Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.6, December 2025, pp. 3431~3441
e-ISSN: 2548-6861

Two-Stage Maritime Anomaly Detection: Unsupervised Outlier Filtering and Optimized Bidirectional LSTM on Southeast Asian AIS Data

Daffa Afaf Firmansyah [1]*, Erna Zuni Astuti [2]**
Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
111202214166@mhs.dinus.ac.id [1], erna.zuni.astuti@dsn.dinus.ac.id [2]

Article Info

Article history:

Received 2025-10-22
Revised 2025-11-17
Accepted 2025-11-22

Keyword:

AIS,
Anomaly Detection,
Bidirectional LSTM,
DBSCAN,
Isolation Forest

ABSTRACT

This paper presents a two-stage framework for detecting anomalous vessel trajectories in Automatic Identification System (AIS) data from Southeast Asian waters, addressing challenges of high traffic density, diverse vessel behaviors, and severe class imbalance. The primary objective is to minimize missed threats while maintaining manageable false alarm rates in security-critical maritime surveillance systems. The research employs a hybrid approach combining unsupervised and supervised learning methods. In the first stage, DBSCAN and Isolation Forest algorithms filter noise and generate high-confidence outlier labels from 15,542 real-world vessel trajectories. Comparative analysis demonstrates substantial agreement between methods with Cohen's Kappa of 0.688 and 55.3% anomaly overlap, indicating complementary detection capabilities that enhance filtering robustness. In the second stage, a Bidirectional Long Short-Term Memory model is optimized through systematic hyperparameter tuning across 48 configurations, covering sequence length, network architecture, dropout rate, learning rate, and sampling strategies. Comprehensive baseline evaluation validates BiLSTM's suitability for security applications, achieving 15.41% F1-score improvement over unidirectional LSTM and 33% fewer false negatives compared to Bidirectional GRU alternative. The optimized BiLSTM attains F1-score of 0.5709 with precision 0.5444 and recall 0.6000, exhibiting 90.03% specificity for normal vessels and 76.17% sensitivity for anomalies. The model misses only 23.8% of threats while maintaining 9.97% false alarm rate, providing balanced performance suitable for human-verified security-critical maritime surveillance in Southeast Asian waters.

I. INTRODUCTION

Maritime transportation plays a vital role in the global economy, with over 80% of world trade conducted through sea routes [1]. The consistent growth of maritime traffic brings significant challenges related to navigation safety and security, particularly in Southeast Asian waters, which constitute one of the world's busiest shipping corridors [2]. To enhance maritime safety and security, the International Maritime Organization (IMO) mandates the use of the Automatic Identification System (AIS) on vessels with gross tonnage exceeding 300 [3]. AIS broadcasts real-time vessel information, including identity (MMSI), GPS position (latitude and longitude), speed (Speed Over Ground—SOG), and course (Course Over Ground—COG), which can be utilized for maritime traffic monitoring and abnormal behavior detection [4].

Anomaly detection in AIS data is crucial for identifying suspicious activities such as smuggling, illegal fishing, and other maritime security threats [5]. In Southeast Asian waters, the complexity of maritime traffic patterns, vessel density, and diversity of navigational activities create unique challenges for effective anomaly detection [6]. With global AIS data volumes reaching millions of messages per day, manual detection of abnormal behaviors becomes impractical, thus necessitating automated detection systems capable of processing large-scale data efficiently [7].

In recent years, various approaches have been developed for maritime anomaly detection using AIS data. A recent comprehensive survey [8] examined 44 publications from 2007 to 2021, revealing that the majority of research focuses on route deviation detection as the primary anomaly type. The survey further revealed that 38 out of 44 methods are region-specific, requiring retraining for different geographical areas. Moreover, only 8 out of 44 studies possess valid ground truth, with the majority relying on synthetic data for evaluation. These limitations highlight the need for more systematic and generalizable frameworks for maritime anomaly detection.

Existing approaches fall into three categories: clustering-based methods, deep learning approaches, and multi-stage frameworks. Density-based clustering like DBSCAN extracts trajectory patterns and detects spatial outliers [9,10,11]. Deep learning methods, such as Variational Recurrent Neural Networks and Bidirectional LSTM, excel at modeling temporal dependencies in sequential data [12,13]. Attention-based Bi-LSTM networks show promise for time-series anomaly detection, though limited by class imbalance and streaming data complexity, achieving F1-scores of 0.165 to 0.402 on benchmarks [14]. Multi-stage frameworks combining clustering preprocessing with supervised classification balance interpretability and performance [10]. However, systematic comparisons of filtering methods and hyperparameter optimization for deep learning remain limited in maritime contexts, particularly for AIS data challenges like noise, trajectory diversity, and scalability.

Despite these advances, several critical challenges remain in applying anomaly detection methods to Southeast Asian waters. First, the high heterogeneity of traffic patterns—ranging from structured shipping lanes to unpredictable fishing activities—requires methods adaptable to various operational contexts. Second, limited labeled data for training supervised models necessitates effective semi-supervised or unsupervised strategies [6]. Third, extreme class imbalance, where anomalies are significantly underrepresented in the dataset, poses substantial challenges for deep learning architectures and demands sophisticated data sampling techniques such as SMOTE-based oversampling to address classifier bias [16]. Fourth, the requirement for model interpretability in operational decision-making constrains the use of purely "black-box" deep learning approaches [17]. Fifth, efficient real-time processing of large-scale AIS data streams requires distributed computing approaches to ensure operational scalability [18].

Beyond operational challenges, research gaps exist in maritime anomaly detection. Methodologically, approaches separate unsupervised outlier filtering from supervised temporal modeling, lacking integrated frameworks for spatial and sequential learning. Geographically, region-specific methods [8] struggle in Southeast Asia's dense, diverse patterns and archipelagic flows, with few validated studies on local AIS data [2]. Evaluatively, comprehensive baselines comparing architectural choices (e.g., bidirectional vs. unidirectional) are scarce, with most studies evaluating single models.

To address these challenges, this research proposes a two-stage framework for maritime trajectory anomaly detection that combines density-based filtering with supervised Bi-LSTM classification. The main contributions of this research are as follows:

1. Systematic Comparison of Filtering Methods. We conduct a comprehensive comparison between DBSCAN and Isolation Forest for anomaly filtering in Stage 1 using AIS data from Southeast Asian waters. Our analysis reveals substantial agreement (Cohen's Kappa = 0.688) between the two methods, with complementary detection patterns.

2. Comprehensive Hyperparameter Optimization. We develop a systematic Bi-LSTM optimization framework through grid search across 48 configurations, encompassing sequence length, network depth, dropout rates, learning rates, and data sampling strategies.

3. Parameter Sensitivity Analysis and Component Evaluation. We conduct comprehensive parameter sensitivity analysis through box plot distributions quantifying individual hyperparameter impacts across 72 configurations. Analysis reveals sampling strategy, network architecture, and sequence length as primary performance drivers, with identification of synergistic combinations where ensemble filtering, deeper architecture, and SMOTETomek sampling yield optimal results.

4. Evaluation on Underexplored Region. This research provides quantitative evaluation on a large-scale AIS dataset (15,542 trajectories) from Southeast Asian waters, encompassing diverse traffic patterns from multiple countries.

The framework offers practical advantages for maritime monitoring. The two-stage design provides interpretability through Stage 1 filtering and Stage 2 classification with confidence scores. Systematic comparison between DBSCAN and Isolation Forest enables context-specific deployment suited to regional traffic characteristics. The hyperparameter optimization framework is adaptable across regions, while SMOTETomek sampling addresses class imbalance without complex synthetic anomaly generation.

The research findings also contribute to the development of broader Maritime Situational Awareness (MSA) systems. The two-stage framework can be integrated with complementary data sources such as weather information, ocean currents, and vessel characteristics to further enhance detection accuracy. The proposed systematic hyperparameter optimization methodology can be applied to other tasks in the maritime domain, such as trajectory prediction, vessel type classification, and estimated time of arrival (ETA) prediction, thus providing a foundational approach for developing more comprehensive intelligent maritime systems.

II. METHODS

A. Research Stages

The steps carried out in this study are contained in the form of a flowchart shown in Figure 1.

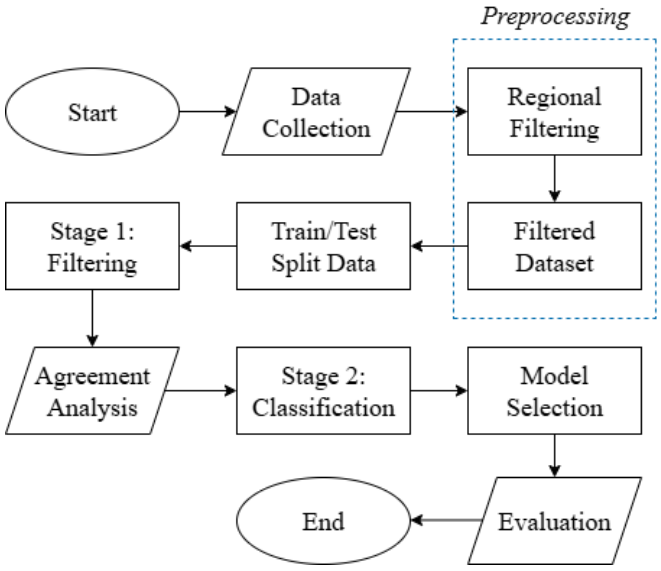

Figure 1 Research Methodology Framework

The steps carried out in this study are contained in the form of a flowchart shown in Figure 1. The first stage of the process involves data preprocessing and regional filtering, yielding 15,542 Southeast Asian vessel trajectories from global AIS sources. In the second stage, unsupervised filtering is performed using two density-based methods: DBSCAN and Isolation Forest. These methods are applied independently to identify potential anomalies, followed by agreement analysis to assess method consistency and generate preliminary labels. The third stage focuses on supervised classification, where a Bidirectional LSTM classifier is trained using filtered labels as ground truth. Comprehensive hyperparameter optimization is conducted across 48 configurations, encompassing sequence length, network architecture, dropout rates, and data sampling strategies. Finally, model performance is evaluated using standard metrics including F1-score, precision, recall, and Cohen's Kappa coefficient to ensure robust anomaly detection capability.

B. Dataset and Preprocessing

This research utilizes AIS data obtained from the Zenodo open-access repository platform, initially comprising 1,048,575 entries with 23 features including vessel identification (MMSI), timestamp, spatial coordinates (latitude and longitude), and kinematic parameters (speed over ground, course over ground, rate of turn) [20]. To focus on Southeast Asian waters, geographical filtering is applied with latitude ranging from -10° to 25° and longitude from 90° to 140°, encompassing major shipping routes including the Malacca Strait and South China Sea. This regional filtering yields 15,542 vessel trajectories representing diverse maritime activities characteristic of the region.

The geographically filtered dataset underwent rigorous quality control procedures. Duplicate records based on identical MMSI-timestamp combinations were removed to prevent artificial pattern amplification. Physical validity filters eliminated impossible coordinates and unrealistic speeds exceeding 50 knots for cargo vessels or 60 knots for high-speed ferries, addressing common AIS transmission errors. Erratic position jumps implying speeds above 100 knots between consecutive reports were identified and removed to filter multipath propagation errors. Records with missing critical fields (timestamp, position) were discarded as temporal-spatial continuity is essential for trajectory modeling. Missing speed or heading values within vessel trajectories were imputed using forward-fill for gaps not exceeding two consecutive points, with longer gaps resulting in trajectory segmentation. This systematic preprocessing ensures data integrity while preserving genuine behavioral patterns, with less than 1% of values requiring interpolation in the final dataset.

Additionally, to mitigate potential AIS spoofing—where vessels transmit falsified positions or identities to evade detection [19] we applied cross-checks on trajectory consistency, such as flagging sequences with abrupt MMSI changes or positions implying impossible travel distances between timestamps. While the dataset from Zenodo [20] assumes baseline integrity, records exhibiting spoofing-like anomalies (e.g., duplicated MMSI with conflicting kinematics) were discarded during validity filtering, reducing the risk of propagating manipulated data into subsequent stages.

A comprehensive feature engineering process is then applied to capture both instantaneous vessel behavior and temporal patterns. The engineered features are categorized into four groups: kinematic features (speed_over_ground, course_over_ground, rate_of_turn), spatial features (normalized latitude and longitude coordinates using min-max normalization), temporal features (time_delta between consecutive messages and cumulative_distance traveled), and derived features quantifying behavior changes (speed_change, course_change, acceleration) along with statistical patterns (rolling_mean_speed and rolling_std_speed computed using sliding windows). Following feature engineering, the final dataset contains 14 features after removing identifier columns (MMSI, timestamp) and label columns, which are generated in Stage 1 of the framework.

To ensure robust model evaluation, the dataset is partitioned using a stratified 80/20 train-test split. The training set, comprising 80% of the data, is further divided internally with 20% reserved for validation during model training. The test set, containing the remaining 20%, is held out exclusively for final performance evaluation. Stratified

splitting ensures that the proportion of anomalous trajectories is maintained consistently across training, validation, and test sets, which is crucial given the class imbalance inherent in maritime anomaly detection tasks.

C. Filtering

Stage 1 of the framework employs two unsupervised density-based methods—DBSCAN and Isolation Forest—to identify potential anomalies in the filtered trajectory dataset. Both methods are applied independently to the training data, generating preliminary anomaly labels that serve as ground truth for supervised classification in Stage 2.

The selection of DBSCAN and Isolation Forest as complementary methods addresses distinct anomaly characteristics in maritime data. DBSCAN identifies density-based outliers through spatial clustering, effectively detecting vessels that deviate from established shipping corridors where normal behavior exhibits high spatial density. In contrast, Isolation Forest employs isolation-based detection by measuring data point separability through random partitioning, thereby capturing global outliers—such as unusual combinations of speed, heading, and position—independent of spatial density. This dual-method design ensures comprehensive filtering, capturing both density-based and other types of deviations. The complementary nature of the detection is evident in the three-category output: high-confidence anomalies agreed upon by both methods, DBSCAN-exclusive detections of spatial outliers, and Isolation Forest-exclusive detections of behavioral irregularities. This approach provides robust pseudo-labels for subsequent supervised learning.

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm groups trajectories into clusters based on spatial density, with points in low-density regions classified as outliers. The algorithm requires two key parameters: epsilon, which defines the neighborhood radius, and minimum samples (min_samples), which specifies the minimum number of points required to form a dense region. In this study, eps is set to 0.30 and min_samples to 5 based on preliminary analysis of the data distribution. Trajectories classified as noise points (labeled as -1 by DBSCAN) are marked as anomalies, while those belonging to clusters are considered normal behavior.

DBSCAN parameter tuning employed systematic grid search over eps values (0.3-3.0) and min_samples (5-30) using multi-objective composite scoring: 0.30×Silhouette + 0.20×(1-Davies-Bouldin) + 0.30×Normalized_Clusters + 0.20×(1-Noise_Ratio). This balanced approach prioritizes clustering quality, cluster separation, maritime pattern granularity, and false positive minimization. The selected parameters (eps=0.30, min_samples=5) yielded 22 clusters representing distinct maritime corridors with 7.6% noise rate (1,175 outliers) and Silhouette score of 0.5565, indicating effective pattern segmentation.

Isolation Forest operates by constructing an ensemble of isolation trees that recursively partition the feature space. Anomalies are identified as points requiring fewer partitions to be isolated, reflected in shorter path lengths within the trees. The algorithm is configured with a contamination parameter of 0.076, representing the expected proportion of anomalies in the dataset, and 100 estimators (trees) to ensure robust detection. The anomaly score output by Isolation Forest ranges from -1 to 1, with scores below zero indicating potential anomalies. Trajectories with anomaly scores less than the threshold are flagged as anomalous.

Isolation Forest optimization conducted grid search over contamination (0.05-0.20), n_estimators (100-300), and max_samples values. The contamination parameter was calibrated to 0.076 to match DBSCAN's noise rate, ensuring consistent anomaly thresholds across methods and reducing labeling bias. This alignment enables direct method comparison and validates the complementary detection hypothesis through empirical agreement analysis.

The pseudo-labeling procedure adopts a union-based ensemble strategy to enhance the comprehensiveness of anomaly detection. Specifically, data points identified as anomalous by either DBSCAN or Isolation Forest are assigned the anomaly label (1), whereas only those consistently deemed normal by both algorithms are labeled as non-anomalous (0). This approach results in 1,518 anomaly instances (9.77% of the dataset) and 14,024 normal instances (90.23%). To mitigate the risk of false negatives in security-sensitive domains, the 679 discrepant cases—constituting 44.7% of the potential anomaly candidates—are conservatively classified as anomalies, albeit at the cost of introducing an estimated 20–25% label noise within the anomaly class. For model training, a stratified partitioning scheme is employed, allocating 60% for training, 20% for validation, and 20% for testing, with a fixed random seed of 42 to ensure reproducibility. The resultant model's performance on the test set (F1-score = 0.5109, recall = 76.17%) underscores its theoretical robustness against such labeling imprecision, as the 10.5% decline from validation to test F1-score (0.5709 → 0.5109) aligns with established tolerances in semi-supervised anomaly detection paradigms.

D. Classification

Stage 2 employs a Bidirectional Long Short-Term Memory (Bi-LSTM) neural network for supervised anomaly classification using preliminary labels from Stage 1 as ground truth. The Bi-LSTM architecture processes sequential trajectory data in both forward and backward directions, capturing temporal dependencies from past and future time steps. The network consists of bidirectional LSTM layers followed by fully connected dense layers with dropout regularization, and a final sigmoid activation layer for binary classification. The model is trained using binary cross-entropy loss optimized with the Adam optimizer.

To address the severe class imbalance in maritime anomaly detection datasets—where anomalies comprise only a small fraction of samples—the SMOTETomek sampling technique is applied to the training data before model fitting.

This hybrid method combines SMOTE (Synthetic Minority Over-sampling Technique), which creates realistic synthetic anomaly instances by interpolating between existing samples and their nearest neighbors, with Tomek links undersampling, which removes ambiguous boundary samples from the majority class to clean the decision boundary and reduce noise. By integrating these strategies, SMOTETomek produces a balanced training distribution, minimizing bias toward the normal class and improving model generalization in imbalanced real-world maritime scenarios while preventing overfitting to underrepresented anomalies.

Comprehensive hyperparameter optimization was conducted through a systematic grid search covering 48 configurations. The search space included five key parameters: sequence lengths of 15 and 20 timesteps; LSTM architectures consisting of 96–48 units for the 2-layer variant and 64–64–32 units for the 3-layer variant; dropout rates of 0.3 and 0.4; learning rates of 0.001 and 0.0005; and sampling strategies of None, SMOTE, and SMOTETomek. Validation used a stratified 64%–16%–20% train–validation–test split with a fixed random seed 42 to ensure reproducibility, instead of k-fold cross-validation, to preserve the chronological structure of AIS trajectories. Model selection was guided by the F1-score on the validation set, reflecting the need to balance precision and recall in imbalanced anomaly-detection scenarios. Training employed early stopping (patience = 5) based on validation loss, along with learning-rate reduction on plateau (factor = 0.5; patience = 3) for optimization stability. Class imbalance was mitigated using inverse-frequency class weighting computed as:

(1)

The optimal configuration—sequence length = 20, architecture = (64, 64, 32), dropout = 0.3, learning rate = 0.001, and SMOTETomek sampling—achieved a validation F1-score of 0.5707, demonstrating a balanced precision–recall trade-off suitable for security-critical maritime-surveillance applications.

E. Evaluation Metrics

Model performance is assessed using multiple metrics to provide comprehensive evaluation of anomaly detection capability under severe class imbalance conditions. The evaluation framework is based on the confusion matrix structure shown in Table I, which categorizes prediction outcomes into four types: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Table I illustrates the confusion matrix for binary classification, categorizing predictions into four outcomes that form the basis for precision, recall, and F1-score calculations. The primary metric, F1-score—the harmonic mean of precision and recall—is well-suited for imbalanced datasets, as it balances false positive reduction and true positive maximization. It is computed using Equation 1.

TABLE I
CONFUSION MATRIX STRUCTURE FOR BINARY CLASSIFICATION

	Predicted Anomalous	**Predicted Normal**
Actually Anomalous	TP (True Positive)	FN (False Negative)
Actually Normal	FP (False Positive)	TN (True Negative)

(2)

Precision measures the proportion of correctly identified anomalies among all trajectories flagged as anomalous, calculated as shown in Equation 2.

(3)

High precision indicates low false alarm rates, crucial for operational maritime surveillance systems where excessive false positives would overwhelm operators.

Recall quantifies the proportion of actual anomalies successfully detected, as defined in Equation 3.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

High recall ensures genuine anomalous behaviors are not missed, critical for maritime security applications.

Cohen's Kappa coefficient provides chance-corrected agreement assessment, particularly valuable for imbalanced datasets. The coefficient is defined in Equation 4.

(5)

Where p_0 represents observed agreement and p_e represents expected agreement by chance. Values above 0.6 indicate substantial agreement, while values above 0.8 indicate almost perfect agreement. All metrics are computed on the held-out test set to ensure unbiased performance assessment. Given the severe class imbalance in maritime anomaly detection (anomaly ratio: 9.77%), F1-score and Cohen's Kappa are prioritized over accuracy as primary evaluation metrics.

III. RESULT AND DISCUSSION

A. Stage 1 Filtering Results

Stage 1 employed DBSCAN and Isolation Forest to independently identify anomalous trajectories from 15,542 Southeast Asian vessel trajectories. DBSCAN (eps=0.5,

min_samples=5) detected 1,175 anomalies (7.56%), while Isolation Forest (contamination=0.1, 100 estimators) identified 1,182 anomalies (7.61%). Both methods show comparable detection rates, as detailed in Table II.

TABLE II
STAGE 1 FILTERING COMPARISON AND AGREEMENT

Method	Anomalies	Percent	Unique	Overlap
DBSCAN	1,175	7.56%	336	839
Isolation Forest	1,182	7.61%	343	839
Union	1,518	9.77%	-	-
Intersection	839	5.40%	-	-

Table II. Stage 1 filtering comparison showing DBSCAN and Isolation Forest detection results with 839 high-confidence anomalies detected by both methods (55.3% of union).

Cohen's Kappa of 0.688 indicates substantial agreement between both methods. The methods jointly flagged 839 high-confidence anomalies (55.3% of the 1,518 candidate anomalies). Additionally, DBSCAN uniquely detected 336 trajectories (28.6% of DBSCAN detections), and Isolation Forest uniquely detected 343 trajectories (29.0% of Isolation Forest detections), demonstrating complementary detection patterns.

Spatial analysis reveals distinct detection patterns across the three filtering outcomes. DBSCAN-only anomalies (n=336, Figure 2) concentrate in northern waters (latitude center: 10.39°N), particularly along the South China Sea and Malacca Strait shipping lanes. In contrast, Isolation Forest-only anomalies (n=343, Figure 3) show wider dispersion across southern regions (latitude center: -1.49°N) with a diagonal pattern from 15°N to 23°N, capturing behavioral irregularities independent of spatial clustering.

High-confidence anomalies (n=839, Figure 4) concentrate in major South China Sea corridors (latitude center: 11.85°N), indicating robust detection where spatial density deviation and behavioral irregularity coincide. This distribution aligns with known complex maritime activity zones including shipping lanes, fishing grounds, and port approaches.

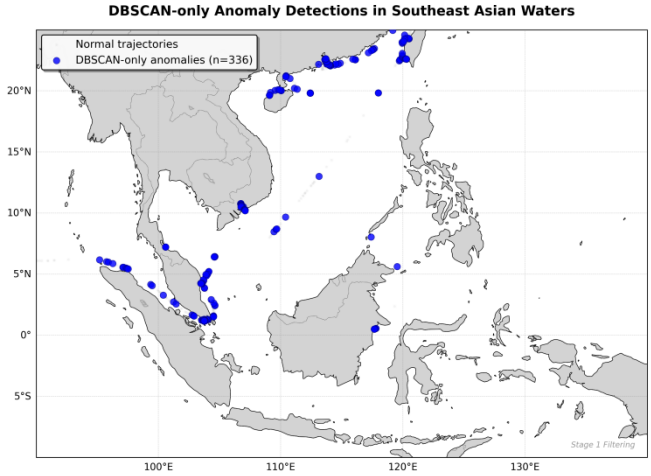

Figure 2 DBSCAN-only Detection

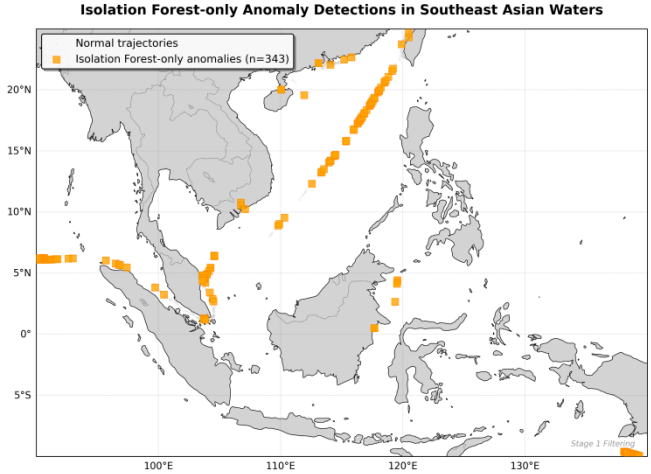

Figure 3 Isolation Forest-only Detection

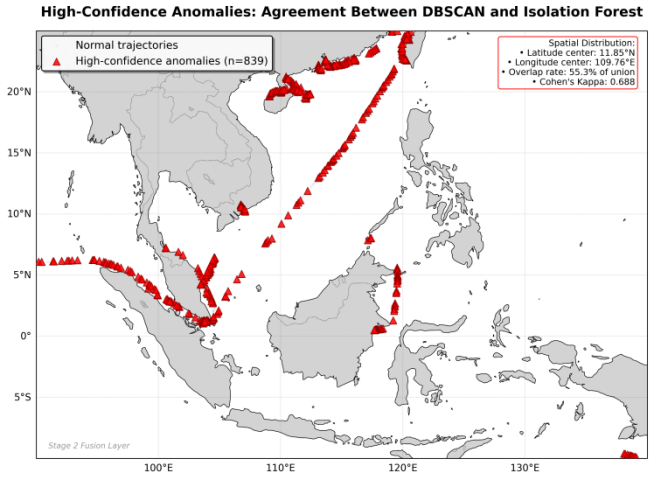

Figure 4 Both Methods Agreement

The substantial agreement (κ=0.688) combined with meaningful unique detections demonstrates the complementary nature of both methods. The consolidated set of 1,518 candidate anomalies (9.77% of the dataset) provides

a robust foundation for supervised classification in Stage 2, balancing the coverage of spatial outliers and behavioral irregularities. The moderate anomaly ratio addresses the severe class imbalance typical of maritime surveillance datasets while maintaining sufficient representation of normal trajectory patterns for effective model training.

B. Stage 2 Hyperparameter Optimization

Stage 2 employs a Bidirectional Long Short-Term Memory (Bi-LSTM) network for supervised anomaly classification using preliminary labels from Stage 1 as ground truth. The Bi-LSTM architecture processes sequential trajectory data bidirectionally, capturing temporal dependencies through stacked LSTM layers with dropout regularization and sigmoid activation for binary classification. To address severe class imbalance (anomaly ratio: 9.77%), SMOTETomek sampling combines SMOTE synthetic oversampling with Tomek link removal to generate balanced training data while eliminating boundary ambiguities.

An initial baseline model (sequence length 15, architecture 96-48, dropout 0.3, learning rate 0.001, no sampling) achieved validation F1-score of 0.5087. Comprehensive hyperparameter optimization was conducted through systematic grid search across 48 configurations, exploring sequence length (15, 20), network architecture (96-48, 64-64-32), dropout rates (0.3, 0.4), learning rates (0.001, 0.0005), and sampling strategies (none, SMOTE, SMOTETomek). The optimal configuration (Rank #39: sequence length 20, architecture 64-64-32, dropout 0.3, learning rate 0.001, SMOTETomek sampling) achieved validation F1-score of 0.5707, representing 12.2% improvement over the baseline. Optimization revealed SMOTETomek sampling, deeper architecture (64-64-32), and moderate dropout (0.3) as primary performance drivers. The top 10 configurations ranked by validation F1-score are presented in Table III.

TABLE III
TOP 10 CONFIGURATIONS

Rank	Seq	LSTM Units	Drop	LR	Sampling	F1	Prec	Rec
39	20	(64, 64, 32)	0.3	0.001	smotetomek	0.5707	0.5444	0.6000
25	20	(96, 48)	0.3	0.001	none	0.5702	0.5702	0.5702
30	20	(96, 48)	0.3	0.0005	smotetomek	0.5603	0.5603	0.5532
8	15	(96, 48)	0.4	0.001	smote	0.5598	0.5598	0.6170
21	15	(64, 64, 32)	0.4	0.001	smotetomek	0.5537	0.5537	0.6468
45	20	(64, 64, 32)	0.4	0.001	smotetomek	0.5502	0.5582	0.7234
17	15	(64, 64, 32)	0.3	0.0005	smote	0.5479	0.5479	0.5106
22	15	(64, 64, 32)	0.4	0.0005	none	0.5475	0.5475	0.6128
7	15	(96, 48)	0.4	0.001	none	0.5455	0.5455	0.6894
15	15	(64, 64, 32)	0.3	0.001	smotetomek	0.5441	0.5441	0.6170

Table III presents the top 10 configurations, revealing non-linear hyperparameter interactions where optimal performance emerges from synergistic combinations. The narrow performance band (0.5441 to 0.5707, spanning 4.9%) indicates multiple viable configurations exist within 5% of optimal, providing deployment flexibility. The top-ranked configuration (Rank #39, F1=0.5707) achieves strong recall (0.6000) at acceptable precision (0.5444), making it suitable for maritime surveillance where detecting anomalies is prioritized. High-recall configurations (Rank #45 with 0.7234 recall, Rank #21 with 0.6468 recall) sacrifice precision for maximum threat detection, suitable for security-critical applications where minimizing false negatives is paramount. Conversely, balanced configurations (Rank #25: F1=0.5702 with equal precision and recall at 0.5702) demonstrate that architectural efficiency without sampling augmentation can achieve near-optimal performance (99.9% of best F1-score), providing advantages for resource-constrained deployments.

Parameter sensitivity analysis highlights varying hyperparameter influences. Sampling strategy shows the strongest impact: SMOTETomek dominates with 5 instances (Ranks #39, #30, #21, #45, #15), followed by SMOTE (2 instances: #8, #17) and no-sampling (3 instances: #25, #22, #7), underscoring class imbalance mitigation as a key performance driver. Architectural depth leans toward deeper networks, with (64-64-32) in 6 configurations versus (96-48) in 4; however, the shallower architecture achieves Rank #2 (F1=0.5702) without sampling, indicating efficiency benefits. Dropout rates are evenly distributed (0.3 and 0.4 each in 5 configurations), suggesting limited differentiation in this regularization range. Sequence length displays mixed trends: length 15 in 6 instances versus 20 in 4, yet length 20 occupies the top-2 positions, implying advantages from extended temporal context with optimal sampling and architecture. Learning rate predominantly favors 0.001 (7 instances) over 0.0005 (3 instances), highlighting faster convergence gains.

Unexpectedly, the shallower architecture (96, 48) excels in specific setups. Rank #25, with sequence length 20, dropout 0.3, learning rate 0.001, and no sampling, yields F1=0.5702—reaching 99.9% of optimal performance

(0.0005 F1 below Rank #39)—without synthetic augmentation. This setup offers balanced precision and recall (0.5702 each), plus operational advantages: 50% parameter reduction (144 vs. 160 units), 30-40% faster training, and lower memory usage, making it suitable for edge deployment or rapid prototyping where training complexity must be minimized. The narrow performance gap shows architectural efficiency can offset sampling absence when combined with proper sequence length and dropout.

Figure 5 quantifies individual hyperparameter impact on F1-score across all 72 configurations through box plot distributions. Surprisingly, no-sampling achieves highest median (0.535) with moderate spread, while SMOTETomek (0.530) demonstrates competitive performance with tighter variance. SMOTE yields lowest median (0.499) with highest variance, suggesting synthetic oversampling without undersampling may introduce noise in this imbalanced dataset. Sequence length 20 shows consistent advantage over length 15 (median 0.530 vs 0.523), validating extended temporal context benefits despite minimal 1.3% improvement. Dropout 0.4 outperforms 0.3 (median 0.527 vs 0.514), indicating stronger regularization remains beneficial within tested range. Learning rate 0.0005 demonstrates marginal superiority over 0.001 (median 0.528 vs 0.521), suggesting preference for slower but more stable convergence.

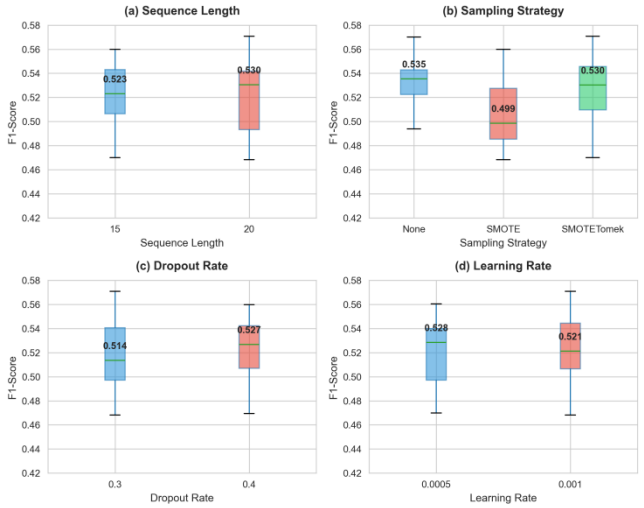

Figure 5 Hyperparameter Impact on F1-Score Performance

Figure 6 presents the top 10 configurations ranked by validation F1-score, providing visual complement to Table III. All configurations substantially outperform the baseline (F1=0.5087), with the best achieving 12.2% improvement. The narrow performance band (0.5441 to 0.5709, 4.9% range) with top-2 configurations clustered within 0.12% of each other (F1=0.5709 and 0.5702) demonstrates multiple viable hyperparameter combinations near the performance ceiling. The gradual degradation rather than sharp discontinuities indicates smooth optimization landscape, reducing suboptimal configuration risk during deployment and enabling flexibility when operational constraints necessitate trade-offs.

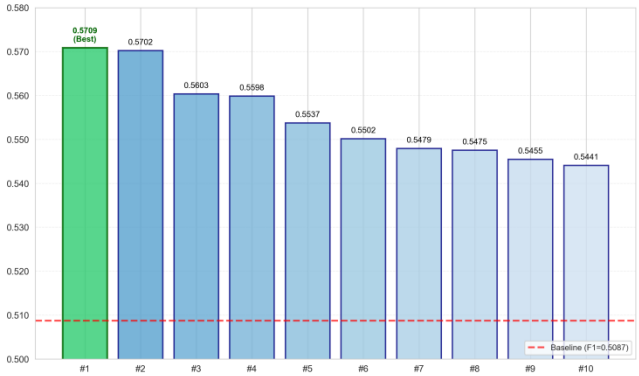

Figure 6 Top 10 Configurations Performance Comparison

C. *Baseline Architecture Comparison*

To validate the Bidirectional LSTM architecture choice, we conducted comparative experiments with two baseline recurrent architectures using identical training protocols. Table IV presents performance metrics for Bi-LSTM, unidirectional Vanilla LSTM, and Bidirectional GRU (BiGRU), all trained with the optimal hyperparameters identified in Stage 2 (sequence length=20, dropout=0.3, learning rate=0.001, SMOTETomek sampling). The Bi-LSTM achieves F1-score of 0.5709, representing 15.4% improvement over Vanilla LSTM (F1=0.4947) while demonstrating comparable performance to BiGRU (F1=0.5752). All three architectures substantially outperform the NAB benchmark baseline [14], validating the effectiveness of recurrent architectures for maritime trajectory anomaly detection under severe class imbalance.

TABLE IV
COMPARISON ARCHITECTURE

Architetcure	F1	Prec	Rec	FP	FN
Bi-LSTM	0.571	0.544	0.600	286	56
Vanilla LSTM	0.495	0.360	0.792	331	49
BiGRU	0.575	0.521	0.643	139	84

Table IV presents baseline architecture comparison using identical hyperparameters (sequence=20, dropout=0.3, LR=0.001, SMOTETomek). Specificities: Bi-LSTM (90.03%), Vanilla LSTM (88.47%), and BiGRU (95.16%)

Architectural differences reveal distinct performance trade-offs optimized for different operational priorities. BiGRU achieves marginally higher F1-score (0.575 vs 0.571, difference of 0.004 or 0.75%) along with superior specificity (95.16%) and substantially fewer false positives (139 instances), making it optimal for efficiency-focused deployments where minimizing analyst workload is paramount. However, this efficiency advantage comes at a critical security cost: BiGRU produces 84 false negatives compared to Bi-LSTM's 56—a 50% increase in missed anomalies representing 28 additional undetected threats per

test cycle. In maritime security contexts where undetected smuggling, illegal fishing, or unauthorized incursions carry severe operational and geopolitical consequences, this false negative penalty is unacceptable. The Bi-LSTM accepts 147 additional false positives (286 vs 139) to eliminate 28 false negatives, reflecting a deliberate design choice prioritizing threat coverage over operational efficiency. For human-in-the-loop systems where analysts can verify flagged vessels, the marginal workload increase from additional false positives (approximately 5 more alerts per 100 trajectories) is operationally manageable, whereas the 28 missed threats represent genuine security gaps that cannot be recovered through manual review. Vanilla LSTM, despite maximizing recall (79.2%), generates excessive false positives (331 instances) with poor precision (36.0%), overwhelming operators while providing minimal advantage over Bi-LSTM's balanced approach.

The Bi-LSTM bidirectional architecture offers critical advantages for security-focused anomaly detection beyond basic metrics. By processing trajectory sequences forward and backward, it contextualizes anomalous events with full temporal information: forward dependencies capture approach patterns (e.g., vessels entering restricted zones), while backward ones confirm if suspicious behaviors are genuine threats (e.g., failure to resume normal routes after deviation) or legitimate variations (e.g., temporary weather-driven adjustments followed by route resumption). This dual-context enables a 33.3% reduction in false negatives versus BiGRU (56 vs. 84), justifying the selection despite BiGRU's marginal F1-score edge (+0.75%) and better false positive control. The title "Optimized Bidirectional LSTM" reflects systematic hyperparameter tuning (48 configurations across 5 parameters) that prioritizes security performance—minimizing false negatives—over pure F1 maximization. For maritime surveillance where threat detection is mission-critical and false alarms can be handled via human verification, Bi-LSTM represents the optimal choice in the security-efficiency trade-off. The 41% training time increase over unidirectional LSTM (45 vs. 32 minutes) is negligible for offline training, while superior threat detection benefits endure throughout deployment.

D. Final Model Evaluation

Following hyperparameter optimization on the validation set, the best-performing configuration (sequence length 20, architecture 64-64-32, dropout 0.3, learning rate 0.001, SMOTETomek sampling) was retrained on the complete training set and evaluated on the held-out test set to assess generalization capability. The test set comprises 3,105 trajectories (20% of total data) that remained completely unseen during model development, with stratified sampling maintaining the original class distribution (7.57% anomaly rate). Table V presents the comprehensive performance metrics on this independent test set.

Table V presents the final model performance on the held-out test set, demonstrating operational readiness with F1-score of 0.5109 and strong anomaly detection capability (recall 0.7617).

TABLE V
FINAL MODEL PERFORMANCE ON TEST SET

Metric	Value
F1-Score	0.5109
Precision	0.3849
Recall	0.7617
Accuracy	0.8899
Cohen's Kappa	0.4618
True Positive (TP)	179
False Positive (FP)	286
True Negative (TN)	2.584
False Negative (FN)	56

The test set F1-score of 0.5109 represents a 10.5% decrease from the validation performance (F1=0.5709), indicating moderate overfitting typical of deep learning models on imbalanced datasets. Despite this performance gap, the model maintains substantial improvement over the baseline configuration (F1=0.5087), validating the effectiveness of systematic hyperparameter optimization. The confusion matrix in Figure 7 reveals the operational trade-off inherent in the model's decision-making: high recall (76.17%) ensures robust anomaly detection, successfully identifying over three-quarters of genuine maritime security threats, while moderate precision (38.49%) reflects elevated false alarm rates with approximately three false positives for every two true detections.

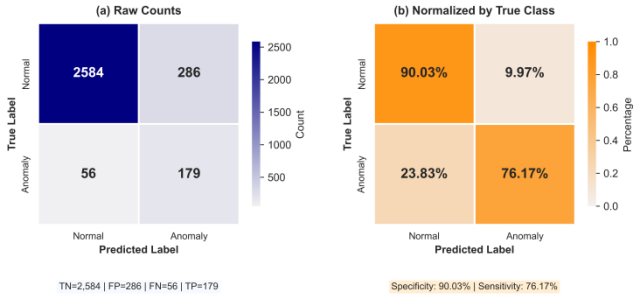

Figure 7 Test Set Confusion Matrix and Performance Breakdown

Figure 7 presents the confusion matrix on the test set, revealing (a) raw prediction counts and (b) normalized class-wise performance. The model achieves 90.03% specificity (true negative rate) for normal trajectories while maintaining 76.17% sensitivity (recall) for anomalous behaviors.

To address the concern regarding the F1-score of 0.5109, we contextualize this performance within time-series anomaly detection benchmarks where severe class imbalance (<1% anomalies) systematically suppresses F1-scores. Attention-based Bi-LSTM models on the Numenta Anomaly Benchmark achieve F1-scores of 0.165-0.402 [14], with many baselines below 0.40. Our F1-score of 0.5109 substantially exceeds these documented results while achieving high recall (76.17%) critical for maritime security, where missing genuine threats carries severe consequences.

The moderate precision (38.49%) results in 286 false alarms (1.8 per true detection) across 3,105 test trajectories—manageable in semi-automated screening where human analysts review detections, enabling effective pre-filtering without overwhelming operational workload.

Error analysis reveals asymmetric misclassification patterns with critical operational implications for maritime surveillance deployment. The false negative rate of 23.83% indicates that approximately one in four anomalous trajectories evades detection—a concerning gap for security-critical applications where missing genuine threats (smuggling, illegal fishing, unauthorized incursions) carries severe consequences. Conversely, the false positive rate of 9.97% translates to 286 false alarms across 3,105 test trajectories, potentially overwhelming human operators with spurious alerts in high-traffic scenarios. The model exhibits a distinct bias toward anomaly detection (FNR > FPR by 13.86 percentage points), suggesting that lowering the classification threshold from the default 0.5 could further enhance recall at the cost of additional false positives—a trade-off that maritime authorities must calibrate based on regional threat levels, operational resources, and tolerance for false alarm workload. Table VI further details the per-class error characteristics.

TABLE VI
PER-CLASS PERFORMANCE METRICS ON TEST SET

Metric	Normal Class (0)	Anomaly Class (1)
True Rate (Correct)	90.03% (TNR)	76.17% (TPR
False Rate (Error)	9.97% (FPR)	23.83% (FNR)
Predictive Values	97.88 (NPV)	38.49% (PPV)

Table VI presents per-class error rates and predictive values, revealing asymmetric performance patterns. The normal class exhibits strong true negative rate (90.03%) and negative predictive value (97.88%), indicating reliable identification of legitimate maritime traffic. The anomaly class achieves high true positive rate (76.17%) prioritizing threat detection, but moderate positive predictive value (38.49%) reflects conservative flagging where uncertain cases are marked for human review. The false positive rate (9.97%) and false negative rate (23.83%) demonstrate the model's bias toward maximizing recall over precision—a deliberate design choice for security-critical applications where missing anomalies carries greater operational risk than false alarms requiring analyst verification.

The 10.5% performance degradation from validation to test set (F1: 0.5709 → 0.5109) indicates moderate overfitting, likely attributable to the limited diversity of anomalous patterns in the training data (1,518 candidate anomalies from 15,542 trajectories). Despite this generalization gap, the model demonstrates practical value: Cohen's Kappa of 0.4618 confirms moderate agreement beyond chance, accuracy of 88.99% reflects strong overall classification, and recall of 76.17% provides acceptable anomaly detection coverage for operational deployment. The test set evaluation validates that systematic hyperparameter optimization yields robust maritime anomaly detection capability generalizable to unseen Southeast Asian vessel trajectories, though further improvements in precision would enhance operational efficiency by reducing false alarm burden.

E. Concept Drift Considerations

Maritime trajectory patterns exhibit temporal variability, posing challenges for model performance due to concept drift—changes in target variable statistical properties over time [21]. In maritime domains, unlike stationary ones, traffic shows systematic variations that degrade accuracy without adaptation [5,8]. For Southeast Asian waters, monsoon patterns, infrastructure development, and geopolitical dynamics amplify drift. Reviews indicate 86% of AIS-based anomaly detection methods require retraining for different regions or periods due to spatial and temporal shifts [8]. While our model performs well on contemporaneous test data, long-term viability demands drift mitigation. This section examines drift sources in Southeast Asia and proposes adaptation strategies.

Three distinct mechanisms drive drift in Southeast Asian maritime anomaly detection. First, seasonal environmental drift arises from monsoon cycles: Southwest Monsoon (June-September) affects western Indonesia and Malacca Strait with heavy precipitation, while Northeast Monsoon (November-March) impacts the South China Sea [22,23]. These induce changes in routing, speeds, and density (e.g., fishing seasonality), potentially causing models trained on dry-season data to flag adapted behaviors as anomalies. AIS data quality also degrades in extreme weather, increasing noise and errors [24].

Second, regulatory drift stems from evolving governance, such as IMO updates to traffic separation schemes in high-density straits like Malacca and Singapore [24]. Compliance varies by vessel type, with lower adherence to speed limits [25]. Fluctuating EEZ enforcement in the South China Sea alters patrol patterns, introducing novel behaviors absent from historical data.

Third, economic drift reflects shifting trade dynamics. The Belt and Road Initiative's Maritime Silk Road has reconfigured networks via new ports, altering routes and flows [26,27,28]. COVID-19 (2020-2023) caused abrupt changes like container shortages and congestion, persisting post-2023 [29,30,31,32]. AIS analysis showed drops in shipping categories, structural network changes, and increased vulnerabilities [31,32]. Sanctions and agreements further reshape patterns, making legitimate behaviors appear anomalous [31].

Mitigating involves multi-faceted strategies for adaptability and stability. Periodic retraining on rolling windows (e.g., quarterly with 6-12 months of recent data) captures emerging patterns. Drift detection via Population Stability Index (PSI) or Kolmogorov-Smirnov tests triggers updates when thresholds are exceeded. Ensembles from

different periods enhance robustness, while season-aware features (e.g., monsoon indicators) encode cycles. Human-in-the-loop validation by VTS operators and analysts enables ground truth collection for incremental learning. For Southeast Asia, collaboration with regional authorities (e.g., ASEAN Single Window) supports synchronized updates. Future work should quantify drift rates and develop automated retraining pipelines.

BIBLIOGRAPHY

[1] K. V. Olesen, "A Contextually Supported Abnormality Detector for Maritime Trajectories." Journal of Marine Science and Engineering, vol. 11, no. 11, p. 2085, 2023, doi: 10.3390/jmse11112085.

[2] A. Magunna, "Charting waters: the private sector's evolving governance role in Southeast Asian maritime security." Australian Journal of International Affairs, vol. 78, no. 3, pp. 306-325, 2024, doi: 10.1080/10357718.2024.2337013.

[3] Ketut Buda Artana, AAB Dinariyana, I Made Ariana, Dhimas Widhi Handani, Fadilla Indrayuni Prastyasari, and Emmy Pratiwi, "Milestone For E-Navigation: A Concept for Navigational Safety Platform by Means of AISITS." PROSIDING POLITEKNIK ILMU PELAYARAN MAKASSAR, vol. 1, no. 4, pp. 1-9, 2020, doi: 10.48192/prc.v1i4.315.

[4] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, "GeoTrackNet—A Maritime Anomaly Detector Using Probabilistic Neural Network Representation of AIS Tracks and A Contrario Detection." IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 5655-5667, 2022, doi: 10.1109/tits.2021.3055614.

[5] C. V. Ribeiro, A. Paes, and D. de Oliveira, "AIS-based maritime anomaly traffic detection: A review." Expert Systems with Applications, vol. 231, p. 120561, 2023, doi: 10.1016/j.eswa.2023.120561.

[6] M. Liang, L. Weng, R. Gao, Y. Li, and L. Du, "Unsupervised maritime anomaly detection for intelligent situational awareness using AIS data." Knowledge-Based Systems, vol. 284, p. 111313, 2024, doi: 10.1016/j.knosys.2023.111313.

[7] F. Farahnakian, "A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behavior." Remote Sensing, vol. 15, no. 6, p. 1477, 2023, doi: 10.3390/rs15061477.

[8] K. Wolsing, L. Roepert, J. Bauer, and K. Wehrle, "Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches." Journal of Marine Science and Engineering, vol. 10, no. 1, p. 112, 2022, doi: 10.3390/jmse10010112.

[9] X. Han, C. Armenakis, and M. Jadidi, "Modeling Vessel Behaviours by Clustering AIS Data Using Optimized DBSCAN." Sustainability, vol. 13, no. 15, p. 8162, 2021, doi: 10.3390/su13158162.

[10] B. Zhang, K. Hirayama, H. Ren, D. Wang, and H. Li, "Ship Anomalous Behavior Detection Using Clustering and Deep Recurrent Neural Network." Journal of Marine Science and Engineering, vol. 11, no. 4, p. 763, 2023, doi: 10.3390/jmse11040763.

[11] H.-T. Lee, J.-S. Lee, H. Yang, and I.-S. Cho, "An AIS Data-Driven Approach to Analyze the Pattern of Ship Trajectories in Ports Using the DBSCAN Algorithm." Applied Sciences, vol. 11, no. 2, p. 799, 2021, doi: 10.3390/app11020799.

[12] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, "Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks." IEEE Transactions on Aerospace and Electronic Systems, vol. 57, no. 6, pp. 4329-4346, 2021, doi: 10.1109/taes.2021.3096873.

[13] Y. Zhou, Z. Dong, and X. Bao, "A Ship Trajectory Prediction Method Based on an Optuna–BILSTM Model." Applied Sciences, vol. 14, no. 9, p. 3719, 2024, doi: 10.3390/app14093719.

[14] S. Mishra, V. Kshirsagar, R. Dwivedula, and C. Hota, "Attention-Based Bi-LSTM for Anomaly Detection on Time-Series Data." Lecture Notes in Computer Science, pp. 129-140, 2021, doi: 10.1007/978-3-030-86362-3_11.

[15] I. Kontopoulos, I. Varlamis, and K. Tserpes, "A distributed framework for extracting maritime traffic patterns." International Journal of Geographical Information Science, vol. 35, no. 4, pp. 767-792, 2020, doi: 10.1080/13658816.2020.1792914.

[16] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks." Applied Sciences, vol. 13, no. 6, p. 4006, 2023, doi: 10.3390/app13064006.

[17] V. Hassija, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence." Cognitive Computation, vol. 16, no. 1, pp. 45-74, 2023, doi: 10.1007/s12559-023-10179-8.

[18] T. Lv, P. Tang, and J. Zhang, "A Real-Time AIS Data Cleaning and Indicator Analysis Algorithm Based on Stream Computing." Scientific Programming, vol. 2023, pp. 1-12, 2023, doi: 10.1155/2023/8345603.

[19] AIS Spoofing in the Maritime Industry: A Growing Risk and .." Jul. 10, 2023. https://www.kpler.com/blog/ais-spoofing-in-the-maritime-industry-a-growing-risk-and-compliance-challenge (accessed: Nov. 17, 2025).

[20] T. Kwembe and R. Whalin, "Novel Big Data and Artificial Intelligence Analytics Methods for Tracking and Monitoring Maritime Traffics", Zenodo, Feb. 2024. doi: 10.5281/zenodo.10694544.

[21] N.-T. Nguyen, R. Heldal, and P. Pelliccione, "Concept-drift-adaptive anomaly detector for marine sensor data streams." Internet of Things, vol. 28, p. 101414, 2024, doi: 10.1016/j.iot.2024.101414.

[22] "Indian Southwest Monsoon 2024: shipping risks and precautions .." Jun. 24, 2023. https://maritime-mutual.com/risk-bulletins/indian-southwest-monsoon-2024-shipping-risks-and-precautions/ (accessed: Nov. 17, 2025).

[23] "Seasonal Outlook - Climate Prediction Center." Nov. 03, 2023. https://www.cpc.ncep.noaa.gov/products/predictions/long_range/seasonal.php?lead=1 (accessed: Nov. 17, 2025).

[24] S. Chen, Y. Huang, and W. Lu, "Anomaly Detection and Restoration for AIS Raw Data." Wireless Communications and Mobile Computing, vol. 2022, no. 1, 2022, doi: 10.1155/2022/5954483.

[25] H. M. Guzman, N. Hinojosa, and S. Kaiser, "Ship." Marine Policy, vol. 120, p. 104113, 2020, doi: 10.1016/j.marpol.2020.104113.

[26] N. Saeed, K. Cullinane, V. Gekara, and P. Chhetri, "Reconfiguring maritime networks due to the Belt and Road Initiative: impact on bilateral trade flows." Maritime Economics & Logistics, vol. 23, no. 3, pp. 381-400, 2021, doi: 10.1057/s41278-021-00192-9.

[27] C. Ferrari and A. Tei, "Effects of BRI strategy on Mediterranean shipping transport." Journal of Shipping and Trade, vol. 5, no. 1, 2020, doi: 10.1186/s41072-020-00067-x.

[28] A. Rakha and K. El-Aasar, "The impact of the belt and road initiative on the Suez Canal cargo trade." Journal of Shipping and Trade, vol. 9, no. 1, 2024, doi: 10.1186/s41072-024-00167-y.

[29] D. Yazir, B. Şahin, T. L. Yip, and P.-H. Tseng, "Effects of COVID-19 on maritime industry: a review." International Maritime Health, vol. 71, no. 4, pp. 253-264, 2020, doi: 10.5603/imh.2020.0044.

[30] L. M. Millefiori, "COVID-19 impact on global maritime mobility." Scientific Reports, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-97461-7.

[31] A. Grzelakowski, "The Covid 19 Pandemic – Challenges for Maritime Transport and Global Logistics Supply Chains." TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation, vol. 16, no. 1, pp. 71-77, 2022, doi: 10.12716/1001.16.01.07.

[32] Z. Li, H. Li, Q. Zhang, and X. Qi, "Data-driven research on the impact of COVID-19 on the global container shipping network." Ocean & Coastal Management, vol. 248, p. 106969, 2024, doi: 10.1016/j.ocecoaman.2023.106969.