Transportation Research Part E 207 (2026) 104599

Contents lists available at ScienceDirect

Transportation Research Part E

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

LLM4STP: A large language model-driven multi-feature fusion method for ship trajectory prediction

Hang Jiao [a], Jincheng Gong [b], Huanhuan Li [c,d,*], Jasmine Siu Lee Lam [e], Yaqing Shu [d], Jin Wang [d], Zaili Yang [d]

[a] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
[b] School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan, China
[c] School of Engineering, University of Southampton, Southampton, UK
[d] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK
[e] Department of Technology, Management and Economics, Technical University of Denmark, Denmark

A R T I C L E I N F O

Keywords:
Large language model
Ship trajectory prediction
Situational awareness
Hierarchical temporal modeling
Trajectory uncertainty
GeoHash encoding

A B S T R A C T

Ship trajectory prediction (STP) is a critical research focus for enhancing maritime traffic situational awareness and supporting navigational decision-making in intelligent transportation systems. The accuracy and robustness of prediction models significantly affect maritime safety and shipping efficiency. Despite advances driven by Automatic Identification System (AIS) data and deep learning techniques, key challenges remain unresolved, including dynamic multi-ship interaction modelling in complex marine environments, multi-scale temporal dependency reasoning, trajectory uncertainty quantification, and effective integration of maritime domain knowledge. Existing methods based on Large Language Models (LLMs) improve generalisation through pretrained knowledge but fall short in real-time interaction topology modelling, geospatial semantic representation, and uncertainty estimation. To address these limitations, this paper proposes LLM4STP, a novel LLM-driven multi-feature fusion method for STP. LLM4STP establishes a new paradigm by deeply integrating LLMs with maritime domain knowledge to collaboratively predict ship trajectories. The model features an adaptive graph-masked Transformer to dynamically capture ship interaction topologies, hierarchical temporal reasoning to jointly model local manoeuvring behaviours and macroscopic navigational intent, and an innovative fusion of Gaussian probability distribution heatmaps with GeoHash-based geospatial encoding to quantify trajectory uncertainty while preserving semantic continuity. Experiments on three representative water areas demonstrate that LLM4STP consistently outperforms nine state-of-the-art (SOTA) methods, as validated by key metrics including Average Displacement Error (ADE), FrÃ©chet Distance (FD), and Final Displacement Error (FDE). Moreover, the few-shot learning experiments demonstrate that LLM4STP can match the performance of models trained on the full dataset using only 20 % of the training data, highlighting its efficiency and strong adaptability in data-scarce environments. The ablation studies empirically validate the significance and distinct contribution of each component within the proposed model architecture. This study integrates LLM into maritime traffic scenarios, making significant contributions to enhancing the robustness, accuracy, and interpretability of STP in high-interference environments. The source code is openly accessible at https://github.com/Joker-hang/LLM4STP.

* Corresponding author.
E-mail address: huanhuan.li@soton.ac.uk (H. Li).

https://doi.org/10.1016/j.tre.2025.104599
Received 25 May 2025; Received in revised form 28 September 2025; Accepted 5 December 2025
Available online 22 December 2025
1366-5545/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ship trajectory prediction (STP) serves as a foundational component of intelligent shipping systems (Zhang et al., 2022; Bai et al., 2024), with significant applications in maritime traffic situational awareness (Liu et al., 2025; Li et al., 2023b), collision risk assessment (Xu et al., 2025), and navigational path optimisation (Zhao and Zhao, 2024). While model-based methods effectively capture the physical characteristics of ship motion, they often struggle to adapt to complex and dynamic marine environments (Jiao et al., 2021). The widespread application of the Automatic Identification System (AIS) and the digital transformation of the global shipping industry (Chen et al., 2024; Shu et al., 2025) have generated vast volumes of dynamic maritime data. This has accelerated the advancement and transformative application of deep learning-based methods in maritime research and operations (Bi et al., 2024; Yao et al., 2024). The dynamic information recorded by AIS, such as ship position, speed, and heading, enables models like Long Short-Term Memory (LSTM) Networks (Li et al., 2024b,c) and Convolutional Neural Networks (CNNs) (Lee and Park, 2024) to capture temporal dependencies in ship movement patterns. However, challenges persist in real-world marine scenarios, including frequent and dynamic multi-ship interactions (Cao et al., 2024), significant trajectory uncertainty (Li et al., 2024a), and difficulties in integrating multi-source heterogeneous data such as weather (Kim et al., 2023), navigation regulations (Yang et al., 2024c), and operational semantics (Yu et al., 2021). These factors cause bottlenecks for traditional deep learning models, often leading to extractions of decoupled spatio-temporal characteristics and inadequate modelling of long-term temporal dependencies (Yang and Pei, 2023; Mo et al., 2024).

Recently, Large Language Models (LLMs) have emerged as a promising paradigm for trajectory prediction, leveraging their powerful contextual reasoning capabilities and extensive pre-trained knowledge (Peng et al., 2025). Several studies have begun exploring their potential in the maritime domain. For example, Nguyen et al. (2024) proposed TM-LLM, which incorporates token and motion embedding layers to capture both local and global patterns in sequential AIS data. These embeddings are subsequently fused through a convolutional layer to ensure compatibility with the LLM input. Similarly, Chen et al. (2025) utilised LLMs to interpret ship navigation behaviours and infer multiple navigation intentions semantically. By leveraging a Transformer architecture, the framework effectively associates key waypoints with inferred intent information, enabling more consistent and accurate long-term trajectory forecasting. However, current LLM-driven approaches still face several critical challenges: (1) They often fail to incorporate rich semantic information, such as spatial interactions between vessels, probabilistic trajectory distributions, and geographic context, thereby limiting the full potential of LLMs in capturing maritime dynamics. (2) Most methods do not provide an end-to-end prediction framework, using LLMs primarily for intention inference while relying on separate transformer-based modules for trajectory prediction. This fragmented design reduces efficiency and model coherence. (3) Few studies have explored the few-shot learning capabilities of LLMs in STP, overlooking their potential to achieve high performance and adaptability even with limited training data.

To address these challenges, a novel LLM-driven multi-feature fusion method, LLM4STP, is proposed for STP. LLM4STP systematically tackles critical issues such as dynamic interaction modelling, multi-scale temporal processing, and uncertainty quantification through a unified and deeply integrated architecture. Specifically, the prompt project encodes maritime domain knowledge (e.g., waterway restrictions and weather alerts) into structured embeddings. To accurately capture dynamic ship interactions, a graph-masked spatial Transformer mechanism is developed, employing adaptive distance thresholds to construct real-time interaction topologies. To simultaneously preserve local manoeuvring details and global navigational patterns, a hierarchical temporal modelling paradigm is proposed, combining local convolutional processing with global LLM-based reasoning. To handle trajectory uncertainty and maintain geographic semantic continuity, a dual modelling system using Gaussian heatmaps and Geohash-based hierarchical encoding is introduced, converting physical space perturbations into learnable feature representations. Through a five-fold feature embedding mechanism, LLM4STP achieves the first deep integration of semantic, spatial, temporal, probabilistic, and geographic information for trajectory prediction.

The proposed model is evaluated on three representative AIS datasets: Chengshan Jiao Promontory area (CSJ), Caofeidian water area (CFD), and Tianjin Port water area (TJP). These regions were chosen for their distinct characteristics and operational scenarios, encompassing complex geographical features, high-intensity maritime traffic, and diverse navigational challenges (Li et al., 2023a).

Comparative experimental results demonstrate that LLM4STP consistently outperforms nine state-of-the-art (SOTA) methods in key evaluation metrics, including Final Displacement Error (FDE), Average Displacement Error (ADE), and FrÃ©chet Distance (FD). Furthermore, the discussion of LLM4STP's performance under few-shot learning settings has been strengthened. Experimental results reveal that LLM4STP achieves performance comparable to models trained on the full dataset even when using only 20 % of the training data, highlighting its strong generalisation ability and data efficiency. Additionally, ablation studies confirm the effectiveness and contributions of each module within LLM4STP.

The innovative contributions of this paper can be summarised as follows:

(1) Develop a spatial Transformer for dynamic relationship modelling. A graph-masked spatial attention mechanism is developed to dynamically construct ship interaction topologies based on adaptive distance thresholds. By integrating multi-head self-attention, the method enables accurate quantification of complex and evolving spatial interactions, effectively overcoming the limitations of traditional approaches in capturing dynamic ship interactions.
(2) Introduce a hierarchical temporal modelling approach. A "Local Convolution-Global LLM" hybrid module is introduced for time-series processing. By using a patching operation, the model achieves hierarchical temporal representation, preserving local motion details while simultaneously leveraging the global temporal reasoning capabilities of pre-trained LLMs to capture long-term navigation patterns.

(3) Establish a dual modelling system for probability uncertainty and geographic semantic continuity. A probabilistic modelling network using 2D Gaussian heatmaps is developed, combined with a Geohash-based hierarchical encoding mechanism. This framework transforms positional uncertainty into learnable feature vectors, enhancing model robustness in complex marine environments while maintaining the continuity of geographic semantics.
(4) Propose LLM4STP, a LLM-driven multi-feature fusion framework for STP. A novel multi-feature fusion framework is proposed, integrating semantic, probabilistic, geohash, local, and spatial embeddings to comprehensively represent multi-source heterogeneous data. This pioneering fusion achieves the first deep coupling between large language models and STP, addressing long-standing limitations in traditional models and advancing the application of LLMs in maritime intelligent systems.

The remainder of this paper is organised as follows: Section 2 provides a systematic review of current trajectory prediction studies based on deep learning and LLMs and summarises the main research gaps. Section 3 presents the model framework and the design of multiple embedding modules. Section 4 describes the experimental details and analyses the results. Section 5 discusses the implications of this study, and Section 6 concludes this paper.

2. Literature review

Deep learning-based STP has become a mainstream research direction in intelligent transportation systems. Architectures such as CNNs, LSTMs, and Gated Recurrent Units (GRUs) have proven effective in capturing spatiotemporal features and long-term dependencies within trajectory data, particularly for scenarios with relatively regular ship movement patterns (Li et al., 2023a). However, with the rise of LLMs, their powerful capabilities in contextual reasoning and complex sequential data processing have gradually driven their application in STP (Liu et al., 2024b). In contrast to conventional deep learning models, LLMs serialise trajectory information into tokenised sequences, enabling them to capture complex contextual dependencies and model trajectory uncertainties more effectively, particularly in highly dynamic marine environments.

To reflect the evolving landscape of STP, this paper categorises the development of trajectory prediction into two major categories: deep learning-based trajectory prediction and LLM-based trajectory prediction, focusing on differences in model architectures, data processing strategies, optimisation methods, and application scenarios.

2.1. A systematic review of deep learning approaches

Deep learning has significantly advanced STP, contributing to enhanced maritime traffic safety, improved shipping efficiency, and the development of intelligent transportation systems. It is crucial in scenarios including collision avoidance, route planning, and maritime monitoring.

To investigate research trends, this study retrieved related publications from the Web of Science (WoS) Core Collection (Mongeon and Paul-Hus, 2016) using the search terms: (“TS = ship trajectory prediction or TS = vessel trajectory prediction) and (TS = deep learning”). As of May 11, 2025, a total of 181 SCI-related papers were retrieved. Among these, 177 highly relevant publications from 2020 to 2025 were analysed using the CiteSpace software (Chen, 2006), resulting in thematic keyword clustering (Fig. 1) and timeline visualisations (Fig. 2).

As shown in Fig. 1, the keywords cluster into eight thematic clusters. Cluster #0 (Trajectory) data forms the foundation of STP, with AIS data serving as the primary data source. AIS data, praised by Yang et al. (2019) for its transformative potential, has become the foundation of the digital era in shipping. Beyond AIS data, Cluster #4 (maritime track association) reflects efforts to integrate multi-source trajectory data for enhanced predictive capabilities. For instance, Gülsoylu et al. (2024) combined AIS data with maritime images, ship speed, and heading information to establish baselines for ship detection and STP. Xiao et al. (2024) developed an adaptive multimodal prediction model that integrates AIS and environmental data, improving the precision of the prediction through the fusion of historical and environmental characteristics.

Clusters #3 (STP), #6 (long-sequence forecasting), #7 (Long short-term memory) and #3 (attention mechanism) focus on methodological developments. Long-sequence forecasting (#6) can effectively capture long-term dependencies in ship trajectory data, significantly improving the forecasting of future trajectory trends. Xiong et al. (2024) has employed the Informer model for long-term trajectory prediction.

Attention mechanisms (#2) are widely used. For example, Xue et al. (2024) proposed the G-Trans technique based on the Transformer architecture and feature clustering, and Yoo et al. (2024) combined GRU with attention mechanisms for fishing vessel trajectory prediction. Lin et al. (2023) proposed TTCN-Attention-GRU by embedding attention layers within hierarchical temporal CNNs. Some researchers have also combined attention mechanisms with LSTM (#7) to predict STP (Zhang et al., 2021; Zhou et al., 2024).

Clusters #1 (maritime safety) and #5 (anomaly detection) represent the application scenarios for STP tasks. Zhang et al. (2022) highlighted that STP is one of the most important topics for safe, intelligent, and efficient maritime transportation. Zaman et al. (2024) laid the groundwork for deep learning’s adoption in trajectory prediction aimed at enhancing maritime traffic safety. Additionally, numerous researchers have applied STP for anomaly detection (#5). For instance, Xie et al. (2024) integrated DBSCAN-based trajectory clustering with attention-based trajectory prediction to effectively identify anomalous behaviours in ship movements. Similarly, Hu et al. (2022) employed deep reinforcement learning algorithms to train trajectory anomaly detection models.

As depicted in Fig. 2, AIS trajectory (#0) data emerged as a pivotal foundation for STP. By recording dynamic ship information, such as position, speed, and heading, AIS data provides essential inputs for understanding vessel movement patterns and forecasting future trajectories. Yang et al. (2019) highlighted that AIS data has immensely enriched maritime research, serving as a fundamental

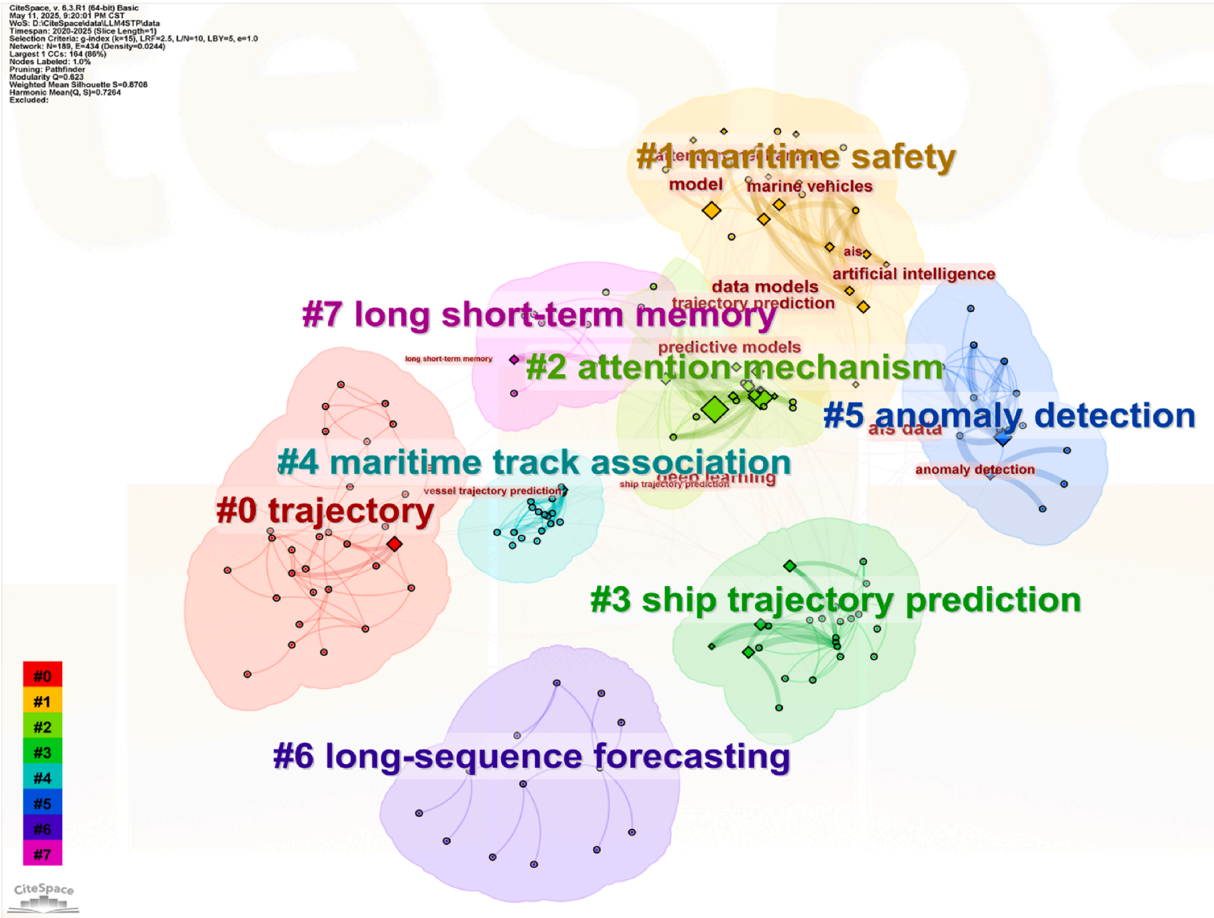

Fig. 1. Keyword clustering of STP papers based on deep learning.

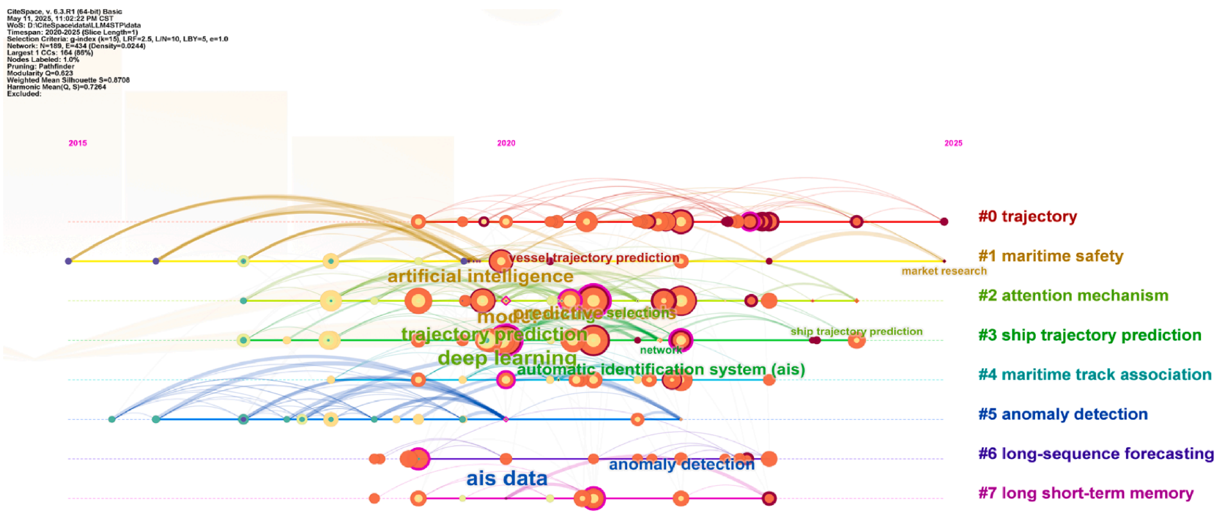

Fig. 2. Keyword timeline of STP literature based on deep learning.

data pillar of the field. Tu et al. (2017) further delved into AIS data sources and their correlation with ship navigation, underscoring their significance in trajectory prediction and their broad applicability in tasks such as anomaly detection, collision prediction, route estimation, and path planning.

With technological advancements, the research focus has shifted from merely collecting AIS data to developing data models for extracting trajectory (#0) information for more accurate trajectory prediction. In recent years, particularly since 2023, AI-driven models have been widely adopted for trajectory prediction. Deep learning, a key branch of AI, has rapidly gained prominence in this domain. Its strong capabilities in feature extraction and complex data pattern modelling have unlocked new possibilities for STP. Many deep learning techniques, including LSTM (#7), GRU, and attention mechanisms, have been incorporated into this field. For instance, Wang et al. (2025) introduced a switching input LSTM model to enhance prediction flexibility. Chen et al. (2022) integrated forward GRU and backwards Bi-GRU networks to improve trajectory fitting accuracy through bidirectional scaling. Shin and Yang (2025) systematically compared Bi-GRU, Bi-LSTM, and Transformer models for STP in port environments.

Among deep learning innovations, attention mechanisms (#2) have attracted particular interest in maritime trajectory forecasting. Inspired by the human visual attention system, attention mechanisms dynamically allocate computational focus to the most informative segments of input data, improving model sensitivity and efficiency. Building on conventional recurrent architectures such as LSTM and GRU, researchers have introduced attention modules to better capture key features and dynamic variations within vessel trajectories.

Recent advancements demonstrate integrated strategies combining deep learning models with attention mechanisms. Li et al. (2024b) developed an Attribute Correlation Attention (ACoAtt)-LSTM framework, embedding AIS data with attribute-aware attention modules for improved prediction performance. Zhao et al. (2024) proposed a dual-attention end-to-end neural network architecture that synergises LSTM units with attention mechanisms. Liu et al. (2024d) innovatively combined CNN architectures with attention-enhanced GRU networks, achieving not only higher predictive accuracy but also improved model interpretability, offering new perspectives for understanding decision-making mechanisms in maritime trajectory systems.

While conventional temporal models often struggle to capture inter-vessel interactions in multi-ship scenarios, spatiotemporal feature fusion networks have emerged as a promising solution. These architectures enable comprehensive representation by integrating temporal evolution patterns with spatial inter-ship interaction mechanisms, thus better reflecting the dynamics of complex maritime environments. Notable implementations demonstrate progressive methodological innovations: Zhang et al. (2023) developed a Spatiotemporal-Aware Graph Attention Network (ST-GAT) based on LSTM, incorporating spatiotemporal attention to dynamically weigh critical time phases and spatial relationships, thereby enhancing predictive accuracy and robustness. Liang et al. (2022) applied Spatiotemporal Graph Convolutional Networks (ST-GCN) to jointly model spatial relationships through graph operations and temporal dependencies. Zeng et al. (2024) advanced this paradigm by combining graph convolutional networks with GRU and attention mechanisms, particularly excelling in long-sequence processing and complex interaction scenarios. Liu et al. (2024c) introduced a hybrid model combining graph attention networks with Transformer architecture, enabling probabilistic trajectory distribution prediction and uncertainty quantification, thereby enhancing operational reliability in practical applications. These developments have propelled rapid progress in spatiotemporal fusion-based prediction technologies, enabling not only more accurate trajectory forecasting but also a deeper understanding of ship interaction patterns, offering novel technical solutions for intelligent shipping systems, traffic management, and maritime safety (#1).

In conclusion, the maritime trajectory prediction domain has undergone transformative advancements. The integration of AIS data with advances in deep learning and the emergence of novel spatiotemporal fusion networks have substantially enhanced both prediction accuracy and operational efficiency. Furthermore, they provide robust technical support for sustainable development in maritime safety (#1), anomaly detection (#5), shipping optimisation, and marine resource management (Zhang et al., 2022; Capobianco et al., 2022). However, trajectory prediction methods based on deep learning models still fail to integrate knowledge from fields such as meteorology and navigation rules.

2.2. LLM-based trajectory prediction methods

As LLMs advance rapidly, scholars have explored their application to trajectory prediction, injecting new vitality into this traditional research domain. Fig. 3 illustrates the parallel development of LLMs and their application to trajectory prediction tasks over time.

The upper part of Fig. 3 delineates the evolutionary trajectory of LLMs development. The emergence of LLMs has achieved remarkable breakthroughs in two critical dimensions: parameter scaling and model capability. The proposal of the Transformer architecture (Vaswani et al., 2017) established a foundational framework for subsequent breakthroughs by enabling efficient capture of long-range dependencies through the self-attention mechanism. In 2019, Generative Pre-trained Transformer (GPT)-2 (Radford et al., 2019) demonstrated the substantial potential of LLMs in multi-task generalisation, marking the first step toward broad cross-domain applicability. GPT-3 (Brown et al., 2020), whose unprecedented 175-billion parameters successfully overcame few-shot learning limitations, empirically validates the pivotal role of scaling law in performance enhancement. This breakthrough solidified the widely recognised positive correlation between parameter scale and model efficacy. By 2022, GPT-3.5 achieved a paradigm shift from pure text generation to marking the first step toward broad cross-domain applicability.

Between 2023 and 2024, technological competition intensified, leading to dual paths of multimodal integration and domain-specific optimisation. Models such as Qwen (Bai et al., 2023) and its successor Qwen 2 (Yang et al., 2024a) demonstrated superior performance in Chinese linguistic optimisation, advancing Natural Language Processing (NLP) applications in Sinophone contexts. Concurrently, Large Language Model Meta AI (LLaMA) 2 (Touvron et al., 2023) and LLaMA 3 (Dubey et al., 2024) achieved enhanced

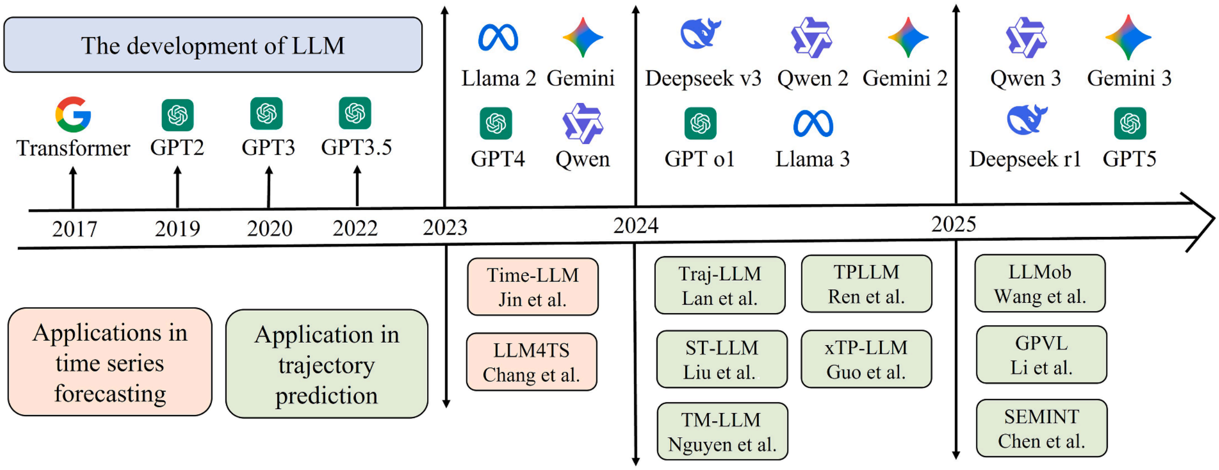

Fig. 3. Timeline of LLM development and its applications in trajectory prediction.

training efficiency and performance through innovative architectural designs that transcend traditional parameter constraints. GPT-4 (Achiam et al., 2023) expanded perceptual dimensionality via cross-modal integration, while DeepSeek-V3 (Liu et al., 2024a) pioneered trillion-parameter modelling through sparse training methodologies. The year 2024 marked the inception of the reasoning era with GPT o1 (Wu et al., 2024), introducing Chain of Thought (CoT) mechanisms to enhance logical reasoning. By 2025, models such as Qwen 3 (Yang et al., 2025), Deepseek r1 (DeepSeek-AI et al., 2025), Gemini 3 and GPT-5 had further improved task performance, inference speed, and model transparency, democratising access to high-performance LLMs through cost-efficient open-source initiatives.

The lower part of Fig. 3 details the evolution of LLMs applications in trajectory prediction. In 2023, researchers began exploring LLMs for time-series forecasting, laying the groundwork for trajectory prediction advancements. For instance, Time-LLM (Jin et al., 2023) proposed reprogramming time-series inputs as textual prototypes to align with frozen LLMs, leveraging their inherent sequence modelling strengths. Meanwhile, LLM4TS developed a dual-stage aggregation framework to integrate multi-scale temporal patterns into pre-trained LLMs, significantly enhancing model interpretability for time-specific patterns and enabling more precise forecasting architectures. The year 2024 witnessed exponential growth in LLM-driven trajectory prediction research. Traj-LLM (Lan et al., 2024) pioneered LLM-driven trajectory generation by capturing historical motion and environmental semantics without explicit prompt engineering. Spatial-Temporal (ST)-LLM (Liu et al., 2024b) introduced spatiotemporal embeddings, representing sensor data as semantic tokens and capturing both spatial and temporal dependencies in traffic prediction. Traffic Prediction LLM (TPLLM) (Ren et al., 2024) advanced this domain through a hybrid architecture combining CNN-based sequence embedding layers and GCN-based graph embedding layers, optimising inputs for LLMs. Meanwhile, the Traffic flow Prediction model based on LLM (xTP-LLM) (Guo et al., 2024) introduced an explainable framework by converting multimodal traffic data into natural language descriptors, enabling LLMs to account for external factors and provide human-interpretable traffic predictions. In addition, Nguyen et al. (2024) proposed a new model called TM-LLM, designed for STP using historical AIS data.

By 2025, LLM-based trajectory prediction methods had diversified further. LLM for personal Mobility generation (LLMob) (JIAWEI et al., 2025) proposed an agent-based LLM framework for personalised mobility generation, promoting user-specific transportation service innovations. Generative Planning with 3D-vision language pre-training (GPVL) (Li et al., 2025) applied 3D vision-language pre-trained models for autonomous driving planning, demonstrating the practical deployment potential of LLMs in complex dynamic environments. These developments highlight LLMs' growing capability to simultaneously address key challenges in trajectory prediction, including improving forecast accuracy, enhancing computational efficiency, and enabling real-world deployment in intelligent transportation systems. Furthermore, Chen et al. (2025) develops a new framework that integrates semantic cognition with context-aware intention modelling, thereby enhancing the accuracy of long-term STP.

Overall, the evolution from early Transformer-based LLMs to current multimodal, domain-specialised models demonstrates their transformative potential across both natural language processing and trajectory prediction domains. Specifically, innovations such as sparse encoding, spatiotemporal reasoning, and multimodal data fusion have significantly expanded LLMs' applicability for intelligent traffic management, autonomous navigation, and personalised mobility solutions. Continued refinement and innovation in LLM technologies are poised to drive further expansion in trajectory prediction and beyond, unlocking new opportunities for industrial transformation, societal advancement, and the resolution of complex real-world challenges through intelligent computational frameworks.

2.3. Research gaps

Despite significant advancements in STP research, the detailed review reveals four critical research gaps requiring urgent resolution:

(1) Limitations in dynamic interaction topology modelling. Most existing trajectory prediction methods rely on static attention mechanisms or fixed adjacency matrices to model spatial relationships between ships. These approaches exhibit limited adaptability to real-time, evolving interaction patterns in maritime traffic. In dynamic scenarios, such as abrupt speed changes or emergency collision avoidance, static models fail to capture the changing nature of ship encounters accurately. This fundamental mismatch between static modelling and the inherently dynamic maritime environment significantly restricts model generalisation in complex scenarios.
(2) Deficiency in multi-scale temporal feature decoupling. Current data-driven models often process ship trajectory sequences at a single temporal scale, lacking effective coordination between local motion details and broader navigation patterns. This single-scale approach fails to reflect the multi-level temporal characteristics of maritime navigation, which spans microscopic manoeuvres, mesoscopic route planning, and macroscopic navigation intentions. As a result, existing models suffer from cumulative prediction errors manifesting as both local distortions and global deviations, ultimately degrading overall prediction accuracy.
(3) Lack of trajectory uncertainty quantification. Prevailing methods predominantly adopt deterministic prediction frameworks, offering limited means for quantitatively characterising the uncertainties introduced by variable marine environmental factors such as weather conditions and ocean currents. The absence of uncertainty quantification impairs models' ability to assess prediction reliability and limits their effectiveness in supporting risk-aware maritime decision-making.
(4) Insufficient maritime domain knowledge transfer. Existing research demonstrates inadequate transfer of structured maritime transportation knowledge into predictive models. In particular, current existing approaches fail to effectively assimilate multi-modal domain information, such as waterway restrictions, meteorological warnings, and vessel operation semantics. This decoupling between domain-specific semantics and numerical trajectory data impairs models' cognitive reasoning capabilities in complex navigation environments, ultimately constraining prediction robustness and interpretability.

3. Methodology

In the STP task, assuming there are B ships in the water area at time step t, the position of ship i at time step t is denoted as x_i^t. The observation time window is defined as T_{obs}, and the prediction time window is defined as T_{pred}. The observed trajectory, predicted and ground-truth future trajectory are formally defined in Eqs. (1) - (3), respectively:

$$X^{obs} = \{x_i^t \mid i \in \{1, 2, \dots, B\}, t \in \{t_0, t_0 + 1, \dots, t_0 + T_{obs} - 1\}\} \tag{1}$$

$$X^{pred} = \{\hat{x}_i^t \mid i \in \{1, 2, \dots, B\}, t \in \{t_0 + T_{obs}, t_0 + T_{obs} + 1, \dots, t_0 + T_{obs} + T_{pred} - 1\}\} \tag{2}$$

$$X^{gt} = \{x_i^t \mid i \in \{1, 2, \dots, B\}, t \in \{t_0 + T_{obs}, t_0 + T_{obs} + 1, \dots, t_0 + T_{obs} + T_{pred} - 1\}\} \tag{3}$$

where $X^{obs} \in \mathbb{R}^{B\times T_{obs}\times C}$ and $X^{pred} \in \mathbb{R}^{B\times T_{pred}\times C}$ represent the observed trajectories and predicted future trajectories of all B ships in the water area, respectively. C denotes the feature dimension of the trajectories. The goal of STP is to accurately forecast the future path of each ship within a given maritime scene using historical trajectory data. Mathematically, this task involves developing a prediction model $\mathcal{M}$ that can generate a predicted future trajectory $X^{pred} \in \mathbb{R}^{B\times T_{pred}\times C}$ that closely approximates the ground-truth future trajectory X^{gt}. This requires modelling the temporal, spatial, and behavioural dynamics of ships to achieve precise and reliable trajectory prediction.

This section elaborates on LLM4STP, an innovative LLM-driven STP model. By deeply integrating LLMs with maritime domain knowledge, this approach establishes a novel paradigm for STP, leveraging multi-dimensional feature fusion to enhance predictive accuracy. Section 3.1 presents an overview of the model architecture, emphasising its core components and innovative design features. Section 3.2 explores the use of prompt engineering to effectively incorporate diverse maritime data into structured prompts, enabling the model to leverage rich domain knowledge during prediction. Section 3.3 discusses the approach to uncertainty modelling, addressing the inherent unpredictability in ship movements. Section 3.4 introduces a Geohash-based method for capturing the geographical semantic continuity of ship trajectories, enhancing spatial awareness. Section 3.5 examines the spatial interaction modelling among vessels, capturing the influence of nearby ships on trajectory patterns. Section 3.6 details the approach for local temporal modelling, focusing on short-term behavioural patterns in ship trajectories. Section 3.7 outlines the complete feature fusion and decoding process, integrating multiple sources of information for accurate trajectory prediction.

3.1. Overview

STP presents significant challenges due to the high complexity and uncertainty inherent in maritime navigation. The LLM4STP framework, as illustrated in Fig. 4, is designed to enhance prediction accuracy by integrating multi-dimensional data through a structured three-stage approach, including a feature embedding module, a feature fusion and decoding module, and an output layer. The key components are described as follows:

(1) Prompt Embedder. STPs are influenced by diverse textual contexts, such as nautical zone descriptions, meteorological conditions, and waterway topology. Traditional models often struggle to capture the deep semantic relationships within this textual data. The Prompt Embedder addresses this limitation by leveraging pre-trained LLM tokenisers to convert textual descriptions into rich contextual embeddings, laying a robust foundation for semantic comprehension. This approach captures the nuanced impact of contextual factors on ship movements, enhancing overall prediction accuracy.
(2) Uncertainty Embedder. Maritime environments are inherently unpredictable, with trajectory uncertainties arising from factors such as weather conditions, traffic density, and ocean currents. This module systematically encodes these stochastic influences

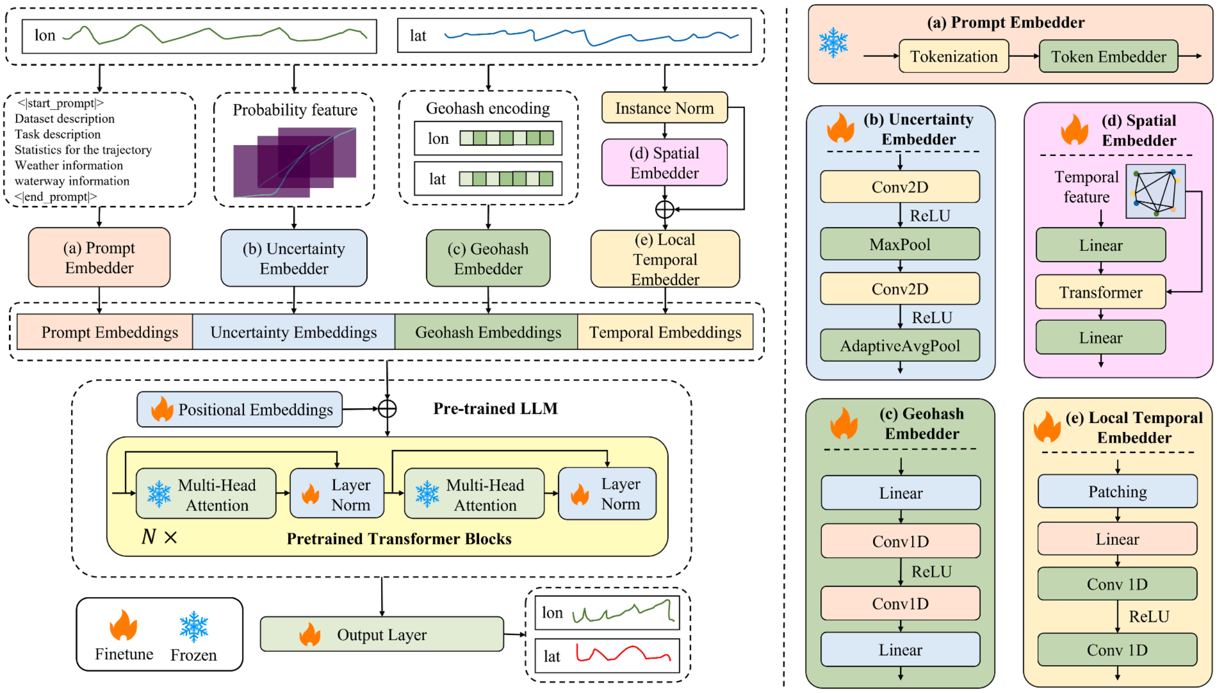

Fig. 4. The proposed LLM4STP framework.

using a CNN to extract features from probabilistic trajectory data. This module generates embedding vectors that explicitly represent uncertainty characteristics, ensuring the model effectively accounts for unpredictable elements in ship trajectories.

(3) Geohash Embedder. Accurate spatial representation is crucial for trajectory prediction. The Geohash Embedder hierarchically encodes geospatial information using Geohash algorithms, preserving locational hierarchies and neighbourhood relationships. By converting Geohash strings into structured embeddings, this module efficiently captures spatial features while resolving challenges associated with direct spatial data processing.

(4) Local Temporal Embedder. To capture fine-grained temporal dynamics, the Local Temporal Embedder partitions input sequences into localised patches, generating embeddings that reflect short-term, high-resolution temporal patterns. This localised processing enhances the model's sensitivity to transient spatiotemporal features, significantly improving its ability to predict rapid changes in ship movements.

(5) Spatial Embedder. Spatial dynamics, including relative ship positions, headings, and collision-avoidance behaviours, are captured using a Spatial Transformer network. This component learns complex spatial interactions and dependencies between vessels, enabling the model to account for interactive behaviours that are often overlooked in traditional approaches.

The concatenated embeddings from these modules are processed by the pre-trained LLM, which performs global trajectory modelling and cross-modal feature integration. The LLM's multi-head attention mechanisms remain frozen to preserve extensive language understanding, while fine-tuning is applied to adapt the model for the specific task of STP. This approach effectively integrates multimodal inputs, leveraging the LLM's powerful semantic comprehension and contextual modelling capabilities to produce refined trajectory representations. Finally, the output layer transforms these refined embeddings into accurate time-series predictions, completing the prediction pipeline.

3.2. Prompt embedder

Prompt embeddings play a crucial role in enhancing STP by providing the model with comprehensive contextual information about the dataset and prediction task. This approach integrates key details such as dataset descriptions, task specifications, weather conditions, channel characteristics, and statistical features of input trajectories, including minimum and maximum positions, median position, average speed, average acceleration, and acceleration trends. This structured information guides the model to consider richer semantic context during prediction, as illustrated in Fig. 5.

The prompt embedding process involves converting this diverse textual input into a structured numerical format that pre-trained LLMs can effectively process. Initially, the input text P is tokenised to generate a sequence of tokens $\{p_1, p_2, \ldots, p_n\}$, where n is the total number of tokens. Each token p_i is mapped to a d-dimensional embedding vector e_i through the pre-trained Token Embedder, forming an embedding matrix X_P as shown in Eq. (4).

$$X_P = TokenEmbedder(Tokenisation(P)) \quad (4)$$

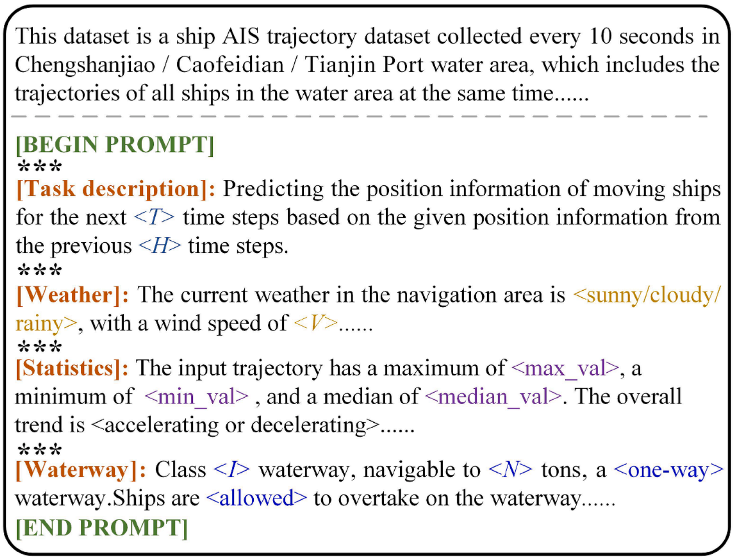

Fig. 5. Prompt example: < > is based on task-specific configuration.

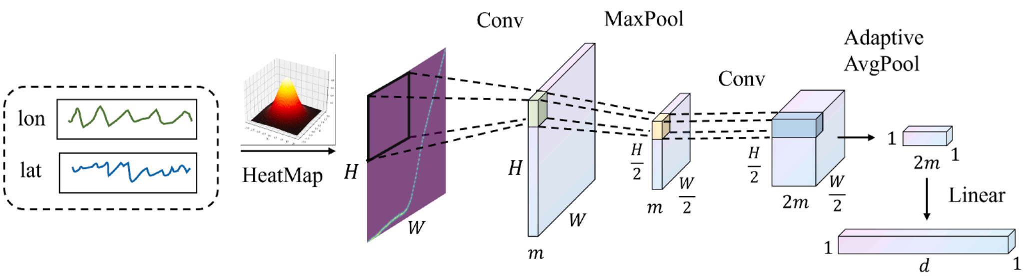

Fig. 6. Trajectory probability modelling.

This transformation converts the input text P into a structured numerical representation $X_P \in \mathbb{R}^{B\times 1\times d}$, where d is the dimension of the hidden layer of the pre-trained Transformer block. This process ensures that the model receives semantically enriched, contextually aware input features, enhancing its ability to capture the complex dynamics of ship trajectories.

3.3. Uncertainty embedder

In dynamic maritime environments, ship trajectories are subject to significant uncertainty due to factors like ocean currents, sudden wind shifts, and variable weather conditions. These stochastic influences can significantly impact prediction accuracy if not properly modelled. To address this challenge, the Uncertainty Embedder uses probabilistic modelling techniques to explicitly quantify trajectory uncertainties, providing a more robust foundation for trajectory prediction (Gong et al., 2025).

Typically, a two-dimensional Gaussian distribution is used to represent the probability density of a ship's position at a specific moment, capturing both the mean positions and the associated uncertainties in spatial coordinates, as expressed in Eq. (5).

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right) \tag{5}$$

where x and y is the longitude and latitude of ship. μ_x and μ_y are the means of the ship positions, and σ_x and σ_y represent the standard deviations.

This probabilistic representation is then processed through a CNN to generate uncertainty embeddings $X_U \in \mathbb{R}^{B\times 1\times d}$, effectively capturing the probabilistic nature of ship movements. The complete modelling process is outlined in Table 1 and illustrated in Fig. 6.

By analysing these probability distributions, the model gains a more nuanced understanding of the uncertainties inherent in ship trajectories, improving its ability to predict future paths under uncertain conditions.

Table 1
Trajectory uncertainty modelling network calculation process.

Layer	Input Size	Output Size	Description
Input	$T_{obs} \times C$	$H \times W \times 1$	Input heatmap from lon and lat
Conv2d	$H \times W \times 1$	$H \times W \times m$	Kernel size = 3×3, stride = 1, padding = 1
MaxPool	$H \times W \times m$	$H/2 \times W/2 \times 2m$	Max pooling with kernel size 2×2
Conv2d	$H/2 \times W/2 \times 2m$	$H/2 \times W/2 \times 2m$	Kernel size = 3×3, stride = 1, padding = 1
Adaptive AvgPool	$H/2 \times W/2 \times 2m$	$1 \times 1 \times 2m$	Adaptive average pooling to reduce size to 1×1
Linear Layer	$1 \times 1 \times 2m$	$1 \times 1 \times d$	Projects features to the output dimension

Note: H and W are the length and width of the heatmap, respectively. m is the number of channels.

Table 2
Geohash modelling network structure.

Layer	Input Size	Output Size	Description
Linear Layer	$B \times T_{obs} \times Geosize$	$B \times T_{obs} \times M$	Transforms input features to a hidden representation
Conv1D Layer	$B \times T_{obs} \times M$	$B \times 32 \times M$	Applies 1D convolution to capture local patterns
ReLU Activation	$B \times 32 \times M$	$B \times 32 \times M$	Introduces non-linearity to the model
Conv1D Layer	$B \times 32 \times M$	$B \times 1 \times M$	Further processes features with another convolution
Linear Layer	$B \times 1 \times M$	$B \times 1 \times d$	Projects features to the output dimension

Note: M is the hidden size and d is the output size.

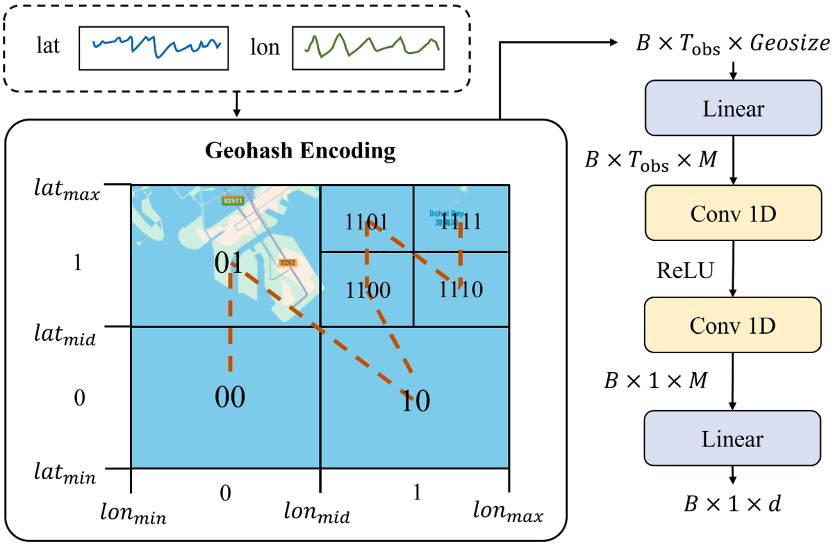

Fig. 7. Geohash modelling.

3.4. Geohash embedder

Geospatial data is a critical component in ship trajectory analysis, as it captures the locational context necessary for accurate prediction. The Geohash Embedder converts continuous geographic coordinates into discrete, grid-based representations using the Geohash encoding system. This approach effectively partitions the Earth's surface into grid cells, each represented by a unique alphanumeric code, facilitating efficient spatial data processing. A schematic diagram of the GeoHash modelling process is illustrated in Fig. 7.

The Geohash encoding process captures hierarchical spatial relationships by progressively subdividing latitude and longitude coordinates, generating variable-length strings where character count reflects spatial resolution. The encoded data $X_{Geo} \in \mathbb{R}^{B \times T_{\mathrm{obs}} \times Geosize}$ is processed using a multi-stage network architecture, as described in Algorithm 1 and detailed in Table 2. This architecture includes a linear projection layer for high-dimensional feature extraction, followed by 1D convolutional layers paired with ReLU activation functions to capture fine-grained spatiotemporal patterns.

This network is designed to perform multi-level feature extraction and transformation, converting raw Geohash-encoded trajectories into spatiotemporally rich representations, denoted as $X_G \in \mathbb{R}^{B \times 1 \times d}$, which are subsequently fed into the LLM as formulated

Algorithm 1 Geohash encoding for trajectory data.

Input: Trajectory data $traj \in \mathbb{R}^{B\times T_{obs}\times 2}$, precision p
Output: Encoded trajectory tensor $encoded\_traj \in \mathbb{R}^{B\times T_{obs}\times p\times 2}$

Extract latitude and longitude sequences from $traj$
Latitude Encoding:
Initialize minimum and maximum values for latitude: $lat\_\mathrm{min} = 0$, $lat\_\mathrm{max} = 1$
Initialize a Boolean tensor lat_bits to store the binary representation
for each bit position i from 1 to p **do**
    Calculate midpoint $lat\_mid = \frac{lat\_\mathrm{min}+lat\_\mathrm{max}}{2}$
    Create a mask where latitude values are greater than lat_mid
    Update $lat\_bits[i]$ with the mask
    Update lat_min and lat_max based on the mask
end for
Longitude Encoding:
Initialize minimum and maximum values for longitude: $lon\_\mathrm{min} = 0$, $lon\_\mathrm{max} = 1$
Initialize a Boolean tensor lon_bits to store the binary representation
for each bit position i from 1 to p **do**
    Calculate midpoint $lon\_mid = \frac{lon\_\mathrm{min}+lon\_\mathrm{max}}{2}$
    Create a mask where longitude values are greater than lon_mid
    Update $lon\_bits[i]$ with the mask
    Update lon_min and lon_max based on the mask
end for
Combine the binary representations: $encoded\_traj = \mathrm{stack}([lat\_bits, lon\_bits], dim = -1)$
return $encoded\_traj$

in Eq. (6). This approach effectively captures navigation patterns and vessel behavioural dynamics, significantly improving prediction accuracy and reliability in maritime trajectory analysis. The hierarchical architecture systematically abstracts both fine-grained spatial variations and long-term temporal dependencies, providing a robust foundation for intelligent maritime surveillance and decision-making systems.

$$X_G = Conv1D(ReLU(Conv1D(Linear(X_{Geo})))) \tag{6}$$

3.5. Spatial embedder

Ship navigation involves complex, dynamically evolving spatial interactions between vessels, making it essential to capture these interdependencies for accurate trajectory prediction. The Spatial Embedder addresses this by integrating a spatial Transformer architecture with a Sparse Attention Gated (SAG) module, as depicted in Fig. 8. This module autonomously generates adaptive graph masks to represent ship-to-ship spatial correlations without requiring prior knowledge of inter-ship distances. Its self-attention mechanism dynamically weights pairwise interactions through multi-head attention operations, effectively aggregating multi-dimensional feature information across ship trajectories.

Given an input trajectory X^{obs}, the Query (Q), Key (K) and Value (V) matrix are calculated from X^{obs} by linear transformation.

$$Q, K, V = X^{obs} W_o + b_o \tag{7}$$

where W_o is the learnable weight matric and b_o is the bias term.

The spatial attention score is calculated as Eq. (8):

$$Attn_s = Softmax\left(\frac{QK^T}{\sqrt{d}}\right) \tag{8}$$

where $\sqrt{d}$ is a scaling factor used to stabilise the gradient, and $Attn_s$ is the interaction matrix between ships.

In the maritime traffic scenario, the SAG selectively masks irrelevant interactions in the interaction matrix, reducing unnecessary calculations and improving model efficiency. The SAG calculation formula is as follows:

$$w = ReLU(Attn_s - Sigmoid(Conv(Attn_s))) \tag{9}$$

where w is the output of SAG module.

The interaction matrix $Attn_s$ undergoes convolution to capture regional interaction patterns, with a sigmoid function adaptively generating a threshold for each interaction pair. This threshold differentiates significant and minor interactions: pairs with attention scores above the threshold are considered impactful, while those below are suppressed to reduce noise. To preserve the initial sparsity and prevent gradient interference in the zero-value regions, a Norm-Softmax function is applied. This approach locks the sparse

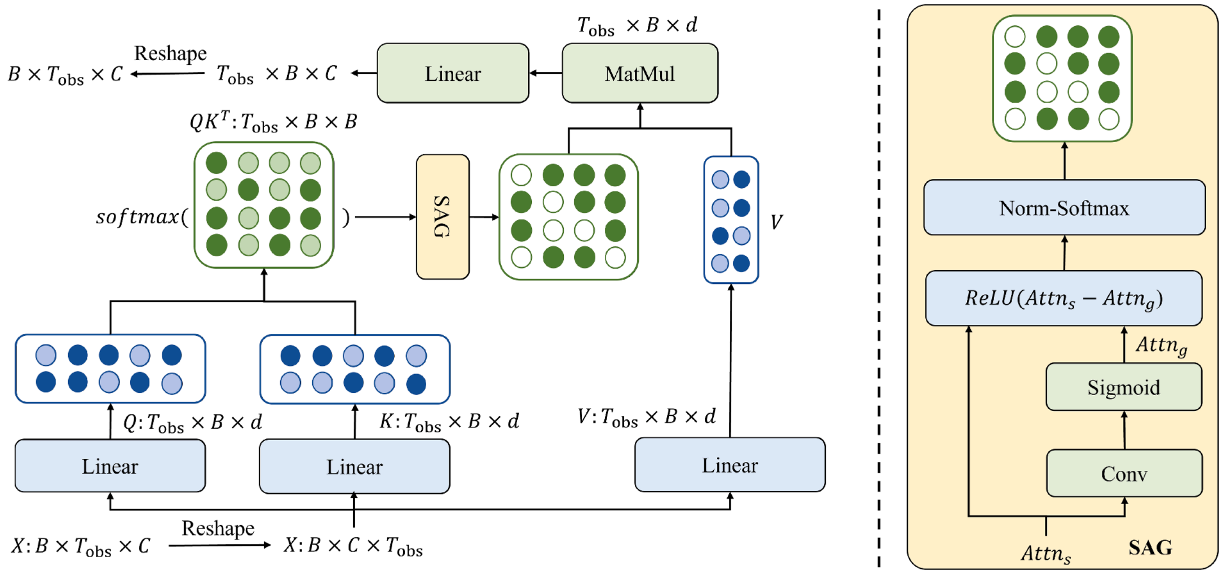

Fig. 8. Spatial embedder.

structure, avoiding attention leakage from numerical perturbations while accurately redistributing the non-zero weights. The specific calculation is as follows:

$$w_n = NormSoftmax(w) = \frac{e_j^w - 1}{\sum_j^N (e_j^w - 1) + \varepsilon} \tag{10}$$

where ε prevents division by 0, N represents the number of ships at the same time. w_n is the output of the Norm-Softmax function.

The Norm-SoftMax function ensures sparsity, and the generated sparse interaction weight matrix is multiplied by V and performs linear projection to obtain the representation of the spatial relationship feature of the fused ship $X_s \in \mathbb{R}^{B \times T_{obs} \times C}$, as shown in Eq. (11).

$$X_s = (w_n V) W_s + b_s \tag{11}$$

where W_s is the weight matric and b_s is the bias term.

3.6. Local temporal embedder

The Local Temporal Embedder is designed to capture short-term temporal dynamics in ship trajectory data while reducing computational complexity. It achieves this by partitioning the time series into fixed-length local windows, which focus on short-term interactions and trends. This approach improves training efficiency and model performance by concentrating on localised temporal variations rather than the entire sequence. The main steps in this process are illustrated in Fig. 9 and described as follows:

(1) Instance normalisation. The input longitude/latitude (lon/lat) time-series data is first processed through instance normalisation to eliminate scale discrepancies and stabilise the training process. This step removes the influence of outliers and ensures consistency in the temporal data, resulting in normalized features, $X_{norm} \in \mathbb{R}^{B \times T_{obs} \times C}$, as shown in Eq. (12).

$$X_{norm} = InstanceNorm(X^{obs}) + X_s \tag{12}$$

(2) Patching for local time window processing. The normalised sequence X_{norm} is then segmented into smaller, fixed-length patches, allowing the model to focus on localised temporal dependencies. The patching operation divides the time series into multiple windows, each of length P, and the number of windows is $w_n = \left\lfloor \frac{L-P}{S} \right\rfloor + 1$. S represents the step size between windows. The specific calculation of X_P is shown in Eq. (13).

$$X_P = Patching(X_{norm}) \tag{13}$$

where $X_P \in \mathbb{R}^{B \times w_n \times Patch}$, $Patch$ represents the length of each patch.

(3) Local temporal feature extraction. To capture the short-term temporal dynamics within each patch, the patch embeddings are processed through two layers of 1D convolution with ReLU activation, as shown in Eq. (14).

$$X_t = Conv1D(ReLU(Conv1D(Linear(X_P)))) \tag{14}$$

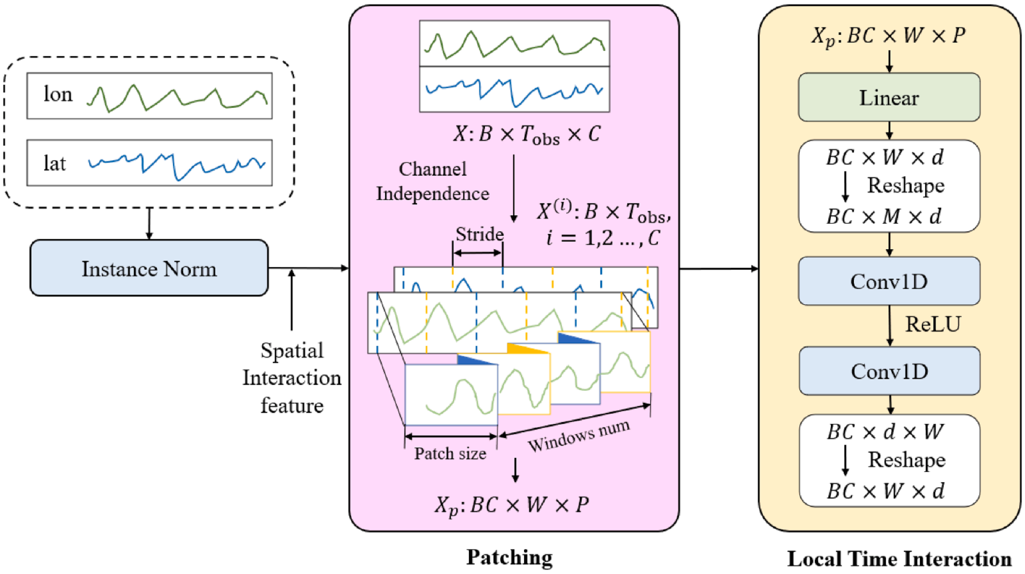

Fig. 9. Local time interaction.

This multi-layer convolution captures the local temporal relationships, generating the final temporal embeddings $X_t \in \mathbb{R}^{B\times w_n\times d}$, effectively reducing the sequence length from L to w_n and enhancing the model's ability to capture both short-term variations and long-term temporal dependencies.

3.7. Feature decoding and prediction

The framework generates four distinct embedding representations: 1) Prompt embeddings X_P encode contextual information, including waterway descriptions, meteorological data, and trajectory background, providing the model with essential semantic context, 2) Uncertainty embeddings X_U contextual information, including waterway descriptions, meteorological data, and trajectory background, providing the model with essential semantic context, 3) Geohash embeddings X_G encode spatial information using hierarchical grid-based encoding, preserving geospatial relationships and spatial semantic continuity, and 4) Local Temporal embeddings X_t focus on short-term temporal patterns, capturing localized, time-dependent interactions within the trajectory sequence. These multimodal embeddings are concatenated along the feature dimension to form the composite embedding $E \in \mathbb{R}^{B\times(w_n+3)\times d}$.

The composite embedding E is then processed by a pre-trained LLM architecture to model global temporal dependencies using multi-head attention (MHA) mechanisms. E is linearly projected into Q, K, and V matrices, as defined in Eq. (15):

$$Q = EW_Q, \quad K = EW_K, \quad V = EW_V \tag{15}$$

where W_Q, W_K, W_V are learnable weight matrices that transform the embeddings into multiple attention heads, allowing parallel processing of spatial and temporal relationships. Each head independently calculates the attention score, capturing different aspects of the input sequence. The outputs from all attention heads are concatenated and linearly transformed to obtain the final result X_m, as shown in Eq. (16).

$$X_m = MHA(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{16}$$

Finally, X_m is processed through the feed-forward network (FFN) to obtain the final prediction result X^{pred}, as shown in Eq. (17).

$$X^{pred} = FFN(X_m) = ReLU(X_m W_1 + b_1)W_2 + b_2 \tag{17}$$

where W_1 and W_2 are the weight matrices and b_1 and b_2 are the bias terms.

To ensure task-specific adaptation without overfitting, the framework freezes most of the parameters in the pre-trained LLM, including the MHA and FFN, while only fine-tuning position embeddings and layer normalisation parameters. This approach maintains the generalisation capabilities of the pre-trained model while adapting it to the unique characteristics of STP.

The model is trained using the Mean Squared Error (MSE) as a loss function, quantifying the average squared difference between the predicted and actual trajectory points. MSE provides a precise evaluation metric for assessing the model's prediction accuracy, as defined in Eq. (18).

$$\mathcal{L}_{MSE} = (X^{pred} - X^{gt})^2 \tag{18}$$

4. Experimental results and analysis

4.1. Experimental datasets and dataset preprocessing

The study leverages AIS data collected in June 2022 from three critical maritime areas in China's Bohai Bay: the CFD water area (longitude 118°25'E–118°92'E, latitude 38°72'N-39°10'N), the CSJ water area (longitude 122°58'E-123°17'E, latitude 37°16'N-37°75'N), and the TJP water area (longitude 117°7'E-118°7'E, latitude 38°7'N-39°1'N). This study selects the Caofeidian (CFD), Chengshan Jiao Promontory (CSJ), and Tianjin Port (TJP) water areas in China's Bohai Bay as study regions for STP. The three datasets are visualised in Fig. 10. These areas are chosen for their unique features and significance in maritime transportation. CFD and TJP are major ports with high traffic density and diverse vessel types, creating complex navigation scenarios. CSJ, as a promontory, imposes geographical constraints that shape vessel routes, providing a controlled environment for studying regulated traffic flow. The high-quality AIS data from these areas, after preprocessing, ensures reliable input for model training. Collectively, they cover port operations, coastal navigation, and open-water convergence zones, offering comprehensive datasets to address challenges in maritime safety and efficiency. The raw AIS dataset comprises both dynamic attributes (e.g., latitude/longitude, speed, course) and static attributes (e.g., MMSI number, vessel type), which were refined into a high-reliability trajectory dataset through a multi-stage preprocessing pipeline.

To address inherent AIS data challenges, such as noise, missing values, and spatiotemporal inconsistencies, a comprehensive preprocessing strategy was employed. First, kinematic constraints were applied via a multidimensional threshold detection method. This included a 3σ-criterion to filter out anomalous position data, such as points located beyond port boundaries or on land, and physical thresholds to eliminate unrealistic speed records (e.g., speeds exceeding 50 knots). Additionally, only navigation status

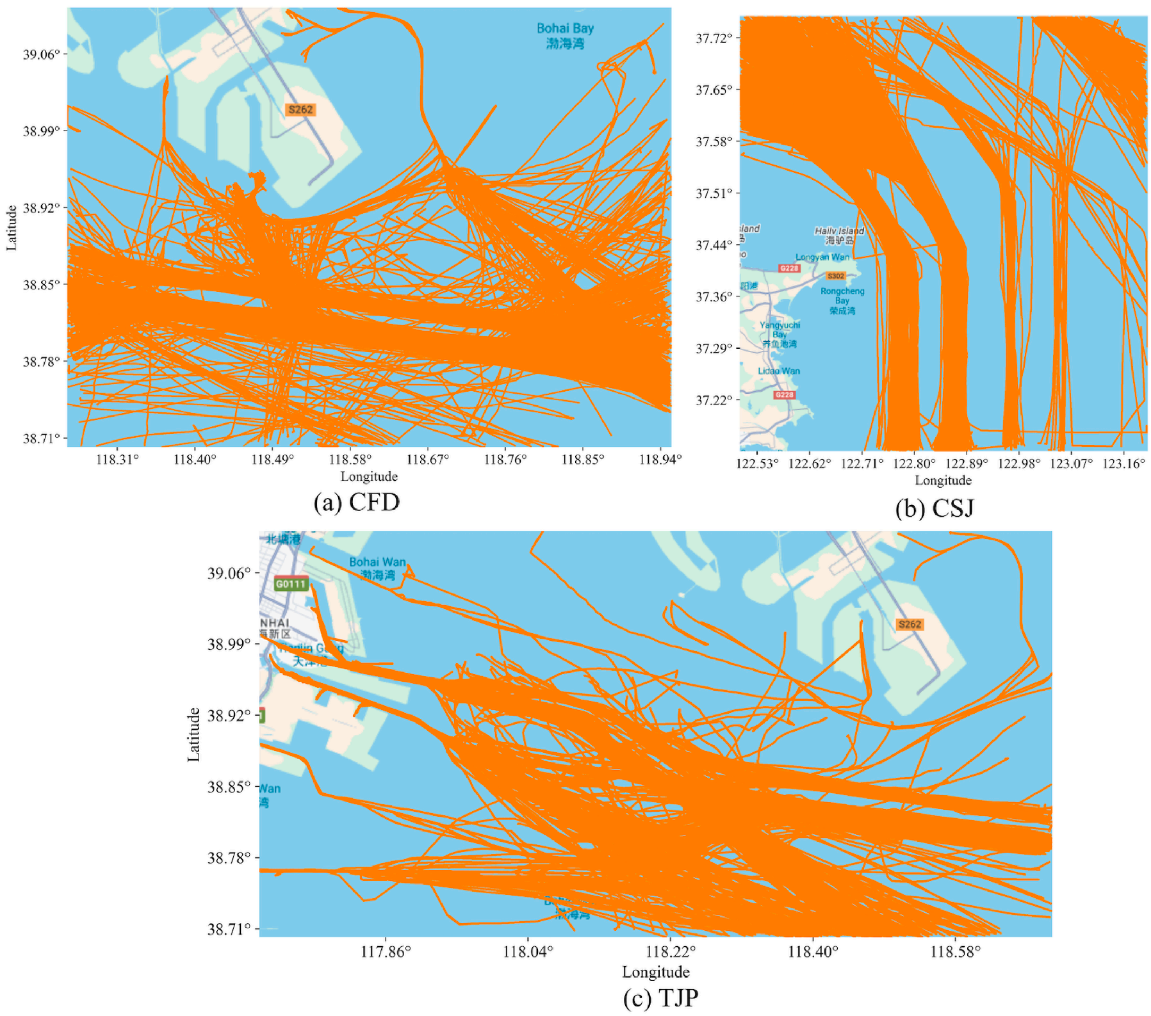

Fig. 10. Dataset visualisation.

Table 3
Baseline Models for STP.

Model	Core Mechanism	Reference
Bi-LSTM	Bidirectional LSTM capture temporal dependencies from both directions.	(Schuster and Paliwal, 1997)
Bi-GRU	Bidirectional GRU mitigate gradient vanishing through gating mechanisms.	(Rana, 2016)
Mamba	Efficient long-sequence modelling via state space models.	(Gu and Dao, 2023)
Transformer	Self-attention mechanism captures global spatiotemporal dependencies.	(Vaswani et al., 2017)
STAR	Spatiotemporal attention dynamically integrates spatiotemporal features.	(Yu et al., 2020)
Social-LSTM	Extends LSTM with social pooling layers to model group interactions.	(Alahi et al., 2016)
Social-STGCNN	Spatiotemporal graph convolution combine social interactions.	(Mohamed et al., 2020)
ST-MGT	Integrates graph convolution and Transformer for dynamic topology learning.	(Liu et al., 2024c)
iTransformer	Enhances Transformer with channel-independent time series modelling.	(Liu et al., 2023)

fields indicating active movement (“Under way using engine” and “Under way sailing”) were retained, excluding stationary data from anchored or moored vessels. Second, trajectories with over 20 % missing values or time gaps exceeding 5 m were classified as low-quality and removed from the dataset. For the remaining high-quality trajectories, cubic spline interpolation was used to mitigate temporal discontinuities caused by inconsistent AIS broadcast frequencies. Finally, the interpolated trajectories were resampled at a uniform 10-s interval to ensure temporal consistency across the dataset, providing a robust foundation for subsequent trajectory prediction analysis.

4.2. Evaluate metrics and baselines

To systematically evaluate the performance of LLM4STP, this study selected 9 baseline methods spanning three categories: classical time-series models, attention-based models, and spatiotemporal graph neural networks, as outlined in Table 3. All baseline models were trained and evaluated on the same dataset under identical experimental conditions, strictly adhering to the parameters and configurations specified in their original papers to ensure a fair and reproducible comparison. The learning rate for the experiment was set to 1e-4, and the Adam optimiser was used during training. The LLM backbone used in LLM4STP is the pre-trained GPT-2 (Small), configured with a hidden size of 768 and three hidden layers. The choice of three hidden layers was made based on experimental results, as detailed in Appendix A. The $Geosize$ value was set to 40.

To ensure a robust and unbiased evaluation, all comparison models were trained on the designated training set, with the validation set used solely for hyperparameter tuning and monitoring convergence, and the test set reserved for final performance assessment. This procedure provides a fair measure of each model’s generalisation capability. To preserve the temporal integrity of vessel movements, the earliest timestamp across all trajectories was taken as the start point and the latest as the endpoint, and the data were partitioned chronologically into training (70 %), validation (20 %), and test (10 %) sets. This sequential split prevents future data from leaking into the training process and is particularly critical for spatiotemporal prediction tasks. By maintaining this natural temporal order, the reliability of the model’s predictive performance is preserved.

The evaluation framework incorporates three key metrics to capture the spatio-temporal consistency and accuracy of the predicted trajectories from multiple perspectives: ADE, FDE, and FD. ADE measures the overall path alignment by calculating the average Euclidean distance between predicted and actual trajectories across all time steps, providing a comprehensive assessment of the model’s path-following accuracy. FDE focuses specifically on the final positioning accuracy, measuring the Euclidean distance between the predicted endpoint and the true target location, making it particularly relevant for applications where precise endpoint estimation is critical. FD serves as a composite metric that jointly considers both spatial proximity and temporal consistency. It evaluates the geometric similarity and motion coherence of the predicted trajectories, offering a more holistic assessment of the model’s ability to capture complex path dynamics. These indicators jointly construct a multilevel evaluation system, providing a quantitative basis for model optimisation and comparative research. The mathematical formulations for these metrics are listed in Eqs. (19) and (21).

$$ADE = \frac{1}{T}\sum_{t=1}^{T}\sqrt{(pred_x^t - gt_x^t)^2 + (pred_y^t - gt_y^t)^2} \tag{19}$$

$$FDE = \sqrt{(pred_x^T - gt_x^T)^2 + (pred_y^T - gt_y^T)^2} \tag{20}$$

$$FD = \min_{\pi\in\Pi_T}\max_{t=1}^{T}\sqrt{(pred_x^t - gt_x^t)^2 + (pred_y^t - gt_y^t)^2} \tag{21}$$

where T denotes the total number of time steps in the predicted trajectory, while $pred$ and gt represent the predicted trajectory generated by the model and ground-truth trajectory, respectively.

4.3. Quantitative analysis

The proposed LLM4STP framework demonstrates superior performance in STP by integrating multimodal feature embedding with pre-trained language models. As demonstrated in Table 4, this model achieves SOTA results on all evaluation metrics (ADE, FDE, and FD) for both CFD and CSJ datasets, while maintaining competitive second-best performance on the TJP dataset. With 65M parameters

Table 4
Comparison of the trajectory prediction results with SOTA methods on three waters for ADE, FDE and FD (in nautical miles).

Model	CFD			CSJ			TJP		
	ADE	FDE	FD	ADE	FDE	FD	ADE	FDE	FD
Bi-LSTM	0.24	0.41	0.41	0.27	0.46	0.46	0.26	0.49	0.50
Bi-GRU	0.23	0.40	0.41	0.24	0.43	0.43	0.27	0.51	0.51
Mamba	0.18	0.36	0.37	0.15	0.30	0.30	0.21	0.38	0.39
Transformer	0.22	0.38	0.38	0.22	0.43	0.45	0.26	0.48	0.48
STAR	0.17	0.33	0.34	0.14	0.28	0.28	0.20	0.44	0.45
Social-LSTM	0.19	0.37	0.37	0.21	0.39	0.40	0.23	0.42	0.43
Social-STGCNN	0.16	0.33	0.34	0.15	0.26	0.27	0.19	0.38	0.39
ST-MGT	0.14	0.30	0.31	0.13	0.27	0.27	0.16	0.36	0.36
iTransformer	<u>0.12</u>	<u>0.28</u>	<u>0.28</u>	<u>0.11</u>	**0.26**	<u>0.27</u>	**0.11**	**0.30**	**0.31**
LLM4STP (Ours)	**0.09**	**0.26**	**0.27**	**0.10**	**0.26**	**0.26**	<u>0.12</u>	<u>0.31</u>	**0.31**

and an average inference time of 32ms (measured on an RTX 3090), the LLM4STP model is optimised for real-time maritime prediction tasks, ensuring both accuracy and efficiency. This performance advantage can be attributed to the following key components:

(1) Prompt Embedder module leverages a pre-trained tokeniser to deeply encode the semantic context of waterway topology and meteorological information. It reduces ADE to 0.09 and 0.10 on the CFD and CSJ datasets, respectively, representing over a 62.5 % error reduction compared to conventional Bi-LSTM models. This result highlights the critical importance of contextual understanding for accurate maritime environment perception.

(2) By employing Gaussian distribution heatmaps combined with CNN architectures, the Uncertainty Embedder module captures the probabilistic distribution of ship trajectories, achieving an FDE of 0.26 on the CFD dataset-an improvement of 6.9 % over the second-best iTransformer model. This design effectively mitigates the impact of external disturbances, such as sudden weather changes, by incorporating robust probabilistic modelling.

(3) Geohash-Spatial Embedder synergistically integrates hierarchical geocoding with spatial Transformer networks, significantly enhancing the model's effectiveness in modelling spatial relationships. It achieves an FD of 0.26 on the CSJ dataset, outperforming the graph convolution-based Social-STGCNN model by 7.1 %, thereby providing a more comprehensive representation of complex ship movement patterns.

Collectively, these innovations address critical gaps in conventional trajectory prediction methods by improving semantic context comprehension, uncertainty quantification, and spatial interaction representation. The LLM4STP framework sets a new benchmark for predictive accuracy in complex marine environments, offering a robust solution for next-generation maritime intelligence systems.

4.4. Few-shot learning analysis

LLMs, trained on vast and heterogeneous corpora, exhibit a pronounced capacity for few-shot learning, enabling generalisation from a small number of training samples. This capability arises from their extensive pre-training, which imparts a rich representation of structural and semantic patterns, allowing accurate inference even under conditions of data scarcity.

To validate the few-shot learning capability of LLM4STP, experiments were conducted on the TJP dataset. The results are shown in Table 5, revealing the following insights:

(1) Monotonic performance improvement: Model accuracy improves consistently across all key metrics (ADE, FDE, FD) as the proportion of training data increases from 5 % to 20 %, reflecting the expected benefit of additional data in capturing trajectory patterns and reducing generalisation error.

(2) Exceptional data efficiency: LLM4STP achieves performance statistically comparable to models trained on the full dataset when utilising only 20 % of the available data. This result highlights its superior capacity for knowledge transfer and efficient utilisation of limited training samples.

The demonstrated few-shot learning ability of LLM4STP is of particular significance in the maritime domain, where trajectory data may be incomplete, sparse, or subject to reporting delays. The capacity to maintain high predictive accuracy under such constraints enables more reliable real-time decision-making, enhances navigational safety, and supports the development of data-efficient intelligent maritime transportation systems.

4.5. Visualisation analysis

The visualisation analysis further supports the effectiveness of the LLM4STP framework in accurately predicting ship trajectories under complex maritime conditions. As demonstrated in Figs. 11 and 13, which cover three distinct waterways, LLM4STP (represented by green lines) consistently generates predicted trajectories that align closely with the actual ground truth paths (represented by red dashed lines). This alignment is particularly evident at critical navigational points, such as waterway intersections, sharp turns, and port approach zones.

Several key observations from the visualisation results reinforce the advantages of the LLM4STP framework:

Table 5
Performance comparison under different missing ratios (5 %, 10 %, 20 %).

Methods	5 %			10 %			20 %		
	ADE	FDE	FD	ADE	FDE	FD	ADE	FDE	FD
Bi-LSTM	0.864	1.321	1.331	0.785	1.287	1.291	0.575	0.874	0.874
Bi-GRU	0.844	1.226	1.229	0.775	1.252	1.291	0.567	0.862	0.869
Mamba	0.721	1.078	1.278	0.679	1.071	1.085	0.472	0.795	0.799
Transformer	1.101	1.612	1.623	0.954	1.304	1.305	0.694	1.047	1.054
STAR	1.021	1.507	1.507	0.794	1.104	1.108	0.554	0.847	0.847
Social-LSTM	0.759	1.085	1.089	0.641	0.957	0.953	0.421	0.638	0.638
Social-STGCNN	0.527	0.831	0.835	0.414	0.687	0.683	0.251	0.478	0.481
ST-MGT	0.692	1.037	1.037	0.597	0.883	0.883	0.404	0.695	0.695
iTransformer	0.534	0.847	0.847	0.515	0.817	0.817	0.247	0.472	0.472
LLM4STP (Ours)	**0.133**	**0.356**	**0.356**	**0.128**	**0.322**	**0.322**	**0.126**	**0.314**	**0.316**

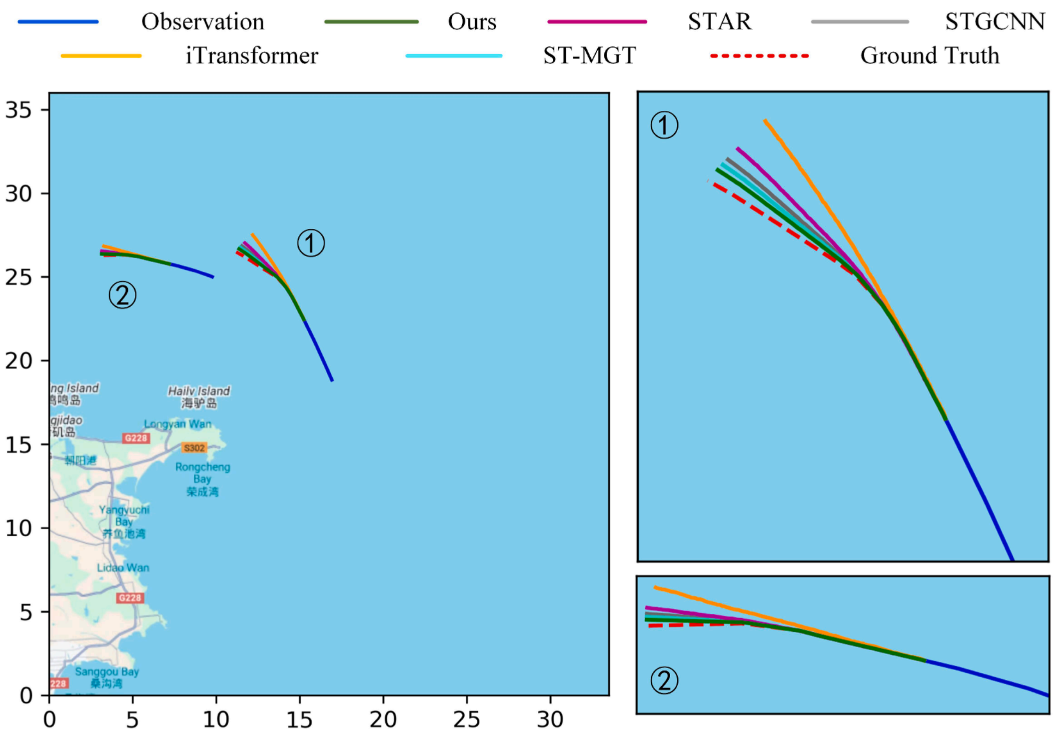

Fig. 11. Visualisation of trajectory prediction results in the CSJ water area dataset.

(1) Superior localisation and spatial awareness. The Geohash Embedder effectively converts latitude-longitude coordinates into hierarchical grid structures, preserving fine-grained spatial relationships. This approach significantly improves localisation accuracy in dense traffic areas, where precise positioning is crucial. For example, in the CSJ (Fig. 11) and TJP (Fig. 13) water areas, LLM4STP demonstrates superior vessel trajectory alignment, reflecting its capacity to accurately capture localised spatial patterns that are often inadequately represented by conventional models.

(2) Dynamic interaction modelling. The Spatial Embedder incorporates self-attention mechanisms that capture complex, dynamic inter-vessel interactions without requiring explicit distance thresholds. This design enables the model to autonomously learn context-specific collision-avoidance protocols, reducing the risk of trajectory overlap and navigational conflicts. In a two-ship interaction scenario within CSJ waters (Fig. 11), LLM4STP produces smoother and more accurate aligned trajectories compared with baseline models, including STGCNN and STAR.

(3) Long-horizon prediction stability. Unlike conventional models, which often exhibit erratic deviations at extended prediction horizons, LLM4STP maintains trajectory continuity beyond 6-nautical-mile forecasts. This stability is critical for long-term route planning and real-time traffic management, offering significant operational benefits for autonomous shipping systems. In contrast, models such as STAR (purple curves) and STGCNN (black curves) display substantial directional drift, indicating weaker long-term spatial coherence and limited adaptability to complex maritime environments.

Overall, these visual comparisons clearly demonstrate that LLM4STP effectively captures both short-term manoeuvring behaviours and long-term navigational intent, providing robust, context-aware trajectory predictions. This capability underscores the value of

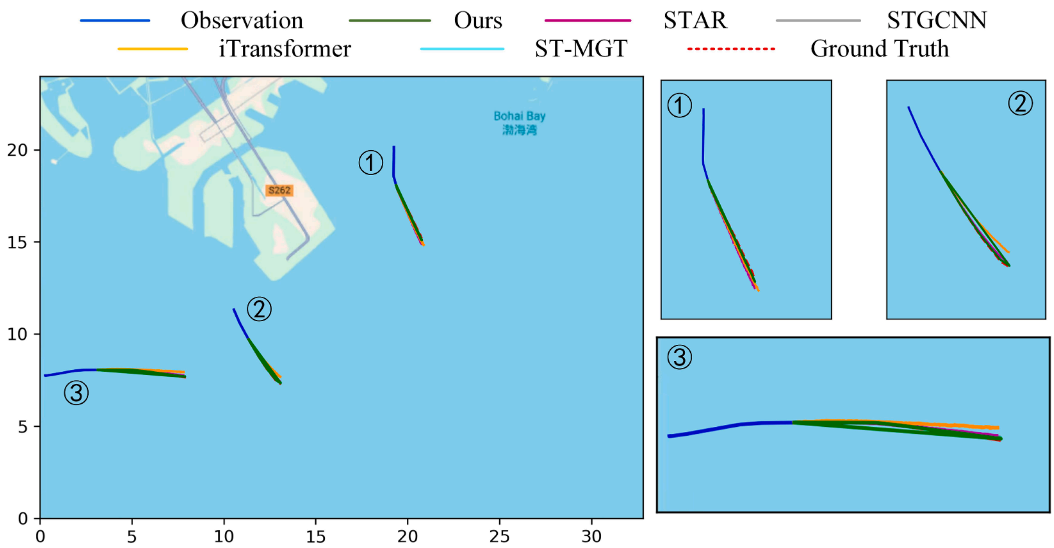

Fig. 12. Visualisation of trajectory prediction results in the CFD water area dataset.

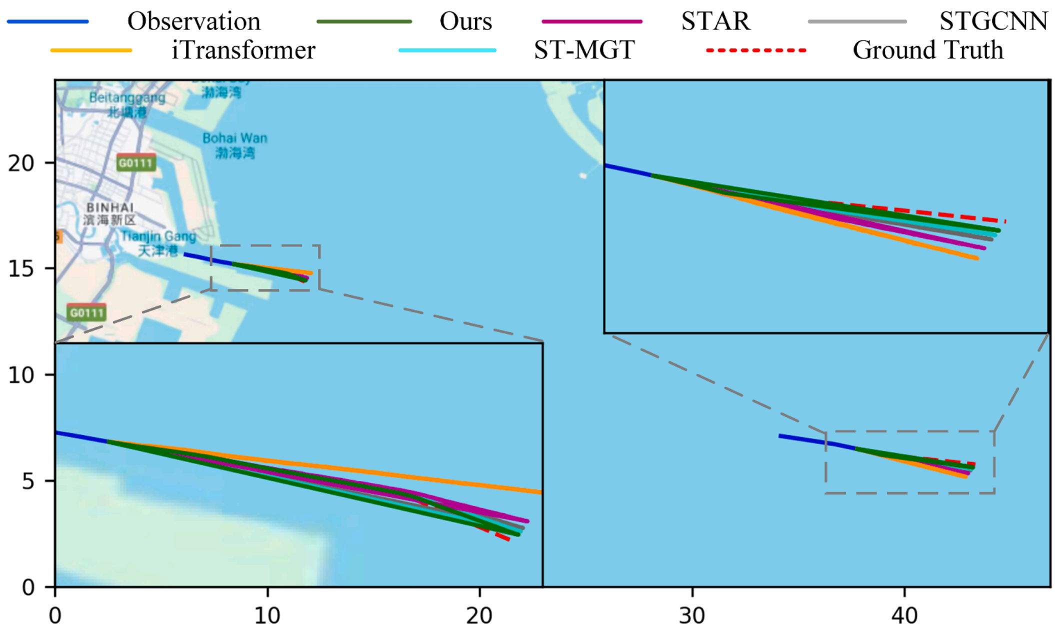

Fig. 13. Visualisation of trajectory prediction results in the TJP water area dataset.

the model's multi-modal feature fusion and pre-trained knowledge transfer mechanisms, positioning LLM4STP as a leading solution for next-generation maritime traffic management and autonomous navigation systems.

4.6. Ablation study

To comprehensively assess the contributions of each module within the LLM4STP framework, an ablation study was conducted, comparing the full model against five distinct variants. These variants were designed by systematically removing one embedding module at a time, allowing for a precise evaluation of each component's impact on overall model performance across multiple maritime datasets. The tested variants include:

(1) LLM4STP-w/o-P: Excludes the Prompt Embedder responsible for integrating contextual semantic information from waterway descriptions, meteorological data, and operational background.
(2) LLM4STP-w/o-U: Excludes the Uncertainty Embedder, which captures probabilistic features and prediction confidence, crucial for handling dynamic and uncertain maritime environments.

Table 6
The result of the ablation study.

Model	CFD			CSJ			TJP		
	ADE	FDE	FD	ADE	FDE	FD	ADE	FDE	FD
LLM4STP-w/o-P	**0.09**	0.27	**0.27**	**0.10**	**0.26**	0.27	**0.12**	0.33	0.33
LLM4STP-w/o-U	0.11	0.29	0.29	0.12	0.30	0.31	0.13	0.34	0.35
LLM4STP-w/o-S	0.10	0.28	0.28	0.11	0.28	0.28	0.14	0.35	0.35
LLM4STP-w/o-G	0.10	0.28	0.28	**0.10**	**0.26**	0.27	0.13	0.33	0.34
LLM4STP-w/o-L	0.13	0.30	0.31	0.13	0.32	0.33	0.15	0.37	0.37
LLM4STP	**0.09**	**0.26**	**0.27**	**0.10**	**0.26**	**0.26**	**0.12**	**0.32**	**0.32**

(3) LLM4STP-w/o-S: Excludes the Spatial Embedder, which models inter-vessel spatial interactions through hierarchical geocoding and self-attention mechanisms.
(4) LLM4STP-w/o-G: Excludes the Geohash Embedder, which encodes geospatial semantic continuity through hierarchical grid encoding, preserving fine-grained spatial relationships.
(5) LLM4STP-w/o-L: Excludes the Local Temporal Embedder, which captures short-term temporal dependencies and vessel manoeuvring dynamics.

The removal of the Prompt Embedder (w/o-P) had a relatively minor impact on the ADE for the CFD and CSJ datasets, reflecting the robustness of the core spatial and temporal encoding mechanisms. However, on the TJP dataset, the absence of the Prompt Embedder increased the Final Displacement Error (FDE) from 0.32 to 0.33 (+3.1 %), indicating that textual semantic cues are critical for accurately capturing complex waterway rules and navigation constraints.

Excluding the Uncertainty Embedder (w/o-U) resulted in the most substantial increase in FDE across all datasets, with an average rise of 9.2 %. The TJP dataset experienced the most pronounced impact, with FDE increasing from 0.32 to 0.35 (+9.4 %). This highlights the critical role of probabilistic uncertainty modelling in maintaining trajectory prediction reliability under variable environmental conditions, such as sudden weather changes and unpredictable vessel movements.

The Spatial Embedder (w/o-S) demonstrated a significant influence on trajectory accuracy, particularly in high-density traffic scenarios. Its removal increased the ADE on the TJP dataset from 0.12 to 0.14 (+16.7 %), suggesting that the spatial Transformer network is more crucial for capturing vessel interaction behaviours than the static geospatial encoding provided by the Geohash Embedder (w/o-G), which only caused an 8.3 % increase in ADE. This indicates that dynamic spatial feature extraction is essential for modelling real-time vessel interactions, where precise collision-avoidance and proximity awareness are critical.

The Local Temporal Embedder (w/o-L) was found to be the most critical module, as its exclusion resulted in the most substantial performance degradation across all metrics. For instance, on the CFD dataset, the FDE increased from 0.27 to 0.31 (+14.8 %), a 2–3 times greater impact than the removal of any other module. This underscores the decisive role of fine-grained temporal segmentation in capturing vessel manoeuvring patterns, particularly in congested or complex maritime environments.

The results demonstrate that each component of the LLM4STP framework plays a distinct yet complementary role in enhancing prediction accuracy. The Prompt Embedder facilitates nuanced contextual understanding, the Uncertainty Embedder ensures robust performance in dynamic settings, the Spatial and Geohash Embedders capture critical spatial relationships, and the Local Temporal Embedder is crucial for short-term behavioural modelling. Together, these modules establish a comprehensive, multi-modal framework that significantly outperforms conventional approaches in maritime trajectory prediction.

5. Implications

The LLM4STP framework introduces a novel technical paradigm for STP, providing significant benefits to a wide range of maritime stakeholders, including ship operators, port authorities, technology developers, regulatory bodies, and researchers. The practical implications can be broadly categorised as follows:

(1) Ship operators and fleet managers. The fusion of semantic, spatial, and probabilistic information significantly strengthens prediction reliability and accuracy under complex maritime conditions. This improvement reduces the risk of incidents arising from human error and provides a more reliable basis for operational decision-making. Enhanced predictive accuracy facilitates dynamic route optimisation, collision risk assessment, and fuel-efficient voyage planning, enabling operators to avoid congestion, lower fuel consumption, and reduce overall operating costs. By leveraging more precise trajectory forecasts, ship operators can make evidence-based navigational decisions, thereby mitigating operational risks and strengthening fleet-wide safety performance.

(2) Port authorities and maritime traffic controllers. The framework's precise trajectory predictions enable better management of port traffic, reducing congestion and optimising berth allocation, leading to more efficient port operations. The ability to assess and predict vessel movements in real time also enhances the effectiveness of emergency response strategies, particularly in cases of mechanical failure, extreme weather, or security threats. Additionally, the integration of multimodal data, including spatial and semantic features, supports comprehensive situational awareness, which is crucial for maintaining safe and efficient maritime operations.

(3) Technology developers and maritime system integrators. The modular architecture and efficient feature fusion mechanism of the LLM4STP framework provide a robust foundation for developing next-generation autonomous maritime systems, including Unmanned Surface Vessels (USVs) and remotely operated ships. The framework's ability to integrate diverse data sources, such as

Geohash encodings and probabilistic embeddings, sets a new benchmark for maritime AI system design. This technical blueprint has the potential to accelerate the commercialisation of intelligent navigation systems, providing scalable designs for autonomous operations.

(4) Regulatory bodies and maritime safety organisations. The framework's focus on uncertainty quantification and long-term sequence prediction supports the development of risk mitigation strategies, aligning with evolving international safety regulations for autonomous ships. Data-driven insights generated from the model can inform regulatory frameworks, ensuring that emerging technologies are effectively integrated into the maritime industry. Additionally, the improved predictive capabilities support the development of stricter safety standards and operational guidelines for autonomous vessels, enhancing overall maritime safety.

(5) Researchers and academic institutions. The proposed framework serves as a valuable reference for ongoing academic studies in the field of maritime AI, spatiotemporal modelling, and multi-modal data integration. The methodologies demonstrated in this study, such as Geohash encoding and uncertainty modelling, can be extended to other transportation systems, including autonomous vehicles and urban traffic management. Furthermore, the framework encourages interdisciplinary collaboration between data scientists, maritime engineers, and AI researchers, fostering innovation in intelligent maritime systems.

(6) Logistics and supply chain managers. Accurate trajectory predictions enable better planning of cargo arrival times, reducing port waiting times and improving overall supply chain efficiency. The framework's emphasis on route optimisation and congestion management can lead to significant fuel savings and reduced operational costs. Additionally, improved reliability and predictability of shipping schedules contribute to more consistent supply chain operations, enhancing customer satisfaction and competitiveness.

(7) Maritime insurers and risk analysts. The detailed uncertainty quantification provided by the framework offers a more accurate assessment of navigational risks, supporting better underwriting and risk pricing. The ability to model high-risk scenarios in real-time can reduce the likelihood of costly maritime incidents, benefiting both insurers and ship operators. Insights from the model can also inform predictive maintenance strategies, reducing unexpected equipment failures and enhancing overall vessel reliability.

6. Conclusions

This paper presents LLM4STP, a groundbreaking framework that addresses critical challenges in STP by integrating LLMs with maritime domain knowledge. The framework systematically overcomes four long-standing limitations in the field: dynamic interaction modelling, multiscale temporal dependency decoupling, trajectory uncertainty quantification, and domain knowledge integration. By doing so, it establishes a new paradigm for intelligent maritime navigation, significantly enhancing both predictive accuracy and operational reliability.

At its core, the framework introduces a graph-masked spatial Transformer that dynamically constructs ship interaction topologies using adaptive distance thresholds. This innovative approach captures real-time spatial interactions among vessels, providing a more accurate representation of complex maritime environments. Building on this spatial foundation, the hierarchical temporal architecture combines local convolutional operations with the global sequence reasoning capabilities of LLMs. This design supports joint inference of localised manoeuvre details and broader navigational intentions, effectively capturing both micro- and macro-scale trajectory dynamics.

To address the inherent uncertainties in maritime environments, LLM4STP incorporates a novel dual-mode uncertainty modelling system, integrating Gaussian heatmaps for probabilistic assessment and GeoHash-based geospatial encoding for precise spatial representation. This approach ensures robust performance even in volatile maritime conditions. Furthermore, the framework pioneers the deep fusion of numerical trajectory data and textual domain knowledge (e.g., weather alerts, waterway regulations) through a comprehensive five-dimensional embedding system. LLMs act as semantic bridges, aligning multimodal features to create a unified, context-rich representation of the maritime environment.

Extensive experimental validation across three benchmark maritime regions demonstrates the framework's superior performance, achieving SOTA results in both prediction accuracy and operational efficiency. Few-shot learning experiments show that LLM4STP significantly outperforms other models in terms of generalisation with limited data. Ablation studies further quantify the individual contributions of each module, highlighting the synergistic benefits of the integrated approach.

Beyond immediate applications in collision avoidance, route optimisation, and traffic management, LLM4STP sets the stage for broader LLM-driven cognitive reasoning in transportation systems. Future extensions could include multimodal sensor fusion, edge-computing deployments, and real-time decision support, positioning this framework as a foundational technology for autonomous maritime operations. By unifying linguistic reasoning with spatiotemporal prediction, LLM4STP redefines the role of foundation models in maritime intelligence, offering both methodological breakthroughs and practical pathways toward fully autonomous shipping ecosystems.

CRediT authorship contribution statement

Hang Jiao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization; **Jincheng Gong:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis; **Huanhuan Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Conceptualization, Methodology, Investigation, Formal analysis; **Jasmine Siu Lee Lam:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis; **Yaqing Shu:** Writing – review & editing, Visualization, Validation, Formal analysis, Data curation; **Jin Wang:** Writing – review & editing, Visualization, Validation, Formal analysis; **Zaili Yang:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition.

Data availability

Data will be made available on request.

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.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 864724).

Appendix A. Performance comparison of stacked transformer layers.

The GPT-2 (Small) model comprises a stack of 12 Transformer layers. The number of layers is a key architectural hyperparameter that directly affects the model's ability to capture complex data dependencies and its computational efficiency. To determine an appropriate network depth for the present task, a comparative experiment was performed by varying the number of stacked Transformer layers.

As reported in Table A.1, models with only one or two layers exhibited unsatisfactory performance, indicating insufficient capacity to learn and generalise from the data. Conversely, models with five or six layers demonstrated a tendency to overfit, reflecting excessive model complexity relative to the available training set. Models with three and four layers achieved comparable prediction accuracy; however, the four-layer configuration incurred noticeably higher computational cost, limiting its suitability for real-time applications. Considering the trade-off between accuracy, generalisation ability, and computational efficiency, a three-layer Transformer architecture was selected as the optimal configuration for this study.

Table A.1
Performance comparison of stacked transformer layers.

Layers	CSJ			ZS			TJP		
	ADE	FDE	FD	ADE	FDE	FD	ADE	FDE	FD
1	0.11	0.28	0.28	0.11	**0.26**	0.27	0.13	0.33	0.34
2	**0.09**	0.27	**0.27**	**0.10**	**0.26**	**0.26**	**0.12**	**0.31**	**0.31**
3	**0.09**	**0.26**	**0.27**	**0.10**	**0.26**	**0.26**	**0.12**	**0.31**	**0.31**
4	**0.09**	**0.26**	**0.27**	**0.10**	**0.26**	**0.26**	**0.12**	**0.31**	**0.31**
5	0.10	0.28	0.29	0.12	0.27	0.27	**0.12**	0.32	0.32
6	0.10	0.28	0.29	0.12	0.27	0.27	**0.12**	0.32	0.32

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al., 2023. Gpt-4 technical report. arXiv:2303.08774.
Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S., 2016. Social lstm: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 961–971.
Bai, C., Jin, S., Jing, J., Yang, C., Yao, W., Rong, D., AlagbÃ©, J.A., 2024. A multimodal data-driven approach for driving risk assessment. Transp. Res. Part E: Logist. Transp. Rev. 189, 103678.
Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et al., 2023. Qwen technical report. arXiv:2309.16609.
Bi, J., Cheng, H., Zhang, W., Bao, K., Wang, P., 2024. Artificial intelligence in ship trajectory prediction. J. Mar. Sci. Eng. 12 (5), 769.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.
Cao, W., Wang, X., Li, J., Zhang, Z., Cao, Y., Feng, Y., 2024. A novel integrated method for heterogeneity analysis of marine accidents involving different ship types. Ocean Eng. 312, 119295.
Capobianco, S., Forti, N., Millefiori, L.M., Braca, P., Willett, P., 2022. Recurrent encoder–decoder networks for vessel trajectory prediction with uncertainty estimation. IEEE Trans. Aerosp. Electron. Syst. 59 (3), 2554–2565.
Chen, C., 2006. Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci. Technol. 57 (3), 359–377.
Chen, J., Chen, H., Zhao, Y., Li, X., 2022. Fb-bigru: a deep learning model for ais-based vessel trajectory curve fitting and analysis. Ocean Eng. 266, 112898.
Chen, J., Zhang, X., Xu, L., Xu, J., 2024. Trends of digitalization, intelligence and greening of global shipping industry based on citespace knowledge graph. Ocean Coastal Manage. 255, 107206.
Chen, N., Yang, A., Wu, H., Chen, L., Xiong, W., Jing, N., 2025. Semint: an LLM-empowered long-term vessel trajectory prediction framework. Int. J. Geographical Inf. Sci. 39, 1938–1972.
DeepSeek-AI, D.G., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al., 2025. Deepseek-r1: incentivizing reasoning capability in llms via reinforcement learning. arXiv:2501.12948.
Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al., 2024. The LLAMA 3 herd of models. arXiv:2407.21783.
Gong, J., Li, H., Jiao, H., Yang, Z., 2025. Uncertainty-aware ship trajectory prediction via spatio-temporal graph transformer. Transp. Res. Part E: Logist. Transp. Rev. 203, 104315.

Gu, A., Dao, T., 2023. Mamba: linear-time sequence modeling with selective state spaces. arXiv:2312.00752.

Gülsoylu, E., Koch, P., Yildiz, M., Constapel, M., Kelm, A.P., 2024. Image and ais data fusion technique for maritime computer vision applications. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 859–868.

Guo, X., Zhang, Q., Peng, M., Zhua, M., et al., 2024. Explainable traffic flow prediction with large language models. arXiv e-prints, arXiv–2404.

Hu, J., Kaur, K., Lin, H., Wang, X., Hassan, M.M., Razzak, I., Hammoudeh, M., 2022. Intelligent anomaly detection of trajectories for iot empowered maritime transportation systems. IEEE Trans. Intell. Transp. Syst. 24 (2), 2382–2391.

Jiao, J., Huang, S., Wang, S., Soares, C.G., 2021. A CFD–FEA two-way coupling method for predicting ship wave loads and hydroelastic responses. Appl. Ocean Res. 117, 102919.

JIAWEI, W., Jiang, R., Yang, C., Wu, Z., Shibasaki, R., Koshizuka, N., Xiao, C., et al., 2025. Large language models as urban residents: an LLM agent framework for personal mobility generation. Adv. Neural Inf. Process. Syst. 37, 124547–124574.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., et al., 2023. Time-llm: time series forecasting by reprogramming large language models. arXiv:2310.01728.

Kim, D., Yim, J., Song, S., Park, J.-B., Kim, J., Yu, Y., Elsherbiny, K., Tezdogan, T., 2023. Path-following control problem for maritime autonomous surface ships (MASS) in adverse weather conditions at low speeds. Ocean Eng. 287, 115860.

Lan, Z., Liu, L., Fan, B., Lv, Y., Ren, Y., Cui, Z., 2024. Traj-LLM: a new exploration for empowering trajectory prediction with pre-trained large language models. IEEE Trans. Intell. Veh. 10, 794–807.

Lee, H.-J., Park, D.-J., 2024. Collision evasive action timing for MASS using CNN–LSTM-based ship trajectory prediction in restricted area. Ocean Eng. 294, 116766.

Li, G., Li, Z., Knoop, V.L., van Lint, H., 2024a. Unravelling uncertainty in trajectory prediction using a non-parametric approach. Transp. Res. Part C: Emerging Technol. 163, 104659.

Li, H., Jiao, H., Yang, Z., 2023. Ais data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. Transp. Res. Part E: Logist. Transp. Rev. 175, 103152.

Li, H., Jiao, H., Yang, Z., 2023. Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis. Eng. Appl. Artif. Intell. 126, 107062.

Li, H., Xing, W., Jiao, H., Yang, Z., Li, Y., 2024. Deep bi-directional information-empowered ship trajectory prediction for maritime autonomous surface ships. Transp. Res. Part E Logist. Transp. Rev. 181, 103367.

Li, M., Li, B., Qi, Z., Li, J., Wu, J., 2024b. Enhancing maritime navigational safety: ship trajectory prediction using ACoatt–LSTM and AIS data. ISPRS Int. J. Geoinf. 13 (3), 85.

Li, T., Wang, H., Li, X., Liao, W., He, T., Peng, P., 2025. Generative planning with 3D-vision language pre-training for end-to-end autonomous driving. arXiv:2501.08861.

Liang, M., Liu, R.W., Zhan, Y., Li, H., Zhu, F., Wang, F.-Y., 2022. Fine-grained vessel traffic flow prediction with a spatio-temporal multigraph convolutional network. IEEE Trans. Intell. Transp. Syst. 23 (12), 23694–23707.

Lin, Z., Yue, W., Huang, J., Wan, J., 2023. Ship trajectory prediction based on the TTCN-attention-GRU model. Electronics 12 (12), 2556.

Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al., 2024a. Deepseek-v3 technical report. arXiv:2412.19437.

Liu, C., Yang, S., Xu, Q., Li, Z., Long, C., Li, Z., Zhao, R., 2024b. Spatial-temporal large language model for traffic prediction. In: 2024 25th IEEE International Conference on Mobile Data Management (MDM). IEEE, pp. 31–40.

Liu, H., Wu, C., Li, B., Zong, Z., Shu, Y., 2025. Research on ship anomaly detection algorithm based on transformer-GSA encoder. IEEE Trans. Intell. Transp. Syst. 26, 8752–8763.

Liu, R.W., Zheng, W., Liang, M., 2024c. Spatio-temporal multi-graph transformer network for joint prediction of multiple vessel trajectories. Eng. Appl. Artif. Intell. 129, 107625.

Liu, W., Cao, Y., Guan, M., Liu, L., 2024d. Research on ship trajectory prediction method based on CNN-RGRU-attention fusion model. IEEE Access 12, 63950 –63957.

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M., 2023. itransformer: inverted transformers are effective for time series forecasting. arXiv e-prints, arXiv–2310.

Mo, X., Xing, Y., Lv, C., 2024. Heterogeneous graph social pooling for interaction-aware vehicle trajectory prediction. Transp. Res. Part E: Logist. Transp. Rev. 191, 103748.

Mohamed, A., Qian, K., Elhoseiny, M., Claudel, C., 2020. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14424–14432.

Mongeon, P., Paul-Hus, A., 2016. The journal coverage of web of science and scopus: a comparative analysis. Scientometrics 106, 213–228.

Nguyen, T. C.A., Tao, T.M., Lee, C., Youn, C.-H., 2024. TM-LLM: token-motion embedding fusion for large language model-based vessel trajectory prediction. In: 2024 15th International Conference on Information and Communication Technology Convergence (ICTC). IEEE, pp. 115–120.

Peng, M., Guo, X., Chen, X., Chen, K., Zhu, M., Chen, L., Wang, F.-Y., 2025. Lc-llm: explainable lane-change intention and trajectory predictions with large language models. Commun. Transp. Res. 5, 100170.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al., 2019. Language models are unsupervised multitask learners. OpenAI blog 1 (8), 9.

Rana, R., 2016. Gated recurrent unit (GRU) for emotion classification from noisy speech. arXiv:1612.07778.

Ren, Y., Chen, Y., Liu, S., Wang, B., Yu, H., Cui, Z., 2024. Tpllm: a traffic prediction framework based on pretrained large language models. arXiv:2403.02221.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.

Shin, G.-h., Yang, H., 2025. Vessel trajectory prediction in harbors: a deep learning approach with maritime-based data preprocessing and berthing side integration. Ocean Eng. 316, 119908.

Shu, Y., Huang, F., Wu, J., Chen, J., Song, L., Gan, L., Yang, Z., 2025. Research on ship following behavior based on data mining in arctic waters. IEEE Trans. Intell. Transp. Syst. 26, 6778–6788.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al., 2023. Llama: open and efficient foundation language models. arXiv:2302.13971.

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., Huang, G.-B., 2017. Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology. IEEE Trans. Intell. Transp. Syst. 19 (5), 1559–1582.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wang, W., Yi, Z., Zhao, L., Jia, P., Kuang, H., 2025. Application of switching-input LSTM network for vessel trajectory prediction. Appl. Intell. 55 (4), 289.

Wu, S., Peng, Z., Du, X., Zheng, T., Liu, M., Wu, J., Ma, J., Li, Y., Yang, J., Zhou, W., et al., 2024. A comparative study on reasoning patterns of openAI's o1 model. arXiv:2410.13639.

Xiao, Y., Hu, Y., Liu, J., Xiao, Y., Liu, Q., 2024. An adaptive multimodal data vessel trajectory prediction model based on a satellite automatic identification system and environmental data. J. Mar. Sci. Eng. 12 (3), 513.

Xie, Z., Bai, X., Xu, X., Xiao, Y., 2024. An anomaly detection method based on ship behavior trajectory. Ocean Eng. 293, 116640.

Xiong, C., Shi, H., Li, J., Wu, X., Gao, R., 2024. Informer-based model for long-term ship trajectory prediction. J. Marine Sci. Eng. 12 (8).

Xu, X., Wu, B., Teixeira, Â.P., Soares, C.G., 2025. Ship collision risk assessment and visualization using radar sequential images for intelligent maritime supervision. IEEE Trans. Intell. Transp. Syst. 26, 9004–9021.

Xue, H., Wang, S., Xia, M., Guo, S., 2024. G-Trans: a hierarchical approach to vessel trajectory prediction with GRU-based transformer. Ocean Eng. 300, 117431.

Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., et al., 2024a. Qwen2 technical report. arXiv:2407.10671.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al., 2025. Qwen3 Technical Report. arXiv:2505.09388.

Yang, C., Pei, Z., 2023. Long-short term spatio-temporal aggregation for trajectory prediction. IEEE Trans. Intell. Transp. Syst. 24 (4), 4114–4126.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K.X., 2019. How big data enriches maritime research–a critical review of automatic identification system (AIS) data applications. Transp. Rev. 39 (6), 755–773.

Yang, X., She, H., Lou, M., Ye, H., Guan, J., Li, J., et al., 2024c. A joint ship detection and waterway segmentation method for environment-aware of USVs in canal waterways. IEEE Trans. Autom. Sci. Eng. 22, 2132–2144.

Yao, H., Li, X., Li, Q., Yu, C., 2024. Safety aware neural network for connected and automated vehicle operations. Transp. Rese. Part E: Logist. Transp. Rev. 192, 103780.

Yoo, S.L., Lee, K., Baek, W.K., Kim, K.I., 2024. Path prediction for fishing boats using attention-based bidirectional gated recurrent unit. Ocean Sci. J. 59 (1), 3.

Yu, C., Ma, X., Ren, J., Zhao, H., Yi, S., 2020. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16. Springer, pp. 507–523.

Yu, H., Fang, Z., Fu, X., Liu, J., Chen, J., 2021. Literature review on emission control-based ship voyage optimization. Transp. Res. Part D: Transp. Environ. 93, 102768.

Zaman, U., Khan, J., Lee, E., Balobaid, A.S., Aburasain, R.Y., Kim, K., 2024. Deep learning innovations in south korean maritime navigation: enhancing vessel trajectories prediction with AIS data. PLoS One 19 (10), e0310385.

Zeng, X., Gao, M., Zhang, A., Zhu, J., Hu, Y., Chen, P., Chen, S., Dong, T., Zhang, S., Shi, P., 2024. Trajectories prediction in multi-ship encounters: utilizing graph convolutional neural networks with GRU and self-attention mechanism. Comput. Electr. Eng. 120, 109679.

Zhang, T., Zheng, X.-Q., Liu, M.-X., 2021. Multiscale attention-based LSTM for ship motion prediction. Ocean Eng. 230, 109066.

Zhang, X., Fu, X., Xiao, Z., Xu, H., Qin, Z., 2022. Vessel trajectory prediction in maritime transportation: current approaches and beyond. IEEE Trans. Intell. Transp. Syst. 23 (11), 19980–19998.

Zhang, X., Liu, J., Gong, P., Chen, C., Han, B., Wu, Z., 2023. Trajectory prediction of seagoing ships in dynamic traffic scenes via a gated spatio-temporal graph aggregation network. Ocean Eng. 287, 115886.

Zhao, L., Zuo, Y., Zhang, W., Li, T., Chen, C.L.P., 2024. End-to-end model-based trajectory prediction for ro-ro ship route using dual-attention mechanism. Front. Comput. Neurosci. 18, 1358437.

Zhao, S., Zhao, S., 2024. Ship global traveling path optimization via a novel non-dominated sorting genetic algorithm. J. Mar. Sci. Eng. 12 (3), 485.

Zhou, Y., Guo, H., Lu, J., Gong, Z., Yu, D., Ding, L., 2024. Using LSTM with trajectory point correlation and temporal pattern attention for ship trajectory prediction. Electronics 13 (23), 4705.