PREDICTION OF RISKY MARITIME ENCOUNTERS IN NARROW AND CONGESTED WATERWAYS VIA CLUSTERING BASED ENSEMBLE MACHINE LEARNING AND SEQUENTIAL DEEP LEARNING

A Thesis

by

Muhammet Furkan Oruç

Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in
Department of Civil Engineering

Özyeğin University
May 2023

Copyright © 2023 by Muhammet Furkan Oruç

PREDICTION OF RISKY MARITIME ENCOUNTERS IN NARROW AND CONGESTED WATERWAYS VIA CLUSTERING BASED ENSEMBLE MACHINE LEARNING AND SEQUENTIAL DEEP LEARNING

Approved by:

Asst. Prof. Dr. Yiğit Can Altan, Advisor
Dept. of Civil Eng.
Özyeğin University

Assoc. Prof. Dr. Bekir O. Bartın
Dept. of Civil Eng.
Özyeğin University

Asst. Prof. Dr. Şirin Özlem
Dept. of Industrial Eng.
MEF University

Date Approved: May 11, 2023

To Elif, for her unwavering and heartfelt belief in my potential.

ABSTRACT

This thesis provides a machine learning framework to predict risky encounters between ships before ships establish visual contact in narrow and congested waterways. There are three parts in this thesis. The first one is clustering for exploration of encounters, the second is ensemble machine learning application to predict risky encounters without distance as a decision factor, and the third one is sequential deep learning based prediction of risky encounters. The Strait of Istanbul (SOI) serves as a case study. Ship–ship interaction database is constructed using historical Automatic Identification System (AIS) messages. Interactions are analyzed via clustering to explore risky encounters using degree of ship domain violation. Findings confirm that ship length and ship speed can serve as reliable indicators to understand the patterns in risky encounters. Results suggest that ship domain violation exists within a nuanced grey zone, requiring cautious consideration instead of rigid categorization. On this basis, clustering based ensemble machine learning framework is developed to predict close encounters and overcome class imbalance. The model successfully predicts each 4 out of 5 risky encounters without the knowledge of distance between two ships. To demonstrate applicability of risky encounter prediction in real time, a practical sequential prediction framework is introduced as the final step. Long Short-Term Memory (LSTM) networks are used to predict risky encounters based on sequential navigation data. The methodology presents an improved encounter model to identify ship-to-ship interactions and classify them as risky, gray-zone, or non-risky encounters based on ship domain violations. The developed approach can be integrated to narrow and congested waterways as a decision support tool to improve maritime safety, and can be a guide to autonomous vessels for safe navigation.

ÖZETÇE

Bu tez, gemiler arasında riskli karşılaşmaları dar ve trafiği yoğun suyollarında görsel temas kurulmadan önce öngörmek için bir yapay öğrenme modeli sunmaktadır. Bu tez üç bölümden oluşmaktadır. İlk bölüm, karşılaşmaların detaylı keşfi için kümelemedir; ikinci bölüm, mesafe faktörünü kullanmaksızın riskli karşılaşmaları tahmin etmek için yapay öğrenme uygulamasıdır; üçüncü bölüm ise riskli karşılaşmaların ardışık derin öğrenmeye dayalı tahminidir. İstanbul Boğazı (İB) bir vaka çalışması olarak kullanılmaktadır. Gemi-gemi karşılaşma veritabanı, geçmiş Otomatik Tanımlama Sistemi (AIS) mesajları kullanılarak oluşturulmuştur. Karşılaşmalar, gemi alan ihlali miktarını kullanarak riskli karşılaşmaları keşfetmek için kümeleme yoluyla analiz edilmiştir. Bulgular, gemi uzunluğu ve gemi hızının riskli karşılaşmaları anlamak için güvenilir göstergeler olarak kullanılabileceğini doğrulamaktadır. Sonuçlar, gemi alan ihlalinin nüanslı gri bir bölgede var olduğunu, katı kategorizasyon yerine dikkatli bir düşünce gerektirdiğini göstermektedir. Bu temelde, kümeleme tabanlı bir yapay öğrenme şeması geliştirilerek yakın karşılaşmaları tahmin etmek ve sınıf dengesizliğini aşmak için kullanılmıştır. Model, iki gemi arasındaki mesafeyi bilmeden riskli karşılaşmaların 5'te 4'ünü başarıyla tahmin etmektedir. Riskli karşılaşma tahmininin gerçek zamanlı uygulanabilirliğini göstermek için son adım olarak pratik bir ardışık tahmin uygulaması tanıtılmıştır. Uzun Kısa Dönemli Hafıza (LSTM) ağları, ardışık navigasyon verilerine dayanarak riskli karşılaşmaları tahmin etmek için kullanılmaktadır. Yöntem, karşılaşmaları riskli, gri bölge veya riskli olmayan karşılaşmalar olarak sınıflandırır. Geliştirilen yaklaşım, deniz güvenliğini iyileştirmek için dar ve yoğun suyollarında bir karar destek aracı olarak entegre edilebilir ve güvenli seyir için otonom gemilere bir rehber olabilir.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Asst. Prof. Yiğit Can Altan, for his unending encouragement and faith in my abilities. His pioneering research on this topic and his persistent personal support made this thesis achievable. I am thankful for his countless hours spent reflecting, reading, and helping me to develop this thesis.

I would like to extend my appreciation to my undergraduate advisor Prof. Emre Otay for introducing me to this topic during my Bachelor's education at Boğaziçi University. He inspired my intellectual curiosity, which led me to devote myself to this thesis.

I would like to express my gratitude to the members of my thesis committee, Assoc. Prof. Bekir Bartın and Asst. Prof. Şirin Özlem, for their support and invaluable advice regarding my thesis and career.

I express my sincere appreciation to Elif for her unwavering encouragement, forbearance, and confidence in my endeavor.

My heartfelt appreciation goes to my mother, father, and my dear brothers, Emir and Ömer, for their endless support in all aspects during this thesis and my career.

I would also like to extend my thanks to MarineTraffic and Stellios Stratidakis for providing research datasets with me. Their support was instrumental in making the final part of this thesis possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

90% of the world trade is being carried by sea, and it's growing at a fast rate [1]. To meet the demand while obtaining optimum efficiency and environmental performance, larger and faster ships are being built each year. This development has led to increased dependence on a relatively limited number of high-capacity vessels, frequently operating at the limits of their performance capabilities. Because of this trend in accordance with the principles of unit economics, risk per trade volume transported is getting higher. With the increase of trade volume and congestion in each vessel, navigation safety is further prioritized. Safety is a prominent concern, especially in narrow and congested waterways such as Singapore Strait, Gulf of Finland, Kattegat in the Baltic Sea, and the Strait of Istanbul (SOI) which constitute a significant risk for ship-ship collisions [2, 3, 4, 5, 6, 7, 8]. Geostrategic and environmental importance of similar waterways, harbor entrances and inland waterways [9, 10] with naturally challenging conditions also leave them vulnerable to safety-related threats especially in potential ship-ship collisions. Ship-ship collision is one of the primary reasons for maritime accidents, with a 16% share of accidents [11]. Thus, it is essential for crews and naval authorities to be aware of potential risks to obtain preventive measures. Also, considering the developments in the maritime engineering, it is critical for the autonomous vessels to decide on the critical encounters before they occur.

Researchers developed concepts from different aspects for improving maritime safety, with a focus on vessel collisions. Stochastic modelling and prediction of maritime accidents have been proposed by Otay et al.[12]. Geometric collision probability analysis [13] and causational collision probability analysis [14] have been conducted. Mazurek

et al. [15] identified collision prone locations for ships via network based IWRAP Mk2 model. Rong et al. [16] proposed a ship collision avoidance model based on ship trajectories. Zhang et al. [17] developed a predictive method to assess inland water complexity. A leading example of inland transportation network is the Strait of Istanbul, which has 933 daily trips spanning to 53 piers and 40 million annual passengers as of 2023 [18]. Another application on maritime traffic has been presented by Cai et al. on Yangtze River with an emphasis on local ferries [19]. Zheng et al. [20] proposed a novel spatio-temporal ship domain to improve risk assessment. Complex nature of maritime accidents [21] have been studied in the scope of globally scalable methodologies' development. Stochasticity of accidents' occurrence and diversity of present local conditions [22, 10] were presented as challenges in the prediction of ship-ship collisions. While, Automatic Identification System (AIS) emerged as a vital source of maritime intelligence [23], where wide range of information can be extracted from vast amount of data.

With the improvements and wide scale adoption of Automatic Identification System (AIS), it has begun to be an increasingly valuable source for maritime accident research [2, 23]. AIS enabled the detection of accidental situations in the historical dataset among non-accidental interactions. Since accidents and ship-ship collisions are infrequent [24], AIS data is significantly prone to imbalance in the scope of maritime accident prediction [25]. To improve benefits obtained from the emerging AIS data and to reach meaningful outcomes about maritime accidents, some researchers focused on predicting non-accidental critical events which potentially lead to ship-ship collisions [26] rather than accidents.

[27] Hanninen and Kujala suggested that there is a certain relation between near misses and accident occurrence, providing that improvement is needed. Non-accidental critical events are defined in different keywords such as potential near miss detection [28] or maritime conflict [29] in the literature. Although still there is an expressed

gap in non-accidental critical events' being a direct sign of maritime accidents [30], the fact that data being more resourceful to extract insights in the scope of maritime conflicts proposes vast opportunities. Initially, implementation of a nearness based understanding for potential maritime conflicts has been conducted by Debnath et al. [31], utilizing the traffic conflict approach. Yoo [32] has presented a near miss density map in South Korea Coast. The work by Zhang et al. [33] demonstrates a deep learning algorithm to classify potential near misses. [34] also conducted machine learning based classification of potential near miss collisions. Su et al. [35] developed a machine learning based approach to predict cruise accidents without navigational data. and Szlapczynski and Szlapczynska [36] studied a novel method to detect near misses via degree of domain violation, relative speed, combination of courses and traffic complexity. These works propose a potential path to machine learning based vessel conflict prediction framework based on AIS data. With the developments of imbalanced data focused research in the machine learning domain [37], AIS data presents an area for imbalanced data based applications to improve safety in maritime traffic research.

Adoption of distance between ships as a risk determination parameter helps to assess the encounter situation with a comprehensive sense, while the caveat is the limitations of distance parameter. Zhang et al. [28] suggests that risk determination methodologies can be helpful to rank cases based on severity and highly ranked cases can be presented for expert judgments to prevent future cases. Considering the varying maneuverability capabilities of ships in short distances and navigational complexity, mitigation of navigational risk in the exact moment of a risky encounter can be challenging. Du et al. [38] have been the first to offer a method for stand-on ship, which demonstrates the issues for late risk determination. To enable preventive navigational action for a risky encounter, maritime authorities need to be warned from time ahead. On this basis, it is suggested that development of risk assessment models independent

of distance between ships offer practically promising risk mitigation planning buffer, both as time and distance.

The main contribution of this research is to develop a novel methodology for predicting non-accidental critical events without distance between two ships as a decision factor. Thus, a forecast can be achieved as risk/non-risky before the beginning of an encounter in a very short amount of time. The outputs can be used as decision support system during the navigation. In this study, this is achieved in three main steps: high dimensional clustering to explore ship domain violation dynamics, integration of clustering to ensemble machine learning prediction pipeline to predict risky encounters and application of deep learning on sequential navigational messages to predict if an encounter will be risk or not. On the basis of combining high dimensional clustering inside the supervised prediction pipeline, a novel proposal to overcome class imbalance is demonstrated.

CHAPTER II

PROBLEM STATEMENT AND APPLICATION AREA

In this chapter, some of the concepts that will be used throughout the work will be introduced, along with the objective of this work.

2.1 Literature Review on Maritime Accident Theory and Applications

According to Qureshi [21], each study in the literature can be classified under three major accident theory subtypes, namely complex linear theory, relational theory and systemic theory based on. In complex linear theory, accidents are explained as a sequence of individual events where cause and effect are present. In relational theory, a pyramid scheme is suggested where changing types of outcome severity are in line with accidents' occurrence frequency. In systemic theory, the emphasis is built on system functions, their relationship and context with environmental conditions. In this theory, each element of the system is modeled to be a part of a feedback loop for dynamic information exchange. While different interpretations can be possible, large majority of historical non-accidental critical events are understood in the scope of relational theory and complex linear theory's combination. Work by [26, 29, 39, 40] belong to complex linear theory and [41, 42, 31, 43, 44, 28, 45] can be exclusively assigned to relational theory. The distinction between studies can be conducted based on the cause and effect relationship, such as determination of risk as a target ship's entrance to the own ship's domain area or one of the ships' velocity is classified in the conflicting ships' risky velocity threshold [26, 40]. In studies which belong in relational accidental theory, the distinctive understanding is through continuous risk parameters, such as variables indicating higher risk with increasing value. For

example, Zhang et al. [45] explains RVCRO algorithm, where increasing value of the corresponding data lead to increasing severity of risk, in accordance with the pyramid scheme. Liu et al. [46] used intervals between ship domains and degree of violation to evaluate collision risk. Wang et al. [47] also developed a real time risk perception model via adopting varying degree of domain violation.

Lu et al. [48] suggest that, in consideration of a maritime conflict between two ships at the exact moment, ship variables, sailing variables and human decision variables are present. Ship length, ship type, ship width, ship speed and ship course are some of the most relevant ship variables, which partly construct The International Regulations for Preventing Collisions at Sea 1972 (COLREGs) [38] as well. Ship position, ship speed and course are the most commonly used variables, followed by ship length in the literature. Distance to closest point of approach (DCPA) and Time to closest point of approach (TCPA) are calculated through position, speed and heading of the ships and used by some researchers [49, 50, 51, 52]. Vessel collision risk operator (VCRO) derived through ships' length, speed, course is also widely adopted by many researchers, [53, 43, 54]. [54] also integrated VCRO's change rate and [43] integrated respective difference between heading of the ships.

In the scope of application method, there are two prominent approaches. The first one is supporting experts to conduct statements about maritime safety and the second one is directly evaluating maritime safety from present conditions. [44, 28] Zhang et al. and Zhang et al. show highest ranked maritime encounters to experts with additional information to determine associated risk with respective encounters. In [52], dangerous encounters are analyzed by experts through taking nearby traffic trajectory into account. While, Watawana et al. [34] proposes a method where near collision assessment can be directly conducted from results. Chen et al. and Yoo [55, 32] provide the count of collision candidates through applied studies. Zhang et al. [45] relies entirely on AIS data to determine risk and determine near-miss

possibility.

Validity of developed methods in determining maritime risk remains open to further evidence contribution [38]. In order to validate a proposed methodology, similarity between measured risk through models and perceived risk by domain experts have been compared by researchers [41, 56, 42]. Wu et al. [10] identified risky hot spots are analyzed with VCRO scores for benchmark purposes. Further, overall validity of assessment of maritime traffic risk through analysis of non-accidental critical events remain an area where investigation is needed, since ideas regarding non-accidental critical events originate from road traffic [38].

2.2 Problem Statements

In this section, problem statements for different applications within the study are presented.

2.2.1 Problem Statement, Clustering

Studies on the analysis of non-accidental critical events [26] have gained attention, and various models are being proposed to identify these risky encounters. While, validity of developed methodologies has not been comprehensively established, researchers suggested that usage of different methods may lead to unreliable results, increasingly [14]. Due to lack of systematic validation, patterns of ships' behaviors during non-accidental critical encounters remain as a research question. Since each method is being developed with a limited number of model variables, specific impact of these variables or other ship properties in the moment of risky encounters are not discovered. Considering the complexity of the relationship between variables with the nature of potential accidents, examining each model variable with respect to others can lead to meaningful outcomes. On this basis, analysis of model variables can help to validate proposed methodologies. As patterns are discovered, models can be improved to adapt the nature of risky encounters. This study proposes a method to

validate risky encounter models through high dimensional clustering based mapping of model variables with respect to risk based segregation with the aid of ship domain. Previous clustering applications showed navigational characteristics and applied risk analysis to model outcomes [57, 42, 53, 36, 58, 45]. Liu et al. [59] presented a conflict detection method using dynamic ship domain. They also used ship domain to detect severity of a conflict. Liu et al. [59] also adopted k-means clustering, though in the spatio-temporal domain to find areas of conflicts as hot-spots. Feng et al. [60] used k-means clustering to quantify collision risk via extracting shipping routes' information entropy. Zhou et al. [61] implemented a ship behavior clustering, which allowed ships' systematic classification based on their characteristics. Wang et al. [62] developed a co-clustering-based methodology for discovering ship trajectory co-occurrence patterns. [63] proposed a ship route design based on AIS trajectory analysis. [64] used clustering for improving AIS device efficiency, where they aimed to eliminate existing outliers resulting from AIS packet collision. [65] estimated collision risk via distance based parameters, such as distance at closest point approach (DCPA) and time to closest point approach (TCPA). [66] presented opportunities and challenges for particularly supervised learning, while outlined issues such as transparency and evaluation were also applicable for unsupervised methodologies presented in this paper. These studies did not use synchronous AIS data during ship-ship interaction for clustering to identify patterns.

The added value of this section is the specific focus being proposed for ship-ship interaction moment. Ships can be defined via their AIS intelligence synchronously in a ship-ship interaction. In this study, the intelligence is being extracted under the hypothesis of proposed model variables, and a large number of interactions are being mapped to a three-dimensional plot. Thus, the aim is to show how ship length and ship speed are distributed and related to each other. These features are also combined with a risk measure in the same plot to identify patterns of these model variables.

The research questions of this study can be summarized as below:

1. With respect to varying degree of risk, how the patterns for risky encounters can be discovered and validated through model variables?

2. Is it possible to detect vessel size and vessel speed as risky encounter parameters in predicting the potential near-miss situations?

3. Can the gray-zones between risk/non-risky encounters be identified?

2.2.2 Problem Statement on Ensemble Machine Learning

This study focuses on prediction of non-accidental critical events in congested waterways via fast, scalable, and interpretable machine learning models. Related works in the scope of near miss or conflict detection conduct classification to assess risk levels [38], implement expert judgment [44] and estimate risk associated with specific examples of interactions. This paper aims to establish a relationship between potential near miss situations and non-distance related variables to predict near miss situations before it takes place. In other words, this work may be understood as rather than being an alternative to before mentioned frameworks, more as a different purposed one which aims to test and validate factors other than distance which result as a maritime conflict, that is later evaluated by distance between two ships [30].

While developing conflict prediction models to assist maritime authorities and captains about a potential near miss situation beforehand, distance has to be excluded since it is the decision criteria in the conflict detection. Although DCPA, TCPA, and Velocity Obstacles (VO) have the capability to predict the conflict detection before it takes places, the main disadvantage of them is the false alarms due to possible changes in the velocity vector which are highly observed in narrow waterways with many maneuver points. To overcome these problems, directly measured variables are needed to be used for the prediction process.

The reasoning for the approach can be supported by relevant works on narrow waterways and their regulations. Narrow waterways are identified with considerably significant risks for transit and local ships [10], especially for ship-ship accidents [4]. Trajectory and speed are predictable variables for ship types and certain time intervals [67]. Although there are some trajectory prediction algorithms, these are not included within the scope of this study, due to their probabilistic results, the outputs increase the uncertainty and decrease the confidence interval of the overall algorithm. Therefore, the main scope of this paper is developed on determining the risky encounters with the distinctive parameters, which depends on observations, with fast machine learning algorithms.

Due to the fact that one of the objectives of this study is predicting a near miss, understanding the impact and relevance of used variables is a prioritized objective. An innovative method's potential applicability as a software to authorities' and maritime navigation systems [68] offer significant challenges. Testing of near miss prediction models has been suggested as a gap in the literature [26]. To contribute testing of the suggested model to real world conditions, model variables are designed to help seafarers to develop further judgments. On this basis, the model is structured in a way that it could be tested without needing a complex computational system or runtime, and with relatively few features. Own ship's length and target ship's speed are used as model variables to enable comprehensive assessment.

This approach serves the needs for the design of a predictive model which would rely on historical data to determine a risky situation before it takes place. The presented work should be understood as a preliminary elimination step where potential conflicts are highlighted to stakeholders before the exact moment of the event which would enable an efficient examination process. The conceptual framework is also in line with the [28] approach while it differs from [31]'s work. Meanwhile, the model is also proposed to be a benchmark for designation of clustering and ensemble-based

classification combined as a predictive approach.

1. Can ship-ship encounters be classified as potential near miss, based on the historical data without distance between ships as a decision factor?

2. How can different models and algorithms be integrated to create a pipeline which would ensure the most accurate, practical, and computationally feasible near miss prediction?

Conceptual Basis for Clustering and Ensemble Machine Learning Approach

The scalable and interpretable nature of a clustering model with few variables would be a basis to pioneer discussions on observational insights based on clustering outcomes. Since one of the objectives of this study is to identify patterns in risky encounters, showing the relevance of used variables has been a prioritized objective. However, challenges are present for an innovative method's potential applicability as a software system to mari-time navigation systems [68]. On this basis, the clustering model is structured so that it can be tested without needing a complex computational system or runtime and with relatively few features. Considering the studies in the literature the main risk parameter is taken as the ship domain violation. Own ship's length and target ship's speed are used as the clustering model variables. In this section, the risk parameter and clustering variables are explained as subsections.

Ship Domain Violation

Fujii and Tanaka [69] suggested that ship domain is an area from which target ships must avoid. On the other than, recent studies show that violation of ship domain may not lead to a risky situation, considering a safe passage in a close distance with complete control [26]. Rawson and Brito [30] demonstrated a critique towards ship

domain-based assessment and risk, and noted potential caveats. While researchers [70, 10] showed that narrow waterways have complex conditions resulting in locally designated safe encounter conditions. A potential result is site-specific conditions based ship domains [71]. Altan and Meijers [72] proposed a ship domain for the SOI in this perspective. The study [72] showed the areas around central ships that surrounding ships tended to avoid in the SOI and mentioned ship domain model is used in this research. The applied ship domain for the proposed model resembles the natural conditions. It signifies those ships violating these naturally identified areas are subject to a violation that is not naturally observed in ship-ship interactions of the SOI. These can be interpreted as potential maritime conflicts with respect to site specific conditions.

Model Variables

Congested waterways bring both transit and local ships together in close proximity. The most distinctive aspect of these types of marine vessels can be mentioned as their length, considering large sizes of cargo ships, bulk carriers and tankers. While, local ships have smaller sizes, which are mainly used as a means for public transportation. Since transit ships are subject to strict speed limit constraints, local ships' speed are found to be more diverse. On these bases, this article aims to highlight how speed of a target ship is determinative in understanding an encounters' risk aspect, considering the size of the own ship as a parameter for target ship's crew. Consequently, target ship's speed and own ship's length are determined as model variables.

Degré and Lefévre [73] has been the first to propose the usage of velocity in the maritime field in the context of determining the collision risk. In addition, Lenart [74] also contributed to the integration of velocity to determine collision danger via Collision Threat Parameter Area (CTPA), based on the idea that if the velocity of the subject ship falls into CTPA. Velocity has been integrated as a basis for Linear

Velocity Obstacle (Linear-VO) models to detect if the subject ship's velocity falls into the dedicated zone of velocity obstacle. This is generated by the model by Chen et al. and Kuwata et al. [55, 75] where VO is integrated with COLREGs (Convention of International Regulations for Preventing Collisions at Sea) to develop a motion planning algorithm for Unmanned Surface Vessels (UAV). Chen et al. [55] suggested a Time Discrete Non-linear Velocity Obstacle (TD-NLVO) method to detect collision candidates. While the Velocity Obstacle (VO) approach is based on a Spatio-temporal relationship among objects, the basis for the proposed model is the utilization of the velocity of the target ship in the closest distance to the own ship. This enables to validate if speed can be a risky encounter parameter in predicting the potential near-miss situation or not.

Zhang et al. [28] has applied ship length as a decision parameter, both directly and indirectly via ship domain approach, and reached the outcome that vessel size is a significant indicator for a possible near-miss situation's detection via classifying ships' size into three clusters. The local and transit maritime traffic characteristics of narrow waterways are distinct. Also, there is a significant difference in local ships' sizes and transit ships' sizes [76]. Due to this, it has been relevant to include ship size as a decision factor for other ships in interaction. Furthermore, this study also describes the understanding of the own ship's length by surrounding vessels. One hypothesis can be outlined as local ships' captains being "more cautious" in certain conditions where transit ships' observable characteristics resemble intimidation. This is also outlined via varying ship domains based on speed, length, vessel type, Course Over Ground (COG), and approaching angles [72]. Including ship length as a feature contributes to detecting the outcome of this hypothesis as well.

2.2.3 Problem Statement on Sequential Deep Learning

Lately, researchers invested more and more to explore capabilities of AIS data to improve maritime safety. [77] used ensemble learning to predict risky encounters via entirely relying on non-distance related variables, indicating the necessity of unbiased approach for practically possible risky encounter predictions. Rong et al. [78] conducted probabilistic prediction of maritime traffic through ship motion pattern. Chen et al. [79] applied PSO-LSTM models to predict collision risk index of encounters and extensively benefited from AIS data. Duan et al. [80] applied a semi-supervised deep learning approach to classify vessel trajectories. [81] analyzed Vessel Collision Risk Assessment (VCRA) via multi-layer perceptrons. Yang et al. [82] used deep learning to predict ship trajectories, Yang et al. [82] also integrated density based clustering. Liu et al. [46] provided a framework to understand ship collision risk with an improved risk model based on Vessel Conflict Risk Operator (VCRO).

In this section, a practical and feasibly applicable framework for predicting risky close encounters is presented in the maritime context. Our methodology involves the development of an extensive encounter model that leverages spatial algorithms to identify historical and significant ship-to-ship interactions in a scalable manner. Utilizing ship domain [36] violations as the foundation for risk labeling, our approach focuses on data transmissions preceding the occurrence of an encounter. For each identified encounter, we record sequential AIS messages leading up to the event, using the information to ascertain whether the encounter is likely to be close and hazardous. Given the sequential nature of navigational data, we employ Recurrent Neural Networks (RNNs) [83], specifically Long Short-Term Memory (LSTM) [84] networks, for this task. The paper elucidates the model architecture and outlines the hyperparameter optimization strategies employed. We present our findings using multi-class classification metrics, such as accuracy, precision, recall, and confusion matrix. The results

indicate that the proposed framework, which relies on historical sequential AIS messages, holds considerable promise for predicting risky maritime encounters, thereby contributing to enhanced navigational safety and improved risk mitigation strategies.

2.2.4 Application Area and Source of Data

In this study, developed approaches have been implemented in long term maritime datasets from the Strait of Istanbul. The Strait of Istanbul is one of the busiest and complex waterways around the world, in a geostrategic location. While Çanakkale Strait [85] also serves the same purpose, the nature of the SOI with its narrowness, sharp turns, complex currents and busy local ferry lines bring further challenges to crews and maritime authorities. Navigational complexity is driven by the combination of local/transit traffic, continuously varying current regime, strict maneuvers and narrow sections of the waterway. These are amplified with the crossing scheme which meets two traffic routes in the busiest section of the straight [76]. With maritime traffic exceeding 300,000 vessels [76], complexity affects ship-ship interactions significantly. Further, considering that more than 16 million people live in Istanbul, maritime traffic density is increasing each year as the local maritime transportation network widens to carry increasing weight of land traffic. Considering fragile marine life near the metropolitan city, along with thousands of years aged historical peninsulas and vibrant life near sea, potential consequences of maritime accidents are serious for life, environment and economy. On this basis, the SOI is an essential benchmark for maritime safety technologies' improvement for research and development purposes.

2.2.5 Source of Data for Clustering and Ensemble Machine Learning Approaches

A long-term AIS dataset is utilized in the scope of this part of the study, which is captured from the Strait of Istanbul (SOI). This part presents a general view on the application area and data duration.

To get the full picture of the waterway traffic, 1 year of AIS data has been analyzed, covering a time range between September 2014 and August 2015. AIS data messages are stored and managed through Structured Query Language (SQL), with a size of 94 Gigabytes (GB). Detailed information about the data collection, initial management and organization process can be found in Altan and Otay [76].

In Figure 1, yearly descriptive statistics about types of ships passing through the SOI are presented. In Figure 2, distribution of length intervals for ships are presented. The largest percentage of the ships are comprised of cargo ships. When the ship length intervals are examined, largest number of ships are present at 100-150 m interval.

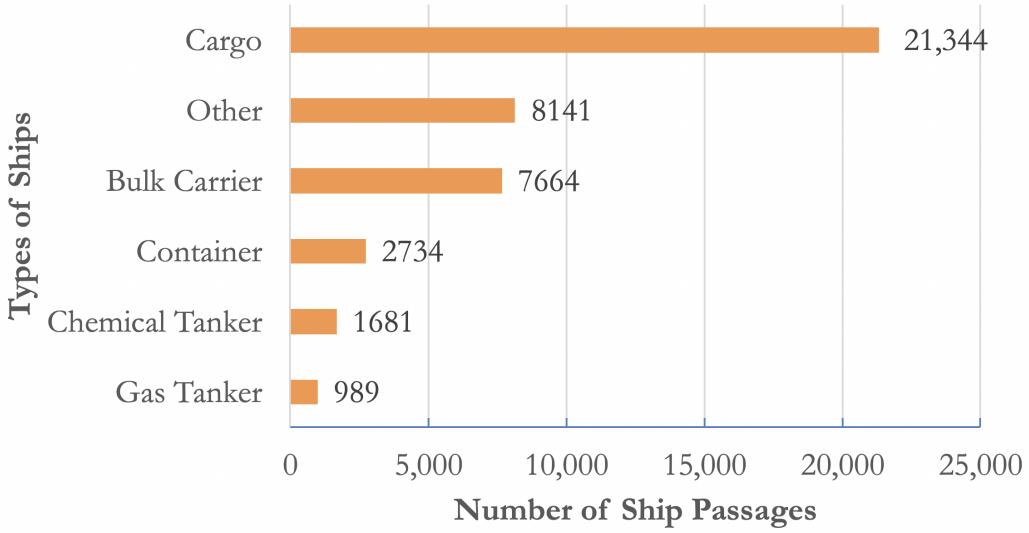

Figure 1: Number of Ship Passages, Types

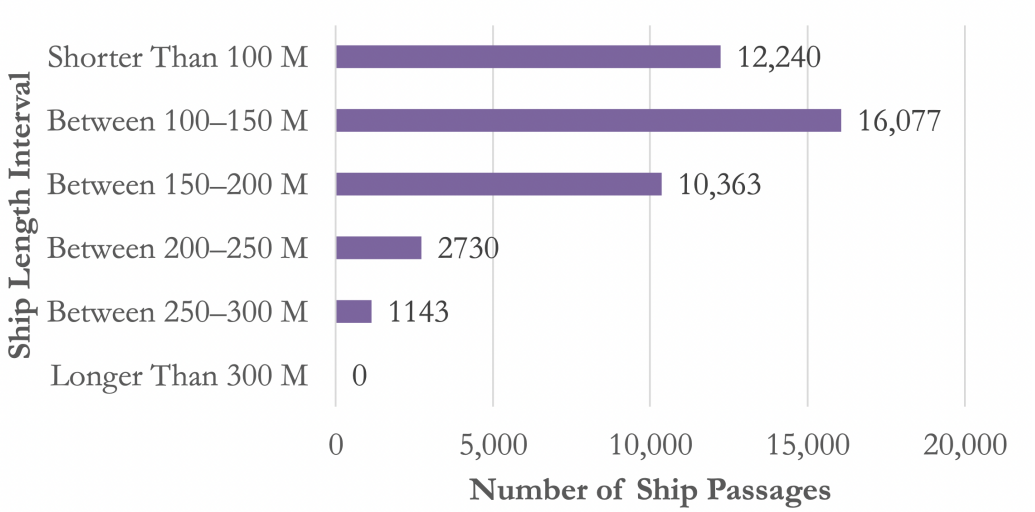

Figure 2: Number of Ship Passages, Lengths

2.2.6 Source of Data for Sequential Deep Learning Approach

In this part of the study, the AIS dataset covered the timeframe between June 2019 to August 2019 in the Istanbul region. Dataset has been provided by MarineTraffic. Due to cost effectiveness requirements, provided AIS messages were downgraded to approximately 2 minutes between each message. To overcome this issue, interpolation has been conducted during historical ship encounter detection processes.

CHAPTER III

METHODOLOGY

3.1 Methodological Approach for Clustering and Ensemble Machine Learning Study

In this section, the framework is summarized, and the methodological approach is outlined. The approach includes curation of the underlying dataset, development of the computational encounter modelling for possible near miss candidate detection in two ships' scenarios, underlying different machine learning algorithms and approaches to successfully predict ship domain violation from the historical dataset, background for combining different algorithms and the conceptual understanding for integrating imbalanced data focused algorithms, which are described as ensemble learning. Also, the underlying basis for the included ship domain approach and variables are described.

3.1.1 Framework of Clustering

Throughout the analysis, different methodologies are combined constructively. In the initial step, an algorithm is developed to transform the raw AIS dataset to an encounter model, where each ship-ship interaction is extracted with respective parameters. Next, based on an extensive literature review and experimental calculations, variables are selected and designed to be introduced to the clustering model. Determination of the used clustering algorithm is also conducted. Based on variables chosen and predetermined algorithms, an optimal number of clusters are calculated via optimality tests, and candidate cluster counts are prepared. In the final step, a three-dimensional clustering application is performed via two distinct ship domain approaches. A high-level diagrammatic representation of the process is provided in

Figure 3.

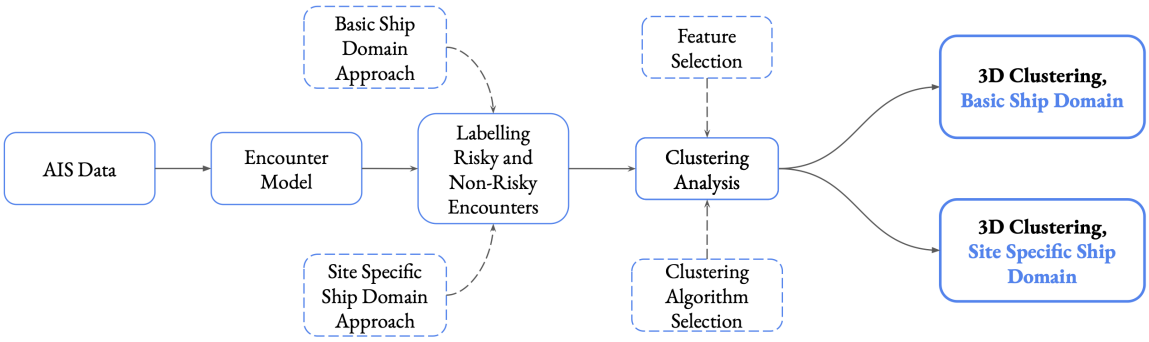

Figure 3: Methodological framework of clustering.

3.1.2 Framework of Ensemble Machine Learning

As the first step, an algorithm is developed which transforms the AIS dataset present in the SQL database to represent ship-ship interactions and information regarding each encounter. Then, 3-dimensional (3D) clustering is applied to the dataset to explore patterns between the following three variables: Own Ship's Length, Target Ship's Speed and Violation Distance per Own Ship's Length. Based on the clustering outcomes, violating and non-violation clusters are determined using ship domain approach. The non-violating clusters are removed from the dataset to improve the balance among classes, which to serve classification algorithms' performance. The violating classes are determined according to class centers. In other words, if a class's center is inside the ship domain, all the member of the class is accepted as violating the ship domain. The procedure helps to reflect the decision of the captains about the ship domain which border is a grey zone during the navigation. In the next step, various classification algorithms are applied in an iterative fashion to predict the violating classes. Results have been validated via 10-Fold Cross Validation for reliability concerns. The overall schematic representation of the framework is given in Figure 4. The developed model is tested with a part of the dataset which has not been used

during the training process. Figure 2 demonstrates the testing procedure. The framework steps are applied for an 80%-20% split version of the dataset for the purpose of conducting testing to 20% of the dataset which is completely left out of any sorts of clustering-based modification procedure, while 80% of the samples are used for training purposes. This designation enabled pure testing by preventing any sort of distance between two ships related intrusion to the test set. While it has been suggested that there aren't certain rules in the determination of this split, this approach has been chosen to obtain a sufficient amount of test and training samples [86].

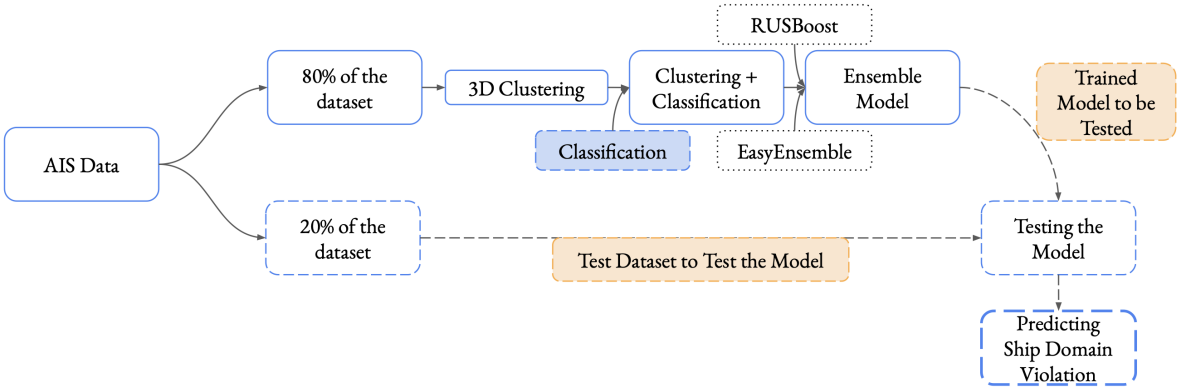

Figure 4: Methodological framework of ensemble machine learning.

3.1.3 Encounter Model

To detect potential near-miss situations, an encounter model is developed based on AIS data. Dataset as the result of the encounter model has been the basis for both clustering and classification algorithms. While detecting potential conflicts, the approach by [87] has been adopted. A circular parameter has been selected as the ship domain, which has been in line with the approach by [88, 70]. Two different ship domain sizes have been used considering the Length overall (LOA) of the own ship and the sizes. Based on the ship domain approach, the safety criteria can be explained as "target ship should not violate the own ship's domain" [89].

In Figure 5, a representation of the encounter searching algorithm is presented as a

sketch in a non-scaled way. Developed algorithm is used to capture the encounter be-tween two ships. Around each own ship, a 1 km meters radius circle is searched. This area is represented with the 1st circle in Figure 5. Ships and their distance inside this radius are recorded. Among interactions, the closest interaction between all ships is extracted. Re-view and elimination of duplicate ship-ship interactions are applied to keep a single interaction between each ship pairs. During the application of the algorithm, the non-regular timely submission of data points is overcome via a linear interpolation process. In this way, each interaction situation is analyzed in a synchronized way.

After determining the ship interactions, a smaller circular area is searched around the own vessels. In this step, radius of the circle is determined through ship domain of the own ship and it is demonstrated as circle 2 in the Figure 5. For each encounter, rate of ship domain violation or non-violation is recorded. Through this approach, relevant en-counters are recorded and dataset is prepared for clustering analysis.

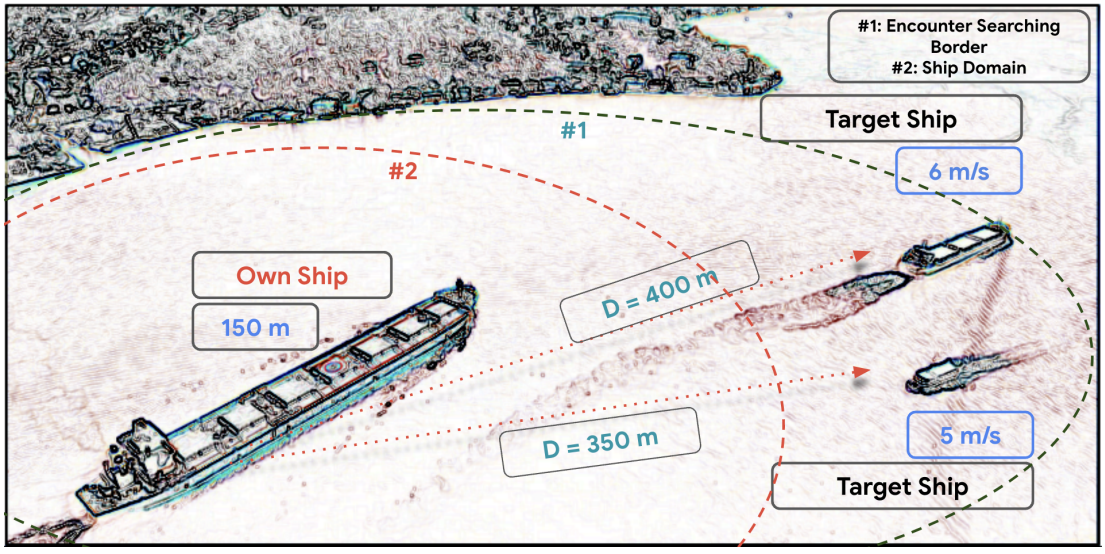

Figure 5: Encounter searching algorithm representation.

Figure 6 shows an example of a conflict, where the target ship is approaching to intrude own ship's domain in the following time interval. Here x_i, y_i and x_j, y_j

indicate own ship's location and target ship's location, respectively. l_i represents the LOA of own ship, v_j represents the velocity of the target ship. D_{ij} is the distance between two ships. r is the ship domain radius and d is the distance from target ship to domain boundary.

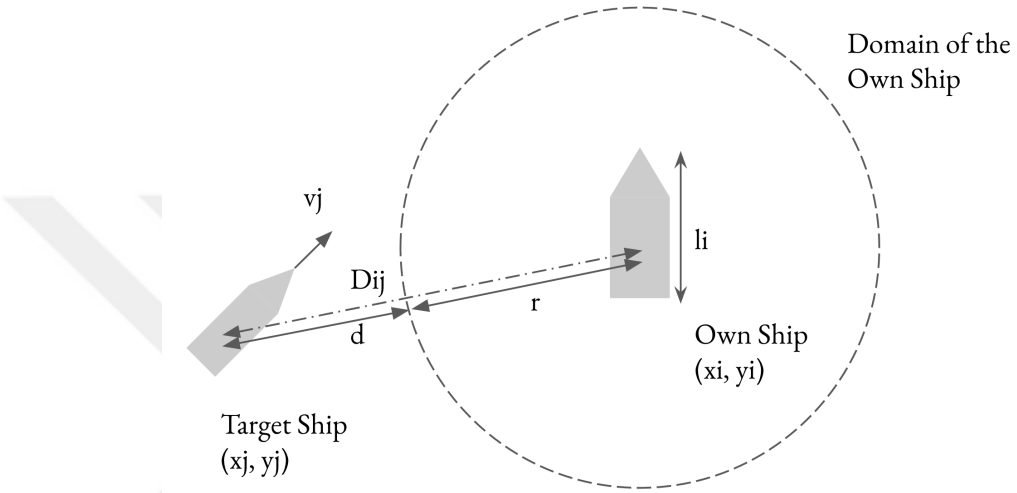

Figure 6: Representation of the encounter and applied safety criteria for two ships.

Distance between the two vessels is calculated based on the Equation 1:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

In Equation (1), (x_i, y_i) denote the position of a specific vessel, namely own ship at a given moment and (x_j, y_j) denote the exact position of the target ship in the same moment.

Table 1 represents different ship domain values. Mou et al. [88] suggested that the majority of domain radii are approximately three times the ship's length. Considering the site-specific conditions and the fact that the SOI gets as narrow as 700 m along the passage, two times the ship's length is represented as the basic ship domain approach. So, the domain radius of a ship for the initial approach can be stated as in Equation (2):

$$r_i = 2 \cdot l_i \tag{2}$$

Altan and Meijers [72] demonstrated two-length groups for the SOI, which are determined as small LOA $\leq$ 157 m and large as LOA $>$ 157 m. According to Altan and Meijers, [72], the naturally occurring ship domain in the SOI for own ship is dependent on ship length. It's suggested that a small length group is observed with comparatively lower ship domain size while a large length group is identified with a greater ship domain size, which changes approximately two to three times, depending on the direction (bow or stern part) of the ship. In order to achieve a representation of site-specific conditions during the ship domain violation, for two different groups, ship domains are defined as in Equation (3) and (4):

$$r_i = 1.75 \cdot l_i \tag{3}$$

$$r_i = 3 \cdot l_i \tag{4}$$

Based on Equations (2), (3), (4), ship domains for different lengths are also provided in Table 1.

Table 1: Ship domain radii for different ship lengths

Ship Length (m)	Basic Ship Domain Radius (m)	Site Specific Ship Domain Radius (m)
Length ($l_i \leq 157$ m)	$2l_i$	$1.75l_i$
Length ($l_i > 157$ m)	$2l_i$	$3l_i$

Implementation of encounter model resulted as a dataset with variables of own ship's length, compared ship's speed, distance between ships, violation distance per own ship's length and violation indicator. Violation distance per own ship's length is calculated via Equation (5):

$$V.D.P.O.S.L. = \frac{(l_i \cdot r) - D_{ij}}{l_i} \tag{5}$$

In equation (5), V.D.P.O.S.L. is the violation distance per own ship's length, where l_i is the own ship's length, r is the ship domain radius for the respective approach, and D_{ij} is the distance between two ships. In addition, the violation indicator is calculated as in Equation (6):

$$f(x) = \begin{cases} 0, & D_{ij} - (x \cdot r) > 0 \\ 1, & D_{ij} - (x \cdot r) \leq 0 \end{cases} \tag{6}$$

Where x is the own ship's length, D_{ij} is the distance between two ships, and r is the respective ship domain radius. In Equation (6), 0 indicates a non-risky interaction and 1 indicates a risky interaction, based on the indicator. Table 2 represents a sample data point which demonstrates a ship-ship interaction.

Table 2: Sample data point demonstrating a ship-ship interaction

Variable	**Value**
Own Ship's Length (m)	157
Target Ship's Speed (m/s)	4.5
Basic Ship Domain (m)	$157 \cdot 2$
Site-specific Ship Domain (m)	$157 \cdot 1.75$
Distance (m)	350
Violation Distance per Own Ship's Length (B.S.D.)	$((157 \cdot 2) - 350)/157$
Violation Distance per Own Ship's Length (S.S.S.D.)	$((157 \cdot 1.75) - 350)/157$
Violation Indicator (0 or 1) (Basic Ship Domain)	0
Violation Indicator (0 or 1) (Site-specific Ship Domain)	1

3.1.4 Clustering Model

Unsupervised learning is generally positioned as the exploratory procedure during an analysis. Also, it's identified to be challenging in the sense that assessing the results of an unsupervised learning implementation can be subjective due to the lack of a clear indicator of success [90]. Clustering can be identified as one of the most fundamental unsupervised learning algorithms, convenient where datasets are not naturally

labeled and patterns are hidden among the hidden layers. It's also defined as the process of finding out what happens naturally in a given dataset [91]. For the case of vast datasets, if labeling is a challenging process, clustering becomes significantly efficient to discover insights [92]. Considering the randomness of the encounters, the K-means algorithm is utilized with its computational feasibility, high dimensional convenience, and robustness [93].

K-means algorithm conducts a center-based grouping of provided data points, and each center would represent their group. It's also referred to as "vector quantization," where the main objective of the algorithm is to apply division to numerous data points for k distinct groups in a way that within-group distances are minimized, and distance between members of different groups are maximized [93]. The algorithm works to iterate itself until the centers of the groups would not be subject to further change with each iteration. Also, a specific limitation of iterations can be included. Previously, for near miss modeling, K-means has been used to model spatial images' clustering to reveal patterns [28]. K-means can be identified as applicable, considering their applicability in visualization and suitability for various shapes to occur due to clustering in the high dimensional space. Since three-dimensional observation is crucial for this research, K-means is suitable to serve.

Since K-means require predetermination of the number of clusters, an initial step for deciding on the optimal number of clusters is needed. Quantitative methodologies to relate the within-cluster sum of distances and the inter-cluster sum of lengths are helpful to conduct judgments. Silhouette Coefficient and Elbow Method are utilized in the scope of this paper. Silhouette can be defined as the measure of within closeness and between apartness of each cluster. This can be expressed as a natural separation assessment for each cluster, in other words, it assesses if the resulting clustering scheme reflects the natural conditions. To calculate Silhouette, the average dissimilarity of a point to its cluster and average dissimilarity to other clusters would be

required metrics. It's calculated for each data point, and each score from a dataset can be averaged to assign a Silhouette score for a given dataset in a given number of clusters status. Secondly, elbow methodology measures the sum of squares of each data point to its cluster center [94]. A higher number of clusters may be potentially more explanatory to identify detailed patterns in a dataset.

However, a lower number of clusters could be more insightful and practical to observe distinct separations. Via the elbow method, the optimal point for these two assessments is compared in a two-dimensional plot, and the number of clusters can be decided. While clustering is an analysis that is highly dependent on the approach and can accommodate subjectivity, visual observation is also helpful to assess optimality in some clusters. For the cases where quantitative methodologies don't propose an obvious outcome, visual exploration and comparison via domain knowledge is a required step, as it's also applied in this paper [95]. In this perspective, the optimum candidate number of clusters is generated and presented in this study rather than selecting a single number of clusters. This can also be identified as a remark on the subjective nature of clustering analysis, especially in higher dimensions [92]. High dimensional clustering is known for its interpretable challenges and lack of visual observational suitability [96]. The third dimension in clustering helps for an additional observational layer for analysis. At the same time, it doesn't suffer from a lack of interpretability, instead, it increases the reveal ability of results, so it is advantageous from multiple perspectives.

3.1.5 Ensemble Prediction Model

Within the scope of this study, the primary purpose is to predict risky encounters via a minimum number of parameters. Due to explainability being a key issue for maritime authorities and industry adoption in safety applications Mazurek2020, achieving predictive performance with limited count of variables have been prioritized. Since

model validation is defined to be a concerning issue by researchers focusing on near misses [26], less contributive parameters have been pruned through training process.

In the candidate variable selection step, historical studies have been analyzed to find the most impactful factors affecting near misses in terms of ships' individual variables, as well as geographical context. [26] introduces that 37 out of 39 near miss models use velocity as a factor and 24 out of 39 models use length as a factor. Further, spatial context in the case of interpreting near misses have been highlighted by [15, 32] to show certain locations can be collision prone or highly dense in terms of near misses. Due to complex nature of the location of case study, where transit traffic meets local traffic while local traffic flows through east-west axis, overall traffic density, average speeds in different directions and average headings in different directions were also included to represent different conditions in 13 sections of the SOI [76]. After variable importance tests, namely LOA of the own ship and Speed over Ground (SOG) of the target ship were selected to be in the final prediction model due to contributions. While LOA of the own ship provides an idea of the ship's size, the SOG of the target ship can provide an idea about the maneuverability of the giveaway ship. Since relied safety criteria is based on target ship's intrusion to own ship's domain area, where speed of the target ship and area around the own ship are primarily relevant, this approach tests a fundamental safety approach within the machine learning framework. The prediction model is developed on the dataset built via the encounter model.

Figure 7 represents the prediction model with its integrated elements. For each approach other than M1, the combination of clustering and classification algorithms is demonstrated. C1, C2, and C3 represent case-by-case implementations of k-means clustering. While C1 represents the approach with the basic ship domain, C2 and C3 are founded on the basis of the site-specific ship domain approach. C3 is highlighted since it is applied only for the testing procedures. M1, M2, and M3 models represent classification models. In each model, different algorithms are constructed

to benchmark with each other. M1 includes K-Nearest Neighbors (K.N.N.), Logistic Regression (L.R.), and Decision Tree (D.T.). M2 and M3 include ensemble meta-algorithms, EasyEnsemble, and RUSBoost. These meta-algorithms also adopt base classification algorithms. In the scope of this methodology, Random Forest and Logistic Regression are implemented as base algorithms, which are demonstrated as benchmarks.

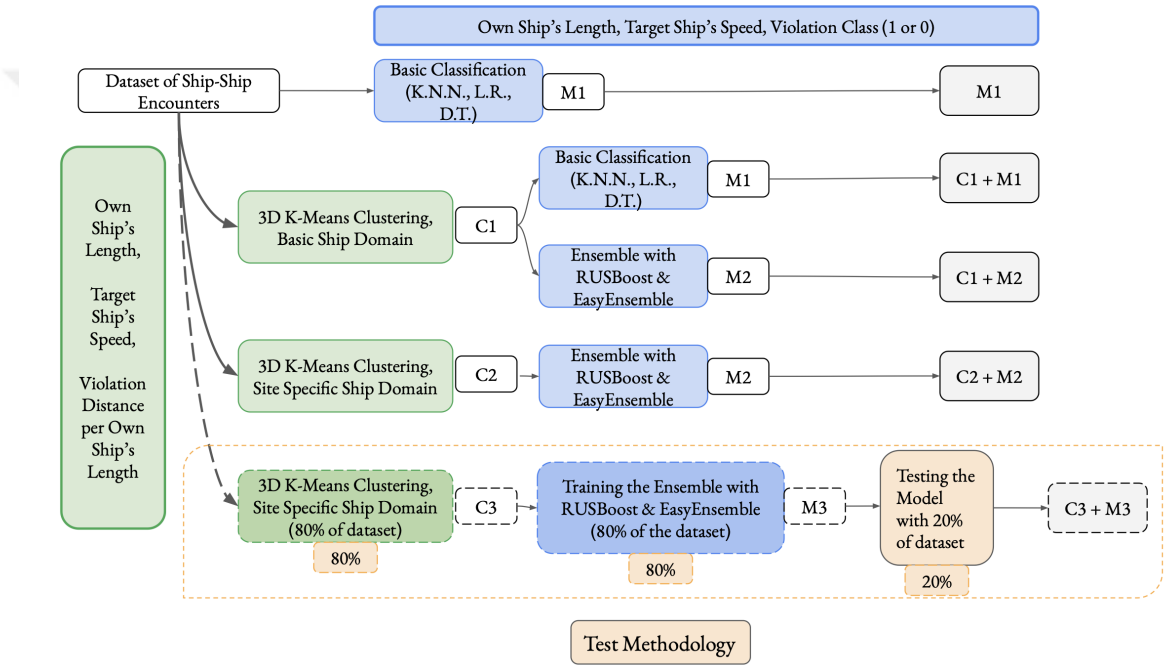

Figure 7: Methodological Structure and Benchmark Models.

Figure 8 outlines the testing procedure in a detailed way. The figure represents the separate integration of training and test sets to relevant algorithms. While the training set has been used for clustering and classification model training purposes, the test set has been only applied via a pretrained model. Since clustering includes the distance between two ships as a parameter, the test set has been separated from clustering-based bias to predict risky encounters. The use of pretrained models is very widespread, especially in the deep learning space [97].

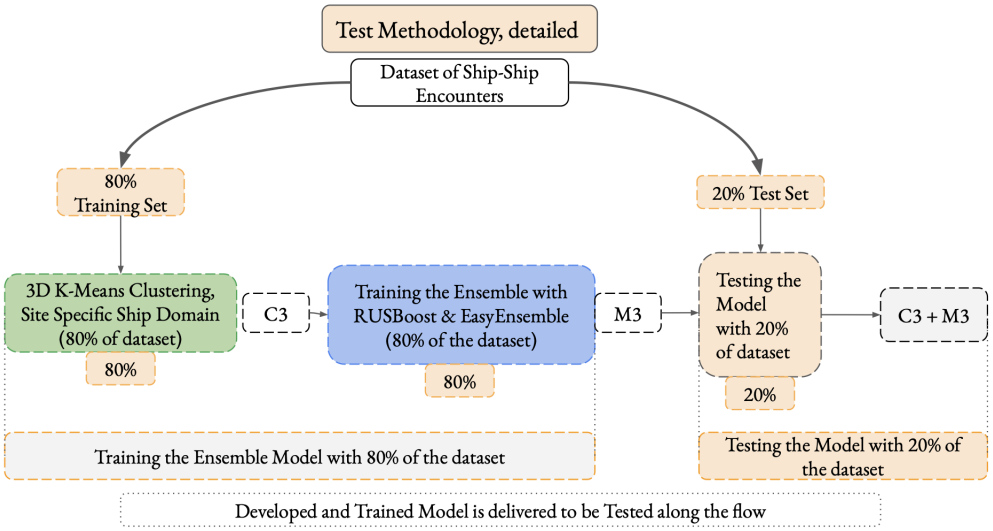

Figure 8: Detailed Representation of the Test Methodology.

The exploratory procedure of the proposed model is provided with clustering analysis. Clustering is one of the most fundamental unsupervised learning algorithms, which is also briefly defined as the approach to find out "what naturally happens" [91] in a dataset. Due to the huge cost of labelling [92] an unlabeled dataset, clustering becomes significantly useful with its leverage to extract hidden meanings, especially when data points are scaled to a significant quantity. In the scope of the provided model, distinct clusters are selected via k-means algorithm and number of clusters are evaluated via elbow method and silhouette score. The clusters are used during the prediction of risky encounters.

The produced clusters are defined as risky/non-risky encounter using the ship domain approach. The produced encounter database is label depending on the cluster label. In the case of a subject historical dataset that includes certain variables combined with a decision or action indication, machine learning models can be trained on this past dataset and can conduct predictions for events that are not concluded yet. Supervised learning can be described as, predicting or estimating an output, either a value or the class of a given data point, based on provided information [86]. In the scope of

this paper, classification of the risky encounter via supervised learning is conducted, which is represented as binary classes, either 0 or 1. K-Nearest Neighbor, Logistic Regression, Decision Tree, and Random Forest algorithms are used and presented as benchmarks. To improve results that are impacted by imbalance, ensemble algorithms are applied. EasyEnsemble and RUSBoost are applied with decision tree and logistic regression as base algorithms.

Imbalance Focused Approach

Imbalanced data is defined as the data which contains some classes many more than certain others. While rare incidents such as maritime collisions [25] or detection of oil spills from radar observations [98] may provide a basis for class imbalance, it is also common in areas such as fraud detection [99] and manufacturing plants [100]. Datasets containing imbalanced distributions challenge traditional machine learning algorithms in various ways, which result in inefficacy to detect minority class members from a given dataset. In concepts like anomaly detection, where some occurrences are dramatically rare, classification rules are found to be prone to inaccurately classify the dominated class, which is the result of the lack of suitability of traditional learning methodologies to class imbalance. Accurate classification of the rare class carries significant importance in many cases, such as finding the distinct parameters which result as a potential maritime conflict [4], which inspired researchers to build algorithms that are more suitable to detect minority class members. In cases where one or more of the classes largely outnumber other class or classes, different sorts of data based, algorithm based and hybrid methodologies comparatively more successfully predict the valuable classes than regular algorithms. Data based methodologies [37] include manipulation of the dataset via arranging the comparative representation of classes, which is based on either increasing the share of minority class via synthesizing artificial members or downsizing the majority class via various techniques. In the scope of this paper, both data based techniques are developed to overcome challenges

provided by imbalance. While 3D clustering served the basis for modification of majority class, RUSBoost [101] and EasyEnsemble [102] also implement undersampling, though, on a random basis. Moreover, algorithm base approaches can be explained as modification of existing algorithms or creation of new ones with designed hyper-parameters, and hybrid methodology includes combination of former two approaches, which are also named as Ensemble.

Application of imbalanced data focused approaches in the maritime traffic research domain to predict potential near miss encounters has not been detected in the review of literature, so this paper serves as a basis for development of methodologies in this sense. What makes imbalanced data-based models to be applicable in this sense can be explained as the result of encounter modelling, which results with a rich dataset of interactions, being labeled in a clear and interpretable manner.

Clustering Based Ensemble

Clustering has been integrated in ensemble approaches via numerous ways in studies. [103] trained different classifiers on subsets of different clusters within a dataset, which showed to be more successful than bagging and boosting. Krawczyk et al. defined this approach as partitioning the space of features to create classifier, which has been similarly approached by Zhou and Zhang [104, 105, 106]. In this paper, clustering has been positioned as a mean to manually downsize the majority class, which is also the basis approach for ensemble meta algorithms. Through 3D clustering, the cluster which is almost entirely belonging to the majority class is observed and it has been manually removed from the training set. Via this removal, the balance between classes have been improved, which enabled the model performance to improve. Clustering has been only implemented through the training process.

EasyEnsemble and RUSBoost

EasyEnsemble and RUSBoost are two meta-algorithms which are intended to improve

prediction performance in the scope of this paper. These meta algorithms [107] are specifically developed to overcome challenges presented by imbalanced datasets and combined with Decision Tree and Logistic Regression as base algorithms. EasyEnsemble [102] can be defined as ensemble of ensembles, which is in fact based on an under sampling methodology. What makes EasyEnsemble relevant in the scope of this paper is its adaptation of an ensemble methodology within its model, which is AdaBoost, short for Adaptive Boosting. AdaBoost is run within each randomly under sampled subsample divided by EasyEnsemble. EasyEnsemble is unique in the sense that it constitutes bagging, boosting and sampling at the same time [108]. RUSBoost also combines boosting and bagging with a random undersampling implementation [101]. With the removal of a certain amount of data points in the majority class, the algorithm provides a competitive speed and performance. It also significantly decreases the time for the model to run, which is also a key benefit in the case of near miss prediction.

Testing the Model

Confusion matrices are one of the most useful methods to measure the outcome of a classification task [90]. A representative of a confusion matrix is provided in table 3. It's also informative in the case of imbalance since it enables an observation for each benchmark's prediction capability of each class. Separate observation of each class's prediction helps accurate assessment of each algorithm. The outcomes are evaluated with confusion matrix, accuracy, precision, recall and ROC-AUC scores.

Here is the confusion matrix in Table 3:

Table 3: Confusion Matrix

	Prediction	
Total Population	**FALSE**	**TRUE**
FALSE	True Negative	False Positive
TRUE	False Negative	True Positive

Accuracy

Accuracy can be defined as the most widespread used metric, as presented in Equation 7. On the other hand, it's strictly prone to misunderstanding while dealing with imbalanced datasets, since a single number does not entirely represent each of the classes' predictability [90]. Due to this fact, scores such as precision, recall and ROC-AUC are prominent measures to better understand outcomes when the data is imbalanced.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Population}} \tag{7}$$

Precision

Precision identifies how selective an algorithm is, while detecting the minority class as defined in Equation 8. In other words, if an algorithm classifies many majority class members as minority class, for the sake of increasing its minority class classification rate, this bias can be detected via precision score.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{8}$$

Recall

Recall indicates how accurately minority class is being detected by an algorithm, as presented in Equation 9. A significantly low recall score may indicate that algorithm is failing to detect and classify minority members.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{9}$$

ROC-AUC Score

An ROC curve can be defined as a two-dimensional visualization of a classifier's performance based on different classifier thresholds. It's built via outcomes of confusion

matrix and one of the most fundamental metrics in imbalanced data based research [37]. ROC-AUC score can be defined as the area under the ROC curve for a certain classifier. This area represents the probability that a minority class member will be classified higher than a majority class member [109].

3.2 Methodological Approach for Sequential Deep Learning Study

In this section, the methodological approach is outlined. Implementation structure has been introduced in three main steps, namely Encounter Model, Risk Labeling and Dataset Preparation, and Deep Learning Prediction Model section. In the Encounter Model section, the process of detection of encounters of ships is introduced. Computational approach adapted in the development of spatial scanning algorithm is explained. In the second part, the encounter dataset is prepared for prediction algorithms through risk labeling. In this section, developed risk approach is provided. Lastly, in the Deep Learning Prediction Model section, application of Long Short Term Memory (LSTM) model is presented.

3.2.1 Framework of Sequential Deep Learning

The purpose of this study is to recreate close encounter scenarios in the most realistic way in the spatio-temporal domain and predict risky encounters relying on the data points which are transmitted before an encounter happens. This approach provides the path to build real-time risky encounter prediction mechanisms via machine learning methodologies. In order to build a real-time prediction system, the most crucial step is the construction of a historical dataset which would include a variety of scenarios in a comprehensive form. Further, it's critical for the historical aggregation of scenarios to be high quality since they will be the training source for models which will prevent maritime collisions. Ensuring quality of ship interactions in the historical dataset relies solely on encounter detection algorithms. AIS data are

transmitted in irregular orders by ships, and lack of certain messages or errors are widespread. On this basis, an encounter model is produced in the scope of this study. The encounter model includes state of the art safety mechanisms, openly transcribed simplifications for reproducibility and limitations for further developments. Considering high dimensionality and scale of maritime datasets, a prioritization basis has been efficiency of code developed to detect encounters. After building an encounter dataset, the second step has been identifying corresponding risk for each encounter and labeling them. A simple approach has been used to identify risk, relying on the state of ship domain violation of ships at the exact moment of encounter. As the next step, a deep learning pipeline has been built to integrate sequential navigation data for predicting encounter's risk based on assigned labels. Recurrent Neural Networks (RNNs) are determined to be the most suitable type of networks based on the nature of sequential AIS data. A specific type of RNNs, Long Short-Term Memory (LSTM) algorithm has been used to predict risky encounters. An example encounter scenario from sequential perspective is visualized in Figure 9.

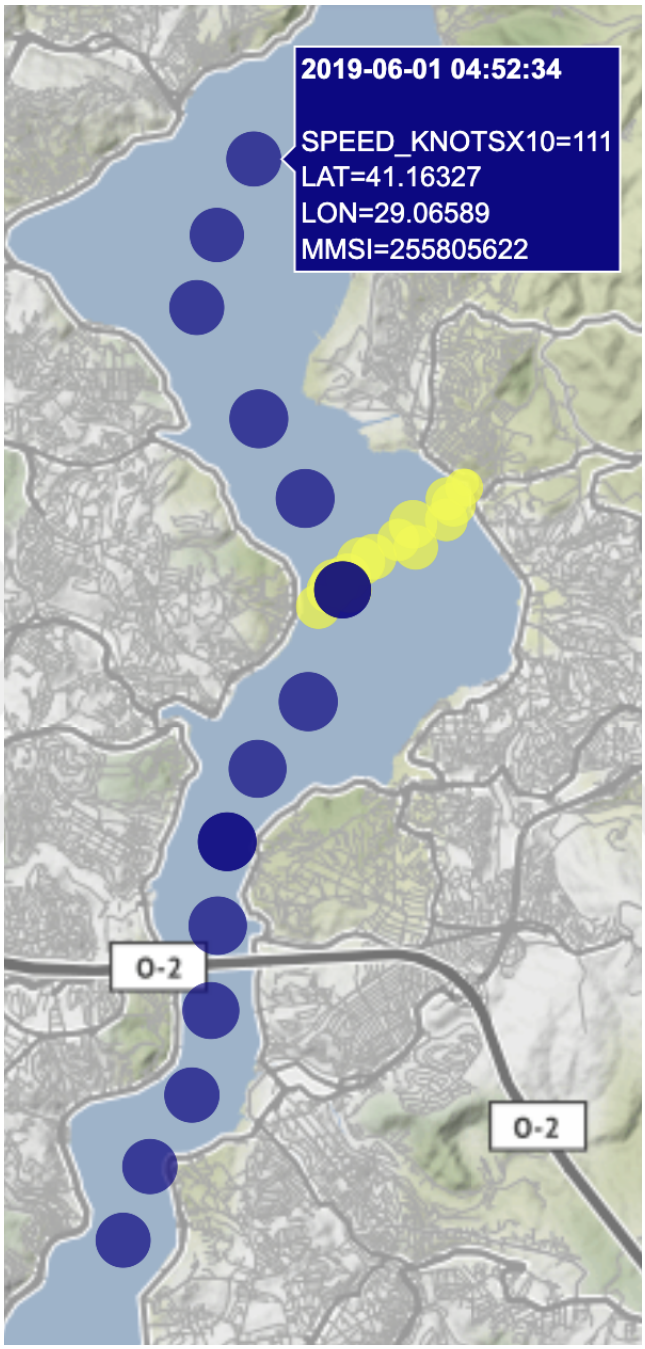

Figure 9: Representation of the encounter scenario with sequential AIS messages the SOI.

The main objective has been adaptability of the proposed prediction mechanism to a real-time potential collision. In this manner, the algorithms rely solely on AIS data, which is scalable, common and transcribes in the same shape for all ships. The

reliance on AIS data presents scalability, cost effectiveness and interpretability in the lens of maritime authorities. The main contribution of this methodological approach is consolidated suitability of the model to be applied in real-world scenarios.

3.2.2 Encounter Model

The raw AIS dataset covering between June 2019 and August 2019 dates of maritime traffic was trans-formed to an encounter model. The resulting dataset has been the input source for the sequential deep learning prediction model, presented in the last section. An extensive scanning algorithm is developed to find all interactions between ships in the SOI during a 2 months period. In the end of the encounter detection process, qualified encounters are classified as risky, gray-zone or non-risky encounters based on spatial proximity of ships with respect to each other.

Encounter detection process is designed to take the sequential nature of ships' journey into account. Each encounter is determined in regards to previous mes-sages of ships leading to a specific encounter. To conduct this, different journeys for each ship are sub grouped with respect to timestamps and coordinates of messages as the initial step. Based on spatiotemporal proximity, encounters of ships are detected and recorded. For each ship, other ships in spatiotemporal proximity are interpolated to the exact message time of the ship, to find the exact distance between the own ship and potential target ship. Since the utilized dataset was a down sampled dataset for efficient storing purposes, our approach in interpolation of the target ship had a significant role in encounter determination. In the initial scanning process, a 1,500 meters radius is used and all ships inside the searching radius are included as potential encounters. Haversine distance is calculated between ships, and the exact distance be-tween AIS transporters of ships are used for proximity calculations. To accurately represent encounters, only ships which have speed more than 3 knots are included, and tugboats are excluded from encounters.

3.2.3 Risk Labeling and Dataset Preparation

Each encounter is labeled based on their risk context for deep learning prediction. On the basis of the novelty of prediction approach, sophistication of risk labeling has been intentionally designed as simple for improving interpretability and reproducibility.

To label encounters as risky, non-risky and gray zone encounters, the principle of "each ship's domain should not be violated by target ships" has been used. Ship domain is defined as a circular area around the own ship, which is the area other ships should intend to keep away from [69, 88, 70, 47]. The approach is also defined as the empirical ship domain by [89].

Previously, [77] provided 3-dimensional clustering results for different encounter scenarios. Their findings suggested that a gray zone is observable for ship-ship encounters, meaning that many of the encounters are actually not risky or non-risky, rather they are observable in a gray zone. This means, there are actually three different risk labels for an encounter: risky encounter, non-risky encounter and gray zone encounters. In the scope of this context, each encounter is labeled based on their proximity with respect to each other. Target ships which have less than own ship's length distance to the own ship are labeled as risky, ships which have between 1 own ship's length and 3 own ship's length are labeled as gray zone and ships which have greater than 3 times own ship's length and less than 4 times own ship's length are labeled as non-risky encounters. Target ships which have greater than 4 times own ship's length are not accepted as an encounter.

3.2.4 Deep Learning Prediction Model

For each encounter, preceding messages of the target ship are used to determine if the target ship will violate its own ship's domain or not. In preceding mes-sages, speed over ground (SOG), course over ground (COG), length, width, coordinates of the target ship in different messages and wind information are used as input variables

for sequential prediction model. Al-so, base ship's information in the exact moment of encounter are used as model input along with sequential navigation information of the target ship.

3.2.4.1 Model Variables

Based on the encounter model, two types of data are stored, namely own ship's static and dynamic in-formation at the exact moment of encounter and the target ship's static and dynamic information during encounter and 5 preceding messages to encounter, ex-cluding the nearest message. Considering that the ap-proximate interval between messages is 3 minutes on average, each encounter's risk is being predicted 6 minutes before a potential incident.

Static information of ships used are length and width of own and target ships. In terms of dynamic information, speed over ground (SOG), course over ground (COG), Latitude and Longitude of ships are utilized. Wind Speed, Wind Angle and Wind Temperature are added as variables for training purposes.

3.2.4.2 LSTM

To predict encounter risk based on navigational messages, Long Short Term Memory (LSTM) models are used. LSTMs are fundamentally designed to be used with sequential data input contexts, and navigational messages present a suitable domain for this model's adaptation. LSTMs are part of a greater family of neural networks, Recurrent Neural Networks (RNNs). RNNs are useful in different tasks, where data format is in lists or sequences. They are successful in predicting the next item in a list of items, as long as information relied on is near the predicted item. While, they have certain practical shortcomings in storing information for long periods of time from previous members of the list, although theoretically RNNs are capable of it. On the other hand, LSTMs are a specific kind of RNNs, where long term dependency is a strong asset.

LSTMs' structure includes repeating compartments, and each compartment hosts four neural network layers. The cell state creates the unique ability to store in-formation along long chains, which is a line passing through compartments with optional interference. Gates work interactively with cell states for managing information flow. Through three gates in each compartment, flow towards the cell state is controlled. Sigmoid layers either allow the full passage of information or prevent the flow in the binary form.

3.2.4.3 Application Details

The applied model consists of two LSTM layers, followed by a dropout layer and a dense output layer. The model is implemented through Tenserflow's Keras Sequential API, and the architecture can be described as follows:

- In the case of this study, each input includes the following features for the own vessel corresponding to sequential AIS messages:
  - Speed
  - Latitude and Longitude
  - Course
  - Length and Width
- For target vessels, data from only the exact moment of encounter are utilized in input sequences. Following features of the target vessel are used:
  - Speed
  - Course
  - Length and Width

1st LSTM Layer: The initial LSTM layer of the model captures temporal dependencies in the input sequences. The model takes sequences of AIS messages as inputs, each with a certain number of static and dynamic features. The layer consists of 64 LSTM hidden units and uses a rectified linear unit (ReLU) activation function. This layer returns full sequences of hidden states and output is in the shape of (32, 5, 64). The first element represents the batch size, the second element represents the number of AIS messages used to train the model in each encounter and the third element represents the number of hidden units. Based on these inputs, considering that each LSTM unit consists of 4 sets of weights and biases, 19200 parameters are included in this step. Number of parameters are calculated as below in Equation (10):

$$\text{Number of parameters} = 4 * ((\text{Number of input units} + 1) * \text{Number of LSTM units} + \text{Number of LSTM units}^2) \tag{10}$$

2nd LSTM Layer: In the second layer, 64 units for each navigational message are taken as input for each member of the batch. Similar to the first layer, this layer also carries LSTM structure, and differently consists of 32 hidden units. This layer also uses the ReLU activation function and unlike the first layer, only the final hidden states of input sequences are returned in this layer. Since LSTMs ensure that each hidden state at each time step is influenced by previous time steps' hidden states, each final hidden state carries full sequence's information. This is called sequence to label mapping and the final hidden state includes all sequences' information, which makes it suitable for classification tasks. Also, reduced dimensionality enables compatibility with remaining layers, considering the final dense layer expects a two dimensional input. The output of this layer includes 32 LSTM units for each member of the batch in (32,32). The total number of trainable parameters of this layer reaches to 12416. Number of parameters are calculated as below in Equation (11):

$$\text{Number of parameters} = 4 * ((64 + 1) * 32 + 32^2) = 12,416 \tag{11}$$

Dropout Layer: This layer maintains regularization with a rate of 0.2 via randomly transforming 20% of inputs to zero and reduces overfitting. This application increases variability within the training dataset. The output shape is (32, 32), where the first element represents the batch size and the second element represents input units from the second LSTM layer. The dropout layer does not have any trainable parameters.

Dense Output Layer: This is a fully connected layer with one output unit and softmax activation. Softmax provides probabilities of outputs representing a summation of 1. The input shape is (32,32), where for each member of the batch, 32 hidden units are present. The output shape is (32,1), where 32 represents the batch size and 1 represents the output value for each member of the batch.

Total number of parameters sums up to 31,649. The model is compiled via Adam optimizer, the loss function used is sparse categorical cross entropy, since outputs are coded in integer units. During the hyperparameter optimization, learning rates between 0.0001 and 0.01 were tested.

Learning rate represents the magnitude of updates conducted to the model's weight function during gradient descent optimization. Through the learning rate's multiplier, steps are taken by the model to minimize the loss. To empirically find the ideal performance, a predetermined range of learning rates are experimented from 0.0001 to 0.01. The aim of this test was measuring the model's convergence behavior and generalization capacity. Holding all other parameters constant, the model is run through Adam optimization for varying learning rates. Results are assessed based on training loss, validation loss and accuracy. The best learning rate in terms of model convergence is determined as 0.001.

CHAPTER IV

RESULTS AND DISCUSSIONS

4.1 Clustering

The presented approach is applied to a comprehensive AIS dataset collected from the Strait of Istanbul. After the extensive encounter model development process, a clustering dataset has been prepared. Three variables are determined to be input for the clustering model, namely, Own Ship's Length (0 to 1), Target Ship's Speed (0 to 1) and Violation Distance per Own Ship's Length (V.D.P.O.S.L.) (-1 to 1). The -1 to 0 range for V.D.P.O.S.L. should be interpreted as the situation of the ship domain being violated. The 0 to 1 range for the same scale should be interpreted as the ship domain not being violated. Moreover, -1 represents the maximum distance between ships with respect to their own ship's length, and $+1$ represents the minimum distance with respect to their own ship's length. Own Ship's Length is scaled between $10, \mathrm{m}$ and $300, \mathrm{m}$. Target Ship's Speed is scaled between 2m/s to 15m/s.

A total of $73,543$ interactions are included in the three-dimensional clustering process. Figure 10 represents a general overview of the process.

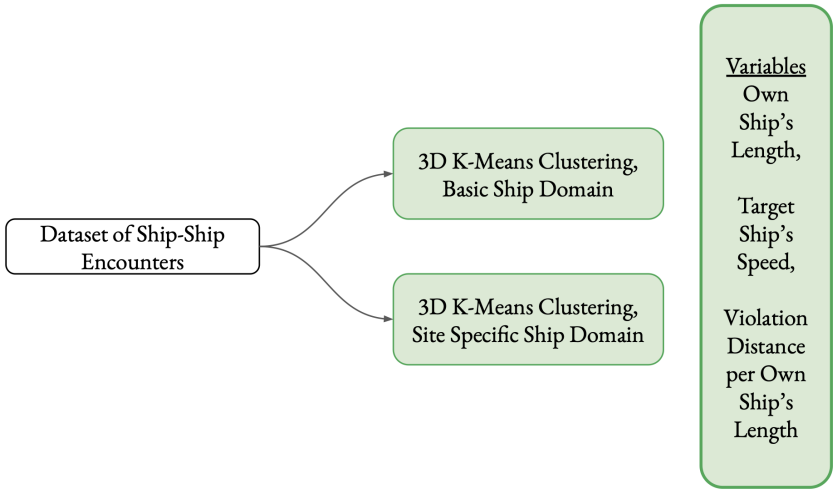

Figure 10: General overview of the clustering process.

In the first clustering model, the basic ship domain approach is utilized, as provided in Equation (2). To detect the optimum number of clusters, Silhouette Coefficient and elbow analyses have been conducted. In the Silhouette analysis, 2 clusters obtained the highest score with 0.3651. It was followed by 3 clusters with the second-highest score of 0.3183 and 4 clusters with 0.3143. Moreover, combined with a visual observation for different clusters and the elbow analysis, local optimums in 3, 5 and 9 are selected as candidate numbers of clusters. Silhouette analysis results can be found in Figure 11. In Figure 12, the elbow method's outcome is presented.

In the second clustering model, site-specific ship domain conditions are used as provided in Equations (3) and (4). Silhouette Coefficient resulted in two possible optimum locations, 3 clusters and 5 clusters, with 0.3703 and 0.3027 scores, respectively. Following an elbow analysis and observational judgment, a number of clusters are designated based on these two alternatives, 3 and 5 clusters. Figure 13 and 14 represent the Silhouette coefficient and elbow model's graphs for site-specific ship domain conditions, respectively.

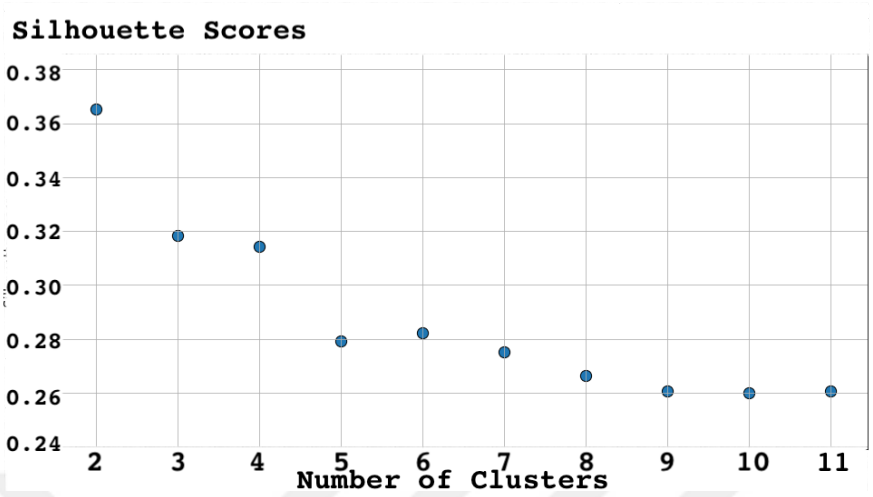

Figure 11: Silhouette Score graph for basic ship domain model.

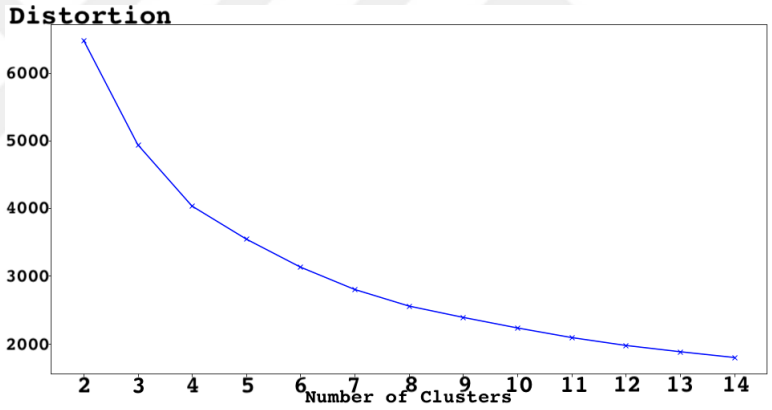

Figure 12: Elbow Method graph for basic ship domain model.

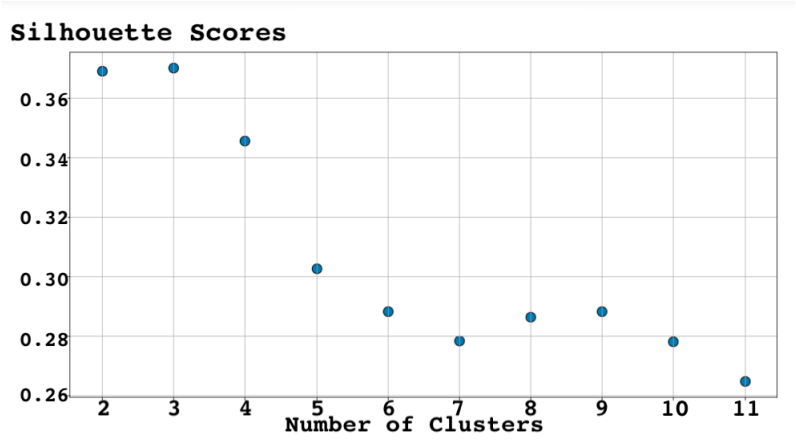

Figure 13: Silhouette Score graph for site-specific ship domain model.

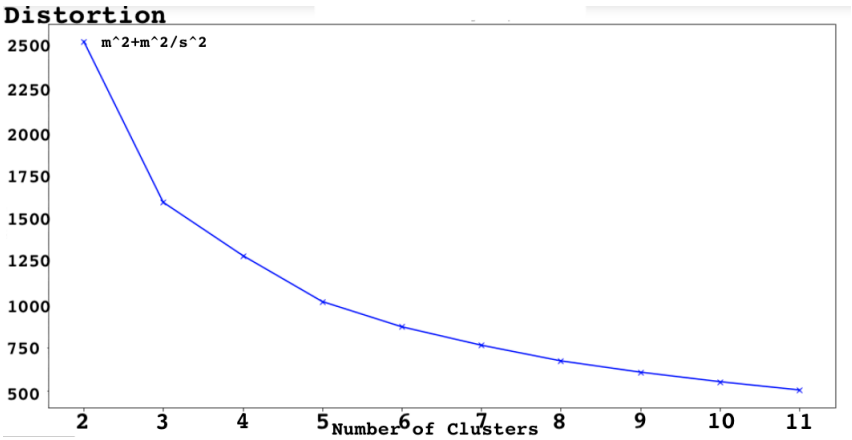

Figure 14: Elbow Method graph for site-specific ship domain model.

Figures 15, 16, 17 and 18, 19 represent three-dimensional clustering visualizations, and in each figure, x, y, z dimensions represent:

1. x = Own Ship's Length, Scaled (0 to 1)
2. y = Target Ship's Speed, Scaled (0 to 1)
3. z = Violation Distance per Own Ship's Length (V.D.P.O.S.L.), Scaled (-1 to 1)

Figures 15, 16, and 17 are the representation of the basic ship domain approach, while Figures 18 and 19 are the representation of the site-specific ship domain approach. In Figures 15 and 18, 3 clusters case is presented for two different ship domain approaches. Figures 16 and 19 represent 5 cluster cases. Finally, 9 cluster cases are presented in Figure 17. In all visuals, colors are randomly assigned and do not have an underlying implication.

When all clustering outcomes are examined, the mutual pattern can be identified as the mid-cluster, where a smooth trend of switching from non-violation to violation is observed. As the z-axis represents the Violation Distance per Own Ship's Length parameter, a certain group of interactions that can be defined as almost violations and

almost non-violations are observed to be carrying similar properties. So, they form a cluster together. This accumulation around point 0 in the z-axis can be interpreted with the indefinite or ambiguous transition between violation and non-violation situations between ships. In other words, this is the grey zone, which is the end-product of captains' judgments about the risk. When a ship domain is applied to label these encounters, some of them are labeled as non-risky, although they are very close to risky encounters. As a result, the classification of each ship–ship encounter is binary in terms of risk can be expressed to be challenged with this result.

Edges of clustering results in the z dimension explain how limiting properties are shaped in ship–ship interactions. While there is an unclear separation between violation and non-violation situations in the 0 line, both edges are distinctly clustered together with some data points in the mid-cloud. Considering the three-dimensional plot, the main distinction between interactions is almost always the Violation Distance per Own Ship's Length, as provided.

In Figures 15, 16, and 17, the dispersion of ship–ship encounters in three dimensions is visualized from the own ship's length perspective. With the increase in the number of clusters for the basic ship domain between Figure 15 and 16, the violation indicating cluster gets smaller. The mid-cloud gets divided, and the non-violation cluster gets divided into two parts: one is in the middle of the violation rate while the other is in the far-right edge. The interpretation of this point is how edge points are not being impacted by the change in the number of clusters and how the middle cluster is being further divided. At this point, since the majority of interactions are in the region where violation to non-violation switch occurs, a further separation within this group is being required by maximization of between clusters and minimization of within-cluster principles. At the same time, with 5 clusters, the far-right cluster gets to be divided into the vertical axis (target ship's speed). This can be explained as the target ship's speed being a more crucial determinant than the own ship's length for

the case of non-risky interactions. In Figure 17, nine clusters case is presented. The occurrence of the far left risky clusters is observed. Moreover, in the nine clusters case, only 8 clusters are visible from the x perspective. This indicates how the own ship's length becomes crucial in the nine clusters case; thus, some of the clusters experience separation in the x perspective.

With the implementation of a site-specific ship domain, more realistic outcomes are attempted, and results are presented in Figure 18 and 19. As a result of the implementation, more data points are found out be belonging to the risky encounter class when compared to the basic ship domain. Figure 18 represents the three clusters case, where increased encounters on the negative z-axis are observed. At the same time, based on a general comparison between basic ship domain and site-specific ship domain approaches for five cluster cases, a slight change towards the negative sign in the z-axis is observed in the cluster centers. This can be understood as the impact of site-specific ship domain and the realistic outcome being riskier in the generous sense when a more accurate ship domain approach is implemented.

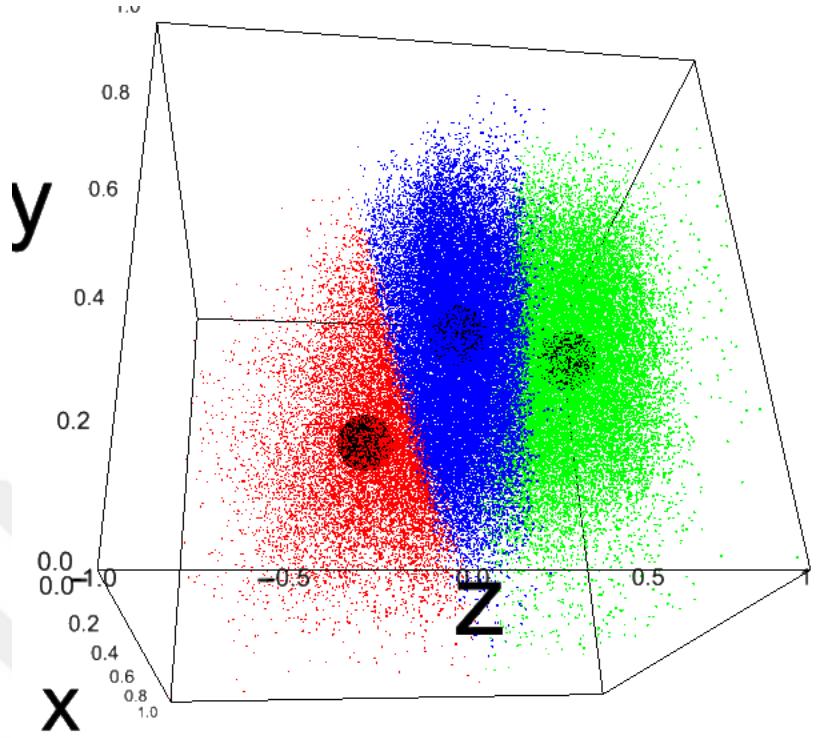

Figure 15: C1: Three-dimensional Clustering with Basic Ship Domain, X perspective: 3 clusters case.

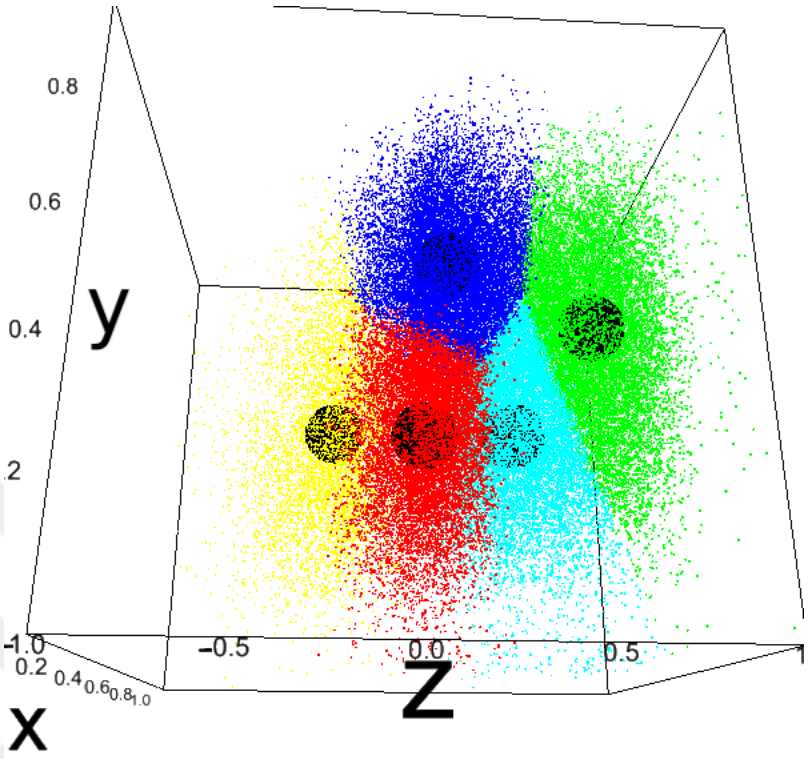

Figure 16: C1: Three-dimensional Clustering with Basic Ship Domain, X perspective: 5 clusters case.

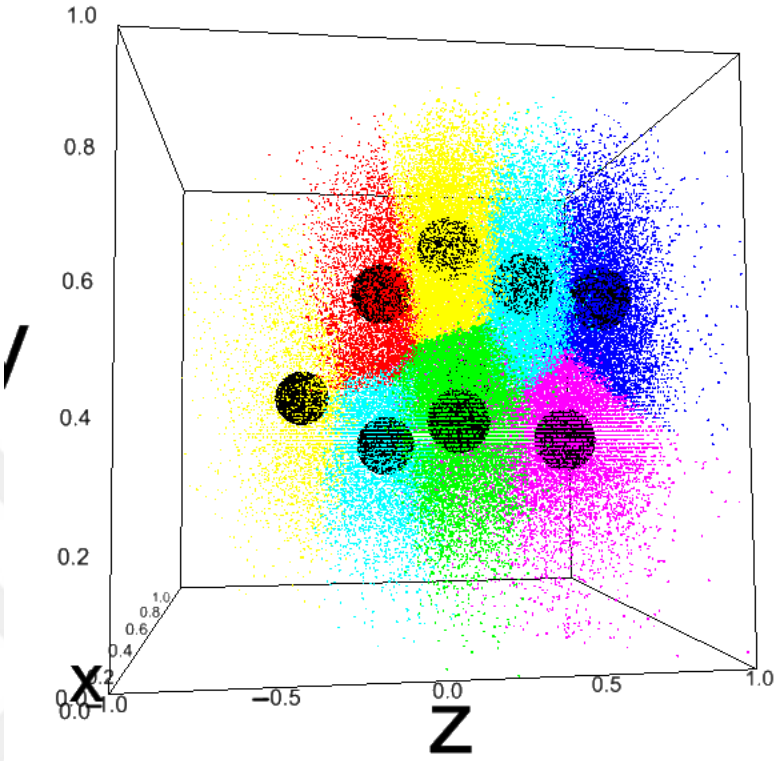

Figure 17: C1: Three-dimensional Clustering with Basic Ship Domain, X perspective: 9 clusters case.

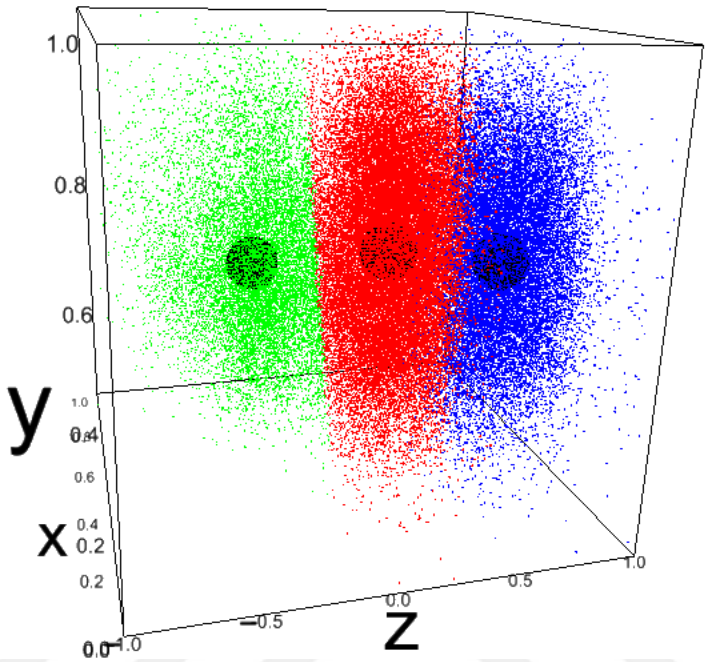

Figure 18: C2: Three-dimensional Clustering with Site Specific Ship Domain, X perspective: 3 clusters case.

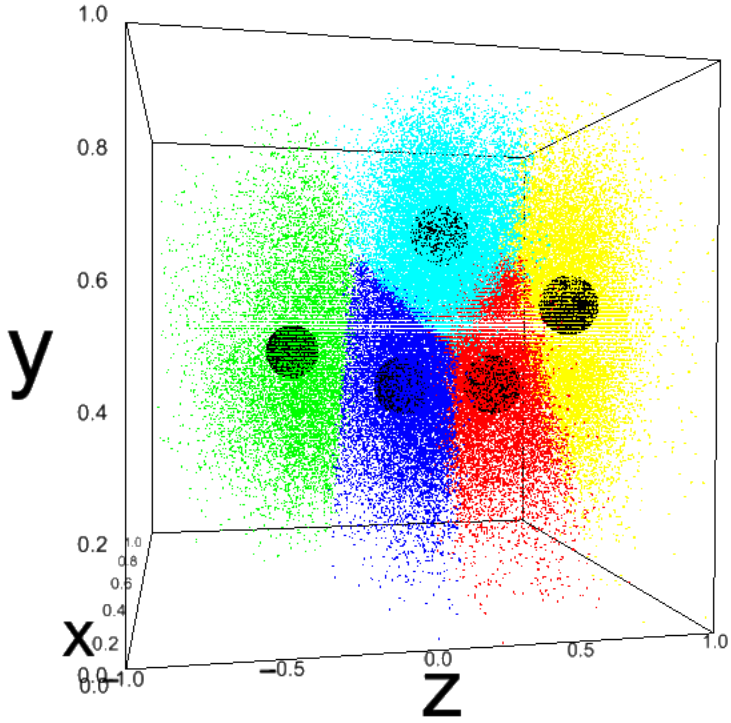

Figure 19: C2: Three-dimensional Clustering with Site Specific Ship Domain, X perspective: 5 clusters case.

In addition to the visual representation of the clustering, results are also shown in Tables 4 and 5 in terms of clusters' centers for each analysis. Fitted centers are demonstrated to detect risky encounters and corresponding model variable values for their fitted centers. In Table 4, basic ship domain approach, the highest target ship speed is observed in the violation occurring cluster's center. While the risky encounter cluster is also represented by the lowest own ship's length value. This indicates that the target ship's speed and own ship's length are important determinants of ship domain violation as they represent two edges for both features. In Table 4, five clusters setting, the risky encounter cluster is also observed with the highest target ship speed and lowest own ship's length. Furthermore, the target ship speed is higher than the three clusters case for this setting. At the same time, in the site-specific

ship domain setting, a different pattern is observed. Violation representing cluster is identified by the highest own ship's length value. This result shows how used decision criteria shapes outcomes, and model variables are representative of risky encounter from different perspectives in varying settings.

Table 4: Basic Ship Domain Clustering Results: Fitted Centers.

Non-Scaled, Basic Ship Domain			
3 Clusters Fitted Centers			
Cluster No	Own Ship's Length (m)	Target Ship's Speed (m/s)	V.D.P.O.S.L.
1	96.872	4.821	-0.203
2	266.912	3.722	0.573
3	139.928	3.562	0.161
5 Clusters Fitted Centers			
Cluster No	Own Ship's Length (m)	Target Ship's Speed (m/s)	V.D.P.O.S.L.
1	229.16	5.421	0.278
2	109.352	3.502	0.055
3	97.808	5.275	-0.278
4	186.728	3.262	0.348
5	284.384	3.489	0.655
9 Clusters Fitted Centers			
Cluster No	Own Ship's Length (m)	Target Ship's Speed (m/s)	V.D.P.O.S.L.
1	129.008	5.035	-0.054
2	298.424	3.262	0.724
3	76.592	3.442	0.08
4	236.024	3.289	0.511
5	91.256	5.654	-0.39
6	117.152	3.249	0.115
7	285.944	5.181	0.493
8	173.624	3.282	0.302
9	205.136	5.315	0.208

Results indicate that the most distinctive axis is the z-axis, which is the defined metric. It would indicate that while length and speed present a diverse distribution, theirs are rather uniform, while V.D.P.O.S.L. metric segments encounter different groups. In other words, the defined metric is a proxy parameter to understand the encounters' nature about involved risk, and occurring clusters along 0 value indicate encounters should not be solely defined as risky or non-risky; rather, the risk is a continuous measure.

Table 5: Site-Specific Ship Domain Clustering Results: Fitted Centers.

Non-Scaled, Site-Specific Ship Domain 3 Clusters Fitted Centers			
Cluster No	Own Ship's Length (m)	Target Ship's Speed (m/s)	V.D.P.O.S.L.
1	124.64	5.055	0.15
2	207.008	4.435	-0.312
3	109.352	4.921	0.604
5 Clusters Fitted Centers			
Cluster No	Own Ship's Length (m)	Target Ship's Speed (m/s)	V.D.P.O.S.L.
1	120.584	4.135	0.384
2	116.216	6.101	0.163
3	105.608	5.181	0.677
4	145.232	4.082	0.052
5	210.44	4.515	-0.359

When the overall results of both the B.S.D and S.S.S.D approaches are analyzed, the clustering approach shows the middle clusters, which are at standing close to the ship domain violation, have members both at the violating and non-violating side of the domain boundary. This outcome proves that captains make their judgments according to visual inspection. As a result, ship domain violation is found to be a grey zone rather than a bold line. This outcome can help while making judgments about the risk level of the encounters.

4.2 Ensemble Machine Learning

The proposed methodology is applied to an extensive dataset which is gathered for the Strait of Istanbul. To set up a well-designed model, an iterative procedure has been followed while different machine learning algorithms and approaches are combined, which are introduced in the Methodology section. This section provides the outcomes of these steps. With the support of results, the relationship of each combination and iteration is shown and flow of improvement via the imbalanced focused approach is demonstrated.

Figure 20 represents the outcomes of the presented machine learning applications with respective evaluation metrics. For each combined model, after training, test

results are obtained and presented in Figure 6. Six different setups are presented with their codes, namely, M1, C1+M1, C1+M2, C2+M2 and C3+M3. All cases are provided for three clusters case. While all applications carry a training and testing procedure, C3+M3 differs with its design. In order to present an unbiased risky encounter prediction, C3+M3's test process has been conducted independent of distance parameter, entirely. This approach enabled a framework for the trained model usage in different straits and canals. In this section, clustering and classification results are presented and discussed with their interpretation. Lastly, the applied testing and evaluation results are presented in a detailed manner.

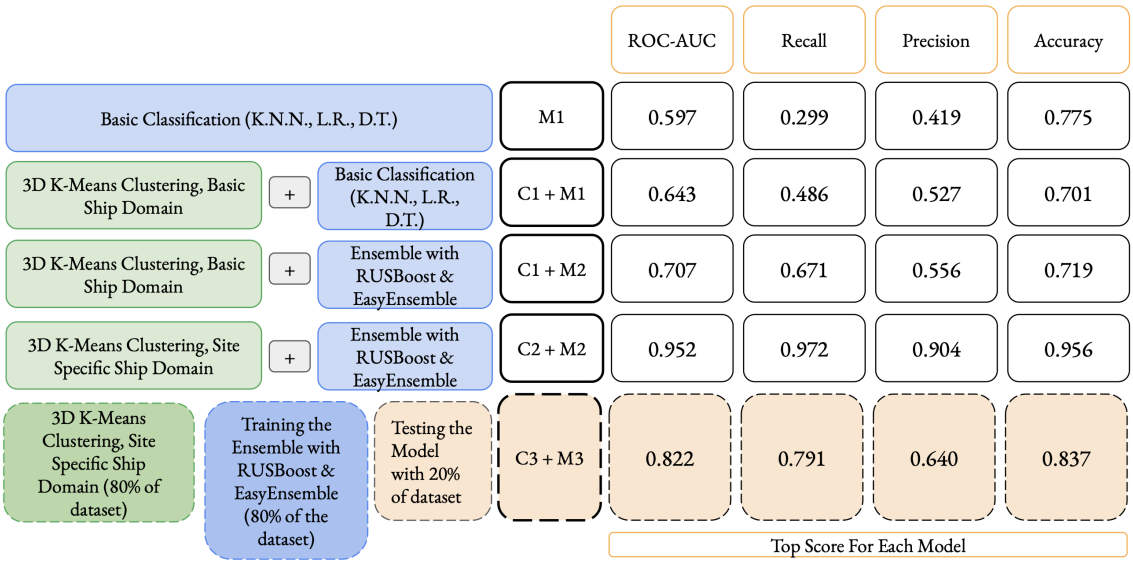

Figure 20: Evaluation Metrics for Benchmark Models.

To accurately predict if a data point is representing a ship domain violation or not, a set of iterative steps have been followed. At each step, it has been aimed to make a better predictor in terms of evaluation metrics. Right after the initial basic classification model denoted as M1, 3 dimensional clustering results have been introduced to classification steps in a way to manually ensemble two models, as also represented in Figure 20.

During the analyses, Stratified 10-Fold Cross Validation has been applied to models which present the clustering analysis and classification under the same dataset, other than the model in the Testing section. Due to the imbalanced nature of the dataset, stratification has been needed to assess the reliability and quality of the proposed model. Separating the training dataset and applying clustering only to the training dataset technically prevented the cross validation to be implemented to this model. As explained in the Methodology section, classification models' outcomes have been reported with four metrics, Precision, Recall, Accuracy and ROC-AUC Score. Due to the significant imbalance observed in the dataset, it's been critical to measure the predictions with correct metrics. Confusion Matrices for related models have been also represented.

In order to select a minimum number of model variables, tree based feature importance test is conducted. Results showed that, Own Ship's Length and Target Ship's Speed have been the top contributor factors, those are selected to be included in the final model. Results of feature importance test have been visualized in Figure 21. Results of decision tree based importance test Model variables and prediction class are represented as:

1. Model Variable 1 = Own Ship's Length, Scaled (0 to 1)
2. Model Variable 2 = Target Ship's Speed, Scaled (0 to 1)
3. Prediction = Violation Class, Indicating if ship domain violation occurred or not (0 or 1)

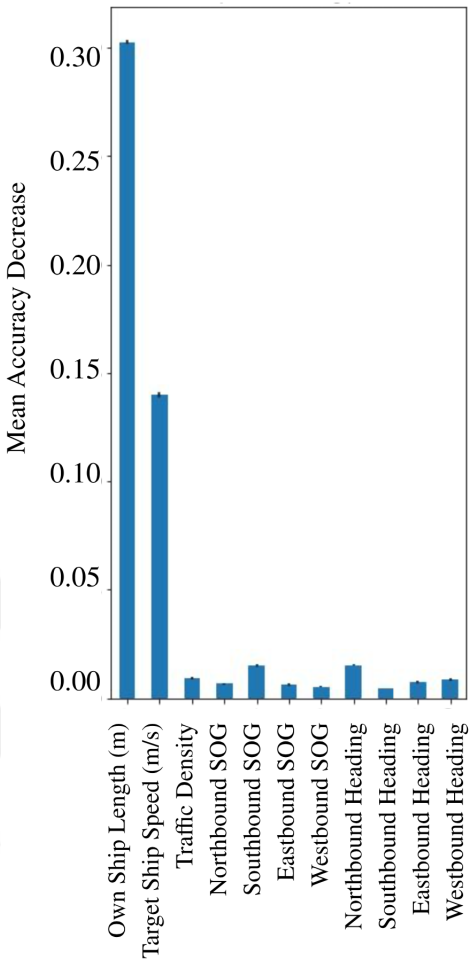

Figure 21: Feature Importance Test Outcomes.

Table 6 and Table 7 and confusion matrices in Table 8 and Table 9 provide outcomes of differently structured machine learning models. The iterative procedure showed that clustering and ensemble algorithms are impactful to conduct more successful predictions. While basic structured machine learning models performed comparatively poorly, integration of clustering results as a preliminary step significantly improved risky encounter predictions. It's important to highlight that ensemble learning methodologies also significantly increased accurate predictions. Meanwhile, a testing set up has been conducted with C3+M3 structure. The significance of testing is presented as its unbiasedness and its suitability for different waterways' with

a pretrained model. Recall scores of the testing structure indicates that pretrained model is able to predict 4 out of 5 risky interactions. Decision Tree and Logistic Regression based EasyEnsemble models provide highest evaluation rates in the scope of this approach. Each model's outcomes are provided and compared. Table 10 and Table 11 represent hyperparameter search outcomes.

Table 6: Classification Outcome for Non-Ensemble Machine Learning Algorithms

	K.N.N.	Random Forest	Logistic Regression	
M1	0.419	0.438	0.529	Precision
	0.299	0.284	0.187	Recall
	0.775	0.782	0.803	Accuracy
	0.597	0.596	0.572	ROC-AUC
C1 + M1	0.527	0.552	0.642	Precision
	0.486	0.473	0.415	Recall
	0.701	0.713	0.743	Accuracy
	0.643	0.648	0.654	ROC-AUC

Table 7: Classification Outcome for Ensemble Machine Learning Algorithms

	EasyEnsemble L.R.	EasyEnsemble D.T.	RUSBoost L.R.	RUSBoost D.T.
C1 + M2 (3 Clusters Case)	0.556	0.517	0.553	0.518
	0.671	0.744	0.671	0.743
	0.719	0.692	0.717	0.692
	0.707	0.706	0.706	0.706
C1 + M2 (5 Clusters Case)	0.470	0.445	0.461	0.444
	0.628	0.693	0.635	0.691
	0.709	0.685	0.702	0.683
	0.683	0.687	0.681	0.686
C2 + M2	0.904	0.906	0.899	0.906
	0.972	0.976	0.953	0.975
	0.956	0.958	0.946	0.959
	0.952	0.954	0.944	0.954
Test: C3 + M3	0.640	0.659	0.581	0.659
	0.791	0.794	0.744	0.791
	0.837	0.846	0.803	0.845
	0.822	0.828	0.783	0.828

Table 8: Confusion Matrix for EasyEnsemble

C2 + M2: EasyEnsemble D.T.

Confusion Matrix				
Truth	0	7745	462	Predicted
	1	112	4725	
		0	1	

Table 9: Confusion Matrix of Prediction Test Results

C3 + M3 (Test): EasyEnsemble D.T.

Confusion Matrix				
Truth	0	9532	1509	Predicted
	1	757	2911	
		0	1	

Table 10: EasyEnsemble Hyperparameter Search

Ratio	Number of Subsets	Replacement
Auto	10 - [10,20,30]	False

Table 11: RUSBoost Hyperparameter Search

Number of Estimators	Learning Rate	Algorithm	Sampling Strategy	Replacement
20 - [20,40,60,80,100]	0.1	SAMME.R	Auto	False

4.2.1 M1: Basic Model

In this model with the most basic structure, predictors are defined as Own ship's Length and Target Ship's Speed. Target Variable is defined as Violation Indicator, which is made of zeros and ones, representing if that specific interaction is a representative of ship domain violation, or not. Target Variable of this model is identified via

the basic ship domain approach, where the ship domain of each interaction is equal to twice of ship length. Three different machine learning algorithms have been adopted, namely, K-Nearest Neighbors (K=3), Random Forest and Logistic Regression. Poor results indicate that a straightforward model is not suitable to detect ship domain violation.

4.2.2 C1 + M1: Clustering Based Ensemble with Basic Ship Domain Approach

The results of three-dimensional clustering showed that one of the clusters is positioned in the far positive side of violation distance per own ship's length axis. This has laid the foundations of an idea to set up a manually ensemble-based model in a way to reduce the impact of class imbalance. This enabled improvement of the model's capacity to correctly classify the test instances. Based on the results of three-dimensional clustering, the cluster which represents non-violation occurring interactions is removed from the dataset. This removal resulted with the decreasing number of 0 classes while it significantly improved the number of 1 classes. It can be also argued that removal of a cluster which contains significant information about non-violation occurring instances with being positioned in the extreme edge of the three-dimensional space, experimental approach proved that this impact did not result as significant as ensuring class imbalance in a more balanced order, as it is observed in the metrics.

As it's provided in table 7, introduction of clustering based assembling to the model significantly contributed to the outcomes. While models differed in Precision and Recall scores, ROC-AUC scores from each provided very similar outcomes. On the other hand, it can be expressed that the model's results are still far from acceptable. In order to improve ensemble models, hyperparameter optimization is conducted through grid search. Due to RUSBoost and EasyEnsemble having highly detailed engineering structure, limited types of hyperparameters are tested and applied. For

base models, logistic regression and decision tree, scikit-learn library's default modes are applied [110]. In tables 10 and 11, grid search for ensemble algorithms and other hyperparameters are presented.

4.2.3 C1 + M2: Clustering Ensemble & EasyEnsemble & RUSBoost with Basic Ship Domain Approach

Although three-dimensional clustering based modification of the dataset improved the outcomes, class imbalance has still not been entirely overcome. In this step, two different models of three-dimensional clustering application, 3 clusters and 5 clusters cases results are demonstrated. In order to achieve better metrics, EasyEnsemble and RUSBoost ensemble methodologies are introduced to the model. After experimentation with various machine learning algorithms, Logistic Regression and Decision Tree are implemented as alternative base algorithms. Introduction of ensemble algorithms significantly improved the prediction metrics, while no significant difference has been observed among base algorithms and ensemble algorithms. Despite the improvement, accuracy in the prediction of the minority class, 1s, could not be identified as adequate as well.

4.2.4 C2 + M2: Clustering Ensemble & EasyEnsemble & RUSBoost with Site Specific Ship Domain Knowledge

Considering the recent work by [72], instead of utilizing a constant multiplier for every sort of ship's length, a changing multiplier based on ship length has been applied during the iterative procedure. This has significantly contributed to the outcomes of the predictions, which is a clear indication of suitability of the length based changing multiplier for the ship domain length to the real world applications. This can be also interpreted as changing the ship domain approach have particularly served the predictors' impact on predicting the classes. The result represents a dramatic increase in the metrics based on the site-specific ship domain model, as in table 8.

4.2.5 Testing

The iterative process besides this section has been based on modifying the main dataset with outcomes of the clustering analysis and splitting the modified dataset to training and testing. These set of iterations provided major insights on integration of different variables and meta algorithms to the model. In this section, by means of removing "distance between two ships" impact on the test set, real-world application of the model is being tested. It could be argued that observing clustering's impact on the test set may not be applied to a real-world model since if the distance between two ships is included to the model in the test set, this might not be the optimum case due to the fact that distance is a direct factor of near miss candidate detection. The approach in this paper adopted the methodology of any method which to predict a near miss candidate needs to conduct this prediction without distance between ships as a model variable, specifically in the test process.

C3 + M3: 3D Clustering Based Test

For three-dimensional clustering of 80% of the dataset, site specific ship domain is practiced. Since a random sample is extracted from the main dataset, ideal candidate number of cluster is accepted as 3, an approach also in line with [96].

In this testing approach, the dataset is split into training and test sets before three-dimensional clustering is implemented. While the training set constituted 80% (58,834 data points) of the entire dataset, the test set has been established as 20% (14,709 data points). Three-dimensional clustering analysis has been applied to the training set. This clustering application constituted the same parameters, where introduced variables are Own Ship's Length, Target Ship's Speed and Violation Distance per Own Ship's Length. Right after applying clustering and removing the non-violation occurring safest cluster in the positive edge of the Z axis, the training set has been assigned to train four different ensemble models. Later, the test set which has not been subject to clustering implementation has been used for validation purposes.

Results of the test are presented in tables 7 and 9. Recall scores indicate that 4 out of 5 of each violation can be predicted accurately.

4.3 Sequential Deep Learning

Developed sequential methodological approach is implemented in a 2 months length AIS dataset collected from the SOI. For each encounter, static and dynamic information are separately stored. In the resulting model, 37.9% of encounters were labeled as non-risky, 55.9% has been labeled as gray zone and 6.1% has been labeled as risky encounters.

Since imbalance within classes pose challenges, gray zone encounter observations are under sampled to ensure balance for the model's learning process. Different numbers of epochs ranging from 10 to 50 have been experimented and in each run, the model with highest accuracy among all epochs is recorded.

Changes in training and validation dataset's loss are shown in Figure 22. A preliminary analysis of this graph could be that the model starts to overfit slowly after around 10 epochs. To determine the comparative contribution of variables in prediction, a permutation importance procedure has been introduced. Feature importance results are presented in Figure 23. The length of both ships, speed, course and target ship's course have been the most crucial variables in the LSTM model's prediction. Performance of the model is presented in Table 12.

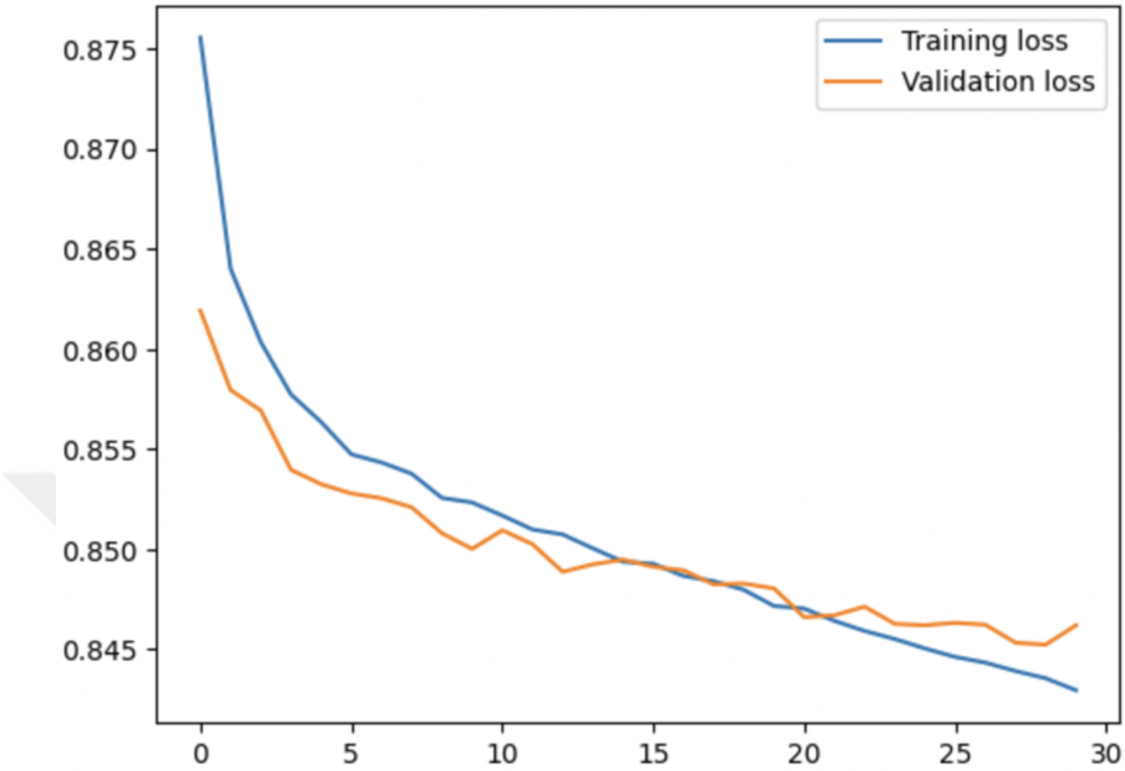

Figure 22: Change of loss for training and validation datasets with increasing epoch.

Table 12: Performance Metrics of the LSTM Model.

Metric	Value
Accuracy	0.563179069
Precision	0.526189077
Recall	0.340983011
F1-score	0.264667268

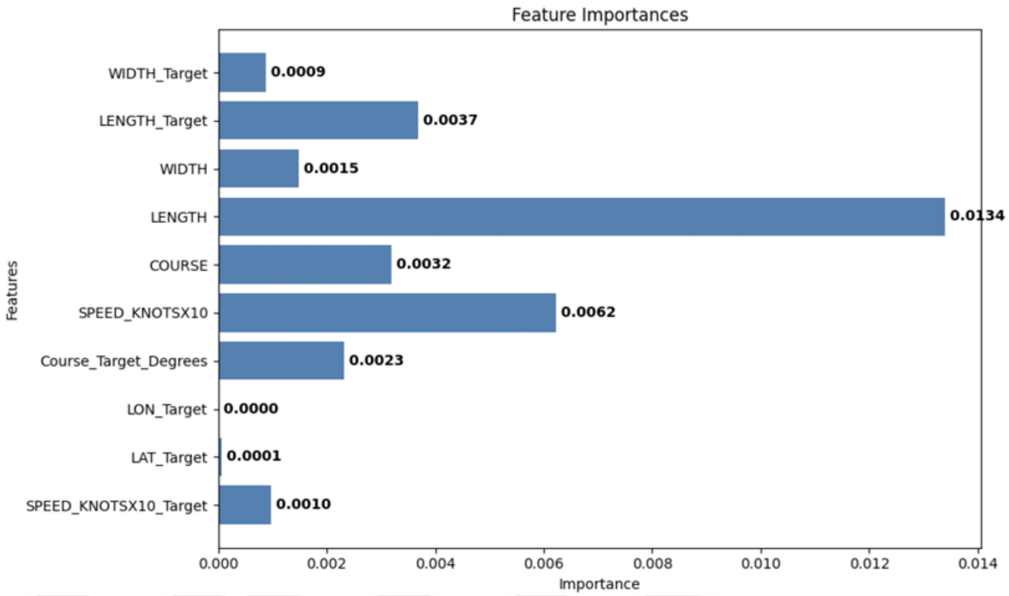

Figure 23: Feature Importance Test Results.

CHAPTER V

CONCLUSIONS

In this study, three novel methodologies are introduced to explore and predict risky encounters between ships in narrow and congested waterways. Initially, three-dimensional clustering analysis is conducted to highlight patterns among risky encounters. Secondly, a novel ensemble machine learning method for predicting risky encounters is introduced. Finally, sequential AIS messages are used to predict risky encounters in real time via deep learning.

Contributions of this study are listed as follows:

1. Previously, maritime safety literature was focused on distance between ships to model risk between ships in a close encounter scenario. In this thesis, the emphasis was on excluding distance between ships from model variables to understand and design safety mechanisms independent of distance between ships. Excluding distance provides the possibility of applying state of the art artificial intelligence approaches to prevent ship accidents before risk occurs.

2. High dimensional clustering approach is presented to highlight violation distance per own ship's length during a risky encounter. Not all risky encounters are alike, and clustering results provide a basis to articulate this.

3. Ensemble machine learning step provides a framework to integrate high dimensional clustering observations to downsample majority class in a supervised prediction pipeline. Further, this novel application provides the basis for predicting potential ship accidents independent of distance between ships, which can enable discovery of drivers of close encounters.

4. Sequential deep learning implementation provides a real time application perspective to prevent risky encounters. While clustering and ensemble machine learning steps provide a basis for applicability of artificial intelligence in potential accident prediction, sequential deep learning process showcases implementation of such a model to ships in real time to support navigational safety. Using sequential nature of AIS messages, potential ship accidents can be predicted via long term historical datasets and risk mitigation measures can be taken before ships establish visual contact.

5.1 Clustering

Three-dimensional clustering analysis is conducted to present patterns among risky encounters. For clustering purposes, the K-means algorithm is used. The number clusters were determined via the elbow and Silhouette method. Visual feature-based mapping for 73,543 encounters is presented, and feature-based separation of risky and non-risky encounters is discussed. The main contribution of this paper is mapping synchronous states of ship speed and ship length in encounter situations together with varying degrees of risk. Ship domain violation per ship's length is presented as a criterion to assess the severity of the risky encounter and integrated into a three-dimensional clustering analysis.

Results show that both the own ship's length and the target ship's speed provide important outcomes in interpreting risky encounters. A cluster of risky encounters forms in each result. For the basic ship domain approach, the lowest own ship's length and highest target ship speed represent the risky cluster. For the site-specific ship domain approach, the highest own ship's length represents a risky encounter cluster. The distinction between risky and non-risky encounters is found to obtain a smooth transition. The Violation Distance per Own Ship's Length (V.D.P.O.S.L.) feature is the most distinctive feature among the presented model features. Basic and site-specific ship domain approaches are used, and both present similar results.

The site-specific ship domain highlights risky encounters and produces a greater risky encounter cluster. In all results, the largest cluster occurs in the middle area, where the smallest Violation Distance per Own Ship's Length values is present. The mid-clusters show that violation is a grey zone rather than a bold line for both basic ship domain and site-specific ship domain applications. Thus, labeling encounters via just ship domain can create over or underestimation of risky encounters.

A limitation of the study can be described as the lack of individual ship-level information due to the analysis of a large number of encounters. On this basis, the study aims to demonstrate the large-scale occurrence of risky patterns and approach encounters independent of their case-by-case nature.

The study can be improved by including additional model variables and cross-comparison of different variable relationships via clustering. Alternative risky en-counter detection frameworks can be applied. The presented model can be applied to different narrow and congested waterways to compare results. Since patterns of risk in terms of model variables are presented, site-specific risk thresholds for model variables can be determined, and these thresholds can be compared between different narrow and congested waterways. In this way, characteristic and quantitative maritime traffic conditions of different sites can be determined. This information can be utilized for maritime traffic regulation through authorities. A ship domain, which has been agreed upon for the given waterway, with all of its irregularities in terms of shape, can be used for further clustering purposes and sensitivity checks of different ship domain sizes can be conducted. Lastly, machine learning-based prediction models can be built to predict risky encounters without including distance by examining ship length and speed as distinctive risk factors. This promising application can help to predict potential risky encounters before vessels enter the waterway.

5.2 Ensemble Machine Learning

Eliminating distance between ships out of decision parameters in predicting risky encounters is achieved via clustering and ensemble machine learning based algorithm training process. Length of the own ship and speed of the target ship are used as model variables. To determine the optimum number of clusters, elbow and silhouette coefficient methods are adapted and 3 clusters occurred as the most distinctive dispersion. Among developed benchmark models, the most successful model resulted with 82.8% ROC-AUC score and 79.4% recall score in an unbiased test process on a highly imbalanced dataset (90%-10%). Ten-fold cross validation has been conducted to improve reliability. In order to ensure the unbiasedness of the validation process, the data set is divided into two. The group with 80% is used for the training process where clustering is applied and the second group with 20% is used for testing purposes.

Since the number of risky encounters are few when compared to whole encounters, a class imbalance problem is observed. It is partially overcome via clustering based segmentation in the algorithm training process. Clustering analysis's results are also presented for visual exploration of the relationship between ships' features considering with the risk assessment of ship domain violation. It's presented that almost violation and slightly violation situations covered the majority of the encounters.

Predictability of risky encounters based on certain features of ships can be interpreted as those features' key part in the occurrence of potential near miss cases. In the scope of this research's outcomes, speed and length of ships are critical variables in the sense of ships' approaching behavior to each other. In other words, speed of a target ship and length of the own ship are important criteria for the severity of involved risk for each ship.

The benchmark studies show that, the accuracy of the prediction is improved with site-specific ship domain values. Therefore, author suggests that during the prediction

process site-specific ship domain should be determined and used for reliable prediction.

The model development and execution are conducted in a historical AIS dataset, thus, the approach can be scaled to test and validate for other waterways. The study can be improved in terms of several ways. One of them is proximity calculations in the ship domain with a parameter which contains maneuverability properties of the ships. This can also help to improve the accuracy of prediction of the risky encounters. In addition to the proximity parameter, different variables of ships' navigation properties, their combinations and given navigational conditions (day-night, visibility, traffic density and etc.) can be examined and their potential impact in the assessment of a ship-ship encounter's risk can be discovered.

5.3 Sequential Deep Learning

In this study, we presented a novel methodological approach for predicting risky maritime encounters using an LSTM model. The approach was implemented on a 2 months length AIS dataset collected from the SOI, including 114,587 encounters. Static and dynamic information were separately stored for each en-counter, and the dataset was divided into non-risky, gray zone, and risky encounters. To address class imbalance, gray zone encounters were under-sampled to ensure a balanced learning process for the model.

The performance of the model was presented with ac-curacy, precision, recall, and F-1 score. Different numbers of epochs were experimented, and the model with the highest accuracy among all epochs was recorded.

Our findings suggest that the LSTM model is effective in predicting risky maritime encounters. The most crucial variables in the model's prediction were the length of both ships, speed, course, and target ship's course. These variables played a significant role in determining the risk associated with a given encounter. While, due to

significant imbalance within encounter types, performance of models have been limited in accordance with the novelty of the approach.

In conclusion, this study demonstrates the potential of LSTM models in predicting risky maritime encounters using AIS data, which can be a valuable tool for maritime authorities and decision-makers in ensuring safer navigation and preventing accidents. Future re-search can extend this approach by incorporating additional factors, improving the model's performance, and exploring its applicability to other maritime scenarios and other busy waterways.

REFERENCES

[1] Z. Wan, J. Chen, A. E. Makhloufi, D. Sperling, and Y. Chen, "Four routes to better maritime governance," *Nature 2016 540:7631*, vol. 540, no. 7631, p. 27–29, 2016.

[2] Y. C. Altan and E. N. Otay, "Spatial mapping of encounter probability in congested waterways using ais," *Ocean Engineering*, vol. 164, p. 263–271, 2018.

[3] L. Gan, Z. Yan, L. Zhang, K. Liu, Y. Zheng, C. Zhou, and Y. Shu, "Ship path planning based on safety potential field in inland rivers," *Ocean Engineering*, vol. 260, p. 111928, 2022.

[4] F. Goerlandt, H. Goite, O. A. Valdez Banda, A. Höglund, P. Ahonen-Rainio, and M. Lensu, "An analysis of wintertime navigational accidents in the northern baltic sea," *Safety Science*, vol. 92, p. 66–84, 2017.

[5] P. Kujala, M. Hänninen, T. Arola, and J. Ylitalo, "Analysis of the marine traffic safety in the gulf of finland," *Reliability Engineering and System Safety*, vol. 94, no. 8, p. 1349–1357, 2009.

[6] X. Qu, Q. Meng, and L. Suyi, "Ship collision risk assessment for the singapore strait," *Accident Analysis and Prevention*, vol. 43, no. 6, p. 2030–2036, 2011.

[7] R. Zhen, Z. Shi, J. Liu, and Z. Shao, "A novel arena-based regional collision risk assessment method of multi-ship encounter situation in complex waters," *Ocean Engineering*, vol. 246, p. 110531, 2022.

[8] M. Oruc and Y. Altan, "Risky maritime encounter prediction via ensemble machine learning," *Trends in Maritime Technology and Engineering Volume 2*, pp. 255–263, 2022.

[9] M. M. Mostafa, "Forecasting the suez canal traffic: A neural network analysis," *Maritime Policy and Management*, vol. 31, no. 2, p. 139–156, 2004.

[10] X. Wu, A. L. Mehta, V. A. Zaloom, and B. N. Craig, "Analysis of waterway transportation in southeast texas waterway based on ais data," *Ocean Engineering*, vol. 121, p. 196–209, 2016.

[11] "Annual overview of marine casualties and incidents 2018," 2018.

[12] E. N. Otay, E. N. Otay, and Özkan, "Stochastic prediction of maritime accidents in the strait of istanbul," in *Proceedings Of The 3rd International Conference On Oil Spills In The Mediterranean And Black Sea Regions*, pp. 92–104, 2003.

[13] A. Mazaheri, J. Montewka, and P. Kujala, "Modeling the risk of ship grounding—a literature review from a risk management perspective," *WMU Journal of Maritime Affairs*, vol. 13, no. 2, p. 269–297, 2014.

[14] F. Goerlandt and P. Kujala, "On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk," *Safety Science*, vol. 62, p. 348–365, 2014.

[15] J. Mazurek, L. Lu, P. Krata, J. Montewka, H. Krata, and P. Kujala, "An updated method identifying collision-prone locations for ships. a case study for oil tankers navigating in the gulf of finland," *Reliability Engineering and System Safety*, vol. 217, p. 108024, 2022.

[16] H. Rong, A. P. Teixeira, and C. Guedes Soares, "Ship collision avoidance behaviour recognition and analysis based on ais data," *Ocean Engineering*, vol. 245, p. 110479, 2022.

[17] M. Zhang, D. Zhang, S. Fu, P. Kujala, and S. Hirdaris, "A predictive analytics method for maritime traffic flow complexity estimation in inland waterways," *Reliability Engineering System Safety*, vol. 220, p. 108317, 2022.

[18] Istanbul Sehir Hatlari, "Sehir Hatlari Official Data, 2023," 2023.

[19] M. Cai, X. Liu, Q. Xu, J. Zhang, and Y. Fang, "Collision risk analysis on ferry ships in jiangsu section of the yangtze river based on ais data," *Reliability Engineering & System Safety*, vol. 215, p. 107901, 2021.

[20] K. Zheng, Y. Jiang, S. Zhou, and Y. Xue, "A comprehensive spatiotemporal metric for ship collision risk assessment," *Ocean Engineering*, vol. 265, p. 112446, 2022.

[21] Z. H. Qureshi, "A review of accident modelling approaches for complex critical sociotechnical systems release limitation approved for public release," 2008.

[22] L. Kang, Q. Meng, and Q. Liu, "Fundamental diagram of ship traffic in the singapore strait," *Ocean Engineering*, vol. 147, no. September 2017, p. 340–354, 2018.

[23] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G. B. Huang, "Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, p. 1559–1582, 2018.

[24] P. Chen, Y. Huang, J. Mou, and P. H. A. J. M. Van Gelder, "Probabilistic risk analysis for ship-ship collision: State-of-the-art," *Safety Science*, vol. 117, no. April, p. 108–122, 2019.

[25] R. J. Bye and A. L. Aalberg, "Maritime navigation accidents and risk indicators: An exploratory statistical analysis using ais data and accident reports," *Reliability Engineering and System Safety*, vol. 176, p. 174–186, 2018.

[26] L. Du, F. Goerlandt, and P. Kujala, "Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data, year=2020, volume=200, number=February, pages=106933, doi=10.1016/j.ress.2020.106933," *Reliability Engineering System Safety*.

[27] M. Hänninen and P. Kujala, "Bayesian network modeling of port state control inspection findings and ship accident involvement," *Expert Systems with Applications*, vol. 41, no. 4 PART 2, p. 1632–1646, 2014.

[28] W. Zhang, F. Goerlandt, P. Kujala, and Y. Wang, "An advanced method for detecting possible near miss ship collisions from ais data," *Ocean Engineering*, vol. 124, pp. 141–156, 2016.

[29] P. R. Lei, T. H. Tsai, Y. T. Wen, and W. C. Peng, "Conflictfinder: Mining maritime traffic conflict from massive ship trajectories," in *Proceedings - 18th IEEE International Conference on Mobile Data Management, MDM 2017*, p. 356–357, 2017.

[30] A. Rawson and M. Brito, "A critique of the use of domain analysis for spatial collision risk assessment," *Ocean Engineering*, vol. 219, p. 108259, 2021.

[31] A. Debnath and H. C. Chin, "Analysis of marine conflicts," in *Proceedings of the 19th KKCNN Symposium On Civil Engineering*, 2006.

[32] S. L. Yoo, "Near-miss density map for safe navigation of ships," *Ocean Engineering*, vol. 163, p. 15–21, 2018.

[33] W. Zhang, X. Feng, F. Goerlandt, and Q. Liu, "Towards a convolutional neural network model for classifying regional ship collision risk levels for waterway risk analysis," *Reliability Engineering and System Safety*, vol. 204, p. 107127, 2020.

[34] T. Watawana and A. Caldera, "Analyse near collision situations of ships using automatic identification system dataset," in *5th International Conference on Soft Computing and Machine Intelligence, ISCMI 2018*, p. 155–162, Institute of Electrical and Electronics Engineers Inc., 2018.

[35] Z. Su, C. Wu, Y. Xiao, and H. He, "Study on the prediction model of accidents and incidents of cruise ship operation based on machine learning," *Ocean Engineering*, vol. 260, p. 111954, 2022.

[36] R. Szlapczynski and J. Szlapczynska, "A ship domain-based model of collision risk for near-miss detection and collision alert systems," *Reliability Engineering and System Safety*, vol. 214, p. 107766, 2021.

[37] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, p. 687–719, 2009.

[38] L. Du, O. A. Valdez Banda, F. Goerlandt, Y. Huang, and P. Kujala, "A colreg-compliant ship collision alert system for stand-on vessels," *Ocean Engineering*, vol. 218, p. 107866, 2020.

[39] H. Nordkvist, "An advanced method for detecting exceptional vessel encounters in open waters from high resolution ais data," 2018.

[40] X. L. Yuan, D. Zhang, J. F. Zhang, M. Y. Zhang, and C. Guedes Soares, "A novel collision risk awareness framework for ships in real-time operating conditions," in *Developments in the Collision and Grounding of Ships and Offshore Structures - Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures, ICCGS 2019*, p. 337–343, 2020.

[41] A. K. Debnath, H. C. Chin, and M. M. Haque, "Modelling port water collision risk using traffic conflicts," *Journal of Navigation*, vol. 64, no. 4, p. 645–655, 2011.

[42] A. Debnath and H. C. Chin, "Analysis of marine conflicts," in *Proceedings of the 19th KKCNN Symposium On Civil Engineering*, 2006.

[43] A. Hörteborn, J. W. Ringsberg, and M. Svanberg, "A comparison of two definitions of ship domain for analysing near ship–ship collisions," in *Developments in the Collision and Grounding of Ships and Offshore Structures - Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures, ICCGS 2019*, p. 308–316, 2020.

[44] W. Zhang, F. Goerlandt, J. Montewka, and P. Kujala, "A method for detecting possible near miss ship collisions from ais data," *Ocean Engineering*, vol. 107, pp. 60–69, 2015.

[45] W. Zhang, X. Feng, Y. Qi, F. Shu, Y. Zhang, and Y. Wang, "Towards a model of regional vessel near-miss collision risk assessment for open waters based on ais data," *The Journal of Navigation*, vol. 72, pp. 1449–1468, 2019.

[46] J. Liu, G. Y. Shi, and K. G. Zhu, "A novel ship collision risk evaluation algorithm based on the maximum interval of two ship domains and the violation degree of two ship domains," *Ocean Engineering*, vol. 255, p. 111431, 2022.

[47] S. Wang, Y. Zhang, R. Huo, and W. Mao, "A real-time ship collision risk perception model derived from domain-based approach parameters," *Ocean Engineering*, vol. 265, p. 112554, 2022.

[48] L. Lu, "Evaluation of selected maritime waterway risk models in light of their potential extension to winter navigation," 2016.

[49] A. K. Debnath and H. C. Chin, "Modelling collision potentials in port anchorages: Application of the navigational traffic conflict technique (ntct)," *Journal of Navigation*, vol. 69, no. 1, p. 183–196, 2016.

[50] S. Li, J. Zhou, and Y. Zhang, "Research of vessel traffic safety in ship routeing precautionary areas based on navigational traffic conflict technique," *Journal of Navigation*, vol. 68, no. 3, p. 589–601, 2015.

[51] E. V. I.-P. o. i. w. o. next and u. 2012, "Detection of hazardous encounters at the north sea from ais data," 2012.

[52] E. van Iperen, "Classifying ship encounters to monitor traffic safety on the north sea from ais data," *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 9, no. 1, p. 51–58, 2015.

[53] Z. Fang, H. Yu, R. Ke, S. L. Shaw, and G. Peng, "Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, p. 534–543, 2019.

[54] W. Zhang, C. Kopca, J. Tang, D. Ma, and Y. Wang, "A systematic approach for collision risk analysis based on ais data," *Journal of Navigation*, vol. 70, pp. 1117–1132, 2017.

[55] P. Chen, Y. Huang, J. Mou, and P. H. A. J. M. van Gelder, "Ship collision candidate detection method: A velocity obstacle approach," *Ocean Engineering*, vol. 170, p. 186–198, 2018.

[56] A. K. Debnath and H. C. Chin, "Navigational traffic conflict technique: A proactive approach to quantitative measurement of collision risks in port waters," *Journal of Navigation*, vol. 63, no. 1, p. 137–152, 2010.

[57] M. F. Oruc and Y. C. Altan, "Risky maritime encounter patterns via clustering," *Journal of Marine Science and Engineering*, vol. 11, no. 5, p. 950, 2023.

[58] J. Weng, S. Liao, B. Wu, and D. Yang, "Exploring effects of ship traffic characteristics and environmental conditions on ship collision frequency," *Maritime Policy & Management*, vol. 47, no. 4, pp. 523–543, 2020.

[59] K. Liu, Z. Yuan, X. Xin, J. Zhang, and W. Wang, "Conflict detection method based on dynamic ship domain model for visualization of collision risk hotspots," *Ocean Engineering*, vol. 242, p. 110143, 2021.

[60] H. Feng, M. Grifoll, Z. Yang, and P. Zheng, "Collision risk assessment for ships' routeing waters: An information entropy approach with automatic identification system (ais) data," *Ocean Coastal Management*, vol. 224, p. 106184, 2022.

[61] Y. Zhou, W. Daamen, T. Vellinga, and S. P. Hoogendoorn, "Ship classification based on ship behavior clustering from ais data," *Ocean Engineering*, vol. 175, pp. 176–187, 2019.

[62] J. Wang, C. Zhu, Y. Zhou, and W. Zhang, "Vessel spatio-temporal knowledge discovery with ais trajectories using co-clustering," *The Journal of Navigation*, vol. 70, no. 6, pp. 1383–1400, 2017.

[63] D. Zhang, Y. Zhang, and C. Zhang, "Data mining approach for automatic ship-route design for coastal seas using ais trajectory clustering analysis," *Ocean Engineering*, vol. 236, p. 109535, 2021.

[64] M. Mieczyńska and I. Czarnowski, "K-means clustering for sat-ais data analysis," *WMU Journal of Maritime Affairs 2021 20:3*, vol. 20, no. 3, pp. 377–400, 2021.

[65] J. Park and J. S. Jeong, "An estimation of ship collision risk based on relevance vector machine," *Journal of Marine Science and Engineering*, vol. 9, no. 5, p. 538, 2021.

[66] A. Rawson and M. Brito, "A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis," *Reliability Engineering & System Safety*, vol. 214, pp. 108–130, 2022.

[67] G. K. D. De Vries and M. Van Someren, "Machine learning for vessel trajectories using compression, alignments and domain knowledge," *Expert Systems with Applications*, vol. 39, no. 18, p. 13426–13439, 2012.

[68] B. Ożoga and J. Montewka, "Towards a decision support system for maritime navigation on heavily trafficked basins," *Ocean Engineering*, vol. 159, p. 88–97, 2018.

[69] Y. Fujii and K. Tanaka, "Traffic capacity," *Journal of Navigation*, vol. 24, no. 4, pp. 543–552, 1971.

[70] J. Weng, Q. Meng, and X. Qu, "Vessel collision frequency estimation in the singapore strait," *Journal of Navigation*, vol. 65, no. 2, pp. 207–221, 2012.

[71] M. G. Hansen, T. K. Jensen, T. Lehn-Schioler, K. Melchild, F. M. Rasmussen, and F. Ennemark, "Empirical ship domain based on ais data," *Journal of Navigation*, vol. 66, no. 6, pp. 931–940, 2013.

[72] Y. Altan and B. M. Meijers, "Ship domain variations in the strait of istanbul," in *WCTRS SIGA2 Conference*, (Antwerp, Belgium), 2021.

[73] T. Degré and X. Lefévre, "A collision avoidance system," *Journal of Navigation*, vol. 34, no. 2, pp. 294–302, 1981.

[74] A. S. Lenart, "Collision threat parameters for a new radar display and plot technique," *Journal of Navigation*, vol. 36, no. 3, pp. 404–410, 1983.

[75] Y. Kuwata, M. T. Wolf, D. Zarzhitsky, and T. L. Huntsberger, "Safe maritime navigation with colregs using velocity obstacles," pp. 4728–4734, 2011.

[76] Y. C. Altan and E. N. Otay, "Maritime traffic analysis of the strait of istanbul based on ais data," *Journal of Navigation*, 2017.

[77] M. F. Oruc and Y. C. Altan, "Predicting the risky encounters without distance knowledge between the ships via machine learning algorithms," *Expert Systems with Applications*, vol. 221, p. 119728, 2023.

[78] H. Rong, A. P. Teixeira, and C. Guedes Soares, "Maritime traffic probabilistic prediction based on ship motion pattern extraction," *Reliability Engineering & System Safety*, vol. 217, p. 108061, 2022.

[79] X. Chen *et al.*, "Automatic identification system (ais) data supported ship trajectory prediction and analysis via a deep learning model," *Journal of Marine Science and Engineering*, vol. 10, no. 9, p. 1314, 2022.

[80] H. Duan *et al.*, "A semi-supervised deep learning approach for vessel trajectory classification based on ais data," *Ocean & Coastal Management*, vol. 218, p. 106015, 2022.

[81] A. Tritsarolis *et al.*, "Vessel collision risk assessment using ais data: A machine learning approach," in *Proceedings - IEEE International Conference on Mobile Data Management*, vol. 2022-June, pp. 425–430, IEEE, 2022.

[82] C. H. Yang *et al.*, "Deep learning for vessel trajectory prediction using clustered ais data," *Mathematics*, vol. 10, no. 16, p. 2936, 2022.

[83] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[84] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[85] H. B. Usluer, A. G. Bora, and C. Gazioglu, "What if the independenta or nassia tanker accidents had happened in the strait of canakkale (dardanelle)?," *Ocean Engineering*, vol. 260, p. 111712, 2022.

[86] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ch. Overview of Supervised Learning, p. 9–41. Springer, 2009.

[87] J. Montewka, T. Hinz, P. Kujala, and J. Matusiak, "Probability modelling of vessel collisions," *Reliability Engineering and System Safety*, vol. 95, no. 5, p. 573–589, 2010.

[88] J. M. Mou, C. v. d. Tak, and H. Ligteringen, "Study on collision avoidance in busy waterways by using ais data," *Ocean Engineering*, vol. 37, no. 5–6, p. 483–490, 2010.

[89] R. Szlapczynski and J. Szlapczynska, "Review of ship safety domains: Models and applications," *Ocean Engineering*, vol. 145, p. 277–289, 2017.

[90] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 7. Springer, 2000.

[91] H. B. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, no. 3, p. 295–311, 1989.

[92] E. Alpaydin, *Introduction to Machine Learning*. MIT press, 2020.

[93] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Applied Statistics*, vol. 28, no. 1, p. 100, 1979.

[94] R. L. Thorndike, "Who belongs in the family?," *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

[95] R. M. Bittmann and R. M. Gelbard, "Decision-making method using a visual approach for cluster analysis problems; indicative classification algorithms and grouping scope," *Expert Systems*, vol. 24, no. 3, pp. 171–187, 2007.

[96] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[97] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 1, p. 30–42, 2012.

[98] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine Learning*, vol. 30, no. 2-3, p. 195–215, 1998.

[99] T. Fawcett and F. Provost, "Adaptive fraud detection," *Data Mining and Knowledge Discovery*, vol. 1, no. 3, p. 291–316, 1997.

[100] P. Riddle, R. Segal, and O. Etzioni, "Representation design and brute-force induction in a boeing manufacturing; domain," *Applied Artificial Intelligence*, vol. 8, no. 1, p. 125–147, 1994.

[101] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: Improving classification performance when training data is skewed," in *Proceedings - International Conference on Pattern Recognition*, Institute of Electrical and Electronics Engineers Inc., 2008.

[102] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 2, p. 539–550, 2009.

[103] A. Rahman and B. Verma, "Cluster-based ensemble of classifiers," *Expert Systems*, vol. 30, no. 3, p. 270–282, 2013.

[104] B. Krawczyk, M. Woźniak, and B. Cyganek, "Clustering-based ensembles for one-class classification," *Information Sciences*, vol. 264, p. 182–195, 2014.

[105] J. Lee and J. H. Lee, "K-means clustering based svm ensemble methods for imbalanced data problem," in *Joint 7th International Conference on Soft Computing and Intelligent Systems, SCIS and 15th International Symposium on Advanced Intelligent Systems, ISIS*, p. 614–617, Institute of Electrical and Electronics Engineers Inc., 2014.

[106] Z. H. Zhou and M. L. Zhang, "Solving multi-instance problems with classifier ensemble based on constructive clustering," *Knowledge and Information Systems*, vol. 11, pp. 155–170, 2007.

[107] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, p. 119–139, 1997.

[108] W. Lu, Z. Li, and J. Chu, "Adaptive ensemble undersampling-boost: A novel learning framework for imbalanced data," *Journal of Systems and Software*, vol. 132, p. 272–282, 2017.

[109] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, p. 861–874, 2006.

[110] F. P. Fabian, V. Michel, O. G. Oliviergrisel, M. Blondel, P. Prettenhofer, R. Weiss, *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 85, p. 2825–2830, 2011.

VITA

M. Furkan Oruc received his Bachelor's of Science in Civil Engineering from Boğaziçi University, İstanbul in 2020. He received a Master of Science degree in Statistics with Computer Science at Georgia State University, Atlanta, Georgia, United States of America in 2022. He received his Master of Science degree in Civil Engineering from Özyeğin University, İstanbul, Türkiye in 2023.