

Article

Channel-Aware Residual BiLSTM for Ship Trajectory Prediction and Collision Risk Assessment in Restricted Waters

Haibo Xie , Zhiqiang Shi , Yujing Jiang and Yangyang Ren

Navigation College, Dalian Maritime University, Dalian 116026, China; xiehaibo@dlmu.edu.cn (H.X.); yu0430aaa@163.com (Y.J.); 17839445884@163.com (Y.R.)

* Correspondence: szq@dlmu.edu.cn

Abstract

Restricted waterways combine curved fairways, limited maneuvering space, and dense encounters, complicating short-term ship trajectory prediction and collision risk assessment. This study proposes a channel-aware residual Bidirectional Long Short-Term Memory (BiLSTM) framework. The Nonlinear Randomly Reuse-based Mutated Whale Optimization Algorithm (NRRMWOA) configures its hyperparameters; a constant-velocity (CV) residual prior, channel look-ahead geometry, global–local fusion, and an Artificial Potential Field–Potential Collision Risk (APF-PCR) risk head are incorporated. Evaluation uses Automatic Identification System (AIS) data from two restricted-water regions. The fused system yields average and final displacement errors (ADE and FDE) of 103.70/217.16 m in Study Area 1 and 116.15/243.67 m in Study Area 2. Compared with the strongest metric-specific model, the ADE/FDE reductions are 5.47%/6.15% and 8.27%/8.07%, respectively; all differences are significant. The alert+ area under the receiver operating characteristic curve (AUC) values are 0.988 and 0.993. Under the tested conditions, the CV residual, channel constraints, and expert fusion reduce trajectory error and supply forecasts for stepwise collision risk assessment.

Keywords: AIS; ship trajectory prediction; BiLSTM; whale optimization algorithm; channel constraint; collision risk assessment; artificial potential field; restricted waters

1. Introduction

Intelligent shipping, Maritime Autonomous Surface Ships (MASS), and Vessel Traffic Service (VTS) systems increasingly rely on data-driven situational awareness and early warning. The Automatic Identification System (AIS) records position, speed, course, ship type, and other vessel information used in behavior analysis and risk assessment [1]. In port fairways, crossing waters, precautionary areas, and narrow waterways, limited maneuvering space and dense traffic make short-term prediction useful for VTS supervision, collision-avoidance assistance, and traffic organization [2].

AIS-based ship trajectory prediction nevertheless faces several challenges. Operational AIS data contain uneven sampling intervals, positioning noise, outliers, missing values, and abrupt changes in speed or course. Direct training on such records can bias the learned motion patterns and increase the accumulated prediction error [1,3]. Moreover, motion in restricted waters is not a constant-velocity straight-line process: turning, deceleration, and collision avoidance can alter the future trajectory. Similar observation histories may lead to different subsequent motion near curved channels, bifurcations, and multi-vessel



Academic Editor: Decheng Wan

Received: 6 July 2026

Revised: 31 July 2026

Accepted: 3 August 2026

Published: 5 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

encounters, where channel geometry, maneuvering inertia, and neighboring vessels act together [4,5].

Ship trajectory predictors include motion models, statistical or machine learning methods, and deep networks. Traditional motion models are interpretable and computationally efficient [1,6–8]. Their assumptions, however, limit adaptation to complex turns and disturbances and can produce drift in restricted waters [1,9,10]. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs) capture temporal dependence in AIS sequences [11,12]. BiLSTM further combines forward and backward context within the observed sequence [13,14]. Section 2 reviews these methods in detail.

Transformer predictors use self-attention to model long-range sequence dependence [15,16], whereas Graph Neural Networks (GNNs) propagate information over vessel nodes and interaction edges [17,18]. Their suitability depends on sample size, input representation, and computational budget. Deep predictors are commonly optimized for displacement error, with physical consistency, channel feasibility, and risk interpretation treated separately. Collision risk methods based on CPA, DCPA, TCPA, ship domains, and learned classifiers are well established [19–23]. Trajectory prediction and risk assessment remain decoupled across much of this work. The experiments therefore use independent labels and feature ablation to separate predicted-trajectory features from rule-based risk variables.

Restricted-water prediction requires a motion reference that can accommodate channel geometry and a means of translating the predicted path into a time-varying risk estimate. These two aspects are examined in waterways with different geometries.

The proposed framework applies NRRMWOA [24] to mixed continuous–discrete hyperparameter tuning for BiLSTM-based ship trajectory prediction. The local predictor combines a causal CV reference with channel look-ahead geometry and learns the residual trajectory. Its output is fused with a global channel expert, and the resulting trajectory is passed to an APF-PCR head for horizon-wise risk assessment.

This study makes four contributions:

- (1) A causal CV trajectory is used as the reference, while the BiLSTM learns its residual correction. Route identity, centerline matching, and six direction-compatible look-ahead points are encoded with the observed AIS sequence and evaluated in two restricted-water regions.
- (2) NRRMWOA is applied to the joint selection of continuous and discrete BiLSTM hyperparameters. The validation objective is based on ADE and gives additional weight to turning samples; the selected regional configuration is fixed before final multi-seed training.
- (3) Horizon-dependent fusion combines the local predictor’s short-term maneuver response with the broader route trend supplied by a global channel expert. The weight at each forecast horizon is selected from validation data.
- (4) The fused trajectory is passed to an APF-PCR risk head that produces a score at each prediction step. Evaluation with an independent geometric label and feature-subset retraining separates the information carried by trajectory-derived features from that of PCR-related inputs.

The remainder of this paper is organized as follows. Section 2 reviews related research on AIS trajectory prediction, physically constrained modeling, multi-vessel interaction, and collision risk assessment. Section 3 introduces the two study areas, data preprocessing, problem formulation, and the proposed NRRMWOA-BiLSTM and APF-PCR risk assessment methods. Section 4 presents comparative experiments, ablation studies, and risk assessment results. Section 5 discusses model performance, the interpretation of the risk

indicators, and methodological limitations. Section 6 concludes the paper and outlines directions for future work.

2. Related Work

2.1. AIS-Based Ship Trajectory Prediction

AIS-based ship trajectory prediction supports maritime traffic monitoring, intelligent navigation, and collision-avoidance assistance. Early studies primarily adopted kinematic models and state estimation. Perera et al. [6] constructed a multi-model extended Kalman filter for different maneuvering states. Semerdjiev and Mihaylova [25] proposed an augmented interacting multiple-model tracking algorithm, and Zhao et al. [8] applied an improved Kalman filter in controlled waterways. These methods are interpretable and efficient, but low-order motion assumptions limit their adaptation to complex turns, variable speed, and uncertain maneuvers [1,9,10].

Machine learning methods mine ship motion patterns from historical AIS data. Virjonen et al. [26] employed k -nearest neighbors to predict target-ship motion from similar historical trajectories, while Liu et al. [9] optimized a support vector regression model with adaptive chaotic differential evolution. Random forests have also been used to forecast ship speed [27] and vessel destinations [28]. These methods provide nonlinear fitting capability but often depend on manual feature engineering and incompletely represent temporal dependence, turning, and accumulated sequence errors.

Deep learning methods have further enhanced temporal representation in AIS trajectories. Representative approaches include interpolation-corrected MP-LSTM [29], hierarchical two-layer LSTM [12], and GRU-based ST-Seq2Seq [30]. BiLSTM has been combined with attention [14] and Optuna-based hyperparameter tuning [31]. These studies support recurrent models for short-term AIS prediction, although their performance remains sensitive to hyperparameter selection, outliers, and accumulated errors. Many also learn primarily from the AIS sequence without explicitly using motion priors or channel geometry.

2.2. Attention, Transformer, and Graph Neural Network Methods

Attention mechanisms help trajectory models identify informative time steps and motion features. Multi-scale convolution has been combined with attention LSTM for ship trajectory prediction [32], and attention BiLSTM has been applied to collision risk warning with AIS-derived relative-motion features [21]. RAGAN instead uses adversarial learning and an interaction gate to represent evolving collision-avoidance effects in multi-ship encounters [33]. Attention and interaction improve feature selection and temporal representation but do not by themselves impose ship kinematics or channel feasibility.

Transformer models form another branch of ship trajectory prediction. TraISformer represents AIS trajectories as sparse discrete sequences and formulates prediction as generative sequence modeling [34]. TRFM-LS combines LSTM temporal features with Transformer multi-head attention [35]. G-Trans integrates trajectory smoothing, latent motion-feature clustering, GRU encoding, and Transformer temporal modeling [16]. These architectures model long-range dependence and multimodal output, usually with more parameters and training data than recurrent alternatives.

Graph neural networks primarily represent multi-vessel interactions. Existing architectures combine graph convolution with LSTM [36], GRU and self-attention [18], or maritime spatio-temporal graph convolution [37]. Other work has addressed graph-driven long-term prediction [17] and heterogeneous graphs with dynamic and static vessel attributes [38]. SNIINet fuses navigation information through position-attention, information-fusion, and global-attention blocks [39]. Their predictions depend on the adjacency definition, neighbor

selection, vessel attributes, and graph construction. In restricted waters, channel geometry, traffic direction, and motion priors remain relevant alongside vessel interactions.

2.3. Physics-Informed and Channel-Constrained Prediction

Physics-informed and hybrid approaches combine interpretable motion models with nonlinear learning. Constant-velocity models and Kalman-filter variants provide stable short-term predictions from the current motion state [1,6,40]. Physics has also been introduced through velocity and turning terms in a neural objective [41], model-driven estimation of difficult-to-express components [42], stacked LSTM with Kalman correction [43], and finite-difference kinematic losses [44]. Physical models can therefore serve as baselines, priors, constraints, or residual anchors for neural predictors.

Channel constraints are particularly important in restricted waters. Spatial grid maps have been combined with trajectory preprocessing and ConvLSTM prediction [45], while data-driven maritime-route extraction supports path planning from large AIS corpora [46]. Work on two-way restricted waterways further emphasizes spatial constraints and traffic organization [47]. Yet many predictors treat AIS trajectories as ordinary time series without explicitly encoding route direction and look-ahead geometry. In curved fairways, this omission can produce off-channel forecasts or conflict with observed maneuvering trends.

2.4. Collision Risk Assessment and Joint Prediction–Risk Modeling

Collision risk assessment is closely related to trajectory prediction. Traditional methods evaluate proximity and collision tendency using CPA, DCPA, and TCPA. Du et al. [19] reviewed AIS-based waterway risk assessment and highlighted the role of CPA-type indicators. Ship-domain models instead describe spatial safety margins around a vessel. Zhang and Meng [23] constructed a probabilistic AIS-based domain from target-position distributions. Im and Luong [20] used kernel density estimation to define continuous risk-level zones. These interpretable methods mostly rely on current relative geometry and use little of the future risk implied by a predicted trajectory.

Deep learning has also been applied to collision risk warning. Attention BiLSTM has been used for supervised risk classification from AIS-derived relative-motion features [21]. A separate framework assesses current risk and forecasts its future evolution [22], whereas multi-task learning has been used to jointly predict trajectories and collision risk [48]. Despite this progress, many risk models remain centered on current geometry or rule-based indicators rather than the full predicted trajectory.

The experiments compare route-aware residual prediction with recent maritime model families in two regions and evaluate the risk features under a label independent of the PCR inputs.

3. Materials and Methods

The method comprises AIS preprocessing, supervised sample construction, local coordinate transformation, CV-residual modeling, channel encoding, NRRMWOA hyperparameter search, global–local trajectory fusion, and per-horizon collision risk assessment. In both areas, the local three-seed ensemble supplies the trajectory used by the risk head. Figure 1 summarizes the workflow.

3.1. Study Areas and AIS Data Preprocessing

Two restricted-water AIS datasets are used in this study. They are denoted as Study Area 1 and Study Area 2 throughout the following sections so that the same preprocessing, prediction, ablation, and risk-assessment stages can be compared directly.

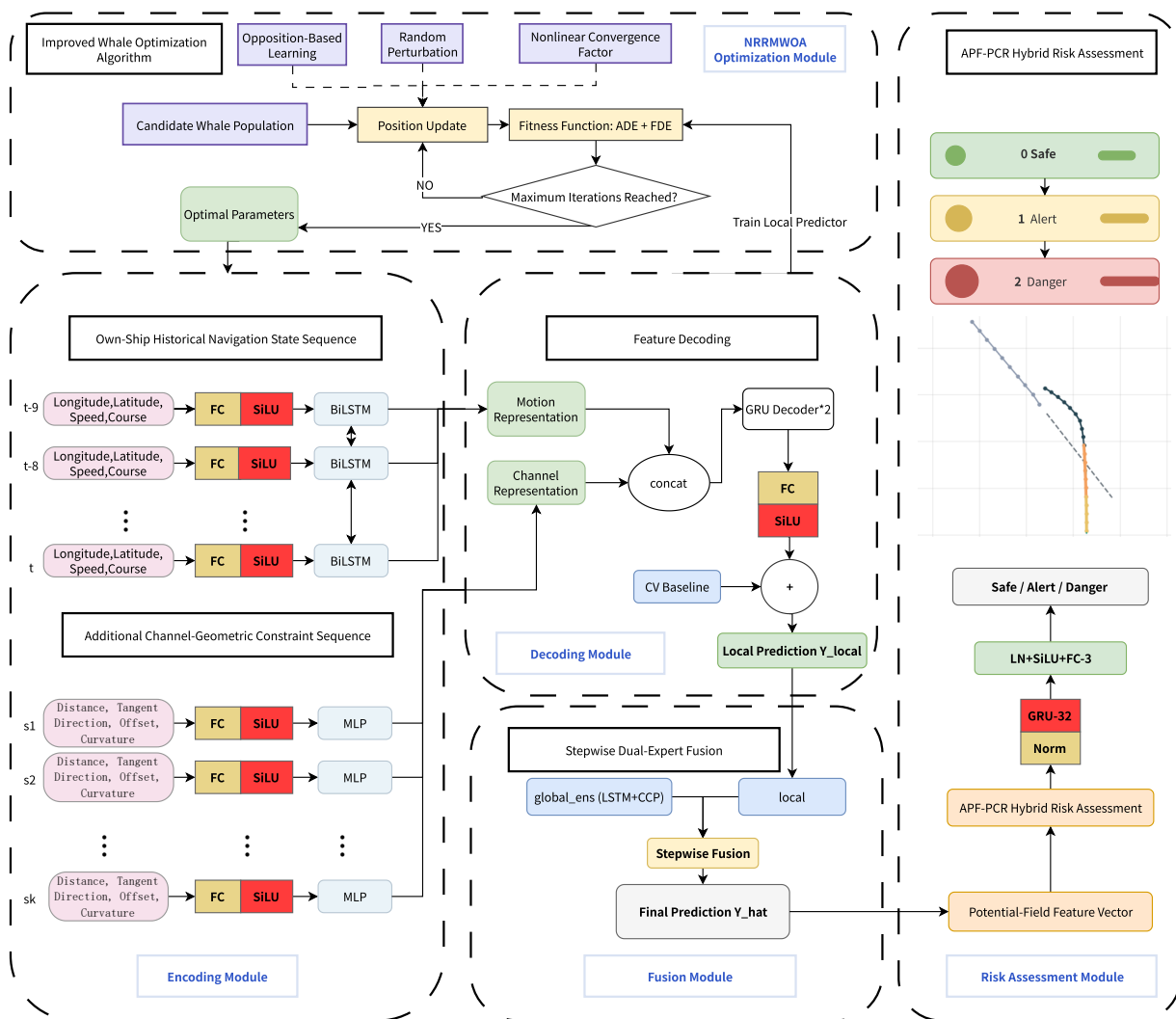


Figure 1. Proposed framework.

3.1.1. Study Area 1

Study Area 1 covers the restricted waters near the Galveston Bay entrance and the Houston Ship Channel in Texas, USA, with a spatial extent of approximately 94.48° W– 94.27° W and 29.04° N– 29.27° N. The raw AIS data include Maritime Mobile Service Identity (MMSI), timestamp, longitude/latitude, speed over ground (SOG), course over ground (COG), and ship length.

3.1.2. Study Area 2

Study Area 2 is located in the waters south of Victoria, British Columbia, Canada, near the eastern entrance of the Strait of Juan de Fuca. It is bounded by 48.192° – 48.330° N and 123.670° – 123.300° W. The source data span 30 December 2024 to 30 December 2025 and contain 647,573 valid AIS messages from 3566 vessels. Spatial filtering retains 620,953 messages from 3506 vessels. After quality filtering, discontinuity splitting, gap-based voyage segmentation, and 1 min resampling, 13,953 continuous voyages and 693,797 resampled points from 3171 MMSIs are retained. Figures 2 and 3 show the regional context and charts; the box marks the study extent, the arrow points north, and the other symbols and colors follow nautical-chart conventions.

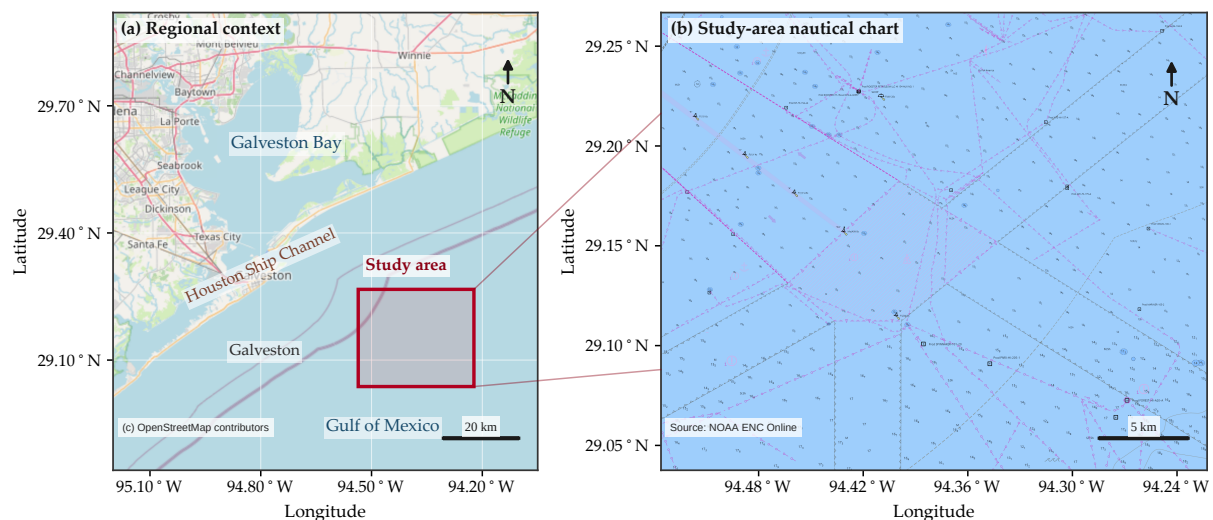


Figure 2. Study Area 1: (a) regional context; (b) nautical chart.

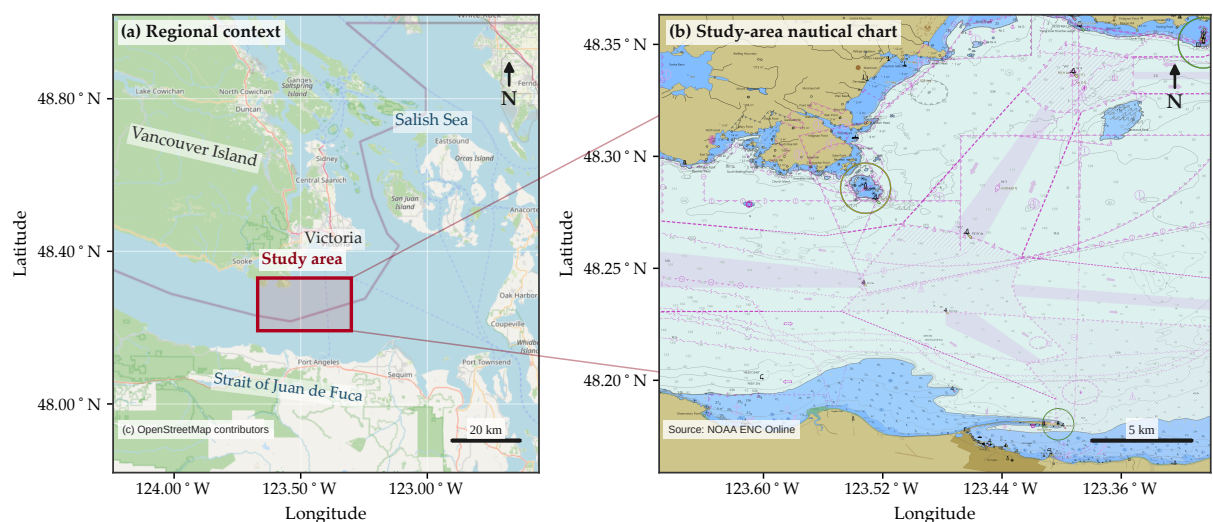


Figure 3. Study Area 2: (a) regional context; (b) nautical chart.

3.1.3. Preprocessing and Sample Construction

Both study areas follow the same processing sequence: spatial filtering, removal of records with missing or physically invalid fields, chronological voyage segmentation, resampling, interpolation, and supervised-window construction. Records are sorted by MMSI and time, and a gap longer than 6 min starts a new voyage. A candidate voyage and each fragment produced by discontinuity splitting must contain at least 20 raw AIS points. After resampling, at least 20 points on the regular 1 min grid are required to support the 10-step history and 10-step future window. Longitude and latitude are interpolated by the shape-preserving piecewise cubic Hermite method (PCHIP), SOG is interpolated linearly, and COG is recomputed from backward differences of the interpolated positions.

In Study Area 1, a manual quality audit removed 310 looping or near-return voyages from 18,219 resampled voyages (1.70%); the audit removed 35,050 of 1,179,753 resampled points (2.97%). The subsequent supervised-sample stage applied no exclusion based on turn angle, acceleration, encounter geometry, or risk level. In Study Area 2, records outside the specified bounds or the 0–50 kn SOG range are removed, gaps over 6 min define new voyages, and nonphysical spatial discontinuities are split before interpolation. The processing retains 13,953 voyages. Route clusters 0, 1, 2, 3, 4, 6, 7, and 8 provide the fixed channel prior; noise cluster −1 and clusters 5, 9, 10, 11, and 12 are excluded. In both areas,

each supervised sample contains 10 historical and 10 future steps, and the windows follow the voyage-disjoint chronological split with a 19 min boundary embargo described below. These stages are shown in Figures 4 and 5.

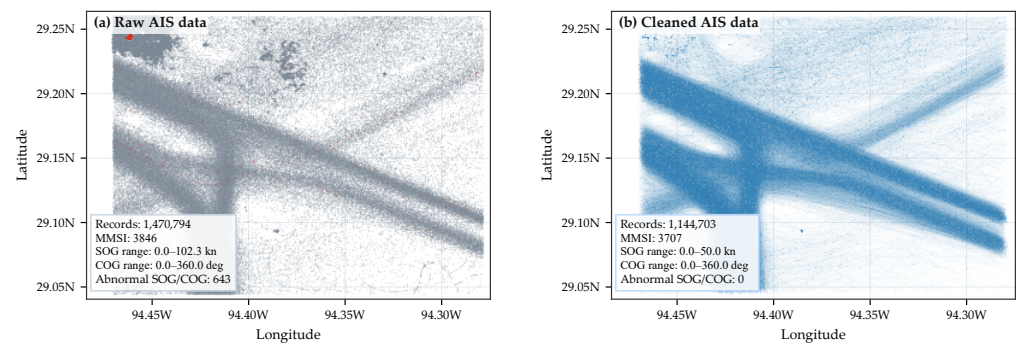


Figure 4. AIS data in Study Area 1: (a) raw; (b) cleaned.

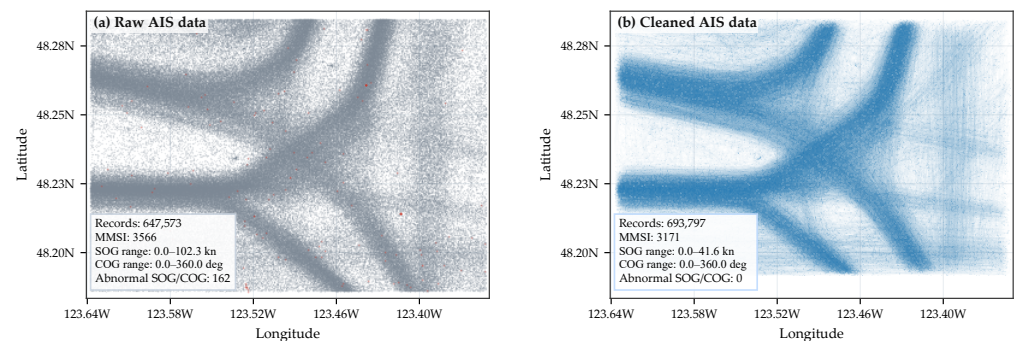


Figure 5. AIS data in Study Area 2: (a) raw; (b) cleaned.

To construct supervised learning samples in both study areas, a sliding window takes the historical H -step AIS states as input and predicts the future T -step ship positions. In this study, $H = 10$ and $T = 10$. The sliding windows are generated and retained independently within each leakage-controlled temporal block; windows crossing a subset boundary are discarded. The model input feature is

$$\mathbf{s}_t = [x_t, y_t, v_t, \psi_t, L], \quad (1)$$

where x_t, y_t are local metric coordinates, v_t is the speed, ψ_t is the course, and L is the ship length. From these raw states, the causal sequence encoder derives the local position, displacement velocity, acceleration, speed, sine and cosine of relative heading, turn rate, and log-scaled ship length. The resulting 11-dimensional feature vector at time t is denoted by $\mathbf{z}_t$.

The trajectory predictor uses fixed physical-reference scaling rather than sample mean–standard deviation normalization. Local positions are divided by 2000 m; the one-minute displacement-velocity components and their magnitude are divided by 400 m; acceleration increments are divided by 100 m; turn rate is divided by π ; and ship length is transformed as $\log(1 + L)/6$. The complete temporal feature vector is clipped to $[-6, 6]$. For the channel input, the six local look-ahead coordinates are divided by 3000 m and clipped to $[-3, 3]$, the anchor coordinates are divided by 20,000 m, and the centerline match score is clipped to $[0, 3000]$ m and divided by 1500 m. These constants are fixed before training and place the physical variables on comparable numerical scales.

3.2. Problem Definition

Given the historical trajectory sequence of a target ship,

$$\mathbf{X}_{t-H+1:t} = [\mathbf{s}_{t-H+1}, \mathbf{s}_{t-H+2}, \dots, \mathbf{s}_t], \quad (2)$$

The short-term trajectory prediction task is to estimate the position sequence over the future T time steps,

$$\hat{\mathbf{Y}}_{t+1:t+T} = [\hat{\mathbf{p}}_{t+1}, \hat{\mathbf{p}}_{t+2}, \dots, \hat{\mathbf{p}}_{t+T}], \quad (3)$$

where

$$\hat{\mathbf{p}}_{t+h} = [\hat{x}_{t+h}, \hat{y}_{t+h}], \quad h = 1, 2, \dots, T. \quad (4)$$

The ground-truth future trajectory is denoted $\mathbf{Y}_{t+1:t+T}$. The training objective is to minimize the error between the predicted and ground-truth positions over the multi-step prediction and to provide reliable future motion states for the subsequent risk assessment.

3.3. Coordinate Transformation and Causal CV Prior

The raw AIS positions are expressed in longitude and latitude. To facilitate the computation of distance errors and motion increments, the longitude/latitude coordinates are converted to local planar coordinates. Let the raw longitude/latitude be (λ_i, φ_i) , the reference longitude/latitude be (λ_0, φ_0) , and the Earth radius be R ; then, the local metric coordinates are

$$x_i = R(\lambda_i - \lambda_0) \frac{\pi}{180} \cos\left(\varphi_0 \frac{\pi}{180}\right), \quad (5)$$

$$y_i = R(\varphi_i - \varphi_0) \frac{\pi}{180}. \quad (6)$$

Short-term ship motion has a certain inertia, so a constant-velocity model is introduced as a causal physical prior. Let the position at the current time be $\mathbf{p}_t$ and let $\bar{\mathbf{v}}_t^{(k)}$ be the mean of the last k backward displacement velocities in the observation window. The value of $k \in \{1, 2, 3, 5\}$ is selected using validation ADE only; both study areas select $k = 1$. The CV prior position at future step h is

$$\mathbf{p}_{t+h}^{CV} = \mathbf{p}_t + h \Delta t \bar{\mathbf{v}}_t^{(k)}. \quad (7)$$

CV extrapolation follows near-uniform motion but accumulates errors during turns, speed changes, and channel-constrained maneuvers. The network predicts a nonlinear correction to the CV reference:

$$\hat{\mathbf{p}}_{t+h}^{loc} = \mathbf{p}_{t+h}^{CV} + s_r \Delta \hat{\mathbf{p}}_{t+h}, \quad (8)$$

where $\Delta \hat{\mathbf{p}}_{t+h}$ is a dimensionless residual output and $s_r = 1000$ m is a fixed output-coordinate scale, not a learned parameter. The scale was fixed before model training so that typical residual targets remain numerically close to unit order; it changes the optimization conditioning but not the physical definition of the reconstructed trajectory. Section 4.4 reports retraining results for 500, 1000, and 1500 m.

3.4. Channel-Constrained NRRMWOA-BiLSTM Predictor

BiLSTM is adopted as the local trajectory prediction backbone. The LSTM unit mitigates the vanishing-gradient problem of the traditional RNN through a gating structure, and its computation is

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_f), \quad (9)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_i), \quad (10)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C[\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_C), \quad (11)$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t, \quad (12)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_o), \quad (13)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t), \quad (14)$$

where $\mathbf{f}_t$, $\mathbf{i}_t$, and $\mathbf{o}_t$ are the forget, input, and output gate vectors, respectively; $\mathbf{C}_t$ is the cell-state vector, $\mathbf{h}_t$ is the hidden-state vector, and $\odot$ denotes element-wise multiplication. A BiLSTM consists of a forward and backward LSTM, and its output is

$$\mathbf{h}_t = [\vec{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t]. \quad (15)$$

Figure 6 shows the BiLSTM structure; the blue arrows indicate backward propagation.

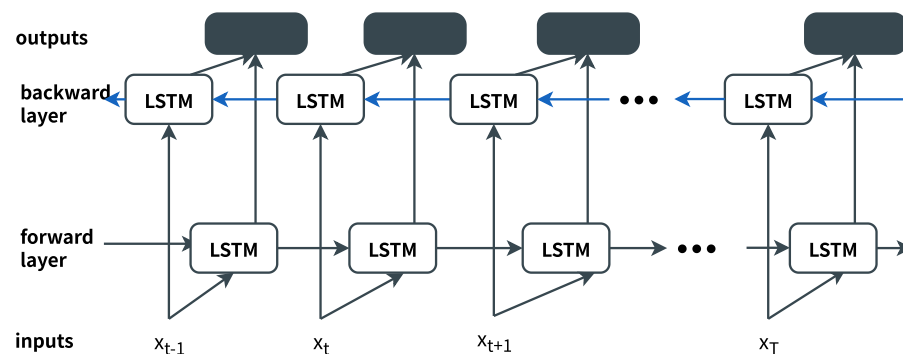


Figure 6. BiLSTM structure.

In the implemented predictor, the 11-dimensional temporal input first passes through a linear projection, LayerNorm, SiLU activation, and dropout. A two-layer bidirectional LSTM then uses $H_d/2$ hidden units in each direction and concatenates the terminal forward and backward states into an H_d -dimensional encoding. The channel vector is encoded by a two-layer MLP with a 64-dimensional hidden representation. After concatenation, a linear–LayerNorm–SiLU layer initializes a two-layer GRU horizon decoder. The decoder is driven by learned 32-dimensional horizon embeddings, and an H_d-96-2 SiLU MLP outputs the two-dimensional residual at each future step. Thus, $H_d = 160$ and 224 correspond to 80 and 112 recurrent units per BiLSTM direction in Study Areas 1 and 2, respectively.

In restricted waters, ship motion is related to both the historical state and the direction of the route ahead. A channel look-ahead representation is therefore introduced. For each route cluster, all retained AIS positions are projected onto the first principal-component axis. A cubic polynomial is fitted by least squares to the transverse coordinate as a function of the principal-axis coordinate. The fitted curve is sampled at 200 ordered points and represented as a piecewise-linear polyline parameterized by cumulative arclength. Let $\Gamma_j(u)$ denote this representation for candidate centerline j , and let $\mathbf{f}_t$ denote the unit heading estimated causally from the observation window. The nearest centerline parameter and its distance are

$$u_j^* = \arg \min_u \|\mathbf{p}_t - \Gamma_j(u)\|_2^2, \quad d_j = \|\mathbf{p}_t - \Gamma_j(u_j^*)\|_2. \quad (16)$$

where $\tau_j(u_j^*)$ denotes the centerline tangent; the direction-compatible centerline is selected by

$$j^* = \arg \min_j \left[d_j + \kappa \left(1 - \left| \tau_j(u_j^*)^\top \mathbf{f}_t \right| \right) \right], \quad \kappa = 500 \text{ m}. \quad (17)$$

The direction sign is $\sigma_t = \text{sign}[\tau_{j^*}(u_{j^*}^*)^\top \mathbf{f}_t]$. Six points are sampled from the selected centerline at signed arclength offsets $\mathcal{A} = \{0, 400, 800, 1200, 1800, 2600\}$ m:

$$\mathbf{c}_a = \Gamma_{j^*}(u_{j^*}^* + \sigma_t a), \quad a \in \mathcal{A}. \quad (18)$$

Their local coordinates, the route one-hot code, the normalized current position, and the centerline match score form the map vector $\mathbf{m}$. The local prediction feature is obtained by concatenating the AIS temporal and map encodings:

$$\mathbf{h}^{fuse} = [\mathbf{h}^{own}; \mathbf{h}^{map}]. \quad (19)$$

Based on $\mathbf{h}^{fuse}$, the decoder outputs the future multi-step residual trajectory, and the local prediction is reconstructed through CV-residual composition. Figure 7 illustrates this process; black circles mark the observed history, and the black line marks the matched centerline.

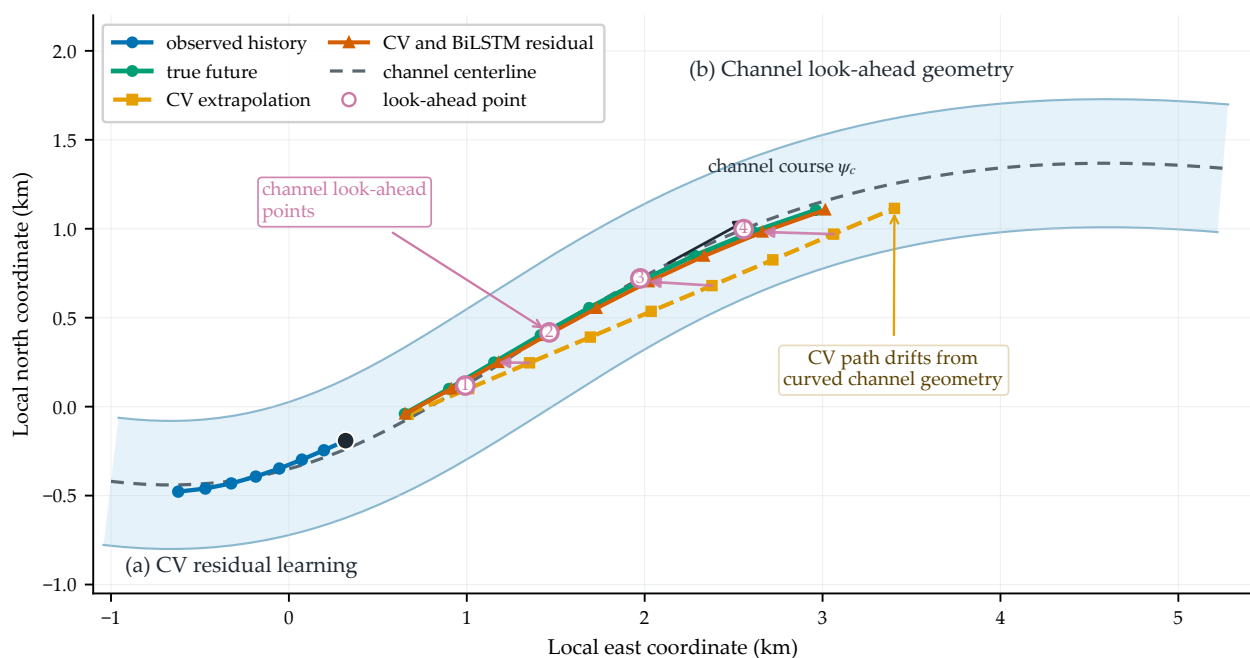


Figure 7. CV-residual prediction and channel look-ahead geometry.

Turning consistency is imposed at both the accumulated-maneuver and stepwise levels. For a future local trajectory $\{\mathbf{p}_h\}_{h=1}^T$ with $\mathbf{p}_0 = \mathbf{0}$, define

$$\mathbf{d}_h = \mathbf{p}_h - \mathbf{p}_{h-1}, \quad \theta_h = \text{atan2}(d_{h,y}, d_{h,x}), \quad (20)$$

Each heading increment is wrapped to $[-\pi, \pi]$:

$$\Delta\theta_h = \text{atan2}[\sin(\theta_h - \theta_{h-1}), \cos(\theta_h - \theta_{h-1})], \quad h = 2, \dots, T. \quad (21)$$

The cumulative turn is $\Theta_1 = 0$ and $\Theta_h = \sum_{j=2}^h \Delta\theta_j$ for $h \geq 2$. With hats denoting predicted quantities, the two turn-related losses are

$$\mathcal{L}_{turn} = \frac{1}{T} \sum_{h=1}^T \ell_{0.04}(\hat{\Theta}_h - \Theta_h), \quad (22)$$

$$\mathcal{L}_{rate} = \frac{1}{T-1} \sum_{h=2}^T \ell_{0.04}(\Delta\hat{\theta}_h - \Delta\theta_h), \quad (23)$$

where $\ell_\delta(e) = e^2/(2\delta)$ for $|e| < \delta$ and $\ell_\delta(e) = |e| - \delta/2$ otherwise. Both terms are averaged over the batch using the same turn-sample weights as the displacement loss. Their contributions to the trajectory objective are $\lambda_{\text{turn}}\mathcal{L}_{\text{turn}}$ and $\lambda_{\text{rate}}\mathcal{L}_{\text{rate}}$. The complete trajectory loss is

$$\mathcal{L}_{\text{traj}} = \mathcal{L}_{\text{pos}} + 0.35\mathcal{L}_{\text{FDE}} + 0.15\mathcal{L}_{\text{vel}} + \lambda_{\text{turn}}\mathcal{L}_{\text{turn}} + \lambda_{\text{rate}}\mathcal{L}_{\text{rate}}, \quad (24)$$

where $\mathcal{L}_{\text{pos}}$ is the horizon-weighted position loss, $\mathcal{L}_{\text{FDE}}$ is the last-step position loss, and $\mathcal{L}_{\text{vel}}$ compares consecutive displacement vectors. The first three terms use SmoothL1 loss after the physical scaling described above.

3.5. Improved Whale Optimization for Hyperparameter Search

BiLSTM prediction changes with the learning rate, hidden dimension, dropout, weight decay, and turn loss weight. NRRMWOA searches these hyperparameters jointly. The Whale Optimization Algorithm (WOA) proposed by Mirjalili and Lewis [49] simulates the encircling-prey, spiral-feeding, and random-search behaviors of humpback whales.

Let the position vector of the i -th whale individual be $\mathbf{X}_i(t)$ and the current best individual be $\mathbf{X}^*(t)$. The update in the encircling-prey stage is

$$\mathbf{D} = |\mathbf{C} \odot \mathbf{X}^*(t) - \mathbf{X}_i(t)|, \quad (25)$$

$$\mathbf{X}_i(t+1) = \mathbf{X}^*(t) - \mathbf{A} \odot \mathbf{D}, \quad (26)$$

where

$$\mathbf{A} = 2a\mathbf{r}_1 - a, \quad \mathbf{C} = 2\mathbf{r}_2, \quad (27)$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors on the interval $[0, 1]$. The spiral-update stage is

$$\mathbf{X}_i(t+1) = \mathbf{D}' e^{bl} \cos(2\pi l) + \mathbf{X}^*(t), \quad (28)$$

$$\mathbf{D}' = |\mathbf{X}^*(t) - \mathbf{X}_i(t)|, \quad (29)$$

where $l \in [-1, 1]$ is a random number and b is a constant defining the shape of the logarithmic spiral. When the algorithm performs a random search, the individual is guided by a random whale $\mathbf{X}_{\text{rand}}(t)$:

$$\mathbf{D} = |\mathbf{C} \odot \mathbf{X}_{\text{rand}}(t) - \mathbf{X}_i(t)|, \quad (30)$$

$$\mathbf{X}_i(t+1) = \mathbf{X}_{\text{rand}}(t) - \mathbf{A} \odot \mathbf{D}. \quad (31)$$

NRRMWOA augments WOA with a nonlinear convergence factor, opposition-based learning, and random perturbation [24]. The nonlinear convergence factor is

$$a(t) = 2 \left[1 - \left(\frac{t}{T_{\max} - 1} \right)^\rho \right], \quad (32)$$

where $T_{\max}$ is the maximum number of iterations, $t = 0, 1, \dots, T_{\max} - 1$, and ρ is the nonlinear control parameter. For a candidate solution in the normalized search space, the opposition-based individual is

$$\mathbf{X}_i^{\text{op}} = 1 - \mathbf{X}_i. \quad (33)$$

Random perturbation is used to maintain population diversity:

$$\tilde{\mathbf{X}}_i = \text{clip}(\mathbf{X}_i + \epsilon, 0, 1), \quad (34)$$

$$\epsilon_j \sim \mathcal{N}(0, \sigma_n^2), \quad (35)$$

where n is the number of completed candidate evaluations. Perturbation is invoked with probability 0.20; 35% of the dimensions are selected independently, with at least one dimension perturbed, and $\sigma_n = 0.08[1 - 0.65(n/N)]$ for an evaluation budget N . Opposition evaluation is invoked with probability 0.35, and the nonlinear exponent is $\rho = 2$. All coordinates are clipped to the normalized interval after updating.

Each whale individual corresponds to a set of model hyperparameters:

$$\Theta = \{\log_{10} \eta, H_d, p_{drop}, \log_{10} \lambda_{wd}, \lambda_{turn}, \lambda_{rate}, \lambda_s, \lambda_{os}, B\}, \quad (36)$$

where η is the learning rate, H_d is the hidden dimension, p_{drop} is the dropout rate, λ_{wd} is the weight decay, λ_{turn} and λ_{rate} are turn-related loss weights, λ_s is the turn-sample weighting coefficient, λ_{os} is the turn-sample oversampling strength, and B is the batch size.

The fitness function integrates the overall validation error and the turning-sample error:

$$\mathcal{J} = 0.75 \text{ ADE} + 0.25 \text{ ADE}_{\text{turn} > 30^\circ}, \quad (37)$$

where $\text{ADE}_{\text{turn} > 30^\circ}$ is the average displacement error for validation samples whose future turning angle exceeds 30° . Configuration selection uses the ADE-based objective above, while FDE is reported as a final-horizon metric. The flowchart of the improved Whale Optimization Algorithm is shown in Figure 8.

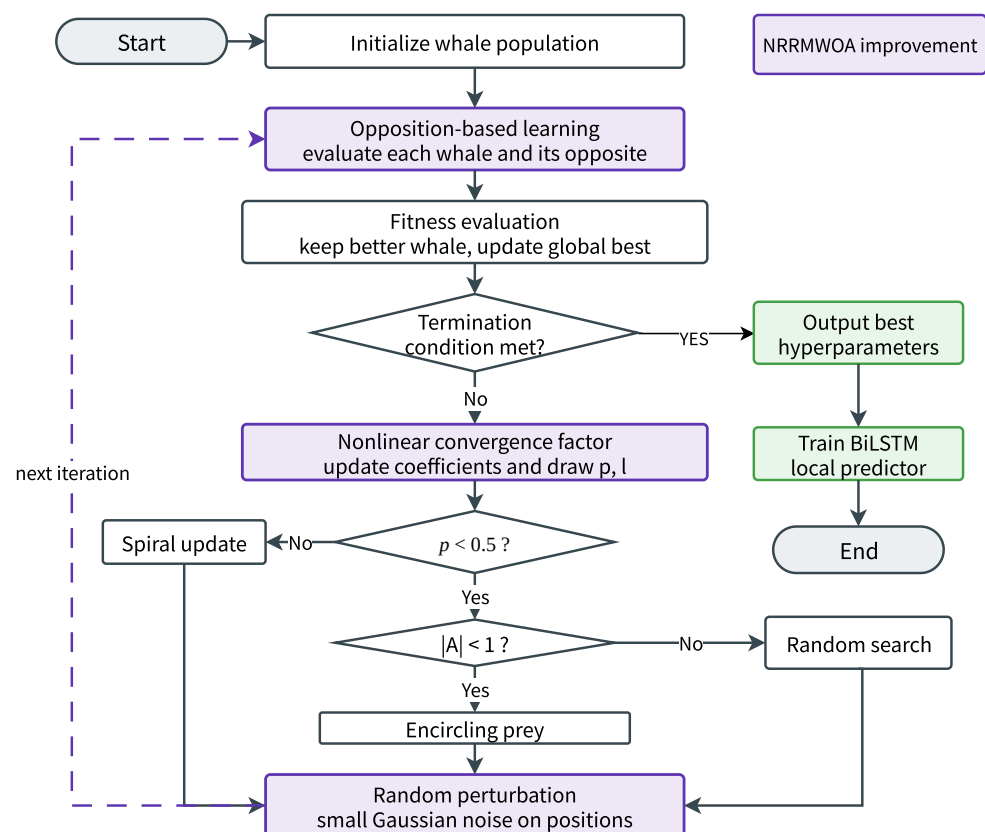


Figure 8. NRRMWOA search procedure.

Combining the search space and fitness design above, Table 1 summarizes the common search ranges and the configurations used in the two study areas.

Table 1. NRRMWOA search space and selected configurations.

Hyperparameter	Symbol	Search Range	Study Area 1 Selected	Study Area 2 Selected
Learning rate	η	$[2.5 \times 10^{-4}, 4.5 \times 10^{-3}]$	1.839×10^{-3}	1.226×10^{-3}
Hidden dimension	H_d	$\{64, 96, 128, 160, 192, 224, 256\}$	160	224
Dropout rate	p_{drop}	$[0.00, 0.30]$	0.000	0.000
Weight decay	λ_{wd}	$[1.0 \times 10^{-6}, 1.0 \times 10^{-3}]$	1.00×10^{-6}	9.53×10^{-4}
Turn loss weight	λ_{turn}	$[0.00, 0.30]$	0.249	0.020
Turn-rate loss weight	λ_{rate}	$[0.00, 0.12]$	0.000	0.000
Turn-sample weight	λ_s	$[0.00, 2.00]$	1.972	0.912
Turn oversampling strength	λ_{os}	$[0.00, 1.00]$	0.021	0.160
Batch size	B	$\{256, 384, 512\}$	256	384

The bounds in Table 1 were set before the searches. Learning rate and weight decay use logarithmic ranges spanning commonly stable AdamW settings for this architecture. The hidden dimensions and batch sizes cover the feasible capacity–memory combinations under the 8 GB graphics processing unit (GPU) limit. Dropout includes the unregularized endpoint and moderate regularization, while the turn-related ranges include their neutral settings and cap the largest weights so that turning samples cannot dominate the total objective. Attention-head count and turn-intent weight are excluded because they are inactive in the selected BiLSTM backbone. Candidate selection uses validation data, and the selected configuration is fixed before final three-seed training.

3.6. Global–Local Expert Fusion

The local NRRMWOA-BiLSTM expert models short-term motion variation, whereas the global channel expert represents the longer-horizon traffic trend along the channel. Their outputs are combined at each horizon. Let the global- and local-expert predictions be $\hat{\mathbf{p}}_{t+h}^g$ and $\hat{\mathbf{p}}_{t+h}^l$, respectively; the fused result is

$$\hat{\mathbf{p}}_{t+h}^{fuse} = w_h \hat{\mathbf{p}}_{t+h}^g + (1 - w_h) \hat{\mathbf{p}}_{t+h}^l, \quad (38)$$

where $w_h \in [0, 1]$ is the weight of the global expert at the h -th prediction step. By adjusting w_h step by step, the model balances short-term local maneuvers against the overall channel constraint.

For each study area, w_h is selected independently at each horizon by minimizing validation ADE over a fixed grid and is then used for test evaluation. The global-expert weights for horizons 1–10 are $[0.05, 0.05, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.40, 0.40]$ in Study Area 1 and $[0.10, 0.25, 0.35, 0.40, 0.45, 0.50, 0.50, 0.55, 0.55, 0.60]$ in Study Area 2; the local weights are their complements.

3.7. APF-PCR Hybrid Collision Risk Assessment

To further use the trajectory prediction result for collision risk assessment, an APF-PCR hybrid risk head is constructed. Let the relative position and relative velocity of the own ship and the target ship at the h -th prediction step be

$$\mathbf{r}_h = \mathbf{q}_h - \mathbf{p}_h, \quad \mathbf{u}_h = \mathbf{v}_h^q - \mathbf{v}_h^p. \quad (39)$$

The time to CPA (TCPA) is

$$\text{TCPA}_h = -\frac{\mathbf{r}_h^\top \mathbf{u}_h}{\max(\|\mathbf{u}_h\|_2^2, \epsilon)}, \quad (40)$$

The corresponding distance to CPA (DCPA) is

$$\text{DCPA}_h = \|\mathbf{r}_h + \max(\text{TCPA}_h, 0)\mathbf{u}_h\|_2. \quad (41)$$

To express the spatio-temporal proximity risk, a time-gating term is defined:

$$g_h = \begin{cases} \exp[-\max(\text{TCPA}_h, 0)/T_0], & \text{TCPA}_h \geq 0 \\ 0, & \text{TCPA}_h < 0 \end{cases} \quad (42)$$

where the time-decay scale is fixed at $T_0 = 15$ prediction steps. The APF potential-field term is

$$U_h^{APF} = \frac{g_h}{\text{DCPA}_h + D_{dom}} + \beta \frac{1}{d_h + D_{dom}}, \quad (43)$$

where $\epsilon = 10^{-6} \text{ m}^2$ per prediction-step² prevents division by a vanishing relative-speed magnitude, $d_h = \|\mathbf{r}_h\|_2$, the own-ship domain scale is $D_{dom} = 6L_{own} + 200 \text{ m}$, and the distance-potential weight is $\beta = 0.3$. These constants are fixed before risk-head training. The length-dependent term scales clearance by vessel size, the 200 m offset avoids very small domains for short vessels, and $\beta < 1$ gives DCPA greater influence than instantaneous distance. Their one-factor sensitivity is reported in Section 4.4; operational thresholds can be calibrated for a specific authority and waterway. PCR characterizes the potential collision risk within the ship domain, and the larger value of the two directional risks is taken as the symmetric risk:

$$\text{PCR}_{sym} = \max[\text{PCR}(\text{own} \rightarrow \text{tgt}), \text{PCR}(\text{tgt} \rightarrow \text{own})]. \quad (44)$$

Finally, the risk feature vector at each prediction step can be written as

$$\mathbf{r}_h^{risk} = [U_h^{pred}, U_h^{CV}, \text{PCR}_{sym,h}, d_h, (d_h + D_{dom})^{-1}, \text{DCPA}_h, g_h, \mathbf{c}^{cur}, \mathbf{n}^{APF}, v_{own}, L_{own}, h/T], \quad (45)$$

where U_h^{pred} is the APF feature computed from the predicted trajectory of this paper, U_h^{CV} is the APF feature computed from the CV-extrapolated trajectory, $\mathbf{c}^{cur}$ is the current encounter geometric feature, and $\mathbf{n}^{APF}$ is the neighbor potential-field aggregation feature.

The risk feature sequence is processed by a lightweight recurrent head:

$$\mathbf{h}_{1:T} = \text{GRU}_{risk}(\text{Norm}(\mathbf{r}_{1:T}^{risk})), \quad \mathbf{o}_h = \mathbf{W}_o \text{SiLU}(\text{LN}(\mathbf{h}_h)) + \mathbf{b}_o, \quad (46)$$

$$\hat{\mathbf{p}}_h^{risk} = \text{softmax}(\mathbf{o}_h). \quad (47)$$

where $\text{Norm}(\cdot)$ applies feature-wise means and standard deviations estimated from valid training steps. The recurrent head is a one-layer unidirectional GRU with 32 hidden units, followed by LayerNorm, SiLU, and a three-unit output layer. It is trained with Adam at a learning rate of 2×10^{-3} , weight decay 10^{-4} , batch size 1024, and cross-entropy label smoothing of 0.02. Class weights are the inverse square roots of the training-class counts, normalized to unit mean and clipped to $[0.25, 6]$. Training is limited to 120 epochs with validation encounter-level macro-F1 checkpoint selection and an early-stopping patience of 20; gradients are clipped at a norm of 5. The PCR-derived categories are safe for $\text{PCR} < 0.01$, alert for $0.01 \leq \text{PCR} < 0.4$, and dangerous for $\text{PCR} \geq 0.4$. For alert+ probability calibration, a single temperature T is fitted separately in each study area by minimizing binary negative log-likelihood on valid validation steps. If p is the raw alert+ probability, the calibrated probability is $\sigma(\text{logit}(p)/T)$. The fitted validation temperature is then applied to test probabilities. This monotone transformation preserves the ROC and precision–recall rankings and leaves the 0.5 alert+ decision unchanged. Independent geometric labels and feature-

ablation experiments are also used to examine the contributions of predicted-trajectory and geometric-proximity features to the risk judgments.

4. Experiments

4.1. Experimental Settings

Experiments are conducted on the cleaned AIS data from both study areas. Each sample uses 10 historical steps to predict 10 future steps, and the input features include local coordinates x, y , speed, course, and ship length. In each area, complete own-vessel voyages are ordered by time and assigned to chronological training, validation, and test periods before sliding-window construction. A 19 min embargo, equal to the span of a complete 10-step-history plus 10-step-future window, is imposed at both temporal boundaries. Windows that cross a boundary are discarded, and a complete own-vessel voyage is assigned to only one subset. Table 2 reports the resulting time ranges, MMSI counts, voyage counts, and window counts.

Table 2. Data partitions.

Study Area	Subset	Complete-Window Time Range	Windows	Own MMSIs	Own Voyages
Study Area 1	Training	2 January 2025 00:01–17 September 2025 10:39	42,000	1697	3111
	Validation	17 September 2025 13:13–26 November 2025 13:51	11,993	644	909
	Test	26 November 2025 14:21–30 December 2025 21:52	5990	356	456
Study Area 2	Training	30 December 2024 00:07–4 September 2025 08:50	42,000	1546	3660
	Validation	4 September 2025 12:04–16 November 2025 06:29	11,982	613	1017
	Test	16 November 2025 06:49–30 December 2025 23:05	5991	381	543

Seventeen boundary-crossing windows are removed in Study Area 1. In Study Area 2, 27 boundary-crossing windows are removed and no cross-subset-voyage windows occur. The retained test sets contain 56,554 and 54,117 valid per-horizon risk labels, respectively. Table 3 separates the raw-message, cleaned-voyage, and supervised-window populations.

Table 3. Cleaning audit.

Population	Study Area 1	Study Area 2
Raw valid AIS messages	1,470,794	647,573
Raw MMSIs	3846	3566
Gap-segmented candidate voyages	43,456	20,959
Cleaned 1 min voyages	17,909	13,953
Cleaned MMSIs	3707	3171
Own-vessel voyages used for windows	4476	5220
Final supervised windows	59,983	59,973

Table 4 compares vessel-type distributions before and after cleaning. The largest absolute change is 1.17 percentage points in Study Area 1 and 6.64 percentage points in Study Area 2. Future-turn, acceleration, encounter, and risk strata are calculated from complete 20 min windows after resampling and supervised-window construction. Table 5 reports the resulting supervised population, which includes sharp turns, rapid speed changes, crossing and head-on geometries, and the minority dangerous class.

Table 4. Vessel types before and after cleaning.

Vessel Type	Study Area 1 (%)		Study Area 2 (%)	
	Raw	Cleaned	Raw	Cleaned
Cargo	35.99	36.90	54.54	61.18
Tanker	50.55	51.71	9.93	11.13
Passenger	1.25	1.13	3.03	2.84
Fishing	4.89	4.37	7.99	7.54
Tug/towing	2.42	2.40	4.21	4.42
Sailing/pleasure	1.92	1.29	15.23	9.11
Other/service	1.72	1.62	3.45	2.68
Unknown	1.27	0.57	1.63	1.10

Table 5. Final-window composition.

Dimension	Stratum	Study Area 1, <i>n</i> (%)	Study Area 2, <i>n</i> (%)
Future turn angle (°)	0–20	45,190 (75.34)	49,482 (82.51)
	20–40	9356 (15.60)	7663 (12.78)
	40–60	4116 (6.86)	1195 (1.99)
	60–90	1028 (1.71)	587 (0.98)
	90–180	293 (0.49)	1046 (1.74)
Maximum acceleration (m/s ²)	<0.02	48,364 (80.63)	44,269 (73.81)
	0.02–0.05	10,829 (18.05)	11,976 (19.97)
	0.05–0.10	429 (0.72)	2400 (4.00)
	≥0.10	361 (0.60)	1328 (2.21)
Encounter geometry	Same-course/overtaking	20,982 (34.98)	23,326 (38.89)
	Crossing	26,348 (43.93)	23,107 (38.53)
	Head-on	12,653 (21.09)	13,540 (22.58)
Maximum valid risk	Safe	52,284 (87.16)	48,551 (80.95)
	Alert	6821 (11.37)	7156 (11.93)
	Dangerous	303 (0.51)	3347 (5.58)
	Unavailable	575 (0.96)	919 (1.53)

Table 6 gives additional retained-data statistics. Voyage length is reported in three complementary forms: resampled points, elapsed minutes, and traveled distance. The mean number of supervised windows is calculated over unique own-ship MMSIs across all three subsets.

Table 6. Retained-data profile.

Study Area	Own Voyages	Points Per Voyage	Duration (min)	Path Length (km)	Windows Per Own MMSI
Study Area 1	4476	54.99	53.99	17.81	27.45
Study Area 2	5220	50.35	49.35	15.14	28.93

The unique own-vessel populations comprise 2185 MMSIs in Study Area 1 and 2073 in Study Area 2. Their AIS-type distributions are, respectively: tanker 58.95%/13.46%, cargo 35.24%/68.36%, tug or towing 2.15%/3.38%, fishing 1.42%/4.20%, passenger 0.78%/3.47%, other or service 0.96%/2.07%, sailing or pleasure 0.18%/4.10%, and unknown 0.32%/0.96%. Table 7 gives the pairwise overlap audit. The temporal intervals and own-vessel voyages do not overlap. The same MMSI can reappear in a later nonoverlapping period when the vessel undertakes a different voyage; these repeated identifiers are not provided to the prediction models as input features.

Table 7. Partition-overlap audit.

Study Area	Subset Pair	Separation (min)	Shared Own Voyages	Shared MMSIs	Temporal Overlap
Study Area 1	Training–validation	154	0	319	No
	Validation–test	30	0	92	No
	Training–test	101,022	0	159	No
Study Area 2	Training–validation	194	0	284	No
	Validation–test	20	0	112	No
	Training–test	104,999	0	156	No

Within each study area, all methods use the same supervised split and test set. Each learned comparison model is trained independently with random seeds 0, 1, and 2, and its result is reported as the mean and sample standard deviation of the three test runs. The proposed result is the score of the frozen per-sample ensemble. The Transformer, graph, physics-informed, and interaction-aware variants receive the same causal CV-residual anchor and training-derived channel context as the local predictor. Their common training settings are a maximum of 60 epochs, an early-stopping patience of 12, AdamW optimization, and checkpoint selection on validation data using $0.75 \text{ ADE} + 0.25 \text{ ADE}_{\text{turn} > 30^\circ}$. The tuner variants use the same BiLSTM prediction interface and validation criterion.

The NRRMWOA search uses a population size of 8, an eight-iteration ceiling, and a hard budget of 64 candidate evaluations. The nonlinear convergence-factor exponent ρ is 2.0, the opposition-based learning probability is 0.35, and the random-perturbation probability is 0.20, with a perturbation standard deviation of 0.08. Each candidate is trained for 18 epochs on a fixed 60% training subset, with early stopping based on the validation metrics. Opposition candidates count toward the same budget. This search determines the Study Area 1 configuration. Study Area 2 uses the region-specific configuration selected by the same procedure for its temporal experiment. In each area, the selected configuration is fixed for final training with seeds 0, 1, and 2; the search is not repeated for each seed. The final local trajectory expert is trained for at most 60 epochs with an early-stopping patience of 12. AdamW is combined with cosine annealing over the 60-epoch ceiling, with $\eta_{\min} = 0.05\eta$, and gradients are clipped at a norm of 5.

The experiments were run under Windows with Python 3.13.0, PyTorch 2.6.0, CUDA 12.4, NumPy 2.1.3, pandas 2.2.3, SciPy 1.14.1, and scikit-learn 1.6.1 on an NVIDIA GeForce RTX 4060 Laptop GPU with 8 GB memory (NVIDIA Corporation, Santa Clara, CA, USA). The Study Area 1 64-evaluation search required 1789.8 s; the earlier search that supplied the frozen Study Area 2 configuration required 1662.7 s. Final three-seed local-model training took 312.3 s in Study Area 1 (104.1 s per model) and 219.0 s in Study Area 2 (73.0 s per model). The two risk heads took 29.5 and 24.4 s, respectively. For the Study Area 1 local model, batched inference over all 5990 test samples required 0.0442 ± 0.0027 s over ten repetitions, with a peak allocated GPU memory of 121.7 MB.

Trajectory prediction is evaluated using average displacement error (ADE) and final displacement error (FDE). ADE averages the position error over the prediction window, whereas FDE measures the error at the final step. Risk assessment is evaluated as both an alert+ binary task and a three-level classification task, where alert+ combines the alert and dangerous classes. The binary metrics are AUC and average precision (AP); the three-level metrics are accuracy, macro-F1, alert recall, and alert precision. Probability calibration before and after validation-only temperature scaling is measured by the Brier score and expected calibration error (ECE). For ECE, the predicted alert+ probabilities are divided into ten equal-width intervals over $[0, 1]$, and the confidence–event-rate gap is averaged across intervals in proportion to their sample counts.

4.2. Comparison with Representative Recent Prediction Models

The comparison spans four model families relevant to restricted-water prediction: recurrent sequence models, Transformer-based models, graph or interaction-aware models, and physics-informed models. The evaluated set comprises a channel-conditioned LSTM, Optuna-BiLSTM and standard WOA-BiLSTM tuner variants, a G-Trans-style maritime Transformer [16], a maritime ST-GCN [37], a finite-difference physics-informed BiLSTM [44], and an SNIINet-style interaction-aware network [39].

All learned comparison models follow the common input, split, and validation procedure specified above. Each model is trained independently with seeds 0, 1, and 2 and receives the same training-derived channel context and causal CV-residual anchor. The G-Trans, ST-GCN, physics-informed, and SNIINet variants retain the defining mechanisms of their respective model families within this common data interface. Figure 9 shows all eight methods under straight-sailing, channel-branch, and large-turn conditions in Study Area 1.

Table 8 reports the aggregate prediction errors in Study Area 1.

Table 8. Prediction results in Study Area 1.

Method	ADE (m)	FDE (m)
Proposed dual-expert system	103.70	217.16
Channel-conditioned LSTM	111.39 ± 0.53	234.96 ± 1.77
Optuna-BiLSTM	110.69 ± 0.33	234.04 ± 0.65
WOA-BiLSTM	109.69 ± 0.27	231.38 ± 0.67
G-Trans [16]	111.08 ± 3.02	235.18 ± 5.42
PINN [44]	111.91 ± 0.30	235.77 ± 0.79
SNIINet [39]	116.38 ± 0.87	241.37 ± 1.77
ST-GCN [37]	121.47 ± 0.90	253.28 ± 1.75

Study Area 1: selected multi-model trajectory cases

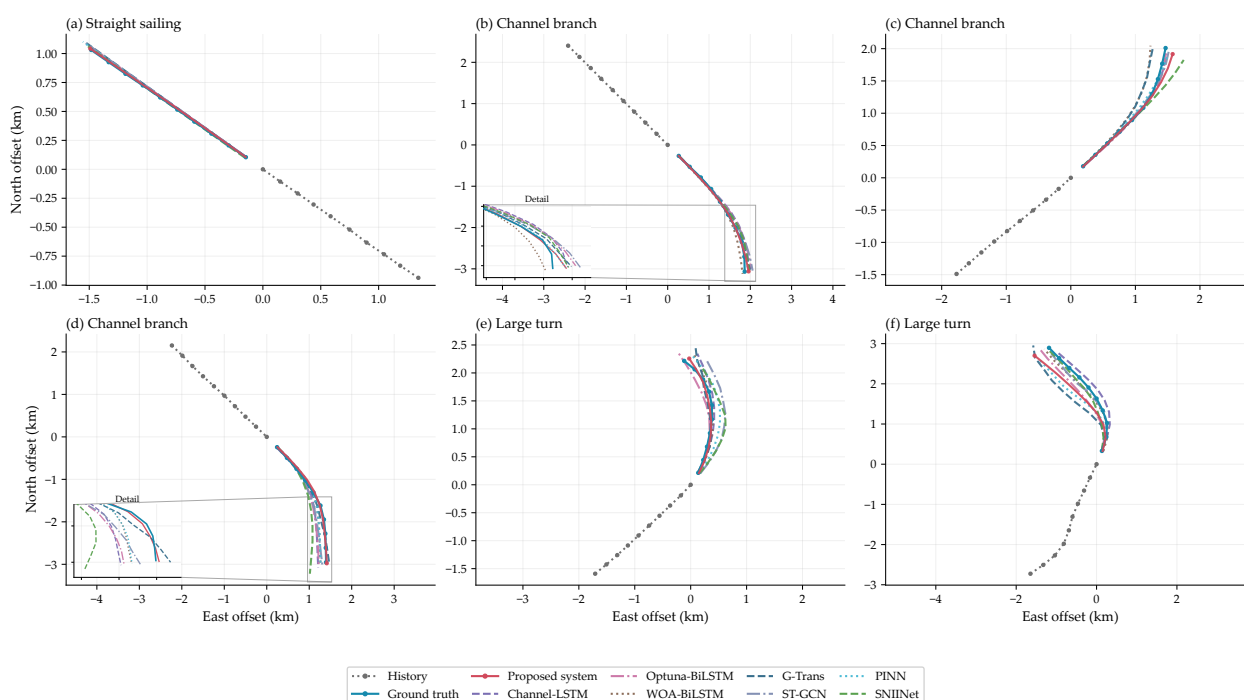


Figure 9. Representative forecasts in Study Area 1.

Figure 10 summarizes Study Area 1: fused-model ADE/FDE is 103.70/217.16 m versus 109.69/231.38 m for WOA-BiLSTM (5.47%/6.15% lower). Purple and yellow denote ADE and FDE throughout; light blue marks the proposed system in aggregate plots.

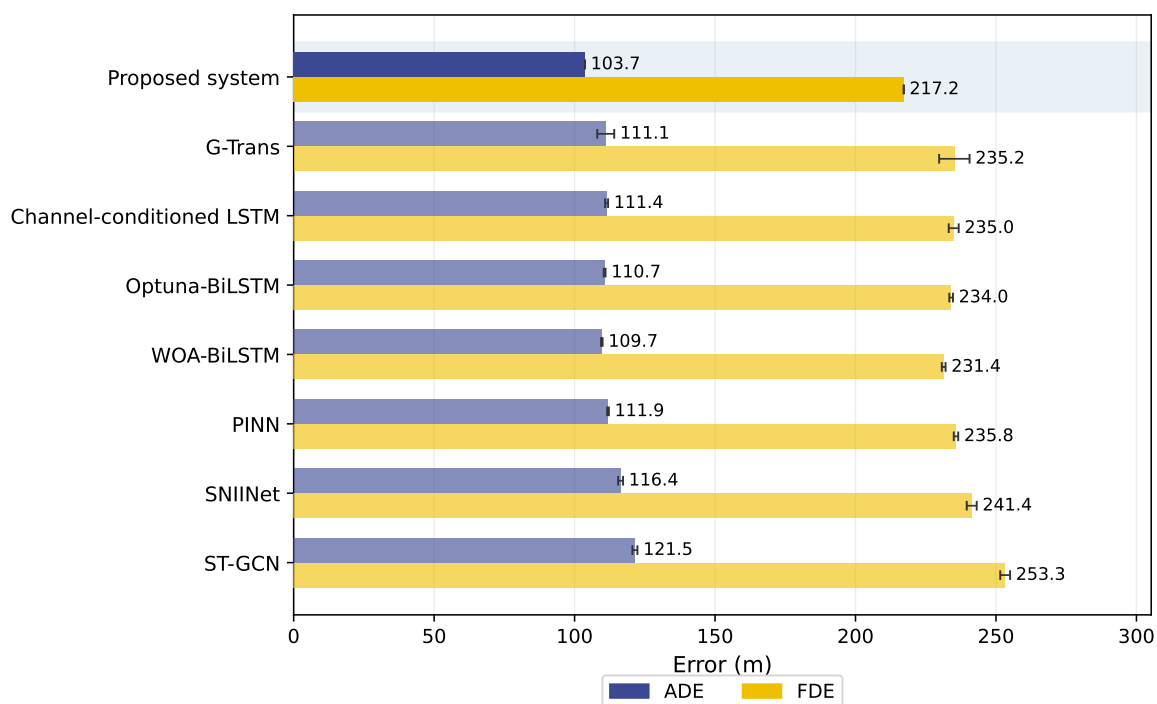


Figure 10. ADE and FDE in Study Area 1.

Figure 11 presents representative Study Area 2 forecasts for straight sailing and gentle, moderate, and large turns. The corresponding aggregate errors are reported in Figure 12 and Table 9.

Study Area 2: selected multi-model trajectory cases

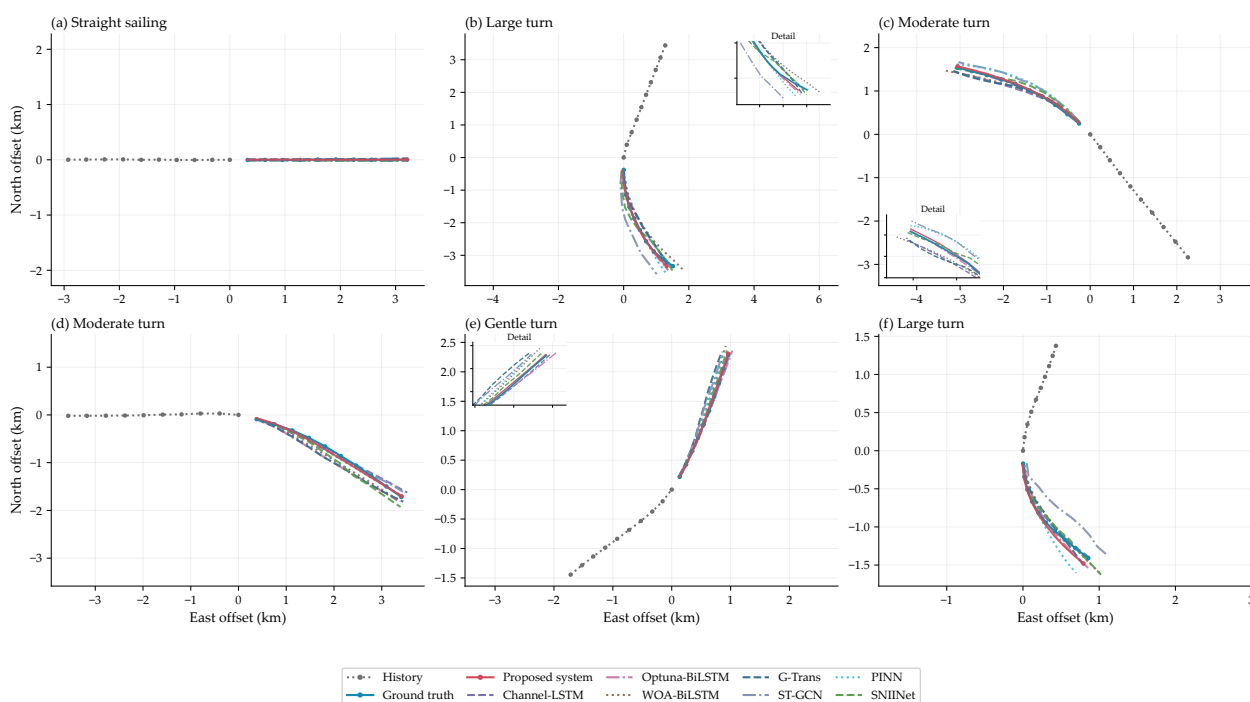


Figure 11. Representative forecasts in Study Area 2.

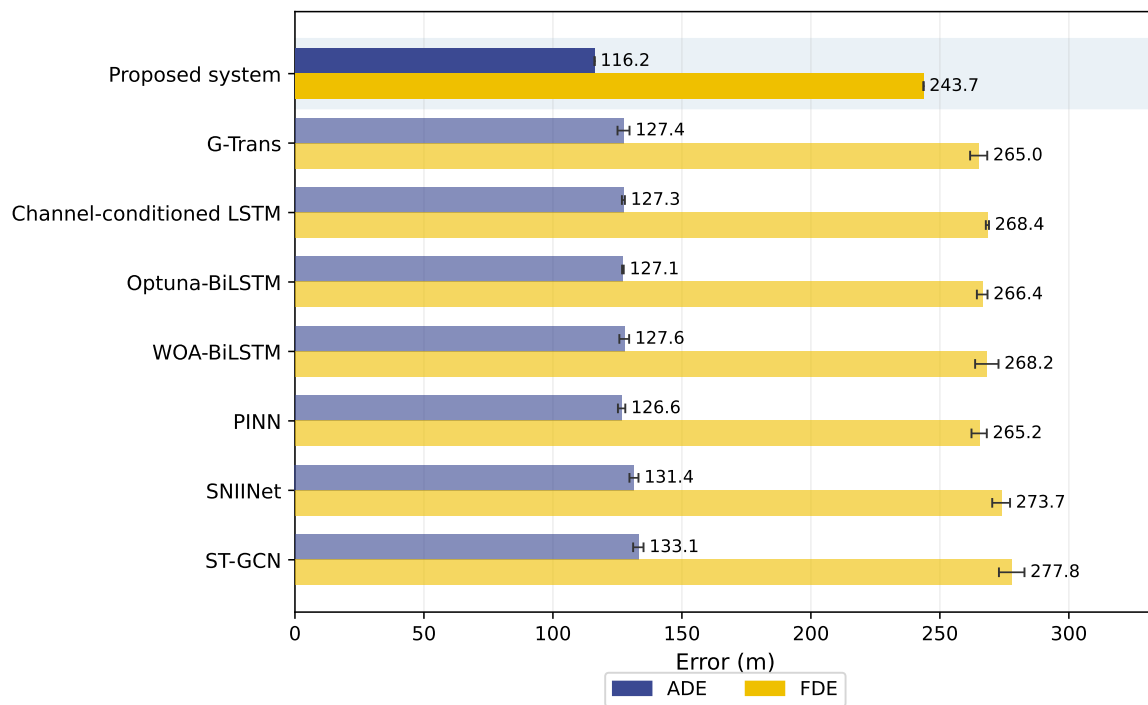


Figure 12. ADE and FDE in Study Area 2.

Table 9. Prediction results in Study Area 2.

Method	ADE (m)	FDE (m)
Proposed dual-expert system	116.15	243.67
Channel-conditioned LSTM	127.32 ± 0.58	268.42 ± 0.61
Optuna-BiLSTM	127.13 ± 0.32	266.40 ± 2.06
WOA-BiLSTM	127.65 ± 1.90	268.19 ± 4.53
G-Trans [16]	127.37 ± 2.33	265.04 ± 3.32
PINN [44]	126.63 ± 1.43	265.22 ± 2.97
SNIINet [39]	131.42 ± 1.75	273.75 ± 3.45
ST-GCN [37]	133.11 ± 2.01	277.84 ± 4.95

In Study Area 2, the proposed dual-expert fusion obtains 116.15/243.67 m. The best comparison ADE is 126.63 m from PINN, and the best comparison FDE is 265.04 m from G-Trans. The corresponding reductions are 8.27% and 8.07%, respectively.

Statistical comparisons preserve dependence among overlapping windows by re-sampling complete own-vessel voyages. For each baseline, the paired window error is first averaged over its independently trained seeds 0, 1, and 2, matching the mean values reported in the comparison tables. A 20,000-repeat voyage-clustered bootstrap provides the 95% confidence interval (CI) of the mean error reduction, and a two-sided Wilcoxon signed-rank test is applied to voyage-level mean differences. Holm correction is performed jointly over all model–metric comparisons. Table 10 reports the strongest baseline in each area.

Table 10. Paired significance tests.

Study Area	Metric/Reference	Reduction (m)	Clustered 95% CI (m)	Holm p
Study Area 1	ADE/WOA-BiLSTM	6.00	[4.78, 7.24]	6.00×10^{-18}
Study Area 1	FDE/WOA-BiLSTM	14.22	[11.25, 17.23]	3.08×10^{-17}
Study Area 2	ADE/PINN	10.48	[8.71, 12.38]	2.90×10^{-32}
Study Area 2	FDE/G-Trans	21.38	[17.19, 25.85]	1.09×10^{-25}

All four paired reductions are statistically significant.

The test periods contain a subset of vessels whose MMSIs are absent from the corresponding training period. Table 11 reports results for this subset. MMSI is used to define the subset but is not supplied to the model.

Table 11. Results for unseen test vessels.

Study Area	Unseen MMSIs	Voyages	Windows	ADE (m)	FDE (m)
Study Area 1	197	241	3296	105.98	226.91
Study Area 2	225	286	3438	107.54	227.09

Because ADE and FDE alone hide when the forecasts diverge, Figure 13 additionally reports displacement error at every minute of the ten-minute prediction range.

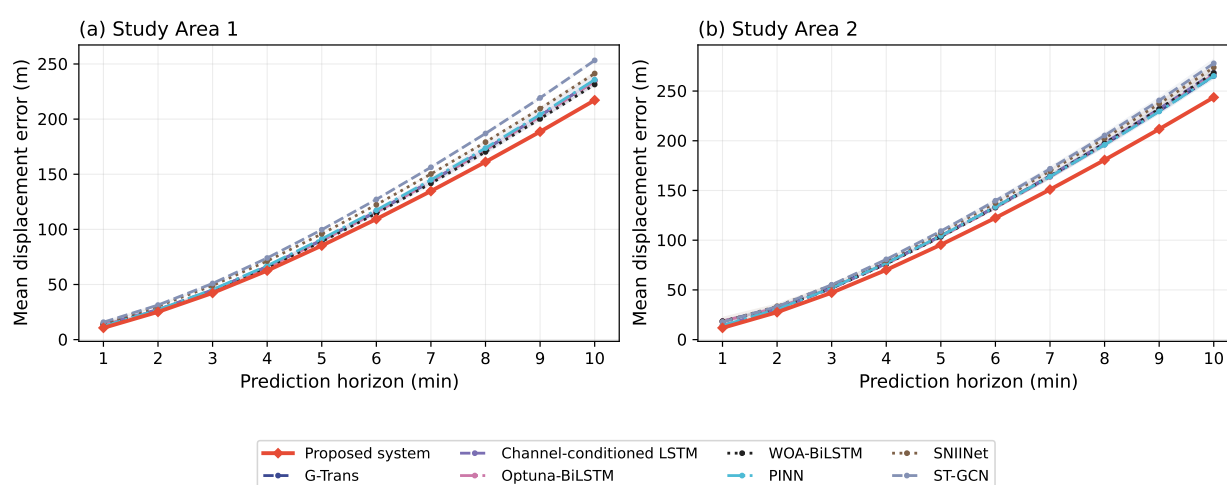


Figure 13. Per-horizon displacement errors.

The parameter counts in Table 12 are per independently trained model. In each study area, the final fusion executes three local NRRMWOA-BiLSTM models and ten frozen global channel-penalized LSTM experts. Each global expert has 1,732,805 registered parameters (1,037,762 trajectory-related). The corresponding 13-expert systems contain 19,570,235 and 21,262,907 registered parameter instances in Study Areas 1 and 2, respectively. For the three tuner-based BiLSTM methods, the common search variables and ranges follow Table 1.

Table 12. Comparison-model settings.

Method	Architecture; Trainable Parameters (Area 1/2)	Optimizer; Learning Rate/Weight Decay (Area 1/2)	Batch; Maximum Epochs/Patience	Search or Model Selection
NRRMWOA-BiLSTM	2-layer BiLSTM encoder, $h = 160/224$; 2-layer GRU horizon decoder; CV-residual/channel encoders; 747,395/1,311,619 per local model	AdamW + cosine; configurations in Table 1; $\eta_{\min} = 0.05\eta$	256/384; 60/12	Area 1: population-8, 64-evaluation search; Area 2: frozen region-specific configuration

Table 12. Cont.

Method	Architecture; Trainable Parameters (Area 1/2)	Optimizer; Learning Rate/Weight Decay (Area 1/2)	Batch; Maximum Epochs/Patience	Search or Model Selection
Channel-conditioned LSTM	2-layer unidirectional LSTM, $h = 160/224$; identical 2-layer GRU residual decoder and channel interface; 886,659/1,549,187	AdamW; Area 1: $\eta = 1.22 \times 10^{-3}$, $\lambda_{wd} = 2.38 \times 10^{-6}$; Area 2: $\eta = 1.23 \times 10^{-3}$, $\lambda_{wd} = 9.53 \times 10^{-4}$	384/384; 60/12	Fixed recurrent control; validation-only checkpoint
Optuna-BiLSTM	2-layer BiLSTM, $h = 160/160$; identical 2-layer GRU residual decoder/channel interface; 747,395/747,587	AdamW; Area 1: $\eta = 1.42 \times 10^{-3}$, $\lambda_{wd} = 3.39 \times 10^{-6}$; Area 2: $\eta = 1.82 \times 10^{-3}$, $\lambda_{wd} = 2.03 \times 10^{-6}$	384/512; 60/12	16/12 trials, each trained for 18 epochs on 60% of the training set (Area 1/2); validation-only selection
WOA-BiLSTM	2-layer BiLSTM, $h = 160/160$; identical 2-layer GRU residual decoder/channel interface; 747,395/747,587	AdamW; Area 1: $\eta = 1.22 \times 10^{-3}$, $\lambda_{wd} = 2.38 \times 10^{-6}$; Area 2: $\eta = 1.22 \times 10^{-3}$, $\lambda_{wd} = 2.38 \times 10^{-6}$	384; 60/12	Standard WOA with 16/12 trained candidates (Area 1/2), 18 candidate epochs, 60% training subset
G-Trans [16]	GRU block ($h = 128$), 2 Transformer layers, 4 heads, cross-attention decoder; 465,442/465,826	AdamW + cosine; $8 \times 10^{-4}/10^{-4}$ in both areas	384; 60/12	Fixed architecture; validation score 0.75 ADE + 0.25 ADE _{turn>30°}
PINN [44]	2-layer BiLSTM ($h = 128$), position/velocity heads, Heun residual; 380,228/380,612	AdamW + cosine; $8 \times 10^{-4}/10^{-4}$	384; 60/12	Fixed; kinematic/velocity loss weights 0.10/0.10; validation-only checkpoint
SNIINet [39]	2-layer own-ship LSTM ($h = 128$), neighbor GRU, position-attention, information-fusion, and global-attention blocks; 615,458/615,842	AdamW + cosine; $8 \times 10^{-4}/10^{-4}$	256; 60/12	Fixed interaction-aware architecture; validation-only checkpoint
ST-GCN [37]	2 graph-convolution layers ($h = 96$), 2-layer GRU ($h = 128$), up to 6 neighbors; 342,498/342,882	AdamW + cosine; $8 \times 10^{-4}/10^{-4}$	256; 60/12	Dynamic-distance graph; fixed architecture; validation-only checkpoint

4.3. Ablation Study

The ablation study is conducted in both study areas. Study Area 1 results are shown in Figure 14 and Table 13, and Study Area 2 results are shown in Figure 15 and Table 14.

The CV-residual variant changes ADE from 173.18 to 171.71 m. The larger change occurs after adding channel features: ADE falls to 112.08 m (34.73%), FDE from 405.73 to 234.36 m (42.23%), and turn30 ADE from 409.02 to 215.56 m (47.30%). Within this ablation, route identity, centerline matching, and look-ahead geometry account for most of the measured improvement.

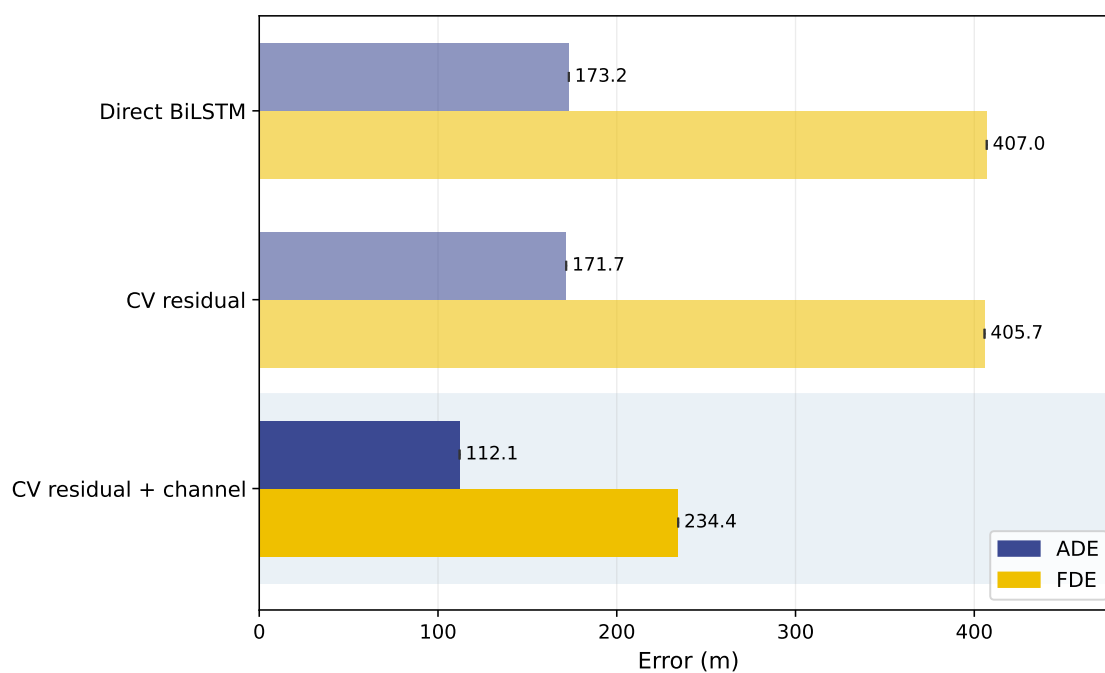


Figure 14. Graphical ablation results in Study Area 1.

Table 13. Numerical ablation results in Study Area 1.

Model Variant	ADE (m)	FDE (m)	turn30 ADE (m)	turn60 ADE (m)
Direct BiLSTM	173.18	406.96	410.90	476.37
CV residual (w/o channel)	171.71	405.73	409.02	470.77
CV residual and channel	112.08	234.36	215.56	294.41

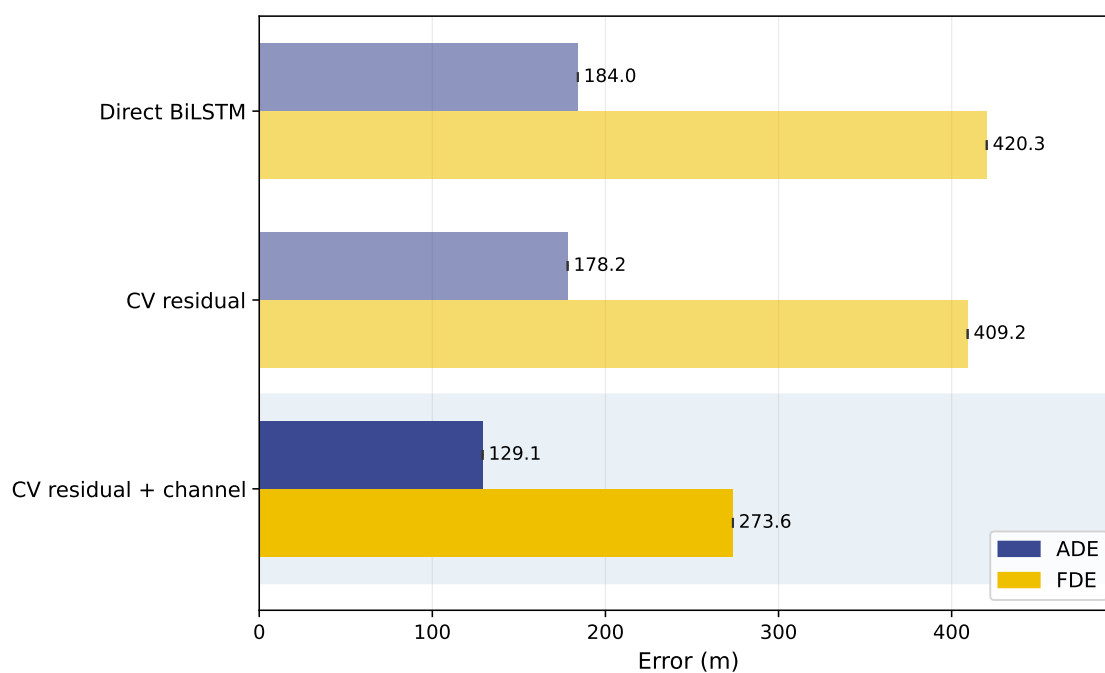


Figure 15. Graphical ablation results in Study Area 2.

Table 14. Numerical ablation results in Study Area 2.

Model Variant	ADE (m)	FDE (m)	turn30 ADE (m)	turn60 ADE (m)
Direct BiLSTM	184.04	420.26	569.22	780.99
CV residual (w/o channel)	178.15	409.24	591.76	730.22
CV residual and channel	129.05	273.64	334.24	720.78

The Study Area 2 ablation leads to the same main interpretation. The CV residual changes ADE/FDE from 184.04/420.26 m to 178.15/409.24 m. Adding channel geometry then reduces ADE to 129.05 m (27.56%) and FDE to 273.64 m (33.14%) relative to the CV-residual variant; turn30 ADE decreases by 43.52%.

4.4. Parameter Sensitivity

The residual output scale is evaluated at 500, 1000, and 1500 m. The 1000 m row uses the final checkpoints; each alternative is retrained from scratch with the same hyperparameters, random seeds 0, 1, and 2, validation checkpointing, and downstream system definition. Table 15 shows that all ADE and FDE changes remain below 1% of the corresponding 1000 m result. The adopted 1000 m configuration remains unchanged in the formal evaluation.

Table 15. Residual-scale sensitivity.

s_r (m)	Study Area 1		Study Area 2	
	ADE (m)	FDE (m)	ADE (m)	FDE (m)
500	103.54	216.68	115.98	243.34
1000	103.70	217.16	116.15	243.67
1500	104.34	217.98	116.64	243.95

Both study areas report the corresponding dual-expert fused system.

Centerline-coordinate sensitivity is evaluated with the trained trajectory models frozen. A fixed standard normal error realization is added to the six local look-ahead coordinate pairs and scaled to standard deviations of 25–200 m; route identity and the centerline match score remain unchanged. Table 16 reports changes relative to the zero-error evaluation on the same device. At a 100 m coordinate error, system ADE increases by 0.79 m in Study Area 1 and 1.22 m in Study Area 2. The larger 200 m perturbation increases ADE by 4.04 and 4.65 m, respectively.

Table 16. Centerline-error sensitivity.

σ (m)	Study Area 1 Change (m)		Study Area 2 Change (m)	
	ADE	FDE	ADE	FDE
0	0.00	0.00	0.00	0.00
25	−0.01	0.03	0.07	0.12
50	0.11	0.42	0.29	0.54
100	0.79	2.34	1.22	2.48
200	4.04	10.98	4.65	9.58

The APF/TCPA constants are examined by changing one factor at a time while keeping the trajectory predictor, feature normalizer, and risk head fixed. The tested values are centered on the adopted defaults. Table 17 shows that varying ϵ from 10^{-8} to 10^{-4} produces no change at the reported precision. Across the wider β , domain, and time-decay

alternatives, AUC remains between 0.9859 and 0.9888 in Study Area 1 and between 0.9926 and 0.9930 in Study Area 2.

Table 17. Risk-parameter sensitivity.

Condition	Study Area 1		Study Area 2	
	AUC	AP	AUC	AP
Default	0.9882	0.9061	0.9930	0.9643
$\epsilon = 10^{-8}$	0.9882	0.9061	0.9930	0.9643
$\epsilon = 10^{-4}$	0.9882	0.9061	0.9930	0.9643
$\beta = 0.15$	0.9885	0.9067	0.9930	0.9644
$\beta = 0.60$	0.9859	0.9030	0.9926	0.9633
$D_{dom} = 4L + 200$ m	0.9869	0.9042	0.9929	0.9636
$D_{dom} = 8L + 200$ m	0.9888	0.9067	0.9930	0.9644
$T_0 = 10$	0.9874	0.9049	0.9930	0.9643
$T_0 = 20$	0.9878	0.9052	0.9930	0.9640

4.5. Collision Risk Assessment Results

The same APF, PCR, and hybrid-GRU risk scoring methods are evaluated on the alert+ task in both study areas. Their AUC and AP values are compared directly in Table 18.

Table 18. APF-PCR risk-assessment results.

Method	Study Area 1		Study Area 2	
	AUC	AP	AUC	AP
Pure APF	0.8471	0.3248	0.8713	0.5067
Pure PCR	0.9826	0.8576	0.9869	0.9375
Hybrid APF-PCR GRU	0.9882	0.9061	0.9930	0.9643

Pure PCR obtains higher AUC and AP than Pure APF in both study areas, and the hybrid head provides the highest AUC/AP pair in each dataset. Specifically, the hybrid head reaches 0.9882/0.9061 in Study Area 1 and 0.9930/0.9643 in Study Area 2. In Study Area 2, step-level accuracy and macro-F1 are 0.9592 and 0.8829, with alert+ recall/precision of 0.9160/0.8700. Encounter-level macro-F1 is 0.8791, and dangerous encounter recall/precision are 0.7500/0.9254.

Figure 16 plots the ROC, precision–recall, and reliability curves for the hybrid head on the same valid test steps used in Table 18. The raw reliability curves show systematic probability overestimation, which is reduced by temperatures fitted only on the validation partitions. Table 19 shows that temperature scaling reduces the Brier score from 0.0263 to 0.0231 and from 0.0245 to 0.0238 in Study Areas 1 and 2, respectively; ECE decreases from 0.0754 to 0.0200 and from 0.0503 to 0.0224. AUC and AP remain unchanged because the transformation is monotone. The original three-level argmax predictions and the confusion matrices below are unchanged.

Table 19. Validation-only temperature scaling.

Study Area	T	Raw Brier	Scaled Brier	Raw ECE	Scaled ECE
Study Area 1	0.5644	0.0263	0.0231	0.0754	0.0200
Study Area 2	0.6894	0.0245	0.0238	0.0503	0.0224

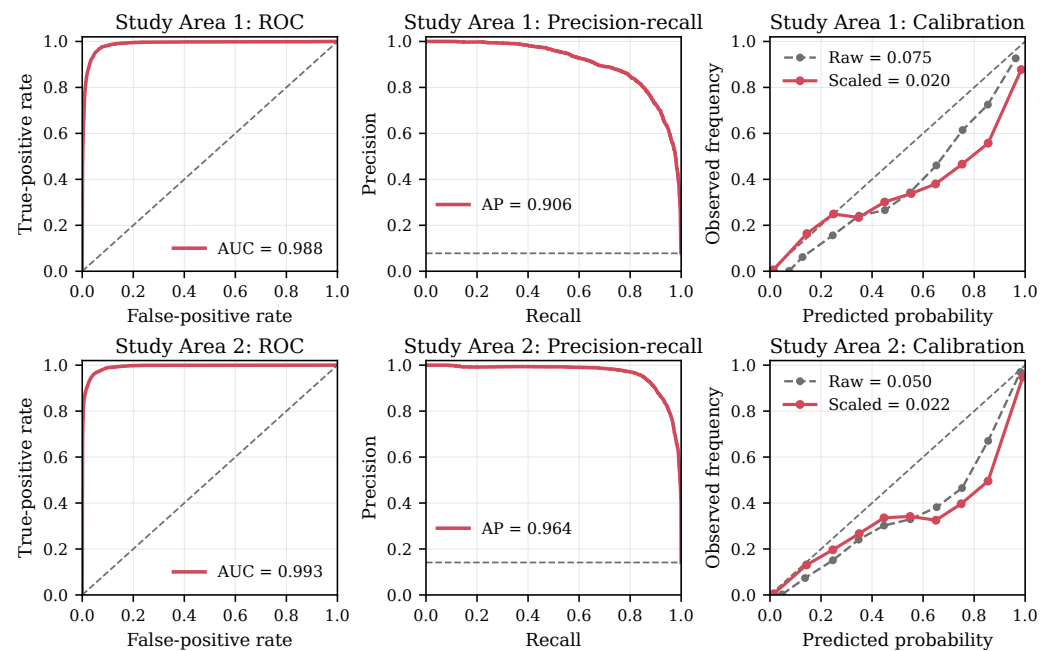


Figure 16. Alert+ discrimination and temperature calibration.

The confusion matrices of the per-sample maximum risk level are shown in Table 20. Accuracy and macro-F1 above are per-prediction-step statistics, whereas Table 20 aggregates each sample by its maximum risk level.

Table 20. Risk-level confusion matrices.

Study Area	True\Pred	Safe	Alert	Dangerous
Study Area 1	Safe	5129	80	0
	Alert	118	572	36
	Dangerous	0	0	7
Study Area 2	Safe	4820	95	0
	Alert	96	630	15
	Dangerous	8	54	186

In Study Area 1, 5129 of 5209 safe samples, 572 of 726 alert samples, and all seven dangerous samples are classified correctly. In Study Area 2, the corresponding diagonal counts are 4820 of 4915, 630 of 741, and 186 of 248.

In both study areas, the risk features are also evaluated with an independent geometric label whose construction does not use the PCR functional form. PCR-related and CV-geometric features are removed progressively, ending with predicted-trajectory potential-field and contextual features. Each feature subset is evaluated with an independently trained 32–32–3 MLP probe (Table 21).

Removing PCR-related inputs reduces AP under the PCR-derived label in both areas. With only predicted-trajectory potential-field and contextual features, the independent-label AUC/AP values are 0.9875/0.8159 in Study Area 1 and 0.9827/0.8040 in Study Area 2.

Table 21. Risk-feature robustness.

Study Area	Feature Subset	AUC (PCR)	AP (PCR)	AUC (Independent)	AP (Independent)
Study Area 1	All features	0.9829	0.8968	0.9716	0.5874
	w/o PCR	0.9676	0.7716	0.9877	0.8211
	w/o PCR and APF-CV	0.9682	0.7649	0.9878	0.8182
	PRED_ONLY	0.9657	0.7552	0.9875	0.8159
Study Area 2	All features	0.9918	0.9564	0.9704	0.6299
	w/o PCR	0.9795	0.9018	0.9847	0.8385
	w/o PCR and APF-CV	0.9805	0.9056	0.9846	0.8398
	PRED_ONLY	0.9788	0.9039	0.9827	0.8040

Notes: The all-feature rows differ slightly from Table 18 because every feature subset is retrained independently. The independent label is based on future geometric proximity and does not use the PCR functional form.

4.6. Case Study

Four own–target encounters are visualized for each study area. Each case is defined at the final observed time. The figures show the observed tracks, predicted own-ship trajectory, target-ship CV trajectory, ship-domain/APF field, and per-step risk evolution. The tables list both vessels' speed, course, and length, together with course difference, encounter type, minimum predicted separation, relative bearing, true and predicted risk levels, and maximum alert+ probability. Study Area 1 cases are summarized in Table 22 and Figures 17 and 18; Study Area 2 cases are summarized in Table 23 and Figures 19 and 20.

Table 22. Case-study encounters in Study Area 1.

No.	Ship	MMSI	Speed (kn)	Course (°)	Length (m)	Δ (°)	Encounter Type	<i>D</i> (m)	γ (°)	True/Pred	Max Alert+
1	Own Target	352004757	9.2	139	256	0	Overtaking	975	257	alert/danger	0.996
		256059000	17.3	139	339						
2	Own Target	538004833	4.9	356	333	35	Crossing	675	84	danger/danger	0.994
		256959000	11.3	321	94						
3	Own Target	354842000	14.3	10	305	103	Crossing	2463	356	alert/alert	0.987
		311000346	14.6	267	225						
4	Own Target	352901000	11.5	175	184	175	Head-on	863	327	alert/alert	0.981
		219626000	10.3	360	185						

Δ : course difference; *D*: minimum distance between the predicted own-ship trajectory and the target-ship CV trajectory over the prediction horizon; γ : relative bearing.

Cases 1 and 2 exhibit high-risk overtaking and crossing geometries, respectively. Case 1 has an alert ground-truth maximum and a dangerous predicted maximum, whereas case 2 is dangerous in both the ground truth and prediction. Their maximum predicted alert+ probabilities are 0.996 and 0.994.

In alert cases 3 and 4, the predicted own-ship trajectory approaches the target-ship path, but the spacing remains larger than in the dangerous cases. The two examples represent crossing and head-on geometries. Their alert+ probabilities rise during the first prediction steps and remain at the alert level over several consecutive steps.

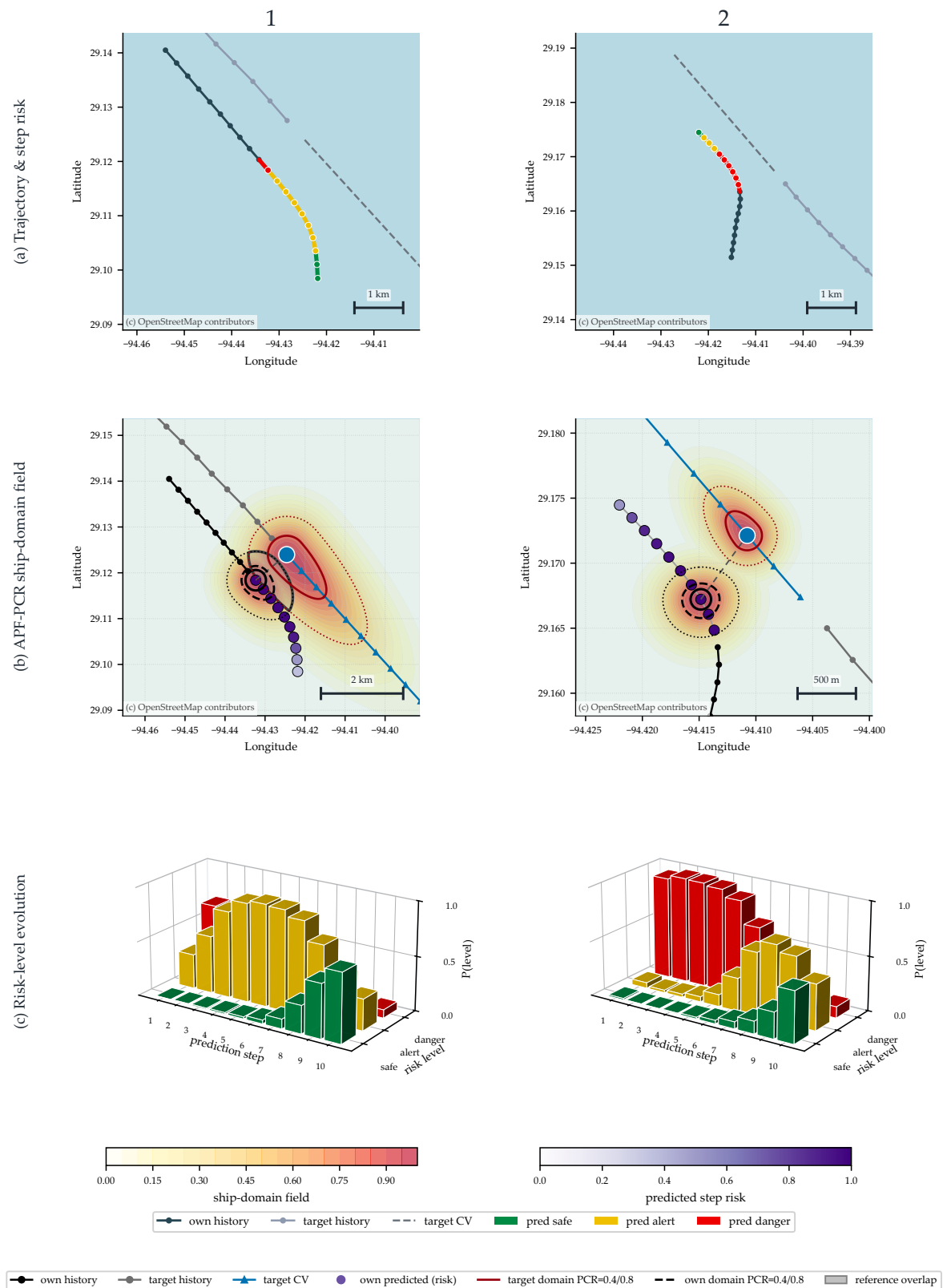


Figure 17. High-risk cases in Study Area 1.

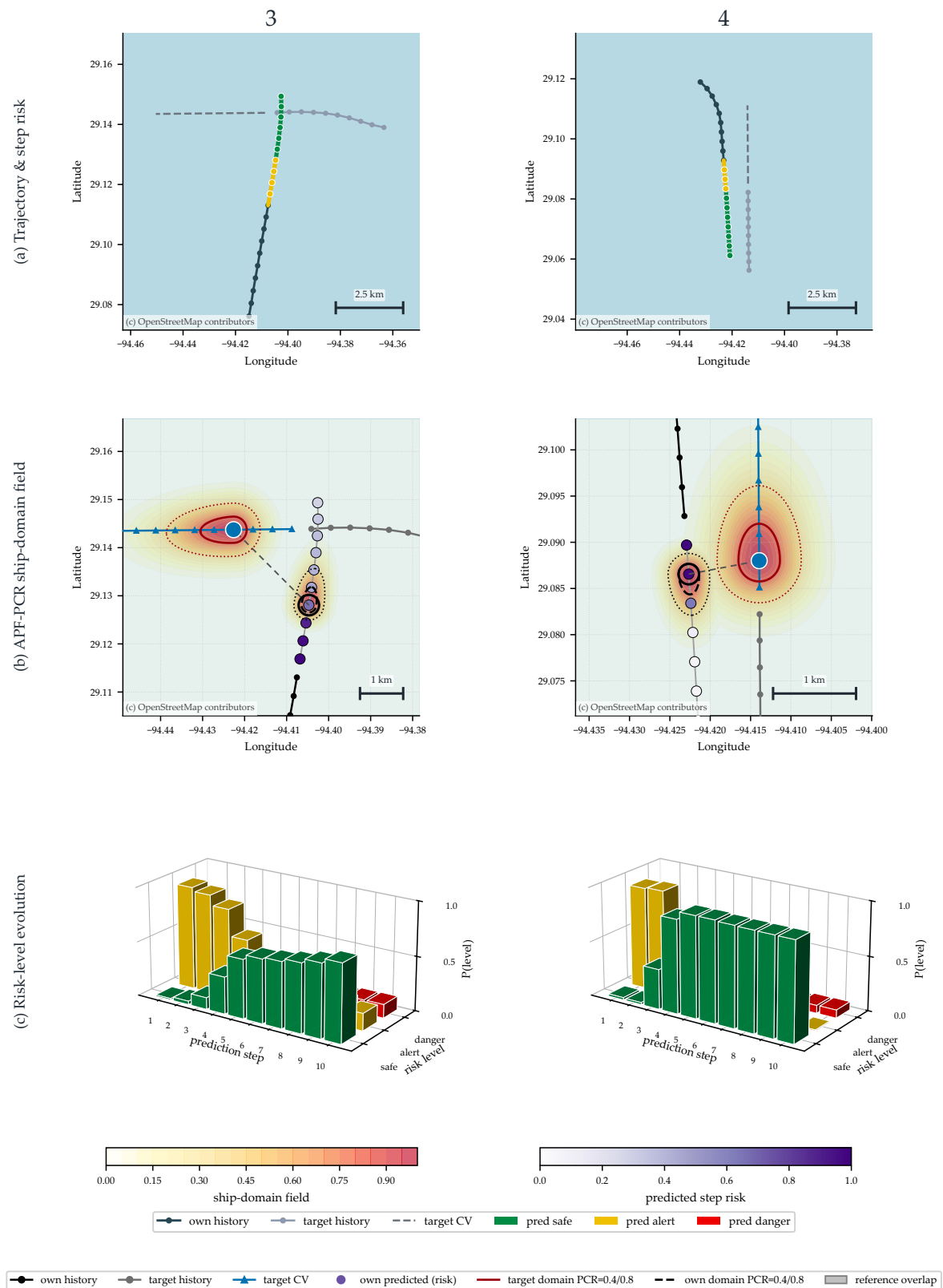


Figure 18. Alert cases in Study Area 1.

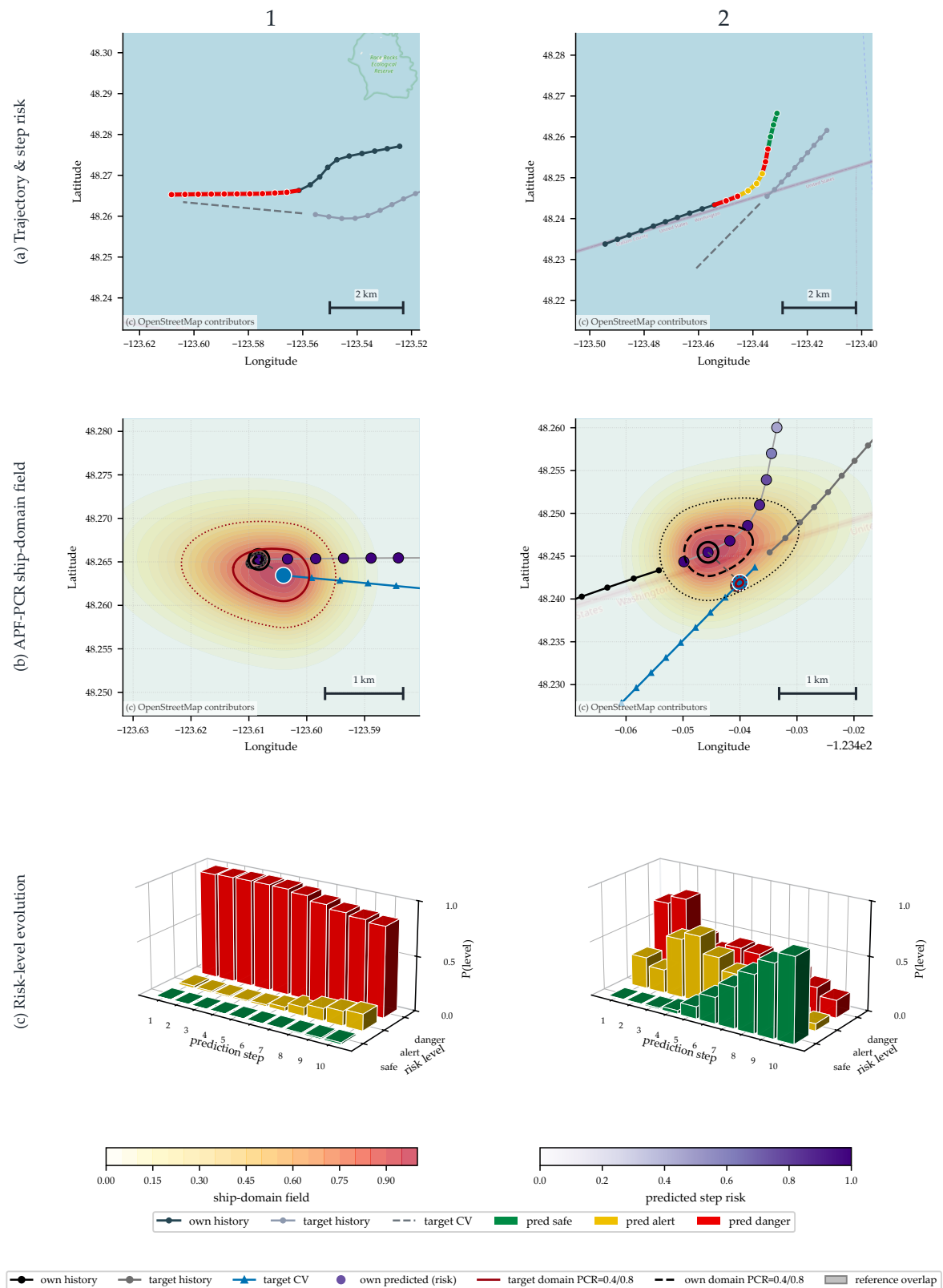


Figure 19. Dangerous crossing and head-on cases in Study Area 2.

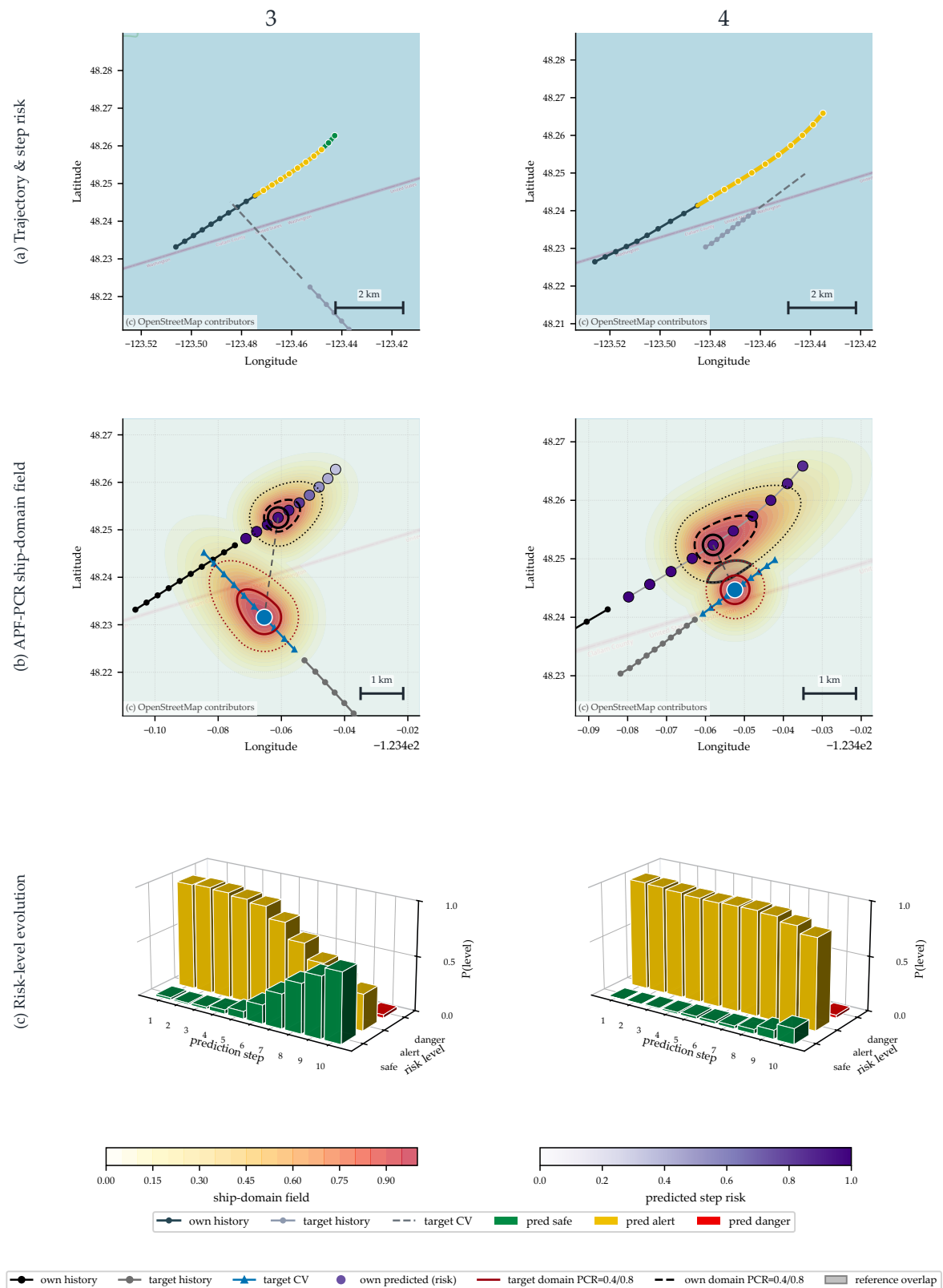


Figure 20. Alert crossing and overtaking cases in Study Area 2.

For Study Area 2, four cases are selected to cover crossing, head-on, and overtaking geometries and both dangerous and alert outcomes. Their predicted minimum distances range from 374 to 2358 m, providing spatial context for the assigned risk levels. The detailed encounter information is listed in Table 23.

Table 23. Case-study encounters in Study Area 2.

No.	Ship	MMSI	Speed (kn)	Course (°)	Length (m)	Δ (°)	Encounter Type	<i>D</i> (m)	γ (°)	True/Pred	Max Alert+
1	Own	316014107	11.0	244	32	35	Crossing	374	262	danger/danger	0.998
	Target	538010103	11.4	279	244						
2	Own	352001284	11.3	72	229	156	Head-on	572	9	danger/danger	0.995
	Target	316003483	8.8	228	31						
3	Own	357449000	10.0	57	245	101	Crossing	2358	92	alert/alert	0.987
	Target	636020252	12.0	316	293						
4	Own	563274700	15.3	60	261	7	Overtaking	948	37	alert/alert	0.998
	Target	371567000	6.1	53	200						

Δ : course difference; *D*: minimum distance between the predicted own-ship trajectory and the target-ship CV trajectory over the prediction horizon; γ : relative bearing.

In Study Area 2, cases 1 and 2 are correctly classified as dangerous crossing and head-on encounters, with maximum alert+ probabilities of 0.998 and 0.995. Cases 3 and 4 are correctly classified alert encounters with crossing and overtaking geometries. In both areas, the stepwise risk sequences vary with predicted proximity and encounter geometry.

5. Discussion

5.1. Overall Prediction Performance

Tables 8 and 9 report the lowest ADE and FDE for the proposed system in both study areas. In Study Area 1, the reductions relative to WOA-BiLSTM are 5.47% in ADE and 6.15% in FDE. In Study Area 2, the ADE reduction relative to PINN is 8.27%, and the FDE reduction relative to G-Trans is 8.07%. Figure 13 gives the corresponding error profiles over all ten prediction steps.

The voyage-clustered paired analysis yields statistically significant reductions over WOA-BiLSTM in Study Area 1 and over the metric-specific best comparisons in Study Area 2.

Recent maritime Transformer, graph, physics-informed, and interaction-aware models emphasize long-range sequence encoding, vessel interaction, or kinematic consistency [16,37,39,44]. Here, the CV reference and direction-compatible channel geometry enter the local prediction pathway. Its output is combined with a channel-trend expert, and the predicted future trajectory is then passed to the risk head. NRRMWOA configures the learning rate, hidden dimension, dropout, weight decay, and turning-sample loss weights. The reported system result includes CV-residual learning, channel-aware representation, automatic configuration, and trajectory-level fusion.

In a VTS setting, the forecasts may be used for traffic-trend monitoring and early warning. International Maritime Organization (IMO) guidance defines VTS as a government-implemented service that can interact with vessel traffic and respond to developing situations, but does not prescribe a universal trajectory-error threshold [50]. Inference over 5990 samples took 0.0442 ± 0.0027 s, below the one-minute sampling interval. At the 10 min horizon, the reported errors describe uncertainty in traffic evolution. The forecasts supplement current radar/AIS observations, local procedures, and operator or shipboard judgment.

5.2. Effects of CV Residual and Channel Constraints

Adding the channel look-ahead features reduces ADE, FDE, and most turning-sample errors. The main gains are associated with route identity, the centerline match score, and the direction of the six look-ahead points, particularly for the more common turning samples.

The channel-constrained representation and global–local fusion act at successive stages. Before decoding, the model encodes the selected route, centerline match score, normalized anchor position, six direction-compatible look-ahead points, and the AIS sequence. At the output stage, w_h combines the channel-consistent global trajectory with the local residual trajectory.

Residual scales from 500 to 1500 m produce small changes in system ADE and FDE. Centerline-coordinate errors up to 100 m have a smaller effect than the 200 m perturbation, which increases ADE by 4.04 m in Study Area 1 and 4.65 m in Study Area 2.

5.3. Interpretation of Collision Risk Results

The Hybrid APF-PCR GRU obtains higher AUC and AP than Pure APF and Pure PCR on the alert+ binary task. Pure APF reflects spatial proximity, while Pure PCR is closely aligned with the PCR-informed label. The recurrent head combines the predicted-trajectory potential field, PCR, DCPA/TCPA, and neighbor-aggregation features over the prediction horizon.

The alert+ AUC values of 0.9882 and 0.9930 are obtained against PCR-informed labels, and PCR variables also appear in the feature set. The effect of this functional overlap is evaluated using an independent geometric label and feature-subset ablations.

In Study Area 2, the independently retrained all-feature probe attains AUC/AP values of 0.9918/0.9564 on the PCR-informed label and 0.9704/0.6299 on the independent geometric label. Removing PCR features raises the independent-label AUC/AP to 0.9847/0.8385. The PRED_ONLY configuration, which retains predicted-trajectory potential-field features and neutral context variables, reaches 0.9827/0.8040 on the independent label. Thus, this feature subset remains discriminative without PCR-derived inputs. Future tests will compare its output with manually annotated expert labels.

5.4. Limitations

The framework is most likely to exhibit degraded performance when the future motion cannot be inferred reliably from the short historical window. Representative conditions include abrupt course alterations, rapid acceleration or deceleration, collision-avoidance maneuvers, and route-choice ambiguity near channel intersections or bifurcations. In these cases, the CV reference may continue the motion trend observed at the end of the historical sequence, while the channel-aware expert may favor the dominant traffic-flow direction. Their fusion can therefore produce a smooth and channel-consistent trajectory that does not fully reproduce an individual vessel's sudden maneuver. The deterministic output also represents only one future trajectory and cannot explicitly describe multiple plausible future routes. The risk case studies extrapolate target-ship motion with a CV model over the ten-minute horizon, so an abrupt target maneuver can alter the estimated CPA and ship-domain field. Wind, current, visibility, and explicit maneuvering commands are not included as model inputs.

The reported results use the stated AIS cleaning, discontinuity splitting, voyage segmentation, resampling, and interpolation protocol. Position noise affects displacement increments and the CV-residual reference, while noisy SOG and COG measurements distort the estimated motion state. Interpolation fills short gaps but may oversmooth genuine turning or acceleration. The Study Area 2 pipeline prohibits interpolation across

detected nonphysical discontinuities. The current experiments do not test explicit position perturbations, reduced sampling frequencies, or simulated packet loss.

Scalability in highly congested waterways also requires further examination. The current implementation limits the number of neighboring vessels retained for each prediction sample, which bounds the input dimension and computational cost but may omit temporarily distant, rapidly approaching vessels or incompletely represent multi-ship interactions. Prediction accuracy, inference time, and memory consumption should be evaluated at different traffic-density levels and compared with dynamic graph- or attention-based neighbor aggregation.

Dangerous samples remain relatively scarce, so the macro-F1 is affected by class imbalance. The risk labels are generated mainly from geometric rules and PCR and lack manual annotations from VTS experts or navigators. The independent-label and feature-ablation analyses reduce the risk of circular interpretation but do not replace evaluation against expert judgments.

6. Conclusions

This paper presents a channel-aware residual BiLSTM workflow for short-term trajectory prediction and collision risk assessment in restricted waters. The fused system records ADE/FDE values of 103.70/217.16 m in Study Area 1 and 116.15/243.67 m in Study Area 2. Voyage-level paired comparisons indicate lower ADE and FDE than the strongest metric-specific baselines in both areas. The hybrid risk head attains alert+ AUC values of 0.9882 and 0.9930 in Study Areas 1 and 2, while predicted-trajectory potential-field features remain informative under the independent geometric label.

Future work will examine multimodal futures in bifurcating channels, collision-avoidance maneuvers, and deceleration or waiting behavior. Manually annotated high-risk encounters will permit comparison with the judgment of VTS experts and navigators. Further tests will cover zero-shot cross-region transfer, traffic-density strata, and original baseline implementations with complete vessel attributes.

Author Contributions: Conceptualization, H.X. and Z.S.; methodology, H.X., Z.S. and Y.J.; software, H.X. and Y.J.; validation, H.X., Y.J. and Y.R.; formal analysis, H.X. and Y.R.; investigation, H.X., Y.J. and Y.R.; resources, Z.S. and Y.R.; data curation, H.X. and Y.J.; writing—original draft preparation, H.X. and Y.J.; writing—review and editing, H.X., Z.S. and Y.R.; visualization, H.X. and Y.R.; supervision, Z.S.; project administration, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Liaoning Provincial Science and Technology Program Joint Plan under Grant No. 2025-MSLH-101.

Data Availability Statement: The AIS data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIS	Automatic Identification System
ADE	Average Displacement Error
APF	Artificial Potential Field
AP	Average Precision
AUC	Area Under the Receiver Operating Characteristic Curve
BiLSTM	Bidirectional Long Short-Term Memory

CI	Confidence Interval
COG	Course Over Ground
CPA	Closest Point of Approach
CUDA	Compute Unified Device Architecture
CV	Constant Velocity
DCPA	Distance to CPA
ECE	Expected Calibration Error
FDE	Final Displacement Error
GNN	Graph Neural Network
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
IMO	International Maritime Organization
LSTM	Long Short-Term Memory
MASS	Maritime Autonomous Surface Ships
MLP	Multilayer Perceptron
MMSI	Maritime Mobile Service Identity
NRRMWOA	Nonlinear Randomly Reuse-based Mutated Whale Optimization Algorithm
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
PCR	Potential Collision Risk
PINN	Physics-Informed Neural Network
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SiLU	Sigmoid Linear Unit
SOG	Speed Over Ground
ST-GCN	Spatio-Temporal Graph Convolutional Network
TCPA	Time to CPA
VTs	Vessel Traffic Service
WOA	Whale Optimization Algorithm

References

1. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.B. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1559–1582. [\[CrossRef\]](#)
2. Zhang, X.; Fu, X.; Xiao, Z.; Xu, H.; Qin, Z. Vessel Trajectory Prediction in Maritime Transportation: Current Approaches and Beyond. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19980–19998. [\[CrossRef\]](#)
3. Li, H.; Xing, W.; Jiao, H.; Yuen, K.F.; Gao, R.; Li, Y.; Matthews, C.; Yang, Z. Bi-directional information fusion-driven deep network for ship trajectory prediction in intelligent transportation systems. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *192*, 103770. [\[CrossRef\]](#)
4. Jia, C.; Ma, J. Conditional temporal GAN for intent-aware vessel trajectory prediction in the precautionary area. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106776. [\[CrossRef\]](#)
5. Li, X.; Wei, Z.; Zhang, Q.; Xing, L.; Han, Z.; Gao, G. Ship trajectory prediction under multi-ship coupling effect based on a generative adversarial network. *Ocean Eng.* **2025**, *342*, 122898. [\[CrossRef\]](#)
6. Perera, L.P.; Oliveira, P.; Guedes Soares, C. Maritime Traffic Monitoring Based on Vessel Detection, Tracking, State Estimation, and Trajectory Prediction. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1188–1200. [\[CrossRef\]](#)
7. Wiig, M.S.; Pettersen, K.Y.; Krogstad, T.R. Collision Avoidance for Underactuated Marine Vehicles Using the Constant Avoidance Angle Algorithm. *IEEE Trans. Control Syst. Technol.* **2020**, *28*, 951–966. [\[CrossRef\]](#)
8. Zhao, S.B.; Tang, C.; Liang, S.; Wang, D.J. Track prediction of vessel in controlled waterway based on improved Kalman filter. *J. Comput. Appl.* **2012**, *32*, 3247–3250. [\[CrossRef\]](#)
9. Liu, J.; Shi, G.; Zhu, K. Vessel Trajectory Prediction Model Based on AIS Sensor Data and Adaptive Chaos Differential Evolution Support Vector Regression (ACDE-SVR). *Appl. Sci.* **2019**, *9*, 2983. [\[CrossRef\]](#)
10. Rong, H.; Teixeira, A.P.; Guedes Soares, C. Ship trajectory uncertainty prediction based on a Gaussian Process model. *Ocean Eng.* **2019**, *182*, 499–511. [\[CrossRef\]](#)
11. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Tang, H.; Yin, Y.; Shen, H. A model for vessel trajectory prediction based on long short-term memory neural network. *J. Mar. Eng. Technol.* **2022**, *21*, 136–145. [\[CrossRef\]](#)

13. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [\[CrossRef\]](#)
14. Xiao, Y.; Li, X.; Yao, W.; Chen, J.; Hu, Y. Bidirectional Data-Driven Trajectory Prediction for Intelligent Maritime Traffic. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1773–1785. [\[CrossRef\]](#)
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. [\[CrossRef\]](#)
16. Xue, H.; Wang, S.; Xia, M.; Guo, S. G-Trans: A hierarchical approach to vessel trajectory prediction with GRU-based transformer. *Ocean Eng.* **2024**, *300*, 117431. [\[CrossRef\]](#)
17. Yang, D.; Yang, K.; Lu, Y.; Liang, M.; Zhao, C. Graph-driven multi-vessel long-term trajectories prediction for route planning under complex waters. *Ocean Eng.* **2024**, *313*, 119511. [\[CrossRef\]](#)
18. Zeng, X.; Gao, M.; Zhang, A.; Zhu, J.; Hu, Y.; Chen, P.; Chen, S.; Dong, T.; Zhang, S.; Shi, P. Trajectories prediction in multi-ship encounters: Utilizing graph convolutional neural networks with GRU and Self-Attention Mechanism. *Comput. Electr. Eng.* **2024**, *120*, 109679. [\[CrossRef\]](#)
19. Du, L.; Goerlandt, F.; Kujala, P. Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. *Reliab. Eng. Syst. Saf.* **2020**, *200*, 106933. [\[CrossRef\]](#)
20. Im, N.; Luong, T.N. Potential risk ship domain as a danger criterion for real-time ship collision risk evaluation. *Ocean Eng.* **2019**, *194*, 106610. [\[CrossRef\]](#)
21. Ma, J.; Jia, C.; Yang, X.; Cheng, X.; Li, W.; Zhang, C. A Data-Driven Approach for Collision Risk Early Warning in Vessel Encounter Situations Using Attention-BiLSTM. *IEEE Access* **2020**, *8*, 188771–188783. [\[CrossRef\]](#)
22. Tritsarolis, A.; Murray, B.; Pelekis, N.; Theodoridis, Y. Collision Risk Assessment and Forecasting on Maritime Data. In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems, Hamburg, Germany, 13–16 November 2023; pp. 1–10. [\[CrossRef\]](#)
23. Zhang, L.; Meng, Q. Probabilistic ship domain with applications to ship collision risk assessment. *Ocean Eng.* **2019**, *186*, 106130. [\[CrossRef\]](#)
24. Wu, L.; Xu, D.; Guo, Q.; Chen, E.; Xiao, W. A nonlinear randomly reuse-based mutated whale optimization algorithm and its application for solving engineering problems. *Appl. Soft Comput.* **2024**, *167*, 112271. [\[CrossRef\]](#)
25. Semerdjiev, E.; Mihaylova, L. Variable- and Fixed-Structure Augmented Interacting Multiple Model Algorithms for Manoeuvring Ship Tracking Based on New Ship Models. *Int. J. Appl. Math. Comput. Sci.* **2000**, *10*, 591–604.
26. Virjonen, P.; Nevalainen, P.; Pahikkala, T.; Heikkonen, J. Ship Movement Prediction Using k-NN Method. In Proceedings of the 2018 Baltic Geodetic Congress (BGC Geomatics), Olsztyn, Poland, 21–23 June 2018; pp. 304–309. [\[CrossRef\]](#)
27. Wang, J.; Guo, Y.; Wang, Y. A Sequential Random Forest for Short-Term Vessel Speed Prediction. *Ocean Eng.* **2022**, *248*, 110691. [\[CrossRef\]](#)
28. Zhang, C.; Bin, J.; Wang, W.; Peng, X.; Wang, R.; Haldearn, R.; Liu, Z. AIS Data Driven General Vessel Destination Prediction: A Random Forest Based Approach. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102729. [\[CrossRef\]](#)
29. Gao, D.W.; Zhu, Y.S.; Zhang, J.F.; He, Y.K.; Yan, K.; Yan, B.R. A novel MP-LSTM method for ship trajectory prediction based on AIS data. *Ocean Eng.* **2021**, *228*, 108956. [\[CrossRef\]](#)
30. You, L.; Xiao, S.; Peng, Q.; Claramunt, C.; Han, X.; Guan, Z.; Zhang, J. ST-Seq2Seq: A Spatio-Temporal Feature-Optimized Seq2Seq Model for Short-Term Vessel Trajectory Prediction. *IEEE Access* **2020**, *8*, 218565–218574. [\[CrossRef\]](#)
31. Zhou, Y.; Dong, Z.; Bao, X. A Ship Trajectory Prediction Method Based on an Optuna-BiLSTM Model. *Appl. Sci.* **2024**, *14*, 3719. [\[CrossRef\]](#)
32. Yin, J.; Yu, Z.; Wu, H. Ship trajectory prediction based on LSTM model with multi-scale convolution and attention mechanism. *Ocean Eng.* **2025**, *338*, 122055. [\[CrossRef\]](#)
33. Jia, C.; Ma, J.; Yang, X.; Lv, X. RAGAN: A Generative Adversarial Network for risk-aware trajectory prediction in multi-ship encounter situations. *Ocean Eng.* **2023**, *289*, 116188. [\[CrossRef\]](#)
34. Nguyen, D.; Fablet, R. A Transformer Network With Sparse Augmented Data Representation and Cross Entropy Loss for AIS-Based Vessel Trajectory Prediction. *IEEE Access* **2024**, *12*, 21596–21609. [\[CrossRef\]](#)
35. Jiang, D.; Shi, G.; Li, N.; Ma, L.; Li, W.; Shi, J. TRFM-LS: Transformer-Based Deep Learning Method for Vessel Trajectory Prediction. *J. Mar. Sci. Eng.* **2023**, *11*, 880. [\[CrossRef\]](#)
36. Zhao, J.; Yan, Z.; Chen, X.; Han, B.; Wu, S.; Ke, R. k-GCN-LSTM: A k-Hop Graph Convolutional Network and Long-Short-Term Memory for Ship Speed Prediction. *Phys. A* **2022**, *606*, 128107. [\[CrossRef\]](#)
37. Wang, S.; Li, Y.; Xing, H.; Zhang, Z. Vessel trajectory prediction based on spatio-temporal graph convolutional network for complex and crowded sea areas. *Ocean Eng.* **2024**, *298*, 117232. [\[CrossRef\]](#)
38. Yu, Y.; Song, Z.; Zhang, H.; Liu, D.; Li, L.; Xie, B. Ship trajectory prediction method based on heterogeneous spatiotemporal graph neural networks. *Appl. Ocean Res.* **2026**, *168*, 104969. [\[CrossRef\]](#)
39. Zhao, L.; Zuo, Y.; Zhang, W.; Li, T.; Chen, C.L.P. SNIINet: Trajectory prediction using ship navigation information interaction-aware neural network. *Ocean Eng.* **2025**, *321*, 120368. [\[CrossRef\]](#)

40. Wang, Z.; Nie, Z.; Sheng, G. Dynamic Position Predicting of Underactuated Surface Vessel with Unscented Kalman Filter. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 4030–4033. [\[CrossRef\]](#)
41. Xu, P.F.; Han, C.B.; Cheng, H.X.; Cheng, C.; Ge, T. A Physics-Informed Neural Network for the Prediction of Unmanned Surface Vehicle Dynamics. *J. Mar. Sci. Eng.* **2022**, *10*, 148. [\[CrossRef\]](#)
42. Forti, N.; Millefiori, L.M.; Braca, P.; Willett, P. Model-based Deep Learning for Maneuvering Target Tracking. In Proceedings of the 2023 26th International Conference on Information Fusion (FUSION), Charleston, SC, USA, 27–30 June 2023; pp. 1–6. [\[CrossRef\]](#)
43. Liu, X.; Yuan, H.; Xiao, C.; Wang, Y.; Yu, Q. Hybrid-driven vessel trajectory prediction based on uncertainty fusion. *Ocean Eng.* **2022**, *248*, 110836. [\[CrossRef\]](#)
44. Alam, M.M.; Soares, A.; Rodrigues, J.F., Jr.; Spadon, G. Physics-Informed Neural Networks for Vessel Trajectory Prediction: Learning Time-Discretized Kinematic Dynamics via Finite Differences. *arXiv* **2025**, arXiv:2506.12029. [\[CrossRef\]](#)
45. Wu, W.; Chen, P.; Chen, L.; Mou, J. Ship Trajectory Prediction: An Integrated Approach Using ConvLSTM-Based Sequence-to-Sequence Model. *J. Mar. Sci. Eng.* **2023**, *11*, 1484. [\[CrossRef\]](#)
46. Zhang, S.K.; Shi, G.Y.; Liu, Z.J.; Zhao, Z.W.; Wu, Z.L. Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Eng.* **2018**, *155*, 240–250. [\[CrossRef\]](#)
47. Zhang, J.; Santos, T.A.; Guedes Soares, C.; Yan, X. Sequential ship traffic scheduling model for restricted two-way waterway transportation. *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.* **2017**, *231*, 86–97. [\[CrossRef\]](#)
48. Liu, T.; Xu, X.; Lei, Z.; Zhang, X.; Sha, M.; Wang, F. A multi-task deep learning model integrating ship trajectory and collision risk prediction. *Ocean Eng.* **2023**, *287*, 115870. [\[CrossRef\]](#)
49. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [\[CrossRef\]](#)
50. International Maritime Organization. *Guidelines for Vessel Traffic Services*; Resolution A.1158(32), Adopted 15 December 2021; International Maritime Organization: London, UK, 2021. Available online: [https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/AssemblyDocuments/A.1158\(32\).pdf](https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/AssemblyDocuments/A.1158(32).pdf) (accessed on 29 July 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.