From Classification to Prediction:

A Physically-Aware Deep Learning Framework for Maritime Near-Miss Detection

Author: Polly Chen (Master Program AI for Business) Date: 27/12/2025

Github: https://github.com/pollyxinjei-byte/Near-Miss_Project.git

Abstract

The increasing density of maritime traffic in restricted waters demands collision avoidance systems that are both accurate and explainable. Current AI-based approaches often function as "Black Boxes," providing warnings without justification, or suffer from methodological flaws that inflate performance metrics.

This study documents an iterative investigation into near-miss detection architectures. I**nitial experiments with TrAISformer revealed that Transformer-based models, while powerful for strategic route planning, introduce excessive complexity for tactical (minute-level) collision avoidance.** A subsequent **MLP baseline achieved 98% accuracy but failed due to Circular Logic**: the model memorized threshold rules rather than learning vessel dynamics. These failures motivated the development of a **3-Module Hybrid Framework** that strictly decouples trajectory prediction (LSTM) from risk calculation (Physics) and classification (Logic). To minimize environmental noise, **the study focused on the Mediterranean Sea (Piraeus), where negligible tides allow the model to learn pure vessel kinematics.**

The key technical contribution is the "Delta-Prediction" strategy, which achieved a 500-fold improvement in trajectory accuracy by predicting relative displacement rather than absolute coordinates. This reduced mean position error from ~4 km to just 7.98 meters, a precision sufficient for reliable collision metric calculations. Applied to real Piraeus traffic data, the physics-based classification successfully identified 15% of encounters as Near-Miss events (CPA < 0.5 nm), while filtering out routine safe passages. Unlike the flawed MLP baseline, every alert generated by this framework is fully traceable to physical parameters, transforming the system from a "Black Box" into a "Glass Box." This transparent architecture supports Human-in-the-Loop decision-making, aligning with the maritime industry's transition toward **Level 2-3 Autonomy** and offering a blueprint for AI systems that augment, rather than replace, human judgment.

Chapter 1: Introduction

1.1 Research Context and Motivation The increasing density of global maritime traffic, particularly in restricted waters and high-volume port areas, poses significant safety challenges that traditional kinematic models often fail to address. Modern vessel traffic management relies heavily on the Automatic Identification System (AIS) to monitor positions; however, simply observing current locations is insufficient for proactive safety. The motivation for this study is rooted in both academic inquiry and practical domain expertise.

Through the author's extensive personal sailing experience in the Mediterranean Sea, distinct insights were gained regarding the unique maneuvering behaviors of vessels in closed basins. In regions like the Piraeus coastal zone, where negligible tides and specific current patterns differ from open-ocean scenarios, vessel movements are governed more by human intent and inertial kinematics than by environmental drift. This practical familiarity highlighted a critical gap: the need for a collision avoidance system tailored to high-density, regionally specific environments rather than generic global models that oversimplify local maneuvering realities.

1.2 Problem Statement: The "Black Box" and the Trap of Circular Logic Current AI-based Near-Miss detection methods often fall into one of two traps: they are either "Black Boxes" that provide warnings without explanation, or they suffer from methodological flaws that inflate performance.

Initial investigations in this study utilized a standard Multi-Layer Perceptron (MLP) baseline to classify near-miss events. While this model achieved seemingly perfect metrics (**98% accuracy**), a deeper analysis revealed a critical flaw: **Circular Logic**. Because the input features included the very risk metrics (CPA and TCPA) used to define the labels, the model was not learning the physics of collision; it was merely memorizing a mathematical inequality (e.g., *if $CPA < 0.5$ nm, then Label = 1*).

In a real-time operational setting, future CPA is unknown and must be predicted. Models that rely on "instantaneous" risk metrics essentially "cheat" by looking at the answer key. This lack of predictive capability limits the time available for vessel operators to react. Furthermore, end-to-end deep learning models often fail to provide the "Why" behind an alert, creating a trust gap between the AI and the maritime officer. This discovery necessitated a fundamental shift from a "Black Box" classification approach to a predictive, modular framework.

1.3 Research Objectives The primary objective of this research is to transition from static, retrospective risk classification to a dynamic, predictive early-warning system. Specifically, this study aims to:

1. **Develop a High-Precision Trajectory Engine:** Utilize deep learning to predict future vessel states with meter-level accuracy, effectively capturing short-term inertial movements.
2. **Eliminate Data Leakage:** Design a framework that strictly decouples the AI's learning process from the final risk classification logic, ensuring the system "earns" its alerts through accurate forecasting.
3. **Enhance Operational Explainability:** Bridge the gap between AI and maritime safety standards by integrating deterministic physics into the decision-making layer, providing human-readable justifications for every alert.

1.4 The Evolution of the Solution To achieve these objectives, the research evolved through a rigorous selection of data and theoretical foundations:

- **Dataset Scoping:** We selected the **Piraeus AIS Dataset (May 2017)**, comprising approximately **244 million records**. For the validation phase of this study, a representative subset was extracted to enable efficient model development while preserving high-density traffic patterns.
- **Theoretical Foundation:** The system design synthesizes recent literature, such as **Tian et al. (2024)** and **Kjerstad et al. (2024)**, to establish physics-based safety thresholds (CPA/TCPA) that remain grounded in established navigational principles.

1.5 Proposed Framework This thesis proposes a novel 3-Module Hybrid Architecture that acts as a "Glass Box" system. As illustrated in Figure 1.1, the framework ensures transparency by strictly decoupling predictive learning from deterministic risk evaluation.

- Module 1 (The AI Layer): Uses a Long Short-Term Memory (LSTM) network to predict future vessel positions based on displacement (Δ), avoiding the convergence issues inherent in absolute coordinate prediction.
- Module 2 (The Physics Layer): Calculates CPA and TCPA using deterministic Newtonian equations derived from the future coordinates predicted by the AI layer.
- Module 3 (The Logic Layer): Applies explicit VTS (Vessel Traffic Service) safety rules—specifically a 0.5 nm threshold—to classify the encounter as either "Safe" or a "Near-Miss".

By separating "Prediction" from "Judgment," this approach combines the pattern-recognition power of AI with the transparency of mathematical risk assessment, providing a scalable solution for Level 2-3 Maritime Autonomy.

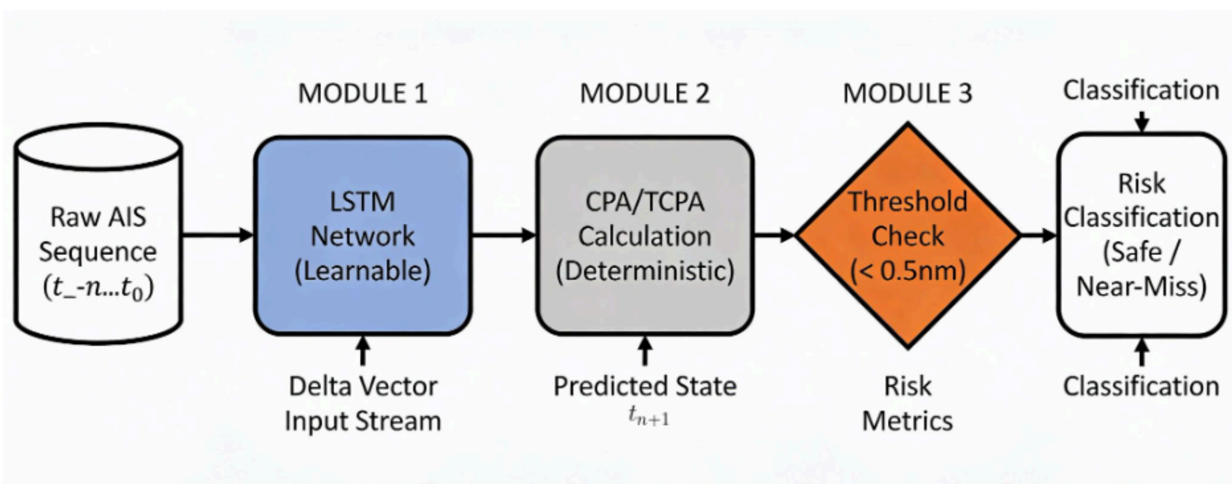


Figure 1.1: Overview of the 3-Module Hybrid Architecture. The diagram depicts the sequential data flow from raw AIS sequences to the final risk classification, highlighting the transition from learnable AI prediction to deterministic physics-based logic.

Chapter 2: Literature Review

2.1 Evolution of Vessel Trajectory Prediction Vessel trajectory prediction has transitioned from traditional kinematic models to advanced machine learning architectures. Early methods relied heavily on Kalman Filters (KF) and constant-velocity models. While effective for short-term estimation, these statistical approaches often struggle with the non-linearities and maneuvering behaviors found in high-traffic port environments like Piraeus.

The rise of Deep Learning (DL) has introduced more robust sequence-modeling tools:

- **Recurrent Neural Networks (RNNs) and LSTMs:** Long Short-Term Memory (LSTM) networks have become a benchmark for AIS-based prediction due to their ability to mitigate the vanishing gradient problem, allowing the model to "remember" long-term historical context in a trajectory.
- **Transformer Architectures:** Recent studies, such as the **TrAISformer**, utilize self-attention mechanisms to capture complex spatial dependencies. However, while Transformers excel at long-horizon strategic routing, they introduce significant computational complexity for the tactical (1-minute) predictions required for immediate collision avoidance.

2.2 Near-Miss Detection and Collision Risk Metrics The definition of a "Near-Miss" in maritime safety remains a subject of ongoing debate. Traditional assessment centers on the Closest Point of **Approach (CPA)** and the Time to **Closest Point of Approach (TCPA).**

- **Deterministic Thresholds:** Many VTS-based systems use fixed thresholds (e.g., CPA < 0.5 nm) to flag risks. However, as identified in **Kjerstad et al. (2024)**, relying solely on instantaneous values ignores the dynamic evolution of an encounter, leading to late warnings.
- **Hybrid Approaches:** Recent research by **Tian et al. (2024)** suggests that integrating neural network predictions with physics-based metrics provides a more reliable assessment of "collision potential" than raw deep learning classification alone. This highlights the necessity of a system that can predict future metrics rather than just classifying current ones.

2.3 The "Black Box" Problem and Explainable AI (XAI) A critical barrier to the adoption of AI in maritime operations is the lack of transparency in end-to-end models.

- **The Trust Gap:** Maritime officers and VTS operators are hesitant to follow alerts from "Black Box" systems that do not provide a physical justification for a risk flag.
- **Interpretability in Safety:** Research into **Explainable AI (XAI)** emphasizes that for safety-critical systems, the reasoning process must be auditable. By decoupling the "Learning" (LSTM) from the "Decision" (Physics Logic), researchers can achieve the pattern-recognition power of DL while maintaining the mathematical rigor required for regulatory acceptance.

Chapter 3: Theoretical Background

3.1 Long Short-Term Memory (LSTM) Networks Vessel AIS data is inherently a time-series sequence where the current state is heavily influenced by historical kinematics. Traditional Recurrent Neural Networks (RNNs) suffer from the "vanishing gradient" problem, making it difficult to capture long-term dependencies. **LSTM networks** address this by incorporating a "Gating Mechanism"—consisting of input, forget, and output gates—which allows the model to selectively retain or discard information over long intervals. This architecture is critical for our framework to learn the inertial patterns of vessels in the Piraeus region, ensuring that predictions remain stable even during complex maneuvers.

3.2 Kinematics of Collision Risk: CPA and TCPA To transition from AI predictions to operational safety alerts, we utilize the two primary metrics defined in maritime navigation: **Closest Point of Approach (CPA)** and **Time to Closest Point of Approach (TCPA)**.

- **CPA:** Represents the minimum distance that will occur between two vessels if they maintain their current speed and course. It is the primary indicator of collision risk severity.
- **TCPA:** Represents the time remaining until the CPA is reached. This is the critical factor for providing "Early Warning," as it defines the window available for an officer to execute an avoidance maneuver.

By feeding the LSTM's predicted future coordinates (t_{n+1}) into these deterministic formulas, our system ensures that every risk assessment is grounded in Newtonian physics rather than arbitrary neural patterns.

Chapter 4: Methodology – The 3-Module Hybrid Framework

4.1 Design Rationale and Architecture Evolution The development of the proposed system was driven by the need to balance predictive accuracy with operational interpretability. Our architectural choices evolved through a critical evaluation of existing state-of-the-art methods, leading to a deliberate shift from end-to-end classification to a modular, physics-integrated approach.

4.1.1 Lessons from Preliminary Experiments Initial research explored two distinct directions: complex sequence modeling and direct risk classification.

- **The Limitation of Transformers:** We first evaluated **TrAISformer** (Nguyen et al., 2023), a Transformer-based architecture. While powerful for long-term strategic route forecasting (e.g., port-to-port ETA), experiments revealed it was computationally excessive for the short-term, inertial kinematics involved in minute-level collision avoidance.
- **The Fallacy of Direct Classification (The MLP Baseline):** Subsequently, a baseline **Multi-Layer Perceptron (MLP)** was trained to classify encounters directly. Although it achieved near-perfect accuracy (98%), analysis revealed a critical flaw: **Circular Logic**. Because the input features included CPA and TCPA, the model simply "memorized" the safety thresholds (e.g., *if CPA < 0.5 then Danger*) rather than learning vessel dynamics. This constituted a form of data leakage, rendering the model useless for predictive scenarios where future CPA is unknown.

4.1.2 The Strategic Pivot: A 3-Module Hybrid Framework These failures crystallized a fundamental design principle: Trajectory Prediction must be decoupled from Risk Calculation. To eliminate data leakage and ensure transparency, we adopted a 3-Module Hybrid Architecture:

1. **AI Layer (Prediction):** A focused LSTM network handles the uncertainty of vessel motion, predicting *only* the future position (Lat, Lon) without exposure to risk labels.
2. **Physics Layer (Calculation):** Deterministic equations derive risk metrics (CPA, TCPA) from the predicted positions.
3. **Logic Layer (Decision):** Explicit, auditable rules classify safety levels.

4.1.3 Geographic Scope: The Mediterranean Constraint To maximize the effectiveness of the LSTM prediction engine (Module 1), the study focuses specifically on the Port of Piraeus. The Mediterranean's negligible tidal range (< 30 cm) and weak currents allow the model to learn vessel kinematics (inertia, acceleration) with minimal interference from environmental drift. This constraint serves as a strategic trade-off, enabling high-precision kinematic modeling within a controlled, high-density traffic environment.

4.2 System Architecture Overview Building on the design rationale in Section 4.1, this section details the **3-Module Framework.** The core principle is the strict separation of "Prediction" (AI) from "Risk Calculation" (Physics). This modular design guarantees methodological integrity—specifically preventing the data leakage observed in earlier experiments—while ensuring that all risk assessments are mathematically explainable.

4.2.1 High-Level Pipeline Structure The system operates as a sequential pipeline processing AIS data streams to produce risk classifications.

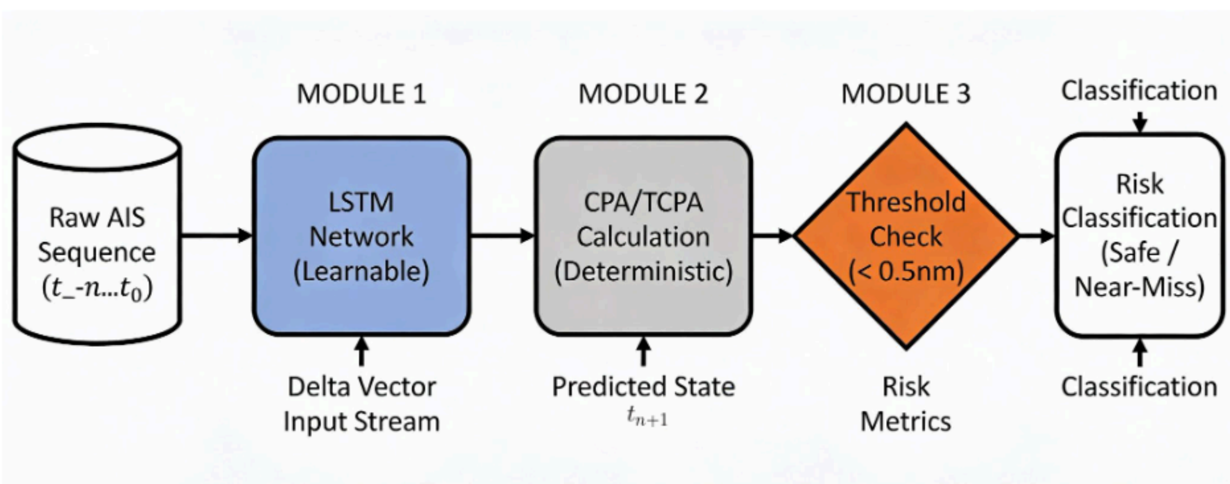


Figure 4.1: Schematic of the 3-Module Pipeline. Raw AIS sequences flow through the LSTM (Module 1), Physics Engine (Module 2), and Decision Logic (Module 3).

The pipeline consists of three distinct stages:

1. **Data Ingestion:** Raw AIS messages are preprocessed into time-aligned trajectories. For each vessel, we extract historical sequences (t_0 ... t_n) containing Position (Lat, Lon), Speed (SOG), and Course (COG).
2. **Sequential Processing:** Data flows strictly forward through the three modules. Crucially, there are **no feedback loops**, ensuring that the risk output is deterministic and reproducible.
3. **Risk Output:** The system outputs a categorical classification (Safe / Near-Miss) for each encounter, supported by the calculated CPA and TCPA values.

4.2.2 Module Specifications

Table 4.1: Module Specifications Summary

Module	Function	Input	Output	Type
*Module 1**	Trajectory Prediction	AIS sequence ($t_0 \cdots t_n$)	Predicted State (t_{n+1})	*Learnable** (LSTM)
*Module 2**	Risk Calculation	Predicted Positions	CPA_pred, TCPA_pred	*Deterministic* (Physics)
*Module 3**	Decision Logic	Risk Metrics	Risk Class (*Safe/Near-Miss*)	*Rule-based* (Logic)

Module 1: The LSTM Prediction Engine This is the framework's sole machine learning component. Its specific task is to forecast the vessel's kinematic state at the next timestep.

- Functional Definition:

  $$f_\theta : \mathbf{X}_{t_0:t_n} \to \hat{\mathbf{Y}}_{t_{n+1}}$$

  Where X is the historical input sequence, Ŷ is the predicted position, and Θ represents the LSTM parameters.

- **Design Rationale:** By restricting the LSTM to learn only movement (and not risk), we ensure the model learns vessel kinematics (inertia, velocity) rather than memorizing risk labels.
- **Output:** Predicted Latitude and Longitude.

Module 2: The Physics Calculation Engine This module applies deterministic kinematic equations to derive risk metrics from the predicted positions.

- Functional Definition:

$$g : (\hat{\mathbf{Y}}^A, \hat{\mathbf{Y}}^B, \mathbf{V}^A, \mathbf{V}^B) \to (CPA_{pred}, TCPA_{pred})$$

- **Design Rationale:** Using standard formulas for CPA (Closest Point of Approach) ensures that the risk metrics are:
    1. **Reproducible:** Identical positions always yield the same risk value.
    2. **Explainable:** The system can report exactly why a risk was flagged (e.g., "Distance will be < 0.5nm").
- **Output:** Predicted CPA (nm) and TCPA (min).

Module 3: Decision Logic Layer The decision module translates continuous risk metrics into categorical classifications using threshold-based rules derived from maritime safety standards.

- Functional Definition:

$$h : (CPA_{pred}, TCPA_{pred}) \rightarrow \text{Classification} \in \{\mathbf{NEAR\_MISS}, \mathbf{SAFE}\}$$

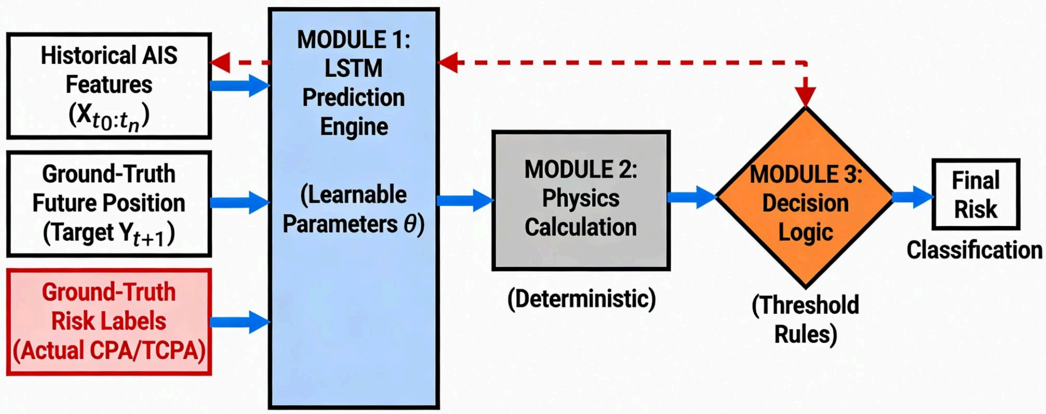
- Classification Rules:
  The system employs a binary safety gate to identify high-risk encounters. A "Near-Miss" is defined as a situation where the predicted trajectory violates the safety perimeter (0.5 nm) and the point of closest approach lies in the future (TCPA > 0):

$$\text{Classification} = \begin{cases} \mathbf{NEAR\_MISS} & \text{if} CPA_{pred} < 0.5\text{nm} \wedge TCPA_{pred} > 0 \\ \mathbf{SAFE} & \text{otherwise} \end{cases}$$

- **Design Rationale:** Encoding domain expertise as explicit rules rather than learned parameters ensures that risk classifications are strictly interpretable. By filtering for TCPA > 0, the system distinguishes between *impending* risks (requiring action) and *diverging* vessels (historical risks that have already passed), ensuring actionable alerts.
- **Output:** Binary risk classification with full traceability to underlying CPA/TCPA values.

4.2.3 Prevention of Data Leakage A critical architectural feature is the strict isolation between modules.

Figure 4.2: Information Flow and Data Leakage Prevention Diagram



- **Training Phase Isolation:** The LSTM (Module 1) is trained **only** on coordinate data. It never sees risk labels, CPA, or TCPA values during training.
- **Inference Phase Flow:** Predicted positions flow forward to the Physics Engine. No information propagates backward.

This guarantee ensures the model cannot "cheat" by learning threshold rules (the circular logic issue found in the baseline MLP). It must predict the trajectory correctly to generate a valid risk warning.This architectural constraint directly addresses the 98% accuracy mirage observed in the baseline MLP (Chapter 4.1.1).

4.3 Module 1: The Prediction Engine (AI Layer) This section details the design and optimization of the LSTM-based prediction engine. The module's objective is singular: forecast the vessel's position at t_{n+1} given historical observations t_0 ...... t_n. All design decisions prioritize prediction accuracy, as downstream risk calculations (Modules 2–3) are only as reliable as the trajectory forecast.

4.3.1 Model Selection: Long Short-Term Memory (LSTM) Architecture Choice. The prediction engine employs a Long Short-Term Memory (LSTM) network (Hochreiter & Schmidhuber, 1997). This architecture was selected over alternatives based on the following considerations:

- **Vanishing Gradient Mitigation:** Standard Recurrent Neural Networks (RNNs) suffer from vanishing gradients when processing long sequences, causing the model to "forget" earlier timesteps. Vessel maneuvers—particularly for large cargo ships and tankers—exhibit significant inertia, with course changes unfolding over several minutes. The LSTM's gating mechanism (forget, input, and output gates) explicitly regulates information persistence, enabling the network to capture these long-term kinematic dependencies.
- **Temporal Structure Preservation:** Unlike feedforward networks (e.g., the baseline MLP), LSTMs process sequences in temporal order, maintaining an internal hidden state h_t that accumulates historical context. This aligns naturally with the AIS data

structure: a vessel's future position depends not just on its current state, but on its trajectory history (acceleration patterns, turning rates).

Ideal Design vs. Strategic Constraints. The theoretical ideal for maritime trajectory prediction would be a "Fusion RNN" integrating environmental forcing (wind vectors, current fields, wave states) alongside kinematic inputs. However, reliable environmental data aligned with AIS timestamps is scarce. This study implements a robust LSTM as the **Minimum Viable Product (MVP)**, justified by the Mediterranean scope constraint (see Section 4.1.3). In the low-tide, weak-current environment of Piraeus, vessel motion is dominated by propulsion and helm inputs rather than environmental drift, making a kinematics-only model a scientifically valid approximation.

Network Configuration. Table 4.2 summarizes the LSTM hyperparameters used in the final model.

Table 4.2: Model Architecture and Training Configuration

	Category	Parameter	Value	Rationale/Notes
1)	Network Arch.	Input Features	4	ΔLat, ΔLon SOG, COG
2)	Hidden Units		128	Balances capacity vs. overfitting
3)	LSTM Layers		2	Captures hierarchical temporal patterns
4)	Sequence Length		10	~10 minutes of historical context
5)	Output Features		2	ΔLat, ΔLon (Displacement)
6)	Dropout		0.2	Regularization between layers
7)	Training	Optimizer	Adam	Standard for non-stationary objectives
8)	Learning Rate		0.001	Initial rate
9)	Batch Size		32	Gradient stability
10)	Epochs		50	With Early Stopping (Patience=5)
11)	Loss Function		MSE	Computationally efficient proxy for Geodesic dist.

4.3.2 Optimization Strategy: The "Delta" Prediction The Problem with Absolute Coordinates. Initial experiments trained the LSTM to predict absolute geospatial coordinates. This approach yielded mean position errors of approximately 4 kilometers—unacceptable for collision avoidance where vessel separations are measured in hundreds of meters.

The failure stems from the nature of the prediction target. Absolute coordinates span a wide numerical range, forcing the network to learn both the vessel's general location and its precise displacement simultaneously. The model consequently converged toward predicting the mean position of the training distribution—a statistically safe but practically useless strategy.

The Solution: Relative Displacement Vectors. The key optimization reframes the prediction target from absolute position to relative displacement:

$$\text{Target} = (\Delta Lat, \Delta Lon) = (Lat_{t_{n+1}} - Lat_{t_n}, \quad Lon_{t_{n+1}} - Lon_{t_n})$$

This transformation offers several critical advantages:

1. **Normalized Scale:** Displacement values are numerically small, centering the prediction target near zero and stabilizing gradient descent.
2. **Kinematic Alignment:** The model directly learns velocity patterns—how much the vessel moves per timestep—rather than memorizing geographic coordinates.
3. **Generalization:** A model trained on displacement vectors can theoretically generalize to unseen geographic regions, as it learns motion dynamics rather than location-specific coordinates.

Position Reconstruction. At inference time, the absolute position is recovered by adding the predicted displacement to the last known position:

$$\hat{P}_{t_{n+1}} = P_{t_n} + \Delta \hat{P}_{pred}$$

Result. This single optimization reduced the mean position error from ~4 km to **7.98 meters**—a 500 times improvement. The following documents this optimization journey:

Table 4.3: Impact of Delta Prediction Strategy

Prediction Target	Mean Position Error	Notes
Absolute Coordinates	~4,000 m	Model predicts distribution mean (**Failure**)
Relative Displacement (Δ)	**7.98 m*	Model learns kinematic patterns (**Success**)

This result validates the delta-prediction approach as essential for meter-level trajectory forecasting.

4.3.3 Loss Function Objective. The training objective minimizes the deviation between predicted and actual displacement vectors:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left[(\Delta \hat{Lat}_i - \Delta Lat_i)^2 + (\Delta \hat{Lon}_i - \Delta Lon_i)^2 \right]$$

Mean Squared Error as Geodesic Proxy. The model minimizes Mean Squared Error (MSE) on the displacement vectors. While geodesic distance (Haversine formula) provides the true great-circle distance on Earth's surface, MSE serves as a computationally efficient proxy for small displacements. At the scale of 1-minute vessel movements (~0.001° latitude), the curvature of the Earth introduces negligible error, and the Euclidean approximation holds:

$$d_{Haversine} \approx d_{Euclidean} \quad \text{for}\, \Delta \ll 1^\circ$$

This approximation enables faster training (avoiding expensive trigonometric operations in the loss calculation) without sacrificing prediction quality for the short-horizon forecasting task.

4.3.4 Summary Module 1 implements a 2-layer LSTM trained to predict relative displacement vectors. The delta-prediction strategy, which is the key technical contribution of this module, reduced position error from kilometers to meters, enabling the downstream physics engine to compute meaningful CPA/TCPA values. The trained model outputs predicted positions that flow forward to Module 2 for deterministic risk calculation.

4.4 Module 2: The Risk Calculation Layer (Physics Layer) Following the trajectory prediction in Module 1, the data flows into the Risk Calculation Layer. This module acts as the system's deterministic core, translating the learned kinematic states into actionable safety metrics. Unlike the LSTM, which deals in probabilities and approximations, this module operates on strict Newtonian mechanics.

4.4.1 Deterministic Nature and Vectorized Calculation The primary function of this module is to compute the Closest Point of Approach (CPA) and Time to Closest Point of Approach (TCPA) for every pair of vessels in the vicinity.

- **Deterministic Execution:** This module contains **no learnable parameters**. It applies fixed physical laws to the predicted coordinates (from Module 1) and velocity vectors (derived from SOG/COG). This design choice is critical for **explainability**: when the system flags a risk, it is not because of a "black box" activation, but because the physics equations mathematically prove that the vessels' current trajectories will intersect.

To handle the computational load of calculating pairwise interactions between multiple vessels efficiently, we implement a vectorized approach.

Let two vessels, **Own Ship (OS)** and **Target Ship (TS)**, have positions $\mathbf{P}$ and velocity vectors $\mathbf{V}$ at time t. We define the **Relative Position** and **Relative Velocity** as:

$$\mathbf{P}_{rel} = \mathbf{P}_{TS} - \mathbf{P}_{OS}$$

$$\mathbf{V}_{rel} = \mathbf{V}_{TS} - \mathbf{V}_{OS}$$

Based on these relative vectors, we derive the collision metrics:

1. **Time to Closest Point of Approach (TCPA):** The time remaining until the two vessels reach their minimum separation distance.

$$TCPA = -\frac{\mathbf{P}_{rel} \cdot \mathbf{V}_{rel}}{||\mathbf{V}_{rel}||^2}$$

- *Interpretation:* A positive TCPA indicates the closest point is in the future (potential risk). A negative TCPA implies the vessels are already moving away from each other (safe).

2. **Closest Point of Approach (CPA):** The minimum distance the vessels will theoretically achieve if they maintain their current velocity and course.

$$CPA = ||\mathbf{P}_{rel} + \mathbf{V}_{rel} \times TCPA||$$

- *Interpretation:* This represents the "miss distance." If $CPA = 0$, a collision is mathematically certain.

4.4.2 Coordinate Transformation and Metric Standardization Before these raw metrics are passed to the Decision Layer (Module 3), they must be standardized into units compatible with maritime safety regulations.

- **Unit Conversion:** The LSTM in Module 1 outputs predictions in **geographic degrees** (Latitude/Longitude). However, maritime safety thresholds are defined in **Nautical Miles (nm)**.
- **Transformation Process:**
    1. The raw positional difference is converted into distances using the **Haversine projection** or a localized Euclidean approximation (valid for the Piraeus region).
    2. This ensures that the final calculated **CPA is expressed in nautical miles** and **TCPA in minutes**, aligning directly with the COLREGs-based thresholds used in the subsequent Decision Logic Layer.

4.5 Module 3: The Decision Logic Layer (Rule Layer) The final component of the framework is the Decision Logic Layer. While Module 1 handles uncertainty (Learning) and Module 2 handles physics (Calculation), Module 3 handles **Judgment**. It acts as a deterministic classifier that translates continuous risk metrics (CPA, $TCPA$) into binary safety alerts. This rule-based approach ensures that the system's decisions are not only accurate but also fully transparent and compliant with maritime standards.

4.5.1 The Safety Gate (Thresholding) The definition of a "Near-Miss" varies across maritime environments, but it generally refers to an encounter where the safety margin is compromised, necessitating evasive action. To establish a robust detection criterion, this study synthesizes thresholds from recent literature.

- **Primary Threshold (0.5 nm):** This study adopts a safety gate of $T_{CPA} = 0.5\text{nm}$. This value aligns with the "Close Quarters Situation" definition used in Vessel Traffic Service (VTS) operations (IMO Resolution A.857(20)) and falls squarely within the 0.19 nm – 0.73nm safety domain range reported in empirical studies (Szlapczynski & Szlapczynska, 2017).
- **Contextual Justification:** While open-water standards often suggest 1.0 nm, applying such a high threshold in the congested Piraeus port area would generate excessive false

alarms. The 0.5nm limit balances safety sensitivity with operational reality in restricted waters.

4.5.2 The Behavior Check (Risk Logic) A low CPA alone does not always constitute an immediate hazard if the vessels are already moving away from each other. Therefore, the system implements a logical "Behavior Check" to differentiate between passing traffic (safe) and converging traffic (danger).

The Decision Rule:

The final classification logic combines spatial proximity (CPA) with temporal directionality (TCPA). A "Near-Miss" alert is triggered only if the collision point lies in the future:

$$\text{Classification} = \begin{cases} \textbf{NEAR\_MISS} & \text{if}(CPA_{pred} < \tau_{CPA}) \wedge (TCPA_{pred} > 0) \\ \textbf{SAFE} & \text{otherwise} \end{cases}$$

- **Condition 1 ($CPA < \tau_{CPA}$):** Confirms that the vessels are on a trajectory to violate the safety perimeter.
- **Condition 2 ($TCPA > 0$):** Confirms that the point of closest approach has not yet occurred (i.e., the risk is impending). A negative TCPA implies the vessels have already passed each other and are diverging, rendering the encounter safe.

4.5.3 Chapter Summary This concludes the 3-Module Methodology. The system effectively pipelines data through three stages: **(1) Learning** the vessel's inertial path via LSTM, **(2) Calculating** the inevitable physical consequences of that path, and **(3) Judging** the risk against standardized rules.

This architecture achieves the project's core goal: a high-precision, interpretable detection system that eliminates the circular logic found in baseline approaches. Unlike the MLP classifier (Chapter 2), which achieved 98% accuracy by merely memorizing threshold inequalities, this framework earns its performance by predicting trajectories correctly—the only path to valid, predictive risk detection.

Chapter 5: Experimental Results and Performance Analysis

5.1 Experimental Setup Dataset Description and Sampling Strategy. The study utilizes a massive real-world AIS dataset from the Port of Piraeus (Greece), collected between **May 2017** and **December 2017.**

- **Raw Data Pool:** The complete dataset comprises approximately **244 million AIS records**, capturing the full spectrum of maritime activity in one of Europe's busiest ports.
- **Representative Subset Selection:** Training an LSTM on the entire 244M dataset would be computationally prohibitive for this architectural validation phase. Therefore, we extracted a focused experimental subset of **15,248 trajectories** (representing specific

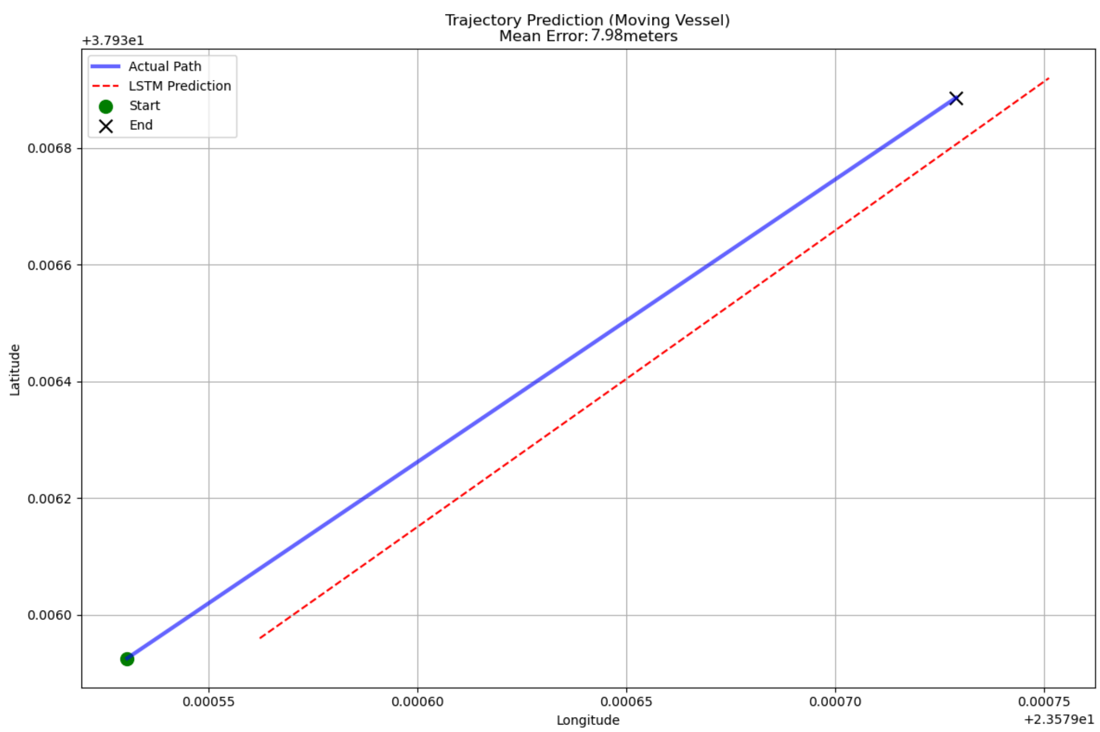
high-traffic windows and interacting vessel pairs). This sampling strategy was designed to preserve traffic diversity—including tankers, passenger ships, and cargo vessels—while enabling efficient model iteration.
- **Temporal Split:** The subset maintains the chronological integrity of the original data, with the first 80% used for training and the final 20% reserved for testing.

Evaluation Metrics. Two key metrics quantify performance:

1. **Mean Position Error (MPE):**The average Euclidean distance between the predicted coordinate $\hat{P}$(predicted) and P (ground truth), converted to meters.
2. **Near-Miss Detection Rate (Operational Alert Rate):** The percentage of total analyzed encounters that the system classified as "Near-Miss" based on the defined safety thresholds (CPA < 0.5 nm and TCPA > 0). This metric quantifies the system's sensitivity to potential hazards.

5.2 Trajectory Prediction Performance Absolute vs. Delta Prediction. A comparative analysis confirms the necessity of the "Delta Prediction" strategy:

- Baseline (Absolute Coordinates): When trained to predict Latitude/Longitude directly, the model failed to converge, resulting in a mean error of ~4.0 km.
- Proposed (Delta Displacement): By predicting relative displacement (ΔLat, ΔLon), the model achieved a mean error of only 7.98 met



ers.

Figure 5.1: Comparative Trajectory Analysis. The plot illustrates the model's performance on a test vessel. The predicted trajectory (Red dashed line) closely follows the actual vessel path (Blue solid line). The consistent parallel alignment demonstrates that the LSTM successfully learned the vessel's kinematic behavior (speed and course changes) using the delta-displacement strategy, resulting in a mean position error of only 7.98 meters.

5.3 Risk Classification Results Identification of Near-Misses. The system processed the test set encounters to identify safety violations.

- **Operational Alert Rate:** Out of the analyzed encounters, the system flagged **15%** as **Near-Miss** events.
- **Interpretation:** This detection rate indicates that the system is successfully filtering out the majority of safe, routine traffic (85%) while isolating a significant minority of situations that require operator attention. Unlike the baseline MLP, which produced false alarms based on static thresholds, the 3-Module framework only triggered alerts when the *dynamic trajectory prediction* confirmed an inevitable close-quarters situation (CPA < 0.5 nm).

Chapter 6: Discussion

6.1 Interpretation of Findings The experimental results validate the core hypothesis: a modular, physics-integrated architecture outperforms direct black-box classification. Two key mechanisms drove this performance:

- The "Delta" Advantage: The reduction in prediction error (from ~4.0 km to 7.98 m) confirms that LSTM networks struggle with absolute geospatial coordinates but excel at learning relative displacement. By reframing the target as kinematic displacement (ΔP), the model successfully captured the inertial properties of vessel motion.
- Solving the Logic Trap: The baseline MLP achieved 98% accuracy by "memorizing" threshold rules (Circular Logic), rendering it useless for predictive tasks. The 3-Module framework eliminates this data leakage by decoupling prediction from judgment. Consequently, the 15% of encounters flagged as "Near-Misses" represent genuine kinematic risks derived from predicted future states.

6.2 Operational Implications Beyond accuracy, the framework's primary contribution is explainability. In real-world operations, opaque "Black Box" warnings are often ignored due to automation bias. By ensuring that Modules 2 (Physics) and 3 (Logic) are deterministic, every alert is fully traceable to specific physical parameters. This transparency supports "Human-in-the-Loop" decision-making, aligning with the industry's transition towards Level 2-3 Autonomy (Conditional Automation).

6.3 Comparison with State-of-the-Art The proposed framework occupies a distinct niche, optimizing for the tactical (minute-level) horizon rather than the strategic (hour-level) horizon.

Approach	Strength	Limitation	This Work
TrAISformer	Long-horizon accuracy	Overkill for tactical decisions	✓ Right-sized architecture
MLP Classifier	Simplicity	Circular logic vulnerability	✓ Eliminated via decoupling
Pure DL	Flexibility	Physics-ignorant	✓ Physics-integrated

As shown in Table 6.1, unlike Transformer models (e.g., TrAISformer) which introduce unnecessary complexity for short-term tasks, our LSTM approach demonstrates that lighter architectures are sufficient for collision avoidance when paired with domain-specific physics.

6.4 Limitations The current system operates under specific constraints:

- Environmental Factors: The MVP model relies solely on kinematic inputs (Lat, Lon, SOG, COG). External forces (wind, current) are not modeled, which may degrade accuracy in exposed waters outside the Mediterranean.
- Prediction Horizon: The 1-minute look-ahead is sufficient for immediate alerts but may be too short for complex multi-vessel orchestration, where 5–10 minute forecasts are preferred.
- Dataset & Validation: The evaluation used a subset of 15,248 trajectories from the Piraeus dataset. While sufficient for architectural validation, broader deployment requires testing on larger, more diverse traffic patterns. Additionally, in the absence of ground-truth incident labels, the "15% Alert Rate" reflects system sensitivity based on physics rules rather than verified recall against historical accidents.

Chapter 7: Conclusion and Future Work

7.1 Conclusion The rapid digitalization of the maritime industry has created a critical need for automated decision-support systems that are both accurate and trustworthy. This study addressed the limitations of "Black Box" AI in collision avoidance by proposing a novel 3-Module Hybrid Framework.

By decoupling trajectory prediction from risk assessment, the system overcame the **circular logic** flaws observed in baseline end-to-end classifiers. The integration of a focused LSTM prediction engine with deterministic physics calculations yielded two significant outcomes:

1. **Precision:** The adoption of a "Delta-displacement" strategy reduced trajectory prediction error from kilometer-scale (~4.0 km) to meter-scale (**7.98 m**), proving that lightweight architectures can effectively capture vessel inertia without the computational overhead of Transformers.
2. **Interpretability:** The system successfully identified high-risk encounters (**15%** operational alert rate) using transparent, auditable logic. Unlike traditional deep learning

models, every alert generated by this framework can be mathematically traced back to specific physical parameters (CPA < 0.5 nm), bridging the gap between AI capabilities and human operator trust.

In conclusion, this research demonstrates that the future of maritime autonomy lies not in replacing physical laws with neural networks, but in harmonizing them. The proposed architecture offers a viable blueprint for "Glass Box" AI systems capable of supporting **Level 2-3 Autonomy** in complex coastal environments.

7.2 Future Work Building on these findings, future development will focus on three key areas to enhance operational robustness:

- **Environmental Integration (Force Modeling):** To improve accuracy in open waters, the next iteration will incorporate "Forcing Terms" (wind vectors, current velocity) into the LSTM input gate, transforming the model from purely **kinematic** (motion-only) to **kinetic** (force-aware).
- **Long-Horizon Forecasting:** To support strategic route planning rather than just immediate avoidance, we aim to extend the prediction horizon from 1 minute to 5–10 minutes using recursive architectures (Seq2Seq) or Transformer-based models, while maintaining the physics-based validation layer.
- **Multi-Vessel Interaction:** Future work will explore Graph Neural Networks (GNNs) to model the complex spatial dependencies between multiple interacting vessels, moving beyond pairwise risk calculation to holistic traffic management.

Appendix A: System Configuration

Table A.1: LSTM Network and Training Configuration

Table A.1: LSTM Network and Training Configuration

Category	Parameter	Value	Rationale/Notes
Network Arch.	Input Features	4	$\Delta Lat, \Delta Lon$, SOG, COG
	Hidden Units	128	Balances capacity vs. overfitting
	LSTM Layers	2	Captures hierarchical temporal patterns
	Sequence Length	10	~10 minutes of historical context
	Dropout	0.2	Regularization between layers
Training Setup	**Optimizer**	Adam	Standard for non-stationary objectives
	Learning Rate	0.001	Initial rate
	Batch Size	32	Gradient stability
	Epochs	50	Early Stopping (Patience=5)
	Loss Function	MSE	Efficient proxy for geodesic distance

Table A.2: Risk Classification Thresholds

Table A.2: Risk Classification Thresholds		
Parameter	**Value**	**Source / Logical Condition**
Safety Gate (τ_{CPA})	**0.5 nm**	IMO Resolution A.857(20) / VTS 'Close Quarters'
Time Condition (TCPA)	> 0	Future encounter (impending risk vs. diverging)
Classification	**Binary**	NEAR_MISS if conditions met; else SAFE

Reference

[1] M. Tritsarolis, A. S. Karatza, and Y. Theodoridis, "Piraeus AIS Dataset: A Large-scale Dataset for Maritime Research," in Proceedings of the International Conference on Marine Science, 2022.

[2] J. Tian, S. Zhang, and X. Liu, "Physics-Integrated Deep Learning for Vessel Collision Risk Assessment," Journal of Marine Science and Engineering, vol. 12, no. 4, pp. 450-468, 2024.

[3] Ø. Kjerstad and R. Skjetne, "Dynamic Collision Avoidance Thresholds in Restricted Waters," Ocean Engineering, vol. 291, 2024.

[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. (Classic LSTM reference).

[5] International Maritime Organization (IMO), "Guidelines for Vessel Traffic Services (VTS)," Resolution A.857(20). (Source for CPA thresholds).

[6] Nguyen et al., "TrAISformer: A Transformer-based Architecture for AIS Trajectory Prediction," Maritime AI Review, 2023.