai), Yangu@zju.edu.cn (Y. Gu), yenan724@163.com (N. Ye).

https://doi.org/10.1016/j.oceaneng.2025.123191
Received 24 June 2025; Received in revised form 16 September 2025; Accepted 13 October 2025
Available online 24 October 2025

to initialization, poor scalability with high-dimensional inputs, and an inability to effectively capture sequential dependencies.

To address these constraints, researchers increasingly adopt recurrent architectures. Zhang (Zhang et al., 2024) introduced an attention-enhanced bidirectional LSTM model, effectively integrating multi-source data to achieve superior performance over standard LSTMs in fuel consumption prediction. Concurrently, Wang (Wang et al., 2023) developed a GA-LSTM model, utilizing genetic algorithms to optimize hyperparameters and enhance prediction accuracy under dynamic operating conditions.

Despite excelling at temporal dependencies, recurrent networks often assume feature independence and lack explicit modeling of structural correlations (Chen et al., 2024). Yet ship parameters, such as speed, wind, draft, and wave dynamics, exhibit strong nonlinear interdependencies, necessitating joint modeling of temporal sequences and pairwise feature interactions for accurate bunker fuel forecasting (Li et al., 2023).

To advance temporal dynamics modeling, studies enhance neural networks' sequential learning capability. Zhao (Zhao and Zhao, 2025) pioneered DeepONet-based coupled models employing branch-trunk decomposition to jointly learn multi-dimensional motion signals for ultra-short-term ship prediction. Mao (Mao et al., 2025) developed the TF-Informer framework, integrating temporal convolutional networks, frequency enhanced channel attention, and informer modules, to capture both transient variations and voyage-scale dependencies. In parallel, recent studies incorporated signal decomposition techniques to address nonlinearity and nonstationarity in maritime time-series forecasting. Zhang et al. (2022) applied wavelet transform with hybrid statistical-neural models, while Hou et al. (2024) demonstrated that ensemble empirical mode decomposition (EEMD) combined with LSTM can effectively exploit multi-scale components for improved predictive accuracy. These innovations demonstrate multiscale attention architectures' efficacy for maritime time-series modeling.

While such improvements enhance temporal learning, the structural dependencies among features remain underexplored. Graph neural networks (GNNs), particularly graph convolutional networks (GCNs), provide a natural solution by representing features as nodes and learning their relationships through message passing. However, most GCN-based approaches rely on static or fully connected graphs, failing to capture the inherent sparsity and evolving nature of feature interactions. This often results in redundant information propagation and oversmoothed node representations (Qureshi et al., 2023). On the other hand, standard LSTM networks face limited representational capacity and lack parallelism, restricting their efficiency in capturing long-range dependencies in complex temporal data.

To overcome the limitations of existing models in capturing the nonlinear, dynamic, and structurally interdependent characteristics of ship fuel consumption, this paper proposes a unified multi-branch prediction framework. The model integrates multi-scale wavelet decomposition, a sparse residual graph convolutional network, and an improved long short-term memory network, with a dynamic attention mechanism enhancing cross-scale feature fusion. Experimental results show that the proposed model achieves a peak R^2 of 0.9667 and a MAPE of 4.051 %, demonstrating superior accuracy and robustness under complex maritime conditions.

2. Methods

We propose a multi-branch ship fuel consumption prediction model integrating multi-scale wavelet decomposition, a sparse residual graph convolutional network, an improved long short-term memory network, and an attention mechanism. The model captures the complex interactions among multiple feature variables and their temporal dependencies. The overall process of the prediction method is shown in Fig. 1.

2.1. Wavelet decomposition

To enhance multi-scale temporal modeling, this paper employs n-layer wavelet decomposition on the original input sequence $x(t)$. This decomposition hierarchically separates the sequence into sub-sequences at distinct frequency levels, capturing both the overall trend and local disturbances. The resultant representation comprises a low-frequency approximation $A_n(t)$ and multiple high-frequency detail components $D_1(t), D_2(t), \ldots, D_n(t)$:

$$x(t) = A_n(t) + \sum_{j=1}^{n} D_j(t) \tag{1}$$

where the low-frequency approximate component $A_n(t)$ describes the slowly varying trend of the time series, while the high-frequency detail component $D_j(t)$ captures rapid fluctuations and abrupt changes at different scales. This decomposition effectively breaks down the original signal into a set of temporally structured, multi-scale features, yielding a more detailed and structured representation for subsequent modeling.

2.2. Residual sparse graph convolutional network

2.2.1. Graph convolutional network

The GCN is a deep learning method specifically designed for graph-structured data, effectively modeling complex node relationships in non-Euclidean space (Bhatti et al., 2023). GCN performs feature aggregation and node representation updates through a propagation mechanism based on the graph adjacency matrix, capturing structural semantics between nodes.

For a given undirected graph $G = (V, E)$, where V represents the node set and E represents the edge set, each node $v_i \in V$ in the graph has an input feature vector $\mathbf{x}_i$. GCN normalizes the adjacency matrix:

$$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} \tag{2}$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops, and $\tilde{D}$ is the degree matrix of $\tilde{A}$. Perform hierarchical propagation and transformation on the input feature matrix X:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) \tag{3}$$

here, $H^{(l)}$ is the node feature matrix of the l-th layer, with the initial layer $H^{(0)} = X$ being the input feature matrix; $W^{(l)}$ is the learnable weight matrix of the l-th layer; and $\sigma(\cdot)$ represents the activation function (such as ReLU). This formula indicates that the feature update of each node is the weighted average of its own and neighboring node features, thereby achieving the fusion and expression of local information.

Although GCNs can expand a node's receptive field by stacking multiple layers, increasing network depth often faces issues like oversmoothing and gradient decay. These problems lead to representational homogeneity among nodes and performance degradation (Qi et al., 2021). To mitigate this issue, this paper adopts ResGCN architecture. This structure incorporates residual connections after each graph convolutional layer. Specifically, the input features from the previous layer are added to the current convolution output, expressed as:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) + H^{(l)} \tag{4}$$

Each layer incorporates residual connections, facilitating cross-layer feature transfer. This mechanism enhances the expressive power and training efficiency of deep GCNs. By preserving low-level features during training, residual connections strengthen the network's representational capacity and training stability, leading to enhanced performance when stacking multiple GCN layers.

2.2.2. Sparse graph convolutional network

In ship fuel consumption prediction, high-dimensional and redundant features can significantly impair model performance. Certain in-

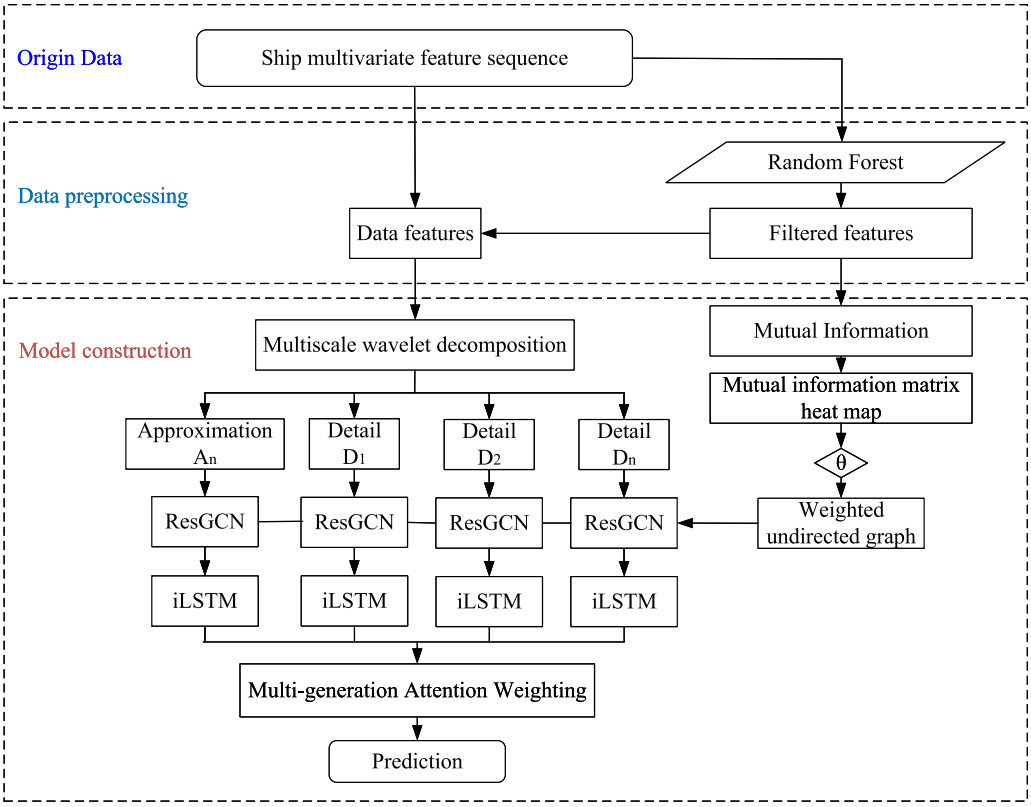

Fig. 1. Overall process of prediction method.

put variables contribute minimally and may even introduce noise, potentially leading to overfitting and computational inefficiency (Zhang et al., 2024). To mitigate these issues, we implement a two-step strategy for constructing an informative yet sparse feature graph: (1) feature screening based on random forest importance, and (2) sparse graph construction leveraging mutual information (see Fig. 2).

Random forest, a widely adopted ensemble learning method, comprises multiple decision trees trained on bootstrapped samples and random feature subsets. It exhibits strong resistance to overfitting and effectively estimates feature relevance (Wen et al., 2022). The importance of a feature f_i is quantified by the aggregate reduction in mean squared error (MSE) it contributes across all splits and trees. Specifically, for a split node utilizing feature f_i, the MSE reduction is computed as follows:

$$\Delta MSE = MSE_{parent} - \left(\frac{N_{left}}{N_{parent}} \cdot MSE_{left} + \frac{N_{right}}{N_{parent}} \cdot MSE_{right}\right) \tag{5}$$

where MSE_{parent} is the MSE of the parent node before splitting. MSE_{left} and MSE_{right} are the MSE of the left and right child nodes after splitting, respectively. N_{parent}, N_{left}, and N_{right} are the number of samples of the parent node and the left and right child nodes, respectively.

By aggregating the MSE reductions for each feature across all trees, we derive a quantitative importance score. This screening step effectively filters out irrelevant or marginally relevant features, thereby enhancing the efficiency and performance of subsequent modeling stages.

Following this feature screening, we construct a sparse graph based on the statistical dependencies among the selected key features. Moving beyond assumptions of predefined or fully-connected relationships, we utilize mutual information to quantify the dependency strength between any pair of features. Given that the feature variables are discrete, the mutual information between two discrete random variables f_i and f_j is

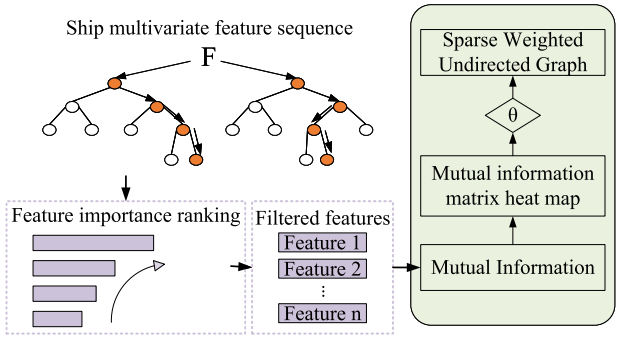

Fig. 2. Feature screening and sparse graph construction.

defined as:

$$\mathrm{MI}(f_i, f_j) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{f_i, f_j}(x, y) \log \left(\frac{p_{f_i, f_j}(x, y)}{p_{f_i}(x), p_{f_j}(y)} \right) \tag{6}$$

where: $\mathcal{X}$ and $\mathcal{Y}$ denote the value spaces of f_i and f_j; $p_{f_i, f_j}(x, y)$ is the empirical joint probability of $f_i = x$ and $f_j = y$; $p_{f_i}(x)$ and $p_{f_j}(y)$ are the marginal probabilities of f_i and f_j, respectively.

Following computation of the mutual information matrix $M \in \mathbb{R}^{n \times n}$, a sparsification procedure is performed using threshold θ. This produces a sparse, data-driven graph topology that selectively retains salient feature dependencies while suppressing noisy or redundant connections. The resultant adjacency matrix A is constructed as follows:

$$A_{ij} = \begin{cases} \mathrm{MI}(f_i, f_j), & \text{if } \mathrm{MI}(f_i, f_j) > \theta \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

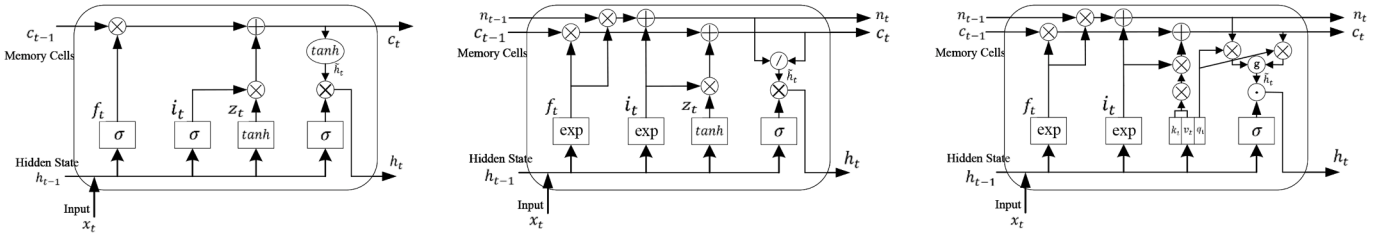

Fig. 3. Structure of LSTM, sLSTM, and mLSTM.

2.3. Improved long short-term memory network

2.3.1. Long short-term memory network

To address the vanishing gradient problem inherent in traditional RNNs for long-sequence modeling, LSTM incorporates three gating mechanisms-forget gate, input gate, and output gate (Hochreiter and Schmidhuber, 1997). The gating operations are formally defined as follows:

$$\{i_t, f_t, o_t\} = \sigma\left(\{\tilde{i}_t, \tilde{f}_t, \tilde{o}_t\}\right) = \sigma\left(W_{(i,f,o)}^{\top} x_t + R_{(i,f,o)} h_{t-1} + b_{(i,f,o)}\right) \tag{8}$$

where i_t, f_t, o_t represent the activation values of the input gate, forget gate and output gate respectively. $\tilde{i}_t, \tilde{f}_t, \tilde{o}_t$ are their corresponding input signals. $W(i,f,o)^{\top}$, $R(i,f,o)$ and $b(i,f,o)$ represent the weight matrix, recurrent weight matrix and bias term respectively. h_{t-1} is the hidden state of the previous moment. σ is the sigmoid function.

The update of memory cell c_t is given by:

$$z_t = \phi(\tilde{z}_t) = \phi(W_z^{\top} x_t + R_z h_{t-1} + b_z) \tag{9}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot z_t \tag{10}$$

where ϕ represents the activation function tanh. c_{t-1} is the memory state of the previous moment. f_t and i_t control the forgetting of old information and the introduction of new information.

The output of the hidden state is modulated by the memory unit:

$$h_t = o_t \tilde{h}_t = o_t \tanh(c_t) \tag{11}$$

While LSTM effectively model long-range sequences, they exhibit inherent limitations including finite storage capacity, difficulty in revising past decisions, and constrained parallelism due to memory mixing effects (Beck et al., 2024). To overcome these constraints, the extended LSTM (xLSTM) architecture was developed. This framework integrates two enhanced variants: scalar-state LSTM (sLSTM) and matrix-state LSTM (mLSTM), which respectively augment storage precision and capacity.

Specifically, sLSTM incorporates an exponential gating mechanism derived from conventional gating structures, replacing the Sigmoid activation function:

$$\{i_t, f_t\} = \exp\left(W_{(i,f)}^{\top} x_t + R_{(i,f)} h_{t-1} + b_{(i,f)}\right) \tag{12}$$

$$n_t = f_t \cdot n_{t-1} + i_t \tag{13}$$

Through state normalization, sLSTM achieves more precise regulation of information retention/forgetting, enhances historical state correction capability, and improves numerical stability. As show in Fig. 3.

mLSTM expands the memory unit from a scalar to a matrix, $c_t \in \mathbb{R}^{d \times d}$. This matrix representation enables storage of multiple state vectors, substantially enhancing model representational capacity and operational flexibility. The gating computations for mLSTM are formalized as follows:

$$\{i_t, f_t\} = \exp\left(\{\tilde{i}_t, \tilde{f}_t\}\right) = \exp\left(W_{(i,f)}^{\top} x_t + b_{(i,f)}\right) \tag{14}$$

$$n_t = f_t \cdot n_{t-1} + i_t \cdot k_t \tag{15}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot v_t \cdot k_t^{\top} \tag{16}$$

where v_t and k_t represent the value vector and key vector. f_t and i_t correspond to the weight decay rate and learning rate, reflecting the effective integration of the fast weight storage mechanism (Schlag et al., 2021).

To stabilize state reading, mLSTM incorporates max-normalization, with its hidden state formally defined as:

$$q_t = W_q x_t + b_p, \quad k_t = \frac{1}{\sqrt{d}} W_k x_t + b_k, \quad v_t = W_v x_t + b_v \tag{17}$$

$$h_t = o_t \odot \tilde{h}_t, \quad \tilde{h}_t = \frac{c_t q_t}{\max\left(|n_t^{\top} q_t|, 1\right)} \tag{18}$$

The output gate subsequently modulates the final cell state to produce the layer output:

$$o_t = \sigma(\tilde{o}_t), \quad \tilde{o}_t = W_o x_t + b_o \tag{19}$$

Relative to LSTM, mLSTM eliminates inter-unit hidden state connections. This architectural modification mitigates sequence dependencies induced by memory mixing effects, thereby enabling efficient parallel computation.

2.3.2. Improved long short-term memory network

To address the architectural imbalance within the xLSTM framework, where sLSTM provides high accuracy but suffers from poor parallelism, while mLSTM supports parallelism but lacks temporal modeling precision, we propose iLSTM. The iLSTM retains mLSTM's parallel-friendly backbone but incorporates targeted architectural modifications to enhance both representational expressiveness and training efficiency.

Unlike sLSTM, which relies on sequential memory updates and parallelism-hindering multi-gated block-diagonal transformations, iLSTM avoids serial bottlenecks by leveraging mLSTM's matrix-based memory mixing.

Compared to mLSTM, the proposed iLSTM introduces several architectural refinements to improve expressiveness while maintaining efficiency. First, we remove the learnable skip connection. This simplifies the information pathway, reduces redundancy, and facilitates more stable optimization. Second, to better encode short-term temporal dependencies, we apply a kernel-size-4 causal convolution prior to projection. This ensures autoregressive consistency while enhancing local context modeling. For projection, we retain block-diagonal transformations for generating queries and keys, but adopt a lightweight up-projection strategy for values. This reduces computational overhead compared to mLSTM's convolution-based value processing. Finally, we replace the original output gate and skip fusion mechanism with a gated multi-layer perceptron (MLP). This MLP, comprising up-projection, GeLU activation, and down-projection, offers superior non-linear modeling capacity at minimal additional cost. Fig. 4 illustrates the complete iLSTM architecture.

2.4. Dynamic weighting attention mechanism

To leverage multi-scale subsequence information from wavelet decomposition, we introduce a dynamic attention mechanism after subband modeling for adaptive multi-frequency feature fusion. Unlike static

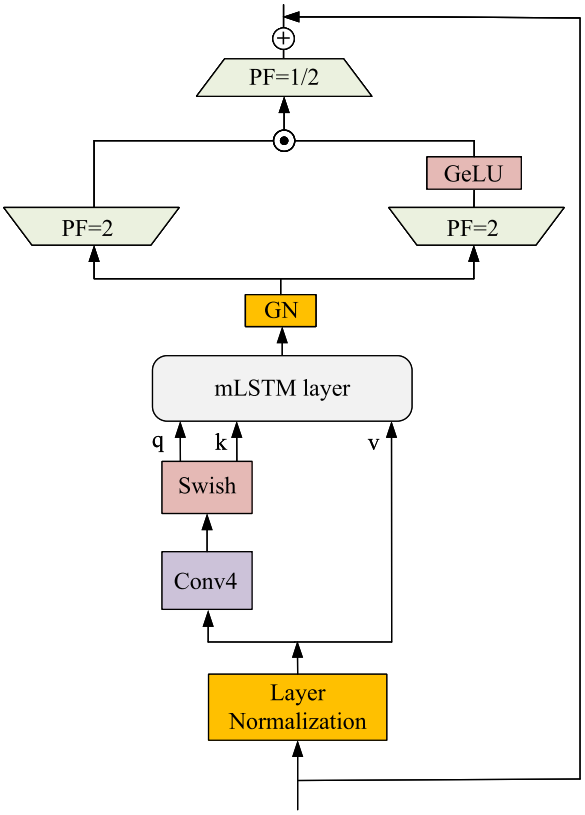

Fig. 4. iLSTM architecture.

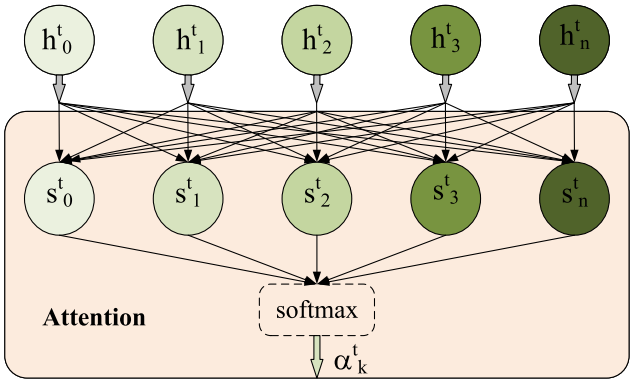

Fig. 5. Dynamic weighting attention mechanism.

weighting approaches, this mechanism's key innovation lies in contextual awareness-dynamically modulating sub-band weights in the final output according to the input state at each prediction time step, as illustrated in Fig. 5.

Formally, given an original time series decomposed into n frequency-specific sub-bands (e.g., high-frequency components $D_1, D_2, \ldots, D_k$ and low-frequency approximation A_n), the attention module computes importance weights $\alpha_k^{(t)}$ at t based on each sub-band model's output representation. These weights determine the contribution of each sub-band to the aggregated prediction.

For the kth subband prediction output $h_k^{(t)}$ at a given time step t, its attention score $s_k^{(t)}$ is calculated by:

$$s_k^{(t)} = \mathbf{w}^\top \mathbf{h}_k^{(t)} + b \tag{20}$$

All scores $\{s_k^{(t)}\}_{k=1}^n$ are normalized by the softmax layer to obtain the attention weight $\alpha_k^{(t)}$:

$$\alpha_k^{(t)} = \frac{e^{s_k^{(t)}}}{\sum_{n=0}^{N} e^{s_n^{(t)}}} \tag{21}$$

Finally, the weighted fusion of multi-subband prediction outputs can be expressed as:

$$y_k^{(t)} = \sum_{k=1}^{n} \alpha_k^{(t)} \cdot \mathbf{h}_k^{(t)}, \quad s.t. \quad \sum_{k=1}^{n} \alpha_k^{(t)} = 1 \tag{22}$$

where $h_k^{(t)}$ represents the predicted output representation of the kth subband at time step t. $\alpha_k^{(t)}$ is the weight coefficient dynamically generated by the attention network.

This mechanism enhances model interpretability: analysis of temporal weight distributions reveals whether the model prioritizes high-frequency disturbances or low-frequency trend information at specific timesteps.

3. Experimental design and model validation

To evaluate the proposed prediction model, this section details the experimental methodology encompassing data preparation, feature selection, model configuration, and evaluation metrics.

3.1. Data preparation

This study utilizes voyage data recorder (VDR) datasets collected at 10-min intervals from a RoPax ferry operating fixed routes in May 2024. The RoPax ferry features a length of 189.5 m, a width of 26.5 m, and a draft of 6.5 m. With a maximum speed of 23 knots, the ship has a total tonnage of 32,729 tonnes.

The dataset consisted of 4328 time-series records with 127 feature columns. Key parameters included ship kinematics (e.g., speed over ground, draft, heading), environmental conditions (e.g., wind, waves), and operating states. The model targeted fuel consumption as the output. Four preprocessing steps were applied to the raw data to ensure reliability:

Voyage segmentation: Retained only cruising-state data based on ship speed thresholds, excluding non-steady phases to ensure the model learns from representative operational conditions. **Missing value removal:** Discarded incomplete records to ensure all variables were consistent. **Outlier filtering:** Removed outliers in key numerical features (e.g., draft, wind speed) to maintain the physical and engineering validity of the data. **Normalization:** Scaled all variables to the $[0, 1]$ using Min-Max normalization to improve model training stability and convergence.

After that, a total of 1840 records were retained for model development, as summarized in Table 1. Adhering to chronological order, the dataset was partitioned into 80 % for training and 20 % for testing. This temporal splitting strategy maintains the natural sequence of operational data and prevents information leakage, thereby improving the validity of the performance evaluation.

To reduce feature redundancy and enhance model interpretability, feature importance was ranked using a Random Forest algorithm. As illustrated in Fig. 6, ship speed was identified as the most influential predictor, followed by water speed and drafts (*MdDf*: bow drafts; *MdDa*: stern drafts). Meteorological factors such as wind direction demonstrated relatively lower importance. Features with negligible contributions-including rudder angle, air temperature (*Ta*), and air pressure (*Pa*)-were subsequently removed from the final model. In this study, we chose the top 9 features as the inputs.

Each training sample comprises a T-step time window, resulting in an input tensor $x \in \mathbb{R}^{T\times D}$, where D denotes the selected input dimension. The output $y \in \mathbb{R}$ corresponds to the specific fuel consumption (MeFoSail, kg/h) at the subsequent timestamp. During training, samples are batched into tensors $x \in \mathbb{R}^{B\times T\times D}$ for mini-batch processing. Algorithm 1 details the data flow with four key stages: multi-scale subband

Table 1
Representative subset of preprocessed data raw values before normalization.

Timestamp	MdDf (m)	MdDa (m)	Wind Dir (°)	Wind Speed (m/s)	Water Speed (m/s)	Ship Speed (kn)	Ship Dir (°)	Wave Dir (°)	Wave Speed (cm)	MeFoSail (kg/h)	…
2024-05-01 10:40	5.945	5.765	255.395	7.873	−0.124	11.010	80.399	141.9	161.1	303.0	…
2024-05-01 10:50	6.159	5.028	269.654	20.324	−0.160	17.517	54.951	192.6	216.5	409.1	…
…	…	…	…	…	…	…	…	…	…	…	…
2024-05-31 11:40	6.000	5.416	54.400	37.179	−0.219	18.432	88.628	219.0	259.3	478.3	…
2024-05-31 11:50	6.000	5.464	42.871	40.031	−0.164	18.338	92.061	223.1	263.3	486.4	…

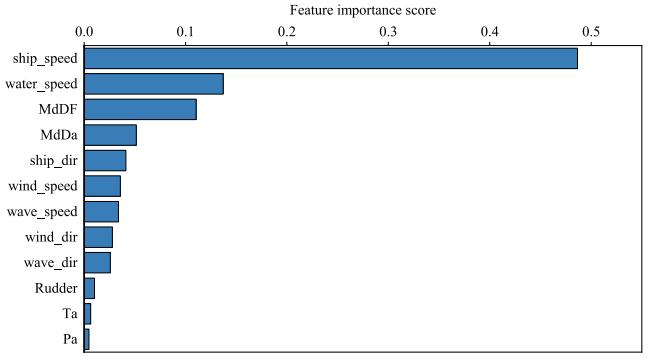

Fig. 6. Feature importance ranking.

decomposition, subband-wise GCN + iLSTM modeling, attention-based fusion, and final prediction.

Algorithm 1 The proposed fuel prediction algorithm.

Require: Input sequence $x \in \mathbb{R}^{B\times T\times D}$, edge index `edge_index`
Ensure: Predicted fuel consumption $\hat{y} \in \mathbb{R}^{B\times 1}$
Step 1: Multi-scale wavelet decomposition
$(Y_L, \{Y_H^{(j)}\}_{j=1}^{J}) \leftarrow \texttt{DWT1DForward}(x)$ ▷ $Y_L \in \mathbb{R}^{B\times D\times L_0}$, $Y_H^{(j)} \in \mathbb{R}^{B\times D\times L_j}$
Step 2: Subband-wise feature extraction using GCN + iLSTM
$H \leftarrow [\,]$ ▷ Initialize subband output list
$Y_L^{\text{seq}} \leftarrow \texttt{Transpose}(Y_L)$
$h_L \leftarrow \texttt{SubbandBranch\_iLSTM}(Y_L^{\text{seq}}, \texttt{edge\_index})$
Append h_L to H
for $j = 1$ to J **do**
  $Y_H^{(j,\text{seq})} \leftarrow \texttt{Transpose}(Y_H^{(j)})$
  $h_H^{(j)} \leftarrow \texttt{SubbandBranch\_iLSTM}(Y_H^{(j,\text{seq})}, \texttt{edge\_index})$
  Append $h_H^{(j)}$ to H
end for
Step 3: Attention-based fusion
$H_{\text{stack}} \leftarrow \texttt{Stack}(H)$
$s \leftarrow \texttt{Softmax}(H_{\text{stack}} \cdot w + b)$
$\mathbf{z}_B \leftarrow \sum_{j=1}^{J+1} s_j \cdot H_j$
Step 4: Final prediction
$\hat{y} \leftarrow \tanh(\texttt{Linear}(\mathbf{z}_B))$
return $\hat{y}$

3.2. Model construction and configuration

3.2.1. Multi-scale feature decomposition

The original data is decomposed using discrete wavelet transform with Daubechies-1 basis to extract features at multiple temporal scales. A 4-level decomposition yields one approximation coefficient A_4 and four detail coefficients D_1, D_2, D_3, and D_4, facilitating hierarchical modeling across low- and high-frequency components. Fig. 7 presents the

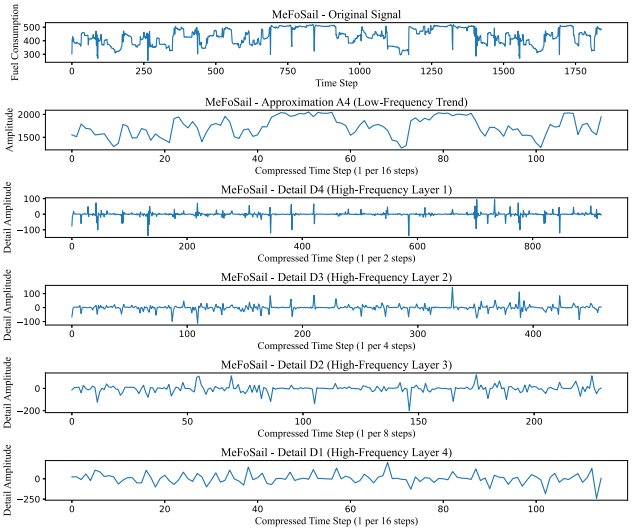

Fig. 7. Wavelet decomposition of fuel consumption sequence.

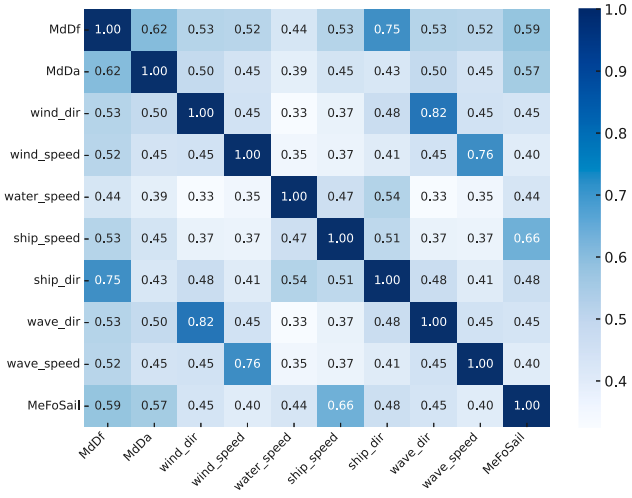

Fig. 8. Mutual information matrix among selected features.

wavelet decomposition of fuel consumption, illustrating the amplitudes of the approximate and detail components at different resolutions.

3.2.2. Residual sparse graph convolutional network

To capture heterogeneous structural dependencies among input features, a ResGCN is established for each sub-band. The graph topology for each sparse GCN is constructed via mutual information among key features, quantifying their nonlinear correlations. As shown in Fig. 8, mutual information values quantify pairwise feature dependencies; a sparsity threshold θ is applied to retain only significant connections, generating sparse adjacency matrices as Fig. 9.

Each sub-band is processed independently by a ResGCN module, aggregating feature interactions through the constructed graph topol-

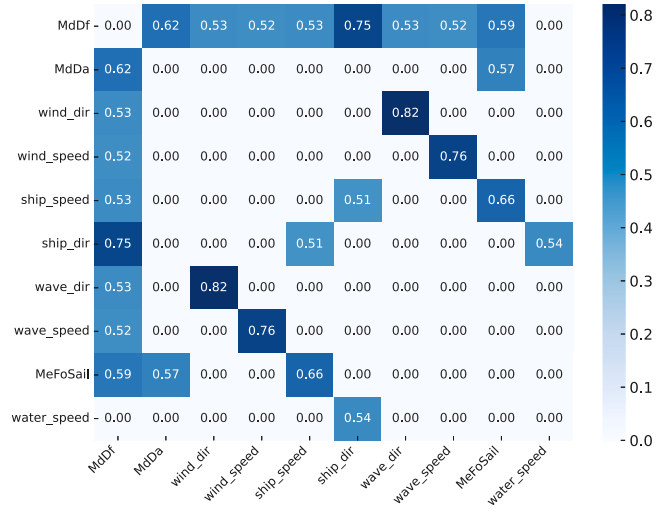

Fig. 9. Sparse adjacency matrix ($\theta = 0.5$).

ogy. Residual connections facilitate gradient propagation across layers, thereby improving model convergence and training stability. The ResGCN output is subsequently fed into the iLSTM module for prediction.

3.2.3. Ilstm and multi-scale fusion

To model temporal dynamics within each frequency sub-band, we employ the iLSTM module as the core unit. Each sub-band-representing either high-frequency detail components or the low-frequency approximation-is processed independently by a parameter-shared iLSTM branch. This modular architecture facilitates efficient parallel modeling of multi-scale temporal features while maintaining inter-scale coherence.

The iLSTM incorporates three key enhancements: (1) causal convolutions for local dependency capture, (2) block-diagonal projection matrices for efficient feature transformation, and (3) multi-head matrix states to augment representational capacity and training stability. Each branch consists of three hierarchically stacked layers with preserved sequence dimensionality.

Following ResGCN and iLSTM processing, sub-band outputs are adaptively fused via a dynamic attention mechanism, as shown in Fig. 10. At each timestep t, the module autonomously adjusts sub-band contributions $\alpha^{(t)}$ based on contextual relevance, enabling synergistic integration of global trends and local variations. The fused representation is then propagated through a fully connected layer for fuel consumption prediction.

This ResGCN and iLSTM multi-branch framework simultaneously captures spatial dependencies and multi-scale temporal patterns, thereby improving prediction accuracy and robustness under complex maritime operating regimes.

3.3. Model evaluation

To rigorously assess model accuracy and stability in ship fuel consumption prediction, this study employs four established regression performance metrics:

Coefficient of determination (R^2) quantifies the proportion of variance explained by the model. Root mean square error (RMSE) measures the standard deviation of prediction errors. Mean absolute error (MAE) represents the average absolute deviation between predictions and true values. Mean absolute percentage error (MAPE) evaluates relative prediction accuracy as a percentage.

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{23}$$

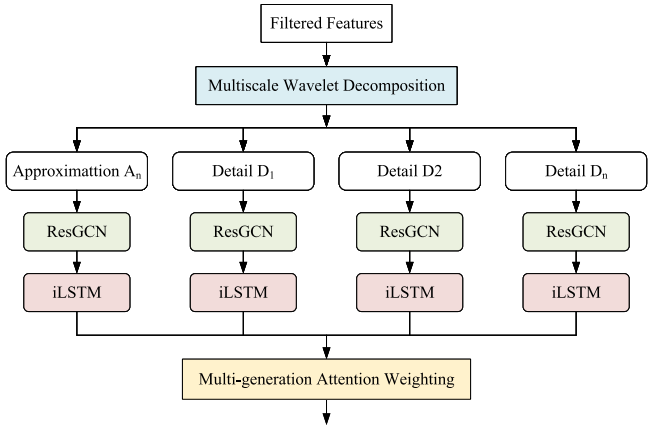

Fig. 10. Multi-subband weighted fusion model.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{24}$$

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_n| \tag{25}$$

$$\text{MAPE} = \frac{100\,\%}{N}\sum_{n=1}^{N}\left|\frac{y_n - \hat{y}_n}{y_n}\right| \tag{26}$$

where y_n denotes the actual value, $\hat{y}_n$ represents the predicted value, and $\bar{y}$ is the mean of the actual values.

4. Experimental results analysis

To validate the efficacy of our hybrid ResGCN and iLSTM framework for ship fuel consumption prediction, we conduct systematic comparisons across three dimensions: model architectures, graph construction strategies, and operational scenarios. The hyperparameters of the proposed model are shown in Table 2.

The proposed MWD-ResGCN-iLSTM achieves state-of-the-art performance across all test scenarios. Under the fully connected graph structure, its R^2 reaches 0.9667. Under $\theta = 0.4$, $\theta = 0.5$, and the empirical graph structure, the R^2 values are 0.9583, 0.9623, and 0.9654, respectively. Notably, MAPE remains below 5 % in all configurations, reaching a minimum of **4.051 %**.

These results demonstrate the model's unique capacity to integrate multi-scale feature decomposition, structural dependency encoding, and temporal dynamics modeling, delivering superior robustness and adaptability for complex spatiotemporal prediction tasks.

4.1. Comparative analysis of performance

Six benchmark models-BP, RBF, LSTM, sLSTM, mLSTM, and iLSTM-were evaluated for ship fuel consumption prediction. Comparative results are presented in Fig. 11.

Traditional neural networks, for BP, R^2 is 0.8119 and MAPE is 18.17 %, for RBF, R^2 is 0.847 and MAPE is 11.54 %. The results demonstrate limited capability in capturing temporal dependencies, resulting in suboptimal predictive performance.

The LSTM baseline achieves moderate performance ($R^2 = 0.8508$, MAPE = 10.21 %), while gated variants show significant improvements: sLSTM reduces MAPE to 6.53 % through enhanced dynamic response, and mLSTM exhibits improved robustness.

Notably, the proposed iLSTM outperforms all benchmarks with $R^2 = 0.9321$, RMSE = 0.0554, and MAPE = 6.81 %. This architecture maintains superior training efficiency while demonstrating enhanced feature modeling capacity and generalization ability.

Table 2
Overall model hyperparameter settings.

Component	Value/Setting	Description
— *Wavelet Decomposition* —		
Wavelet basis	Daubechies-1 (db1)	Used for multi-scale decomposition
Decomposition levels	2, 4, 6	Multi-group experiments with varying temporal resolution
— *Graph Convolution Module* —		
Number of GCN layers	2	Each layer has 32 graph convolution kernels
Graph sparsity thresholds	0, 0.4, 0.5, empirical	Constructing different sparse adjacency matrices
Residual connection	Enabled	Improves stability and gradient flow
— *iLSTM Sequence Model* —		
Input feature dimension	10	Number of features per time step
Input sequence length	10	Sliding window length for modeling
Number of iLSTM layers	3	Stacked temporal modeling blocks
Number of heads	2	Multi-head subspace decomposition
Head size	32	Dimensionality per head
Convolution type	CausalConv1D	Ensures temporal causality
Normalization	GroupNorm	Stabilizes multi-head processing
Activation	GeLU	Used in MLP projection layers
Up/Down Projection	Enabled	Applies dimension transformation with residuals
Parameter sharing	Across subbands	Reduces redundancy in multi-scale modeling
— *Training Configuration* —		
Optimizer	Adam	Adaptive gradient optimization
Learning rate	0.01	Fixed learning rate without decay
Loss function	MSE	Used for regression target
Batch size	256	Samples per iteration
Training epochs	100	Maximum iterations with early stopping

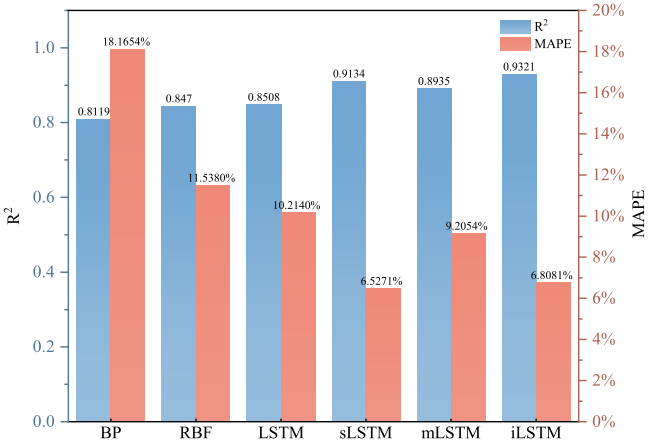

Fig. 11. Comparative analysis of multi-model performance.

4.2. Effectiveness of residual structures in sparse GCN networks

Experimental results in Table 3 demonstrate the impact of residual structures on graph neural network performance. The ResGCN architecture significantly mitigates the over-smoothing problem inherent in standard GCNs while enhancing deep feature modeling capacity and training stability.

For instance, the baseline iLSTM achieves $R^2 = 0.9321$. Notably, integrating standard GCN with iLSTM degrades performance across all graph sparsity levels. At $\theta = 0$ (fully-connected graph), R^2 drops below 0, indicating severe information degradation due to excessive message passing.

In contrast, ResGCN-iLSTM maintains $R^2 > 0.95$ for most graph configurations. Even in the empirical graph structure ($R^2 = 0.9398$), it significantly outperforms non-residual GCN variants by 8.6–15.2 %. This verifies residual connections' critical role in preventing information loss and enabling deep graph representation learning, particularly in densely-connected scenarios.

4.3. Impact of graph structure sparsity

The sparsity of graph structures significantly influences model performance. Table 3 compares prediction results across mutual information thresholds ($\theta = 0$, 0.4, 0.5) and empirical graphs, while Fig. 12 visualizes R^2 and MAPE trends under varying sparsity conditions.

Critically, fully-connected graphs ($\theta = 0$) substantially degrade performance: GCN-iLSTM exhibits negative R^2 (-0.015) and 41.42 % MAPE, indicating that excessive connections introduce redundant noise that disrupts feature extraction.

Conversely, increased sparsity enhances performance. Optimal results emerge at $\theta = 0.5$ and empirical graphs, where MWD-ResGCN-iLSTM achieves peak $R^2 = 0.9623$ with MAPE = 4.05 %, demonstrating superior stability and adaptability.

These findings establish moderately sparse graphs ($\theta = 0.5$ or empirical configurations) as the recommended construction strategy, optimally balancing feature retention against redundancy suppression.

4.4. Effectiveness of multi-scale wavelet decomposition

To augment the model's capacity for capturing multi-scale temporal dynamics, we introduce wavelet decomposition to dissect the original feature sequence into distinct frequency sub-bands. This explicitly represents low-frequency trends and high-frequency disturbances, enhancing feature structure. Table 4 compares prediction performance of MWD-ResGCN variants across wavelet decomposition levels (2, 4, and 6 layers).

Experimental results demonstrate that with graph sparsity fixed at $\theta = 0.5$, MWD substantially enhances model accuracy and robustness. The MWD-ResGCN-iLSTM achieves optimal performance under 4-layer wavelet decomposition: $R^2 = 0.9623$, MAPE = 4.82 %, with significant reductions in both RMSE and MAE. These improvements validate multi-scale modeling's efficacy in feature separation and representational enhancement.

Model performance exhibits significant sensitivity to wavelet decomposition depth. At 2-layer decomposition, acceptable overall accuracy is achieved, in which $R^2 = 0.9447$ and MAPE = 5.93 %, though high-frequency dynamics capture remains limited. The 4-layer configuration

Table 3
Experimental results on test set under different graph structures (4-level wavelet decomposition).

Model	$\theta = 0$ (Fully Connected)				$\theta = 0.4$ (Sparse)				$\theta = 0.5$ (Sparse)				Empirical Graph Structure			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
GCN-LSTM	0.3299	0.1741	0.1437	32.8574 %	0.3904	0.1660	0.1416	32.1103 %	0.8429	0.0843	0.0500	11.6068 %	0.5747	0.1387	0.1146	25.0961 %
ResGCN-LSTM	0.8590	0.0798	0.0431	10.3115 %	0.8615	0.0791	0.0425	10.3518 %	0.8640	0.0784	0.0435	9.7658 %	0.8726	0.0759	0.0429	10.2122 %
MWD-ResGCN-LSTM	0.9305	0.0560	0.0328	6.3106 %	0.9460	0.0494	0.0275	5.1340 %	0.9410	0.0516	0.0318	5.8647 %	0.9411	0.0516	0.0309	5.6625 %
GCN-sLSTM	0.1785	0.1927	0.1557	35.5537 %	0.5736	0.1389	0.1083	24.7676 %	0.9285	0.0568	0.0344	7.3337 %	0.7860	0.0984	0.0668	15.1356 %
ResGCN-sLSTM	0.9604	0.0423	0.0224	4.3341 %	0.9524	0.0464	0.0268	5.3079 %	0.9470	0.0490	0.0251	5.2426 %	0.9264	0.0577	0.0284	6.1297 %
MWD-ResGCN-sLSTM	0.9457	0.0496	0.0256	5.5016 %	0.9489	0.0481	0.0261	4.8716 %	0.9453	0.0497	0.0270	5.4500 %	0.9448	0.0499	0.0276	5.8635 %
GCN-mLSTM	0.2862	0.1797	0.1455	33.0243 %	0.4563	0.1568	0.1294	29.2913 %	0.9063	0.0651	0.0454	8.7317 %	0.7013	0.1162	0.0874	19.6755 %
ResGCN-mLSTM	0.9129	0.0627	0.0412	8.3155 %	0.9273	0.0573	0.0376	7.2965 %	0.9407	0.0518	0.0297	6.1628 %	0.9219	0.0594	0.0394	7.7604 %
MWD-ResGCN-mLSTM	0.9136	0.0625	0.0368	7.1203 %	0.9263	0.0577	0.0408	8.6264 %	0.9565	0.0443	0.0275	5.3456 %	0.9307	0.0560	0.0361	7.2159 %
GCN-iLSTM	−0.0150	0.2142	0.1753	41.4231 %	0.0665	0.2055	0.1689	38.9025 %	0.8839	0.0725	0.0525	10.3275 %	0.7866	0.0982	0.0659	14.5200 %
ResGCN-iLSTM	0.9665	0.0389	0.0206	3.9097 %	0.9585	0.0433	0.0231	4.5092 %	0.9522	0.0465	0.0242	4.8192 %	0.9398	0.0522	0.0303	6.4527 %
MWD-ResGCN-iLSTM	**0.9667**	**0.0388**	**0.0241**	**4.7356 %**	0.9583	0.0434	0.0287	5.7426 %	**0.9623**	**0.0413**	**0.0207**	**4.0510 %**	**0.9654**	**0.0396**	**0.0232**	4.3688 %

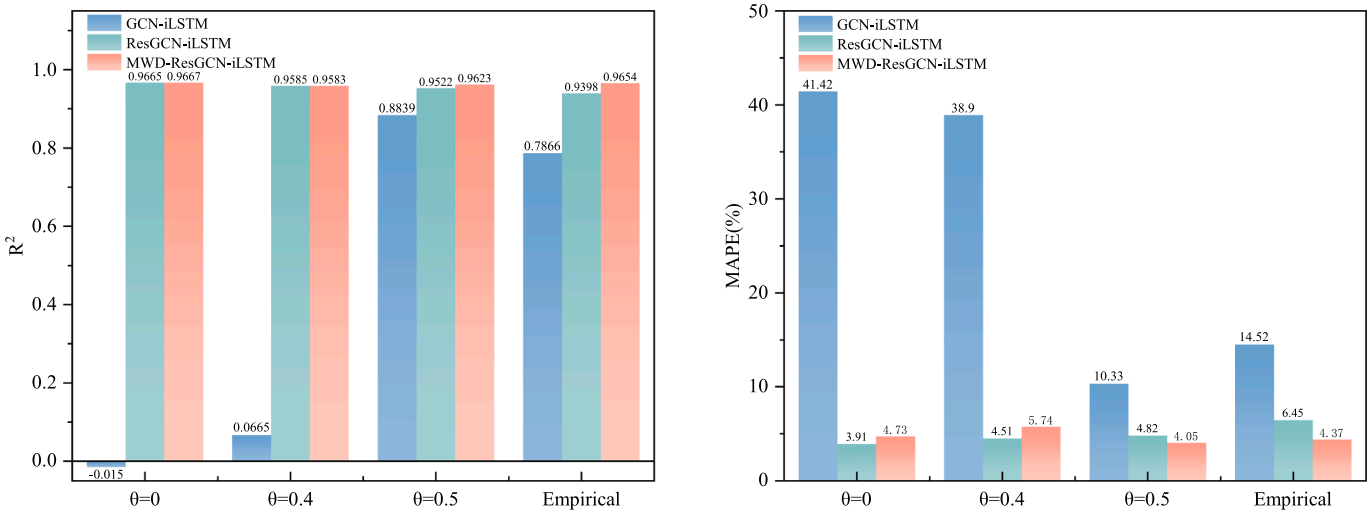

Fig. 12. Performance comparison of models under different graph structures.

Table 4
Performance comparison on test set under different wavelet decomposition levels.

Model	2-level				4-level				6-level			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
MWD-ResGCN-LSTM	0.9040	0.0659	0.0381	7.8167 %	0.9410	0.0516	0.0318	5.8647 %	0.9234	0.0589	0.0316	5.9081 %
MWD-ResGCN-sLSTM	0.9429	0.0508	0.0282	5.4963 %	0.9453	0.0497	0.0270	5.4500 %	0.9346	0.0544	0.0292	6.0441 %
MWD-ResGCN-mLSTM	0.9167	0.0614	0.0396	7.9798 %	0.9565	0.0443	0.0275	5.3456 %	0.9003	0.0672	0.0429	8.5878 %
MWD-ResGCN-iLSTM	0.9447	0.0500	0.0293	5.9318 %	0.9623	0.0413	0.0207	4.0510 %	0.9366	0.0535	0.0256	5.5568 %

yields optimal performance, in which $R^2 = 0.9623$ and MAPE = 4.82 %, demonstrating peak multi-scale modeling efficacy. Conversely, 6-layer decomposition degrades performance with MAPE = 6.21 %, indicating noise introduction from excessive decomposition that disrupts feature extraction.

Optimal decomposition depth is therefore critical for maximizing MWD's time-series modeling quality. The 4-layer benchmark optimally balances time-frequency resolution while preventing spectral leakage from over-decomposition.

Table 5
Prediction performance under noise and abnormal conditions.

Condition	MAE	RMSE	Med. Shift
No disturbance	0.0207	0.0413	−0.0014
GN (std = 1 %)	0.0233	0.0431	−0.0017
GN (std = 3 %)	0.0371	0.0604	−0.0010
GN (std = 5 %)	0.0525	0.0738	−0.0027
GN (std = 10 %)	0.0903	0.1227	−0.0027
Abnormal (× 1.3 wind + speed)	0.0407	0.0657	−0.0025

4.5. Robustness analysis

To evaluate the proposed model's robustness and predictive stability under non-ideal conditions, we conducted two perturbation tests: Gaussian noise injection and abnormal operating condition simulation. The model's performance was evaluated using the testing dataset, and the results are presented in Table 5.

Experimental results in Table 5 and residual distributions in Fig. 13 demonstrate the model's sustained high accuracy and stability across disturbances. Even under maximum perturbation (10 % GN), MAE and RMSE rose marginally to 0.0903 and 0.1227 respectively, with median residuals confined to ±0.003. For abnormal conditions, robustness approached the 5 % noise scenario (MAE = 0.0407, RMSE = 0.0657), demonstrating notable tolerance to disturbances in key variables like wind and ship speed.

In summary, the proposed model demonstrates excellent robustness when faced with input disturbances and abnormal values in key

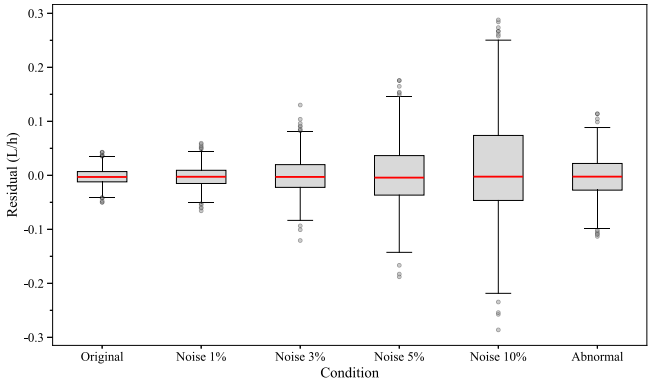

Fig. 13. Residual distribution under different perturbation conditions.

features, supporting its applicability in real-world ship fuel consumption prediction under complex maritime environments.

5. Conclusion

This study proposes a novel multi-branch model, termed MWD-ResGCN-iLSTM, for ship fuel consumption prediction. The model integrates Multi-scale Wavelet Decomposition, a Residual Sparse Graph Convolutional Network, and an improved Long Short-Term Memory network. Key technical innovations are systematically developed and validated through three principal dimensions:

Structural optimization: A mutual information-based weighted graph structure captures implicit feature correlations. Experimental analysis of varying sparsity levels and empirical graphs demonstrates that moderate sparsity effectively mitigates information redundancy and over-smoothing, enhancing feature propagation efficiency. Residual connections in GCNs are employed to mitigate over-smoothing by retaining input features across layers, thereby enhancing representation capacity and training stability.

Sequence modeling: Building upon the traditional LSTM, an enhanced iLSTM incorporating residual mapping is proposed. This architecture achieves an optimal balance between nonlinear modeling capability and computational efficiency while maintaining high prediction accuracy, serving as the core temporal modeling unit.

Feature enhancement: The innovative integration of multi-scale wavelet decomposition decomposes raw signals into distinct frequency subbands. Independent multi-branch pathways model these subbands, followed by attention-based fusion, substantially strengthening the model's ability to capture complex multi-scale temporal patterns.

Extensive experiments confirm that the proposed MWD-ResGCN-iLSTM model achieves state-of-the-art performance across all evaluated graph configurations, attaining a peak R^2 of 0.9667 and a minimal MAPE of 4.0510 %. The model exhibits exceptional generalization capability and robustness, while maintaining a favorable balance between structural complexity and training efficiency. These results underscore its significant potential for practical deployment in multivariate time series forecasting applications.

Though the proposed model exhibits robust empirical performance, we recognize its inherent limitations. It lacks full generalizability across ship types and operating scenarios. Rather than seeking universal solutions, this work deciphers interactions between navigational conditions, environmental factors, and their synergistic effects on fuel efficiency. Future work will expand datasets to include diverse ship types and explore cross-scenario transfer learning, to enhance practical applicability and generalization capacity in maritime settings.

CRediT authorship contribution statement

Weibo Zhong: Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization; **Kang Bai:** Writing – original draft, Validation, Methodology, Formal analysis; **Yang Gu:** Writing – original draft, Software, Methodology; **Nan Ye:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Jiangsu Province Science and Technology Achievement Transformation Special Fund Project (BA2022014) and Jiangsu Province Graduate Research and Practice Innovation Program Project (SJCX25_2507).

References

Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J., Hochreiter, S., 2024. xLSTM: Extended long short-term memory. arXiv:2405.04517

Bhatti, U.A., Tang, H., Wu, G., Marjan, S., Hussain, A., 2023. Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence. Int. J. Intell. Syst. 2023 (1), 8342104. https://doi.org/10.1155/2023/8342104.

Chen, B., Wang, X., Wang, R., Qiu, X., Zhu, Z., Liang, M., 2019. The grey-box based modeling approach research integrating fusion mechanism and data. J. Syst. Simul. 31 (12), 2575–2583.

Chen, Y., Sun, B., Xie, X., Li, X., Li, Y., Zhao, Y., 2024. Short-term forecasting for ship fuel consumption based on deep learning. Ocean Eng. 301, 117398. https://doi.org/10.1016/j.oceaneng.2024.117398.

Fan, A., Yang, J., Yang, L., Wu, D., Vladimir, N., 2022. A review of ship fuel consumption models. Ocean Eng. 264, 112405. https://doi.org/10.1016/j.oceaneng.2022.112405.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hou, G., Wang, J., Fan, Y., 2024. Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction. Energy 286, 129640. https://doi.org/10.1016/j.energy.2023.129640.

Hu, Z., Zhou, T., Zhen, R., Jin, Y., Li, X., Osman, M.T., 2022. A two-step strategy for fuel consumption prediction and optimization of ocean-going ships. Ocean Eng. 249, 110904. https://doi.org/10.1016/j.oceaneng.2022.110904.

Li, Y., Li, Z., Mei, Q., Wang, P., Hu, W., Wang, Z., Xie, W., Yang, Y., Chen, Y., 2023. Research on multi-port ship traffic prediction method based on spatiotemporal graph neural networks. J. Mar. Sci. Eng. 11 (7), 1379. https://doi.org/10.3390/jmse11071379.

Mao, Z., Zhao, Y., Zhao, J., Shao, L., Zuo, H., 2025. A dual-channel ultra-short-term ship motion prediction model using temporal convolutional network, frequency-enhanced channel attention, and informer. Phys. Fluids 37 (2). https://doi.org/10.1063/5.0257465.

Moreira, L., Vettor, R., Guedes Soares, C., 2021. Neural network approach for predicting ship speed and fuel consumption. J. Mar. Sci. Eng. 9 (2), 119. https://doi.org/10.3390/jmse9020119.

Odendaal, K., Alkemade, A., Kana, A.A., 2023. Enhancing early-stage energy consumption predictions using dynamic operational voyage data: a grey-box modelling investigation. Int. J. Nav. Archit. Ocean Eng. 15, 100484. https://doi.org/10.1016/j.ijnaoe.2022.100484.

Orihara, H., Tsujimoto, M., 2018. Performance prediction of full-scale ship and analysis by means of on-board monitoring. Part 2: validation of full-scale performance predictions in actual seas. J. Mar. Sci. Technol. 23 (4), 782–801. https://doi.org/10.1007/s00773-017-0511-5.

Qi, C., Zhang, J., Jia, H., Mao, Q., Wang, L., Song, H., 2021. Deep face clustering using residual graph convolutional network. Knowl. Based Syst. 211, 106561. https://doi.org/10.1016/j.knosys.2020.106561.

Qureshi, S., et al., 2023. Limits of depth: over-smoothing and over-squashing in GNNs. Big Data Min. Anal. 7 (1), 205–216. https://doi.org/10.26599/BDMA.2023.9020019.

Schlag, I., Irie, K., Schmidhuber, J., 2021. Linear transformers are secretly fast weight programmers. In: International Conference on Machine Learning. PMLR, pp. 9355–9366.

Tillig, F., Ringsberg, J.W., 2019. A 4 DOF simulation model developed for fuel consumption prediction of ships at sea. Ships Offshore Struct. 14 (sup1), 112–120. https://doi.org/10.1080/17445302.2018.1559912.

Vinayak, P.P., Prabu, C. S.K., Vishwanath, N., Prakash, S.O., 2021. Numerical simulation of ship navigation in rough seas based on ECMWF data. BRODOGRADNJA: An International Journal of Naval Architecture and Ocean Engineering for Research and Development 72 (1), 19–58. https://doi.org/10.21278/brod72102.

Wang, K., Hua, Y., Huang, L., Guo, X., Liu, X., Ma, Z., Ma, R., Jiang, X., 2023. A novel GA-LSTM-based prediction method of ship energy usage based on the characteristics analysis of operational data. Energy 282, 128910. https://doi.org/10.1016/j.energy.2023.128910.

Wen, Y., Wu, R., Zhou, Z., Zhang, S., Yang, S., Wallington, T.J., Shen, W., Tan, Q., Deng, Y., Wu, Y., 2022. A data-driven method of traffic emissions mapping with land use random forest models. Appl. Energy 305, 117916. https://doi.org/10.1016/j.apenergy.2021.117916.

Xiao, G., Pan, L., Lai, F., 2025. Application, opportunities, and challenges of digital technologies in the decarbonizing shipping industry: a bibliometric analysis. Front. Mar. Sci. 12, 1523267. https://doi.org/10.3389/fmars.2025.1523267.

Yan, R., Wang, S., Du, Y., 2020. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. Transp. Res. Part E Logist. Transp. Rev. 138, 101930. https://dx.doi.org/10.1016/j.tre.2020.101930.

Yan, R., Wang, S., Psaraftis, H.N., 2021. Data analytics for fuel consumption management in maritime transportation: status and perspectives. Transp. Res. Part E Logist. Transp. Rev. 155, 102489. https://doi.org/10.1016/j.tre.2021.102489.

Zhang, M., Tsoulakos, N., Kujala, P., Hirdaris, S., 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. Eng. Appl. Artif. Intell. 130, 107425. https://doi.org/10.1016/j.engappai.2023.107425.

Zhang, W., Lin, Z., Liu, X., 2022. Short-term offshore wind power forecasting-a hybrid model based on discrete wavelet transform (DWT), seasonal autoregressive integrated moving average (SARIMA), and deep-learning-based long short-term memory (LSTM). Renew. Energy 185, 611–628. https://doi.org/10.1016/j.renene.2021.12.100.

Zhao, J., Zhao, Y., 2025. Very short-term forecasting of ship multidimensional motion using two coupled models based on deep operator networks. Ocean Eng. 317, 120044. https://doi.org/10.1016/j.oceaneng.2024.120044.

Zhou, Y., Pazouki, K., Murphy, A.J., Uriondo, Z., Granado, I., Quincoces, I., Fernandes-Salvador, J.A., 2023. Predicting ship fuel consumption using a combination of metocean and on-board data. Ocean Eng. 285, 115509. https://doi.org/10.1016/j.oceaneng.2023.115509.