MALAYSIAN JOURNAL OF
CYBERSECURITY AND APPLICATIONS

CSAM
CYBER SECURITY
ACADEMIA MALAYSIA

ENHANCING MARITIME INTRUSION DETECTION THROUGH A MULTI-STAGE PREPROCESSING AND HYBRID RF–LSTM LEARNING MODEL

Warusia Yassin[1]*, Zulkiflee Muslim[1], Alessandro Guarino[2], Fauzi Adi Rafrastara[3] and Thivya Laxhimi Selvaraja[1]

[*,1]Faculty of Artificial Intelligence and Cyber Security, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]StAG, Gazan Prolongée, Antibes, France
[3]Department of Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia.

Email: s.m.warusia@utem.edu.my[*]

ABSTRACT

The maritime industry is undergoing rapid digital transformation through the implementation of various modern technologies such as the Automatic Identification System (AIS), the Electronic Chart Display Information System (ECDIS), and the Integrated Bridge System (IBS). These new technologies create a much larger attack surface for potentially malicious actors looking to compromise maritime vessels or port facilities. However, the ability of existing Intrusion Detection Systems (IDS) to combat cyber-attacks on the maritime industry is hampered by two main challenges: the first challenge is due to the quality of the datasets (maritime and security) used for training existing IDS models, which results in the datasets being of low quality (noisy, unbalanced and heterogeneous) and limit the detection of a large number of cyber threats with high precision; and the second challenge is that existing machine learning models, which are standalone, depend only on static features (for example, IP addresses, etc.) and do not consider the temporal dynamics embedded in the maritime communication patterns, which results in lower detection performance for sequential and behaviour-based attacks (for example, staging the attack or using multiple transmissions) such as spoofing, staging a coordinated attack, and transmitting sequentially, all three attacks are better detected if the underlying communications between vessels and ports are taken into account. To address these challenges, the present study provides two important contributions: (i) the design of a multi-stage preprocessing module specific to the characteristics of each dataset, which enhances the quality of the training data by filtering out noise, encoding, balancing the classes, and preparing time-series data; and (ii) the development of a hybrid Random Forest (RF) and Long Short-Term Memory (LSTM) Deep Learning framework, which combines the ability of Random Forests to classify based on feature inputs with the ability of LSTM networks to model temporal sequences of input data. The newly proposed framework is thoroughly evaluated against a series of multiple datasets (AIS, CICIDS2017, and Darknet), to ensure it is robust across a variety of maritime and intrusion attack scenarios.

Keywords: Maritime Cybersecurity; Intrusion Detection; Hybrid RF-LSTM; Preprocessing Module; Temporal Anomaly Detection.

1. INTRODUCTION

The maritime sector is of major significance to global logistics, as it supports international commerce as well as many of the key supply chain links in the form of shipping, offshore oil and gas platforms and port operations [1]. The last ten years have seen a massive roll-out of technology in the maritime industry, driven mainly by digitalization and the rise of interconnected cyber-physical systems, including (but not limited to) the Automatic Identification System (AIS), Electronic Chart Display and Information System (ECDIS), Integrated Bridge Systems (IBS), satellite communication networks, and wide-ranging deployment of Internet of Things (IoT) sensors [2]. These systems have enabled vessels to be tracked in real time, allowed vessels to optimise their routes, allowed auto-piloted vessel navigation, and facilitate remote monitoring of ships. With increased reliance on the use of technology, the number of avenues for cyber-attacks (i.e. increased cyber attack surface) has also increased on the maritime industry. The recent increase in GPS spoofing, tampering with the AIS, ransomware and unauthorised access to systems are enough reasons to be alarmed about maritime cyber security [3].

The maritime environment presents unique challenges in the detection of cyber threats despite the significant advances made to date in cybersecurity research [4]. However, the use of legacy security controls, such as signature-based intrusion detection systems (IDS), are ineffective in providing detection for newer or evolving

threats (e.g. zero-day attacks or advanced persistent threats) because they have been designed with a pre-defined threat signature that cannot detect anomalies during operation or an early-stage attack that may follow a non-standard pattern [5]. The introduction of the Internet of Things (IoT) into the maritime industry and the continued use of satellite-based communication systems increases the complexity and attack surface of maritime environments, and currently, security controls are not available to adequately protect the new systems [6]. Other limitations that contribute to poor performance of existing IDS include data characteristics related to cyberspace and the maritime industry (e.g. signal interference, inconsistent devices, and operating conditions), which result in noise, imbalanced data, missing data, and heterogeneous data types [7]. As a result, when data from these factors need to be pre-processed and analysed together, the lack of preprocessing further decreases the accuracy of threat detection. Traditional machine learning (ML) algorithms are incapable of capturing the temporal characteristics associated with maritime-level communications because they are primarily focused on a static representation of features. Furthermore, due to the evolution of many attack types on maritime systems (e.g. spoofing, staged intrusions, and coordinated transmission anomalies) a need exists for a sequence-aware analysis of them [8].

The increasing frequency of maritime cyber incidents and the critical role of navigation and communication systems, coupled with the inadequacy of existing Intrusion Detection Systems (IDS), has created an opportunity to build a hybrid detection system that can exploit both domain awareness and behaviour sensitivity. It is well known that maritime datasets contain characteristics such as signal drift, noise fluctuations, trajectory dependencies, and sequential behaviour that require custom pre-processing and temporal modelling techniques. This lack of capabilities will increase the likelihood that cyber threats will go undetected, threatening ship operations, crew safety, port operations, and the continuity of global trade. Therefore, we are creating a hybrid learning model that can leverage both the static features and the temporal sequences in maritime cyber datasets, supported by a pre-processing pipeline that is specific to the unique characteristics of these types of data. In order to solve these problems, this study will provide two major contributions that address the key problems identified within this study: (i) a multi-stage pre-processing module designed to improve the quality of maritime cyber datasets through noise filtering, class balancing, feature encoding, and temporal formatting corresponding to the feature characteristics of each dataset, and (ii) a hybrid Random Forest-Long Short-Term Memory (RF-LSTM) model that incorporates static feature classification along with deep temporal sequence analysis in order to facilitate better detection of behaviour-based cyber threats.

The main objectives of this research include:

a. The creation of a dataset adaptive preprocessing module that will provide higher reliability for both the maritime and cybersecurity datatypes; and
b. The development of a hybrid RF-LSTM intrusion detection capability that will allow for both static and temporal analysis of attack patterns.

The intent of this article is to demonstrate the value and robustness of an adaptable hybrid intrusion detection framework that is capable of improving the quality of datasets and the way in which attacks can be modeled over time, thereby increasing the accuracy of malware detection and anomaly detection in the modern maritime cyber-physical system.

2. RELATED STUDY

As technology continues to digitize many aspects of vessels' navigation, communication and operational systems, the maritime domain has become a prime target for cyberattacks. The systems onboard vessels today, including Automatic Identification Systems (AIS), Electronic Chart Display Information System (ECDIS), Integrated Bridge Systems (IBS), vessel Radars, satellite communication systems and onboard Internet of Things (IoT) sensors all interact to form a highly developed Cyber-Physical Environmental System (CPES) [5]. The interconnectivity of these tools provides users with increased situational awareness and improves the efficiency of vessel operations, but also opens up new opportunities for the introduction of computer viruses, malware propagation, spoofing and other types of malicious cyber activity that may disrupt vessel operations or put the vessel and its crew at risk. Maritime Cyber Attack Database indicates that maritime cyber attack incidents have increased by over 400% from 2017-2023, creating a rapidly evolving threat environment [9]. AIS spoofing, GPS manipulation, Ransomware attacks on Ports, and Intrusions into Industrial Control Systems (ICS) have resulted in operational disruptions, financial losses, and significant safety risks on vessels operating in waters around the globe [10]. Analysis of Maritime Communications Data (i.e. AIS messages, Navigational Telemetry, and Network

Traffic Logs) frequently demonstrates high levels of "Noise", inconsistencies, missing data, and an imbalanced distribution of both classes and frequencies for this type of data. These characteristics, combined with environmental interferences, multipath Propagation and a multitude of different manufacturers' equipment, create a unique set of challenges requiring Custom Preprocessing and Analytical Methods to be developed for the analysis of Maritime Datasets [11], [12]. If these issues are not addressed prior to utilizing the Maritime Cybersecurity Intrusion Detection System (IDS), the resulting analysis will suffer from decreased levels of accuracy and increased incidence rates of False Positives [13].

2.1 Cyber Threat Detection and AI-Based Hybrid Learning in the Maritime Domain

Detecting cyber-based threats in the maritime environment can be hard due to noisy signal AIS systems, GNSS manipulation, diverse OT/IT systems and limited bandwidth on ships' infrastructure. Traditional intrusion detection systems rely on methods based on signatures (i.e. Deep Packet Inspection, firewalls and VTS rules), which are effective for identifying malware that has already been identified or port scans. However, signature-based detection is ineffective against zero-day (the first time malware is discovered), AIS spoofing (false signals), GNSS manipulation or sophisticated OT malware since ships operate with infrequent signature updates [14]. Consequently, anomaly-based detection learns from historical data on normal vessel behaviour via AIS, telemetry and maritime IoT data and detects abnormal activities — such as the alteration of a vessel's trajectory; inconsistencies in the timing of transmission of an AIS signal; or irregularities in packet bursts. In general, noise in AIS data and sporadic transmission results in high false positive rates unless preprocessing techniques, which takes into account the dataset, are applied [11]. Hybrid IDS attempts to merge the best features of signature-based and anomaly-based approaches; hybrid IDS can leverage static vessel attributes (e.g. ship type, speed range, flag state, route class) and timing of movements (e.g. trajectory changes, synchronised spoofing, campaigns using Darknet and abnormal timing of transmissions) [15].

The use of Artificial Intelligence (AI) has become a focal point for maritime intrusion detection. Machine Learning (ML) techniques like Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN) and Gradient Boosting (GBM), carry many advantages for Maritime Intrusion Detection Systems (MIDS). ML has most successfully identified and classified structured features, operates efficiently, and provides ranked interpretation of unique importance. They all treat inputs to models as 'static tables' and do not incorporate temporal dependencies of inputs making it difficult for these models to identify staged attacks (progressive route manipulation) or type-based attack activities (multi-step malware injection) [8]. Deep Learning (DL) machines have shown better performance in capturing the non-linear and temporal aspects of data collected from both AIS and network logs using LSTM & GRU networks and CNN–RNN hybrid networks. Trajectory prediction, anomalous detection of AIS, and identification of intentional tampering have been performed using these types of DL models. However, due to high levels of sensitivity to noise, missing data, and an imbalance in frequency and timing of these types of data shows that they perform poorly when using smaller computer systems (like ships), and typically requires processing done by larger remote computer systems (like cloud computing) [16]. Therefore, Hybrid Learning Models (ML+DL), such as Random Forest with Long Short-Term Memory (RF-LSTM), CNN with LSTM, AutoEncoder with LSTM, and various Ensemble ML/DL Pipelines have been proposed for detecting and predicting vulnerabilities within Maritime Systems.

Nonetheless, hybrids are not widely studied or validated in the maritime field using a variety of different maritime data sources (including AIS Spoofing, GNSS Manipulation, Port IoT Compromise, and Darknet-Driven Malware). The majority of hybrid study designs are based solely on standard benchmark datasets such as the CICIDS2017 and the UNSW-NB15, and do not incorporate the appropriate dataset adaptive preprocessing based on the characteristics of noisy AIS data and maritime network data such as Class Imbalance, Missing Transfers, and Inconsistent Sampling [7]. There is therefore a lack of a maritime-centred Hybrid RF-LSTM framework that employs a specific Pre-Processing Module to clean and impute noisy AIS data and maritime network data; balance classes and align inconsistent time samples; and co-model both Static Vessel Attributes and Temporal Attack Patterns. Such a framework supports the recent initiatives in the literature for AI based/OT-Centric intrusion detection for Maritime Cyber-Physical Environments and supports the methodological approach taken in this study.

2.2 Previous Work

Over the last ten years, maritime cybersecurity research has matured due to new technologies adopted throughout the maritime industry and the increasing number of cyber attacks targeting ships, ports, and connected equipment

within the maritime ecosystem. Recent research has focused on deep learning techniques for anomaly detection and maritime cybersecurity as shown with the review of the literature in Table 1. The following Table shows how deep learning models demonstrate an increase in research examining new methodologies for using Automatic Identification System (AIS) data to explain and help detect anomalies in vessels crew members maintain. The first investigations from [17] explored the potential for using spatio-temporal neural networks to identify anomalous vessel trajectories. Subsequently, research such as that of [18] began to incorporate Transformer and recurrent architectures for identifying abnormal vessel trajectories (i.e. "vessel shutdowns", or "spoof"), enabling further understanding of behaviours induced by cyber attacks at sea. More importantly, the application of these methodologies has continued to evolve with [19], [20]Tella et al.(2024), who offered real-time anomaly detection approached to support noisy and missing data within their real-time anomaly detection based upon Transformer and autoencoder based methodologies. The works of [11], which focus on both complementary survey and fusion bases, and those of [14], highlight the significance of developing machine-learning pipelines and integrating many different sensors into one single framework, both of which enhance Cyber Threats and Maritime Situational Awareness. Therefore, the studies of AIS Tampering Detection methods (see: [16]) that use an innovative hybrid DNN (deep neural network) approach have revealed new insights into how a hybrid deep neural architecture detects more advanced adversarial maritime threats than previously possible. Taken together, these works indicate that while deep learning methods continue to evolve and revolutionise the process of detecting maritime anomalies, the current framework still has challenges associated with: noise generated from the AIS; concept drift; heterogeneity across various sources; and limited contextual information on cyber threats. These challenges provide a clear motivation for developing an improved, dedicated hybrid maritime-aware anomaly detection framework, which is the focus of this research.

Table 1: Summary of Deep Learning Detection in Maritime Cybersecurity

Author	***DL Method***	***Limitation***	***Future Work***
Nguyen et al. (2022)	GeoTrackNet-style spatio-temporal DL (RNN/CNN encoder–decoder on AIS tracks)	Unsupervised, region-specific model; cannot clearly separate benign route anomalies from deliberate cyber attacks; limited handling of data gaps and spoofed points.	Add cyber-attack labels, multi-region/domain-adaptive training, fusion with radar/GNSS and integration into operational IDS. (arXiv)
Bernabé et al. (2023)	Self-supervised Transformer predicting AIS reception to flag intentional shutdown (sequence model)	Focuses only on intentional shutdown / gaps; does not detect spoofing or message tampering; high computational cost and mainly terrestrial AIS.	Generalise to multiple threat types (shutdown + spoofing + jamming), fuse satellite AIS and radar, optimise for real-time edge deployment. (ACM Digital Library)
Tella et al. (2024)	Transformer-based AIS trajectory model for maritime anomaly detection	Designed mainly for traffic efficiency/safety, not explicit cyber-attack taxonomy; interpretability for operators is limited; no multi-sensor fusion.	Extend labels to cyber-threat categories, add XAI modules for route-risk explanation, and integrate radar / EO data for richer threat context. (arXiv)
Maganaris et al. (2024)	GRU/LSTM autoencoder on AIS streams for unsupervised outlier detection	Sensitive to AIS noise and missing data; single-dataset training leads to poor generalisation across regions; no link to concrete intrusion types.	Introduce multi-dataset AIS preprocessing, domain-adaptive training and mapping between anomaly patterns and specific maritime cyber-attack scenarios.
Yang et al. (2024)	Survey covering ML/DL methods (CNN, LSTM, AE, GNN, etc.) for AIS-driven maritime analysis	Review emphasises traffic modelling and logistics; limited discussion of end-to-end cyber-attack detection and hybrid ML+DL IDS; lacks benchmark comparison for security tasks.	Identify unified benchmarks and metrics for maritime IDS, highlight gaps in AIS-based cyber-threat work, and encourage hybrid + XAI-enabled models for security monitoring. (IDEAS/RePEc)

Potamos et al. (2024)	Multi-modal CNN–LSTM combining AIS with maritime radar / sensor data	Depends on co-registered AIS–radar datasets, which are scarce; training and inference are computationally heavy; evaluated on limited attack types.	Build larger multi-sensor corpora, explore lightweight architectures and knowledge distillation, and test across broader spoofing / jamming / DDoS scenarios.
Lv et al. (2025)	MConLSTM (multi-dimensional CNN-LSTM) for identifying intentional AIS signal tampering in ship trajectories	Trained on data from a limited region; focuses on trajectory-level classification, not full SOC pipeline; still affected by noisy or sparse AIS records.	Extend to multi-region datasets, include AIS preprocessing tailored to spoofing, and combine with traditional ML / rule-based checks for operational maritime surveillance. (MDPI)

In spite of advances in deep learning for the purposes of detecting maritime anomalies and real-time monitoring of cyber-physical systems, there are still some significant gaps that have not been filled, which impacts our ability to operate robustly using the available maritime cyber security technology. To illustrate, while some of the earliest methods using deep learning to detect anomalous trajectories were developed by [17] and demonstrated good detection results, their performance deteriorates markedly when they are confronted with either noise from Aids to Navigation (AIS) systems, irregular sampling issues as well as spoofed data and missing transmissions that occur frequently during an attempted cyber-attack. Current maritime cyber security anomaly detection models that were developed primarily using data from AIS systems have been developed by [18], and have identified several behaviours that can be linked to cyber attacks; however, they do not have the capability to generalise well to other vessel types, sea lanes or different cyber manipulation patterns because of their reliance on either data from a single region or fixed feature sets. On the other hand, more recent models (e.g., Transformer-based models, [19] and RNN models, [20]) are able to achieve significantly better temporal modelling than their predecessors, but also rely heavily on clean AIS data and do not account for the issues of pre-processing GIS data that cause noise, data imbalance, and concept drift inherent in the maritime data communication streams. Due to these issues, studies that focus on survey-oriented approaches using multi-sensor fusion [11], [14] begins to show the need for integrated maritime data pipelines, but there has been little research conducted to quantify the effectiveness of hybrid ML and DL models in the detection of cyber attack threats.

The four limitations discussed collectively identify four areas that remain unaddressed by available literature. Firstly, current literature does not provide a systematic, maritime-aware processing pipeline that includes processing pipelines for addressing the noise of Automatic Identification Systems (AIS), missing values, the erratic nature of sampling intervals and inconsistencies in data from multiple sources. Secondly, there is an insufficient degree of integration between ML-based feature robustness and DL-based temporal modelling wherein the resultant models possess inadequate capabilities to cater both to static vessel characteristics and evolving cyber-attack signature characteristics. Thirdly, very few studies directly address cyber-related threats associated with navigational anomalies. The majority of studies are more focused on the phenomena of navigation anomalies than direct cyber-related threats such as spoofing of the AIS, coordinated trajectory manipulation, and GNSS-based data injection of false information. Finally, there is limited generalisation ability of the majority of currently available AIS models across regions and datasets as models evaluated have typically used only a single AIS or network corpus with no cross-dataset validation. Thus, the identified gaps suggest a strong need for a hybrid framework to facilitate a more effective detection of maritime intrusions through the integration of ML-based feature decay stability (such as Random Forests (RF)), DL-based temporal analysis (such as Long Short-Term Memory (LSTM)) and also include a dataset-aware pre-processing mechanism that addresses the issues related to the complexity of maritime data. This study addresses the aforementioned opportunities for improving the accuracy of detecting intrusions, reducing the occurrence of false positive alerts, and ultimately improving the operational resilience of these systems against real-world maritime cyber-intrusion threats.

3. PROPOSED MODEL

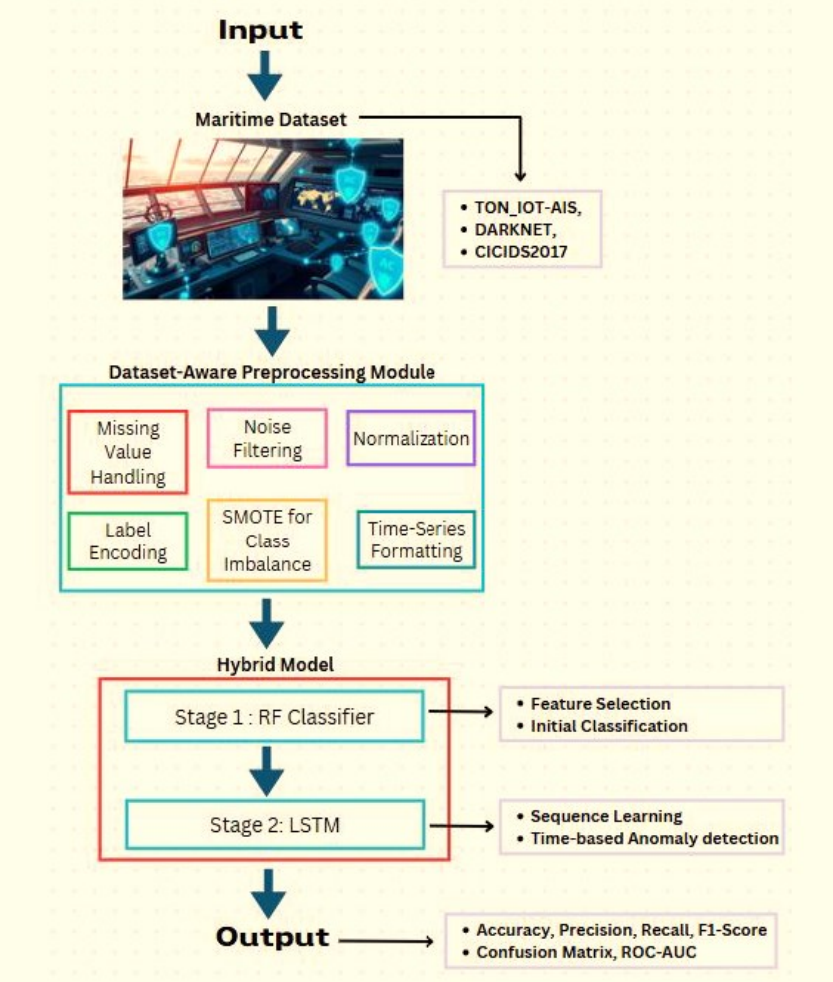

Fig. 1: The Proposed Detection Model

3.1 Module 1: Multi-Preprocessing Module

The proposed model for maritime intrusion detection is comprised of two major modules (illustrated in fig. 1): The Multi-Preprocessing Module and the Hybrid Learning Module (RF-LSTM). These two modules were developed to provide an improvement to the two major shortcomings presented in current maritime cybersecurity research: (1) the heterogeneous nature of maritime datasets, which also display a great deal of noise and variance, and (2) the failure of conventional ML to adequately capture the temporal nature of cyber-attacks. Past work either employed classical ML or deep learning techniques alone; thus, the combination of dataset-aware preprocessing and hybrid temporal/static learning within a single, unified architecture (for datasets from AIS, Darknet, and CICIDS2017) represents a new contribution to the literature. No previous maritime IDS framework contains this combination of elements.

Missing AIS values are reconstructed using linear interpolation to preserve vessel trajectory continuity:

$$\hat{x}_{i,j} = \frac{x_{i-1,j} + x_{i+1,j}}{2}, \tag{1}$$

while CICIDS and Darknet missing values are replaced using robust median imputation. Noise in AIS signals is mitigated using exponential smoothing,

$$x_t^{smooth} = \alpha x_t + (1-\alpha)x_{t-1}, \tag{2}$$

which filters unwanted movement spikes without removing meaningful navigational shifts. Normalisation is applied uniformly across datasets using Min–Max scaling,

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{3}$$

and categorical fields are transformed using One-Hot encoding to preserve semantic relationships. To overcome the severe imbalance found in attack categories, particularly rare AIS spoofing attempts, SMOTE generates synthetic minority samples using

$$x_{\text{new}} = x_i + \lambda(x_{nn} - x_i), \tag{4}$$

strengthening the model's ability to detect minority class behaviours. Finally, all datasets are segmented into sliding temporal windows,

$$S_t = \{x_t, x_{t+1}, \ldots, x_{t+k}\}, \tag{5}$$

enabling downstream temporal analysis.

The preprocessing operations described above collectively define a transformation function $\theta: \mathrm{X} \rightarrow \mathrm{X}'$, where X denotes the raw heterogeneous maritime dataset and X′ represents the cleaned, normalised, and temporally structured feature space. Linear interpolation will help to maintain the continuity of vessel trajectories in areas with sparse AIS sampling; on the other hand, exponential smoothing will eliminate the high-frequency noise present within the data while preserving long-term navigational trends. Min-MAX Normalisation helps to establish a common scale for all of the features that were created during preprocessing and therefore provides numerical stability as the learning proceeds. The SMOTE algorithm will create interpolated instances of the minority class between their nearest neighbours in the feature space, thereby providing a balance to the classifier's bias towards the majority class. Finally, the sliding window formulation of the data allows for mapping static observations as ordered temporal sequences, forming a basis on which to learn subsequent sequences within Module 2.

The initial module is the first to harmonize three datasets involving maritime activities (AIS/Data, DarkNet and CICIDS 2017) using a specialized pre-processing procedure which includes smoothing (for AIS data), balancing of the network flows, encoding of categories, and structuring of time-series. In previous work, each dataset was treated as a separate entity, with no attempt made to integrate disparate domains into a unified framework for developing an IDS to protect the maritime sector from cyberattacks. Dataset-aware pre-processing is an innovative concept that supports the hybrid architecture's ability to take into account and learn from differences in how different types of behaviours are represented within disparate sources of maritime information.

3.2 Module 2: Hybrid Learning Module (RF-LSTM)

Once the data has been pre-processed, it is then fed into a Hybrid Learning Module containing a Random Forest (RF) and Long Short-Term Memory (LSTM) based two-stage learning system. This architecture has been specifically designed to take advantage of the strengths of both RF (Machine Learning) and LSTM (Deep Learning) algorithms. While RF performs well with noisy and heterogeneous tabular data, it is not capable of capturing the sequential nature of attacks. On the other hand, while LSTM captures temporal dependencies between observations and predictions, it does not handle noisy or imbalance datasets effectively. The uniqueness of this hybrid design lies in the sequential nature of RF to LSTM integration, whereby RF first extracts stable, noise-free feature embeddings, and then LSTM models the progression of temporal behaviours based upon those embeddings. This sequential pipeline of RF to LSTM has never before been researched in the context of maritime IDS.

RF models are expressed as an ensemble of decision trees,

$$RF(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x), \tag{6}$$

and produce a feature importance vector based on Gini impurity reduction,

$$\phi_j = \sum_{t=1}^{T} \Delta Gini_{t,j}. \tag{7}$$

This allows the model to emphasise influential maritime features such as AIS speed variations, packet flow durations, or protocol anomalies. The feature importance vector produced by the Random Forest is obtained by aggregating the total reduction in Gini impurity contributed by each feature across all decision trees in the ensemble. The importance of each feature is calculated by summing the amount of impurity that was removed from the nodes when the respective feature was selected for splitting. The mathematical framework developed here provides insight into how the Random Forest model identifies maritime features that are both stable and discriminative while rejecting noise from irregularity in AIS transmissions or from variability caused by the encryption of transmitted messages.

LSTM processes temporal sequences using memory cell operations:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \qquad h_t = o_t \tanh(C_t), \tag{8}$$

producing a temporal embedding E_{LSTM} that captures behavioural transitions such as AIS drift, progressive port scans, or multi-stage command-and-control traffic. Temporal embedding, denoted as $ELSTM$ (representing the time-based encoding of events), corresponds to the hidden state produced by the last time step of input sequences that represent the chronological order in which activities have occurred. This representation provides a summary of all the activities in an input sequence and serves as an input to the Softmax classifier which estimates the posterior probabilities of an intrusion event. This means the Softmax classifier explicitly connects LSTM's memory to the intrusion detection system and uses that memory to define its Class Probabilities.

The embeddings E_{RF} and E_{LSTM} are fused into a unified representation:

$$E = [E_{RF} \| E_{LSTM}], \tag{9}$$

which is classified using a Softmax output layer,

$$\hat{y} = \text{Softmax}(WE + b) \tag{10}$$

The new model being proposed stands out from the previous models outlined in Tables 1-3 and shows how the performance capabilities of various methodologies differ when dealing with maritime data using an IDS approach, such as traditional ML methods used for detecting maritime events. Traditional machine learning based methods have good performance when applied using static, tabulated maritime datasets; however, they continually struggle to detect continuously changing time series traces/activities associated with maritime-based attacks, thus causing a significant drop in accuracy due to the lack of temporal relationships between observations. On the other hand, deep learning methods (especially LSTM & CNN) perform well in detecting temporal correlations, but there are still limitations with these methods when it comes to working with 'Noisy' or 'Imbalanced' Maritime-derived Attack (MDA) classes and 'Heterogeneous' Maritime Feature distributions. Extensive preprocessing is typically required prior to applying these methods on these types of datasets in order to achieve optimal performance. Existing hybrid ML & DL approaches provide some minor enhancements to the model by combining traditional ML methods with deep learning techniques, however, due to the lack of development specifically focused on the unique characteristics of the maritime environment and the need for an integrated approach, these approaches are limited in both the types of datasets that can be accommodated and the types of maritime-related datasets that can

be used in conjunction with one another. In contrast to both previous research methodologies and hybrid approaches, this proposed model combines a dataset adaptive modular preprocessing capability with a Sequential RF→LSTM hybrid architecture that works seamlessly across three unique datasets (i.e., AIS and Darknet and CICIDS), thus addressing all of the limitations of prior works across all previous research studies and providing an all-inclusive, domain-specific IDS solution for maritime intrusion detection that has not been previously developed.

3.3 Proposed Maritime Multi-Preprocessing Hybrid Learning Algorithm

The maritime multi-preprocessing hybrid learning algorithm outlined here illustrates the full process for constructing the proposed maritime intrusion detection model. This model comprises two primary components—the Multi-Preprocessing Component and the Hybrid Learning Component (RF to LSTM)—that provide the foundation for a subsequent Fusion-Based Classification step. The maritime multi-preprocessing hybrid learning algorithm has been developed to address the unique challenges posed by maritime datasets such as TON_IoT-AIS, Darknet, and CICIDS2017. To provide for an accurate maritime intrusion detection capability, each phase within the algorithm will apply to the data as well as provide a basis for creating temporally-stable and adjustable features that can be used to determine what constitutes an intrusion event within the context of the maritime environment.

Input:
X_AIS ← TON_IoT-AIS dataset
X_DK ← Darknet dataset
X_CICIDS ← CICIDS2017 dataset
Output:
y_hat ← Final intrusion prediction (normal / attack type)

Module 1: Multi-Preprocessing Module
For each dataset Xi in {X_AIS, X_DK, X_CICIDS} do
  Identify dataset_type ← DetectType(Xi)
  If dataset_type = AIS then
    Xi ← LinearInterpolation(Xi)
  Else
    Xi ← MedianImputation(Xi)
  If dataset_type = AIS then
    Xi ← ExponentialSmoothing(Xi)
    Xi ← MinMaxNormalize(Xi)
    Xi ← OneHotEncode(Xi)
    Xi ← SMOTE(Xi)
  Si ← CreateSlidingWindows(Xi, window_size=k)
End For

Module 2: Hybrid Learning Module (RF → LSTM)
RF_model ← TrainRandomForest({X_AIS, X_DK, X_CICIDS})
ERF ← RF_model.Transform({X_AIS, X_DK, X_CICIDS})
LSTM_model ← TrainLSTM({S_AIS, S_DK, S_CICIDS})
ELSTM ← LSTM_model.Encode({S_AIS, S_DK, S_CICIDS})

Fusion and Final Classification
E ← Concatenate(ERF, ELSTM)
y_hat ← Softmax(W·E + b)
Return y_hat
End Algorithm

4. EXPERIMENTS AND RESULT DISCUSSION

4.1 Dataset Used

This research will comprehensively evaluate the hybrid intrusion detection system (IDS) model developed by RF-LSTM through the use of three distinct heterogeneous datasets. Each dataset will apply to various facets of maritime cyber threats (e.g., navigating time series data, general intrusion traffic, and a dataset of stream-based cryptographic traffic). Each of these datasets represents both static threats (e.g., operational vessel data), temporal threats (i.e., intrusion attempts), and obfuscated threats. This table provides a summary of each dataset, along with a description of the threat it represents and a description of its respective function within the study of the hybrid IDS system (RF-LSTM). The combined usage of all three datasets allows for an effective assessment of whether the proposed hybrid IDS pipeline (RF-LSTM) performs effectively when confronted with static, temporal, and obfuscated traffic types.

Table 2: Dataset Used in This Study

Dataset	*Data Type*	*Threat Representation*	*Purpose in This Study*
AIS (TON_IoT AIS)	Time-series navigational logs	GPS spoofing, route anomalies, behavioral manipulation	Evaluate temporal anomaly detection + maritime-specific component
CICIDS2017	General intrusion network traffic	DoS, DDoS, brute-force, infiltration, botnets	Benchmark IDS performance and generalization
Darknet2020	Encrypted Tor/VPN flows	Obfuscated malware C2, covert communications	Test robustness under encrypted/noisy conditions

4.2 Evaluation Measurement

The proposed Hybrid RF–LSTM intrusion detection model's performance and robustness are evaluated using the supervised-learning evaluation metrics defined herein, due to their acceptance as standardised metrics in evaluating supervised learning. They provide relevant information regarding the classification performance in the maritime cybersecurity domain; therefore, they are suitable for our current application. The necessity of minimising false alarms while maximising the ability to accurately identify low-frequency, high-consequential attacks makes these metrics particularly relevant for assessing the Hybrid RF–LSTM intrusion detection model. The details contained in Table 3 indicate how each metric is defined and outlines their relevance for use within the area of maritime intrusion detection (Cybersecurity for Maritime Transportation).

Table 3: Evaluation Metrics Used in This Study

Metric	*Definition*	*Purpose in Maritime IDS*
Accuracy	Measures the proportion of correctly classified samples (both benign and malicious) over the total number of samples.	Provides an overall indication of model correctness across AIS, CICIDS2017, and Darknet datasets.
Precision	Measures the proportion of true positives among all predicted positives. Indicates how reliable the model is when it flags malicious samples.	Reduces false alarms in maritime operations, preventing unnecessary navigation or security intervention.
Recall	Measures the proportion of correctly identified malicious samples among all actual malicious cases.	Ensures critical attacks (e.g., GPS spoofing, botnet infiltration) are not missed.
F1-Score	The harmonic mean of precision and recall, balancing false positives and false negatives.	Provides a balanced performance measure under imbalanced datasets (AIS anomalies, Darknet minority classes).
ROC Curve / AUC	ROC plots True Positive Rate vs. False Positive Rate; AUC measures the area under this curve. Higher values indicate better class separability.	Evaluates discriminative ability under noisy or encrypted traffic (Darknet) and ambiguous boundaries (AIS).

4.3 Comparative Synthesis with Existing Baseline Approaches

The performance trends observed in this study are consistent with findings reported in prior maritime and network intrusion detection literature, particularly with respect to the strengths and limitations of static, temporal, and hybrid learning approaches. Random Forest (RF) models have traditionally been shown to perform well on both structured training datasets and benchmark datasets (e.g. based off of CICIDS) as they take advantage of their capability to extract relevant tabular features from various sources of data while maintaining a high level of resilience to error or inconsistency due to their built in statistical nature. The findings from this current study support those of previous studies, which identify this characteristic of RFs as the reason for the strong performance of the RF baseline model on the CICIDS2017 dataset, but lower performance by RFs when tested against higher levels of noise and more erratic datasets that do not exhibit the same level of structure. Furthermore, Sequence-based deep learning architectures, including LSTMs have been well studied for behavior-based anomaly detection in either maritime or network environments. Previous work has found that LSTMs work well when they have strong temporal dependencies and there are consistently strong temporal dependencies. However, once data is missing, sampled at irregular intervals, or has weak temporal continuity, the performance drops sharply. The current results support these prior findings as the LSTM approaches have poor performance on AIS and encrypted Darknet datasets when they contain fragmented sequences and stochastic traffic patterns.

Recent research indicates that creating a hybrid technique for creating models with both machine and deep learning components can reduce the impact of these problems, since they combine the strengths of both static and temporal feature robustness. The framework developed in this study, which is a combination of a Random Forest (RF) and Long Short-Term Memory (LSTM) approach, continues this trend by providing an enhanced way of preparing datasets for analysis and by providing improved temporal/spatial feature stability. On average, the hybrid RF–LSTM method produces a more optimal balance between precision and recall than both of the baseline models on all datasets, highlighting better performance when it comes to predicting outcomes based on data that has conflicting information that may be misleading or inaccurate. In contrast to existing hybrid intrusion detection approaches that are typically evaluated on a single benchmark dataset, the proposed model demonstrates consistent performance across AIS, CICIDS2017, and Darknet datasets. The robustness of the cross-dataset evaluation demonstrates how well the integrated methods of temporal modelling and feature stabilisation are able to complement each other. The results further attests to the suitability of our approach as a framework for real-life maritime cyber systems, where attack features and data quality can vary widely.

4.4 Comparative and Error Analysis

The proposed Hybrid RF – LSTM intrusion detection framework is evaluated with respect to traditional models for identifying intrusions that are frequently referenced in research for intrusion detection, in order to provide an element of comparability (or consistency) between both types of learning (static and temporal) in determining the levels of intrusion detection. Due to its performance on tabular or structured datasets and its ability to process high-feature ordinal values, Random Forests (RF) are chosen to represent an example of a classical machine learning baseline and the static-feature model in this comparison. Therefore, RF can determine the number of maritime intrusions which can be detected without incorporating any explicit modelling of temporal dependencies. In contrast, LSTM represents how to detect intrusions through the use of a sequence-based technique. It provides an example of how to model the temporal dependencies of Network Traffic and Aeronautical Information Systems (AIS) base formats and is often used to identify anomalous behaviour based on the behaviours of people, vehicles and vessels, among others. Therefore, the intrusion detection performance of behavioural models based on time will be evaluated only after the introduction of (a) prior feature stabilization; and (b) all temporal dependencies.

The Hybrid RF-LSTM model proposed combines a sequential approach of LSTM temporal modelling with RF feature learning. The aim of the evaluation process is to separate the effect of the hybridisation, and to find out if the addition of temporal dependency modelling to robust static features results in measurable improvements. For the evaluation of all baseline models, the same datasets (AIS, CICIDS2017, and Darknet) have been used; the same pre-processed pipelines outlined in Section 3, and all baseline models have been evaluated with the same set of evaluation metrics (Accuracy; Precision; Recall; F1-score; and ROC-AUC). This guarantees that any differences in performance observed between any two baseline Model Architectures will be purely due to their individual architecture as opposed to how the data were handled, or which metrics were selected for the evaluation. By reviewing the performance of Random Forest (RF), LSTM, and RFID-LSTM Hybrid models using AIS, CICIDS2017, and CICDarkNet 2020 datasets we see significant differences in the way they classify. The differences in performance can be seen clearly within the confusion matrix of each model. These differences arise

due both to the specific characteristics of individual datasets, as well as due to the advantages and disadvantages present in each model architecture. Furthermore, providing FP, FN, TP, and TN percentages enables a clearer quantitative representation of where the errors arise, thereby enhancing the scientific basis to adopt the hybrid model design. Table 4 presents the quantitative comparison between the proposed Hybrid RF–LSTM model and the baseline models (Random Forest and LSTM) across all evaluated datasets.

Table 4: Comparative Performance of RF, LSTM, and Hybrid RF–LSTM Across Datasets

Dataset	*Model*	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*	*ROC-AUC (%)*
AIS (TON_IoT)	Random Forest	96	77	73	75	97
	LSTM	95	55	38	39	97
	Hybrid RF–LSTM	98	78	75	76	99
CICIDS2017	Random Forest	98	98	97	98	98
	LSTM	98	96	98	98	98
	Hybrid RF–LSTM	99	99	100	100	100
CICDarknet2020	Random Forest	91	82	72	76	91
	LSTM	84	72	44	45	78
	Hybrid RF–LSTM	91	80	78	79	93

Upon analysis of the CICIDS2017 dataset, the Random Forest (RF) classification demonstrated an exceptional analytical performance. The results from the confusion matrix indicated that the RF had identified 244,748 True Negatives (TN) and zero False Positives (FP). In other words, the RF achieved a 100% True Negative Rate (TNR) for this dataset, which is quite rare in the area of Intrusion Detection Systems. In addition, the model identified 36,893 True Positives (TP) and three False Negatives (FN) yielding a False Negative Rate (FNR) of only 0.008%. This reinforces the competency of the RF model in recognising the clearly ordered and highly discriminative features of the CICIDS2017 dataset. The temporal structure of this dataset was shallow, indicating that the data contains short flow patterns rather than long behavioural sequences. Therefore, the RF classifier's decision boundary-driven approach was exceptionally suited to the dataset. The Long Short Term Memory (LSTM) model also performed very well, but was less robust than RF since it was not designed to efficiently detect all of the features and attributes of this nature, rather it was designed to identify specific burst-based behaviours. The Hybrid Modelling outperformed both of the models by attaining a Recall (100%) and an F1 Score (100%) and eliminating even the few FN instances created by the RF model. This indicates that the hybrid model is capable of capturing residual sequence-level dependencies, such as timing variations or inter-packet patterns, which RF alone may not fully exploit. In short, CICIDS2017's structure allows RF to dominate, but the hybrid model refines the detection process to achieve complete separation between benign and malicious samples.

Analysing the AIS dataset shows different results in terms of dynamics. With the application of RF, there were 35,212 TN and 33,736 TP; however, there were also 1,271 FP and 1,452 FN. The FP total results in an approximately 3.48% false positive rate (FPR) and the FN total results in a false negative rate (FNR) of 4.12%. While the FPR and FNR numbers are acceptable within numerous contexts, they point out limitations. The nature of AIS data is that it is inherently noisy, variable and incomplete; moreover, static decision boundaries based on RF are sensitive to abrupt yet legal vessel movements, which can include navigational course or velocity changes due to manning operations or environmental conditions (e.g., wind) or interactions with other vessels. These natural fluctuations generate FP errors because the static RF decision boundaries misinterpret them as anomalies. In contrast, FN errors are generated by slowly changing (i.e., subtle) gradual or low amplitude anomalies due to: (a) Native maritime operational scenarios that create 'drifting' (i.e., vessels moving without propulsion) or (b) Spoofing behaviours that are characterised by low amplitude (i.e., not greatly varying), where the static RF decision boundaries do not have the capacity to remember time to recognise deviations occurring over time.

The LSTM model does very poorly at detecting intrusions in AIS data; based on the multi-class confusion matrix, there is severe confusion between behaviour categories. While the exact counts of behaviours vary between each

class within the AIS data, the overall trend is that there are many more off diagonal entries (i.e., the behaviour categories share this trend) than there are matching diagonal entries for those behaviour categories. So when looking at the cases for intrusion classification in a collapsed format, LSTM performs poorly as evidenced by 38% and 0 for FNR. Based on this confusion, the reason there is a high FN rate for detection (62% of attacks classified as normal) is due to a lack of strong sequenced temporal relationships within the AIS temporal data (e.g., timestamps missing, irregular intervals, incomplete voyages, and heavy interpolation of values); the very discontinuous nature of the dataset prevents the LSTM model from learning how to represent stable sequences. Where there are temporal patterns (particularly with respect to changes in speed or changes in course), they tend to be both very slow and inconsistent in development or evolution, which renders the LSTM ineffective for learning to represent stable sequences of behaviour.

The use of a hybrid approach to improving classification of AIS significantly demonstrates that the hybrid classification model produces strong results indicated by a confusion matrix showing significant increases along the diagonal for the primary classes and significantly fewer misclassifications between adjacent behaviour classes as evidenced by the lower confusion matrices between adjacent behaviours. The utilisation of the hybrid model has resulted in hybrid classification results with approximately 78% precision and 75% recall. If we convert the recall to an error rate, then that means the false-negatives fall to around 25%, which is a very strong improvement, compared to standalone LSTM's false negative rate of about 62%. The hybrid model also demonstrates that the false positive rate of approximately 2.8% is lower than the random forest's false positive rate of 3.48%. The improvement arises, in part, because of the way the random forest produces a stable feature representation that reduces AIS noise and missing values before LSTM is applied to those data. Because LSTM works with temporal ordered representations of the evolution of vessel states, as opposed to raw AIS messages, it provides greater resistance to temporal misalignments and irregularities. It directly addresses the weaknesses identified in Problem Statement 1 of insufficient preprocessing, and Problem Statement 2 of temporal loss of features.

As we look at the CICDarknet2020 dataset, comparing the results from the two datasets highlights issues that have emerged because of encryption. While the LSTM does not perform well in this case, since it shows large numbers of misclassified samples across virtually all classes, most of the benign traffic within the Darknet (ex. browsing with Tor) are misclassified as attack traffic and the FN rates of many different malicious flow types are greater than 50%, primarily due to the random nature of the temporal characteristics of encrypted metadata. When summarised on the binary classification level, LSTM only has a 44% recall, indicating it misses 56% of malicious encrypted flows. The LSTM's confusion matrix also provides evidence of this trend through a high degree of dispersion from the diagonal, which is indicative of both overfitting and an uncertainty level under weak temporal conditions.

RF has a greater level of effectiveness when applied to Darknet than when it is applied to the other datasets (i.e., 82% precision and 72% recall). When translated into error rates, RF's false positive rate is about 9% to 10% and its false negative rate is approximately 28%. These errors occur because of the similarities between benign and malicious encrypted traffic in terms of burstiness and their respective statistics. In contrast to the structural patterns that can be found in both benign and malicious traffic (e.g., differing byte counts and duration ranges), stealth C2 attacks mimic the appearance of normal encrypted flows.

The hybrid model provides the best overall balance by demonstrating 80% precision and 78% recall, which corresponds to a false positive rate of approximately 8% and a false negative rate of 22%. Both of these values are improved compared to RF alone. The hybrid confusion matrix indicates stronger diagonal values and less confusion across multiple attack classes. In addition, the hybrid model captures low-frequency repetition patterns (i.e., periodic keep-alives or typical burst intervals) due to the addition of LSTM after the transformation of the original metadata into structured features using RF. Although encryption of traffic obscures the ability to detect semantically based payloads, the hybrid model is still capable of detecting small-timed variations that would be missed by RF alone.

The three datasets include many scientific results which show that the Random Forest (RF) algorithm produces more false positives than LSTM algorithms. RFs are more prone to producing false positives because they rely on static decision boundaries and they are much more sensitive to patterns that occur in temporal data due to the inherent variability of the time component. Conversely, LSTMs produce a greater number of false negatives when applied to datasets that have weak, inconsistent, or noisy temporal patterns, as evidenced by both AIS (Automated Information System) and Darknet datasets. The hybrid model offers a systematic approach to reducing both types of errors. RFs create stable inputs for LSTMs that result in reductions in false positives. Moreover, the LSTM adds temporal context to RF predictions and results in lower false negatives. This combination of RFs and LSTMs

works effectively with the research goals of minimising the noise associated with both RF and LSTM algorithms as independent models, which is accomplished by reducing the temporal degradation problem exhibited by current machine learning models.

To sum up, analysis based on a Gradient-Boosted Decision Tree Classification illustrates that Remote Fencing has the highest sensitivity in structured settings. The LSTM had good predictive power where temporal patterns are strong and clean. Whereas the Hybrid RF and LSTM have performed well regardless of conditions because they combine both types of predictions from each model into one prediction. Because of the Hybrid RF and LSTM's ability to reduce false positives and false negatives and improve the quality of discernment when looking for maritime intrusions, it represents not only a slight improvement over using a single model but rather an improved and more generalizable method of detecting intrusion into maritime operations.

4.5 Hyperparameters Used in This Study

Table 5: Hyperparameters for RF, LSTM, and Hybrid RF–LSTM

Model	*Hyperparameter*	*Value / Setting*	*Description / Rationale*
Random Forest (RF)	Number of Trees (n_estimators)	100	Ensures stable ensemble behaviour with low variance.
	Maximum Depth	None (expand until pure)	Allows RF to fully explore feature splits for tabular intrusion features.
	Minimum Samples Split	2	Default setting to allow fine-grained node splitting.
	Minimum Samples Leaf	1	Enables deeper tree growth for complex anomaly patterns.
	Criterion	Gini Impurity	Preferred for faster, reliable splitting in high-dimensional datasets.
	Max Features	“sqrt”	Standard choice to reduce correlation between trees.
LSTM	Number of LSTM Units	128	Captures temporal dependencies in AIS and flow sequences.
	Number of Layers	1	A single-layer LSTM chosen to reduce overfitting due to noisy AIS data.
	Dropout	0.2	Prevents overfitting by reducing reliance on specific neurons.
	Learning Rate	0.001	Balanced for stable convergence in time-series learning.
	Optimizer	Adam	Adaptive gradient technique suitable for mixed datasets.
	Loss Function	Binary Cross-Entropy	Appropriate for binary intrusion classification.
	Batch Size	64	Ensures efficient training with stable gradient updates.
	Epochs	20	Sufficient for convergence without overfitting.
Hybrid RF–LSTM	RF Feature Output Dimensions	100 features (post-selection)	RF used as feature selector before temporal modelling.
	Sequence Window Size	10 time steps	Converts selected features into temporal sequences for LSTM.
	LSTM Units	128	Same LSTM size to maintain consistency across datasets.
	Fusion Method	Sequential Pipeline (RF → LSTM)	RF enriches input features; LSTM models temporal structure.
	Final Activation	Sigmoid	Generates probabilistic attack/no-attack output.

For the three heterogeneous datasets used in this study, hyperparameter values were chosen to create a balance between complexity and generalization of models through computational efficiency (i.e., Random Forest, LSTM

& RF-LSTM). Using Random Forest as an example, hyperparameter values of 100 estimators, unlimited tree depth, and Gini criterion were employed to give it robust performance against high-dimensional, structured intrusion events collected from both the CICIDS2017 project and encrypted Darknet flows. Increasing the number of trees will reduce the amount of variability between the predictions produced by the ensemble of trees, which provides for greater stability of outputs; unrestricted tree depth enables greater flexibility in modelling intricate decision boundaries that have been identified within normal multi-featured cyberattack patterns. Adapting the "sqrt" feature selection strategy further reduces the correlation among the trees within the ensemble, limiting the potential for overfitting by this Random Forest model through use of independent sets of attributes.

The hyperparameters for the LSTM component of this research were explicitly designed to work with the "noisy" and "irregular" characteristics of the time series data that compose the majority of AIS communications logs. A single-layer LSTM containing 128 units was eventually selected since larger or deeper architectures are likely to overfit the data due to its limited temporal continuity. Additionally, a dropout rate (0.2) was included to provide regularisation against overfitting the noise contained within the data as opposed to learning the true temporal behaviour of the data itself. The Adam Optimiser was chosen for the model due to its established stability on sequential learning tasks and will allow for efficient convergence while handling the large variations found between the three separate sequences (AIS, CICIDS2017 and DARKNET). The model's training and validation will occur with a batch size of 64 and 20 epochs, which provide adequate gradient stability, but at the same time, do not generate any superfluous overhead.

The Hyperparameters of the Hybrid RF-LSTM Model were designed with the goal of retaining the characteristics of each individual model, as well as preserving the benefits that the models provide to each other. random forest serves as a stabilizer for features, producing about 100 filtered/ranked features (feature reduction), which provide some noise reduction before temporal sequencing. The window size (ten time steps) is the most appropriate for providing sufficient temporal context for an LSTM to learn behaviors such as ais drift or periodicity in darknet command and control, and consequently, avoids the use of very long sequences, which can be noisy and computationally inefficient. Using a Sigmoid Activation Function for the output layer is aligned with the intrusion detection objective, and allows for efficient probability thresholding during model evaluation. Overall, the design of the hyperparameters focuses on interpretability, generalizability, and efficiency in terms of data throughput, thus making the hybrid model viable for real-world maritime environments where computational resources and latency will be factors.

5. CONCLUSION AND FUTURE WORK

This study addressed key challenges in maritime cybersecurity intrusion detection, particularly the difficulty of handling heterogeneous maritime data and the limitations of standalone machine learning or deep learning models in capturing both static and temporal behavioural patterns. To overcome these issues, a hybrid framework combining a multi-preprocessing module and a Random Forest–Long Short-Term Memory (RF-LSTM) architecture was proposed. The main contribution of this work lies in integrating robust feature-based learning with temporal sequence modelling to improve intrusion detection performance across diverse maritime datasets. The proposed hybrid approach was evaluated using AIS navigational data, CICIDS2017 intrusion data, and CICDarknet2020 encrypted traffic datasets, representing different cyber risk characteristics. Experimental results showed that the Hybrid RF-LSTM model consistently outperformed standalone RF and LSTM models by reducing false negatives and improving overall detection reliability, particularly in noisy or weakly structured environments. These findings demonstrate that hybridisation between statistical feature learning and temporal modelling provides a practical and scalable solution for maritime intrusion detection. The framework not only improves detection accuracy but also enhances operational reliability by reducing undetected attacks, which are critical risks in maritime systems. Overall, this research establishes a strong foundation for next-generation maritime cyber defence systems capable of operating across multiple data domains.

Future work should focus on enhancing the proposed framework through several complementary directions, including the integration of advanced temporal modelling techniques such as attention mechanisms and Transformer-based architectures, as well as graph-based learning to better represent relational interactions within maritime environments. Exploring semi-supervised and unsupervised learning approaches is also important to address the limited availability of labelled maritime cyberattack data, particularly for rare or emerging threats. Further research should investigate real-time deployment by optimising the model for resource-constrained maritime systems using methods such as edge inference, model pruning, and quantisation. In addition, integrating multimodal data sources such as AIS, radar imagery, GNSS, onboard sensors, and network traffic can improve

situational awareness and reduce false alerts through multimodal fusion. Enhancing model interpretability using explainability techniques such as SHAP analysis and temporal saliency mapping, together with the use of maritime digital twin environments for simulation and synthetic data generation, will further strengthen the practical applicability and reliability of future maritime intrusion detection systems.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for the institutional support provided throughout this work. Appreciation is also extended to StAG, France, for their expert cybersecurity insights and collaboration, and to Universitas Dian Nuswantoro (UDINUS), Indonesia, for their academic support and valuable contributions to this research.

REFERENCES

[1] M. Christiansen, K. Fagerholt, and D. Pisinger, "Fifty years on maritime transportation," *EURO Journal on Transportation and Logistics*, vol. 14, p. 100148, Jan. 2025, doi: 10.1016/J.EJTL.2024.100148.

[2] G. Liu *et al.*, "Survey and Future Trends for Cybersecurity in Maritime and Port Sectors: A Discrete Event Systems Perspective," *Mathematics*, vol. 13, no. 22, p. 3650, Nov. 2025, doi: 10.3390/math13223650.

[3] M. V. Clavijo Mesa, C. E. Patino-Rodriguez, and F. J. Guevara Carazas, "Cybersecurity at Sea: A Literature Review of Cyber-Attack Impacts and Defenses in Maritime Supply Chains," *Information*, vol. 15, no. 11, p. 710, Nov. 2024, doi: 10.3390/info15110710.

[4] E. Blanco-Davis, S. Loughney, and Z. Yang, "Novel Maritime Techniques and Technologies, and Their Safety," *J Mar Sci Eng*, vol. 14, no. 1, p. 30, Dec. 2025, doi: 10.3390/jmse14010030.

[5] M. Li, J. Zhou, S. Chattopadhyay, and M. Goh, "Maritime Cybersecurity: A Comprehensive Review," Nov. 2024, [Online]. Available: http://arxiv.org/abs/2409.11417

[6] L. Wang and H.-H. Hsu, "IoT technology in maritime logistics management: exploration of data analysis methods," *Discover Internet of Things*, vol. 5, no. 1, p. 66, Jun. 2025, doi: 10.1007/s43926-025-00167-9.

[7] T. Miller, I. Durlik, E. Kostecka, S. Sokołowska, P. Kozlovska, and R. Zwolak, "Artificial Intelligence in Maritime Cybersecurity: A Systematic Review of AI-Driven Threat Detection and Risk Mitigation Strategies," *Electronics (Basel)*, vol. 14, no. 9, p. 1844, Apr. 2025, doi: 10.3390/electronics14091844.

[8] I. Durlik, T. Miller, E. Kostecka, and T. Tuński, "Artificial Intelligence in Maritime Transportation: A Comprehensive Review of Safety and Risk Management Applications," *Applied Sciences*, vol. 14, no. 18, p. 8420, Sep. 2024, doi: 10.3390/app14188420.

[9] J. Pijpker, S. McCombie, S. Johnson, R. Loves, and G. M. Makrakis, "An Open-Source Database of Cyberattacks on the Maritime Transportation System," Oct. 25, 2024. doi: 10.20944/preprints202410.1996.v1.

[10] A. Karas, "Maritime Industry Cybersecurity: A Review of Contemporary Threats," *EUROPEAN RESEARCH STUDIES JOURNAL*, vol. XXVI, no. Issue 4, pp. 921–930, Oct. 2023, doi: 10.35808/ersj/3336.

[11] Y. Yang, Y. Liu, G. Li, Z. Zhang, and Y. Liu, "Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review," *Transp Res E Logist Transp Rev*, vol. 183, p. 103426, Mar. 2024, doi: 10.1016/J.TRE.2024.103426.

[12] L. Chen and J. Liu, "Identification of Shipborne VHF Radio Based on Deep Learning with Feature Extraction," *J Mar Sci Eng*, vol. 12, no. 5, p. 810, May 2024, doi: 10.3390/jmse12050810.

[13] R. Anuja and J. Annrose, "End-to-end deep learning for smart maritime threat detection: an AE–CNN–LSTM-based approach," *Sci Rep*, vol. 15, no. 1, p. 36316, 2025, doi: 10.1038/s41598-025-19450-4.

[14] G. Potamos, E. Stavrou, and S. Stavrou, "Enhancing Maritime Cybersecurity through Operational Technology Sensor Data Fusion: A Comprehensive Survey and Analysis," *Sensors*, vol. 24, no. 11, p. 3458, May 2024, doi: 10.3390/s24113458.

[15] Nitish Raj and Prabhat Kumar, "Vessel Trajectory Route Spoofed Points Detection Using AIS Data: A Bi-LSTM Approach," *Def Sci J*, vol. 75, no. 2, pp. 243–249, Mar. 2025, doi: 10.14429/dsj.20464.

[16] X. Lv, R. Jiang, C. Chang, N. Shu, and T. Wu, "A Deep Learning Approach for Identifying Intentional AIS Signal Tampering in Maritime Trajectories," *J Mar Sci Eng*, vol. 13, no. 4, p. 660, Mar. 2025, doi: 10.3390/jmse13040660.

[17] 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, Jun. 2022, doi: 10.1109/TITS.2021.3055614.

[18] P. Bernabé, A. Gotlieb, B. Legeard, D. Marijan, F. O. Sem-Jacobsen, and H. Spieker, "Detecting Intentional AIS Shutdown in Open Sea Maritime Surveillance Using Self-Supervised Deep Learning," Oct. 2023, doi: 10.1109/TITS.2023.3322690.

[19] D. Tella, C. T. Tiriveedhi, N. Rishe, D. E. Tamir, and J. I. Tamir, "Evaluating Neural Networks for Early Maritime Threat Detection," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.20054

[20] C. Maganaris, E. Protopapadakis, and N. Doulamis, "Outlier detection in maritime environments using AIS data and deep recurrent architectures," Jun. 2024, doi: 10.1145/3652037.3663916.