Sveučilište
u Splitu
University
of Split

FACULTY OF MARITIME STUDIES

Zlatko Boko

INTEGRATING MACHINE LEARNING AND MULTI-CRITERIA ANALYSIS FOR MODELLING SAFETY RISK OF OFFSHORE VESSELS DURING PSC INSPECTION

DOCTORAL DISSERTATION

Split, 2026

Sveučilište u Splitu
University of Split

FACULTY OF MARITIME STUDIES

Zlatko Boko

INTEGRATING MACHINE LEARNING AND MULTI-CRITERIA ANALYSIS FOR MODELLING SAFETY RISK OF OFFSHORE VESSELS DURING PSC INSPECTION

DOCTORAL DISSERTATION

Supervisor: Ivica Skoko, PhD

Split, 2026

IMPRESSUM

This doctoral dissertation was submitted to the University of Split, Faculty of Maritime Studies, in partial fulfilment of the requirements for the academic degree of Doctor of Science (PhD) in the scientific field of technical sciences, the scientific field of Traffic and Transport Technology.

The dissertation has been written in British English.

Supervisor: Assistant Professor Ivica Skoko, PhD

University of Split, Faculty of Maritime Studies, Croatia

COMMITTEES FOR EVALUATION AND DEFENCE OF THE DOCTORAL DISSERTATION

Doctoral Dissertation Evaluation Committee:

1. Chair: Assistant Professor Zaloa Sanchez Varela, PhD, University of Split, Faculty of Maritime Studies, Croatia
2. Member: Associate Professor Dario Medić, PhD, University of Split, Faculty of Maritime Studies, Croatia
3. Member: Associate Professor David Boullosa Falces, PhD, Department of Energy Engineering, Bilbao School of Engineering, University of the Basque Country (UPV/EHU), Spain

Doctoral Dissertation Defence Committee:

1. Chair: Assistant Professor Zaloa Sanchez Varela, PhD, University of Split, Faculty of Maritime Studies, Croatia
2. Member: Associate Professor Dario Medić, PhD, University of Split, Faculty of Maritime Studies, Croatia
3. Member: Associate Professor David Boullosa Falces, PhD, Department of Energy Engineering, Bilbao School of Engineering, University of the Basque Country (UPV/EHU), Spain

ABSTRACT

Port State Control (PSC) inspections, conducted by national authorities on foreign ships in their ports to ensure compliance with international regulations, serve as a frontline defence for maritime safety. However, limited resources force authorities to rely on risk-based prioritisation systems and methods that target vessels for inspection using a variety of indicators. These systems, built on broad indicators and legacy scoring (historical evaluation methods using established factors), often fall short for offshore support vessels (OSVs), a class of ships that provide operational support to offshore installations and have unique operational and technical profiles that set them apart from standard merchant ships. OSVs remain underrepresented in existing risk models. Meanwhile, emerging machine learning (ML) algorithms that automatically learn patterns from data to make predictions offer sharper predictions, yet spark debate among regulators about their transparency and trustworthiness. This doctoral research confronts these challenges by designing a hybrid framework that blends ML with multi-criteria decision making (MCDM), an approach that considers multiple factors for ranking or selection, to sharpen the prioritisation of PSC inspections for OSVs. The aim is to boost the accuracy of detention risk predictions (the likelihood a vessel will be held in port due to deficiencies) while keeping the process transparent and traceable for regulators. Drawing on inspection and vessel data from the European Quality Shipping Information System (EQUASIS), a database that contains information on ships, companies, and inspections, the study treats detention outcomes as key signals of elevated safety risk. A suite of ML models, including decision trees (tree-like models for decision-making), ensemble algorithms (combinations of models for improved accuracy), and neural networks (systems modelled on the human brain to recognise patterns), is trained and tested to forecast detention risk using carefully grouped historical inspection data that reflect the information available to inspectors before boarding. The models are assessed using metrics tailored to imbalanced datasets, where one outcome (such as detention) is much rarer than another. Explainable artificial intelligence (AI) techniques, tools that interpret and clarify how AI models reach their conclusions, spotlight the most influential features and demystify the decision process. To translate predictions into actionable inspection priorities, the risk scores are used in a Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) framework, which ranks options based on their similarity to an ideal scenario, generating clear, structured vessel risk rankings. The findings reveal that ML models outperform traditional approaches in predicting detentions. Insights into the models' decision-making show that their logic closely aligns with established inspection practices. The integrated ML and MCDM framework strikes a strong balance between precision and clarity, empowering inspectors to prioritise effectively while retaining their professional judgment. Sensitivity along with robustness tests, which examine how results change when assumptions or parameter values are varied, confirm the framework's reliability, even as model assumptions or criteria weights shift within reasonable bounds. This research advances maritime safety by introducing one of the first integrated ML and MCDM tools tailored for PSC inspections of OSVs. In real-world use, the framework equips regulators with a flexible, transparent, and compatible decision-support system. The study acknowledges several limitations, including the scope of the data, model generalizability (how well findings apply to data beyond the study sample), and potential inspection bias (systematic deviations in findings due to subjective factors). It also charts a path for future research, such as expanding to more vessel types, incorporating diverse data sources, and exploring the interplay between human judgment and system recommendations in inspection decisions.

Keywords: Port State Control; offshore support vessels; inspection prioritisation; machine learning; explainable artificial intelligence; multi-criteria decision-making; TOPSIS; detention risk prediction;

LIST OF PUBLISHED PAPERS

This doctoral thesis is based on the compilation of the following published papers:

Paper I

Boko, Z., Skoko, I., Sanchez-Varela, Z., & Pincetic, T. (2024). Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making. Journal of Marine Science and Engineering, 12(11), 1905. https://doi.org/10.3390/jmse12111905

Paper II

Boko, Z., Skoko, I., Sanchez Varela, Z., & Milin, V. (2025). Advancing Maritime Safety: A Literature Review on Machine Learning and Multi-Criteria Analysis in PSC Inspections. Journal of Marine Science and Engineering, 13(5), 974. https://doi.org/10.3390/jmse13050974

Paper III

Boko, Z., Stanivuk, T., Radanović, N., & Skoko, I. (2025). Machine Learning-Driven Prediction of Offshore Vessel Detention: The Role of Neural Networks in Port State Control. Journal of Marine Science and Engineering, 13(3), 472. https://doi.org/10.3390/jmse13030472

The papers are reproduced in Appendices A – C and constitute the scientific basis of this doctoral dissertation.

CONTENTS

1. INTRODUCTION

1.1. Background and Research Context

Offshore Support Vessels (OSVs) are critical assets in the offshore energy sector, facilitating operations that demand high technical reliability, operational flexibility, and ongoing coordination between ship systems and crew. Their operational scope encompasses frequent manoeuvring near offshore installations, dynamic positioning (DP), towing, subsea support, and emergency response. These tasks impose sustained and often concurrent demands on propulsion systems, deck machinery, control systems, and human operators, resulting in safety conditions distinct from those on conventional merchant vessels. As offshore activities extend into deeper waters and more complex environments, the operational requirements for OSVs continue to intensify (Hetherington et al., 2006; Schroeder et al., 2015; Lai et al., 2023). This trend has been associated with increasingly complex risk profiles that challenge traditional inspection and safety-assessment approaches (Schröder-Hinrichs et al., 2013).

Concurrently, maritime safety oversight has witnessed considerable evolution due to increasing digitalisation. Inspection histories, vessel particulars, deficiency records, and company-level performance indicators are now accessible via public and semi-public information systems.

Platforms such as the European Quality Shipping Information System (EQUASIS) offer visibility into inspection outcomes and vessel characteristics that were previously inaccessible across geographical areas and inspection regimes (Knapp, 2006; Cariou & Mejia, 2017). These advancements facilitate more in-depth analyses of safety performance over time and across vessel segments.

Despite the increased availability of data, a structural imbalance has emerged. Although inspection datasets are now more comprehensive, the analytical tools used to interpret them have largely remained static. Many safety risk assessment methods continue to employ simplified scoring rules or categorical indicators, treating technical, operational, and organisational factors as independent contributors to risk. Initial exploratory analyses conducted for this doctoral research indicated that inspection outcomes for OSVs are frequently influenced by interactions among multiple variables rather than by isolated deficiencies (Knapp & Franses, 2007).

This observation served as the foundation for the present dissertation. The main challenge identified was not insufficient information but the lack of analytical structures capable of decoding complex inspection data into assessments that accurately portray the operational circumstances of OSVs and remain interpretable and practical within inspection-based safety systems.

The diagram showed in Figure 1.1. illustrates the operational environment for OSVs, highlighting how technical systems, operational tasks, and human operators interact simultaneously. Tasks such as dynamic positioning, cargo handling, towing, and working with offshore installations create tightly coupled working conditions, with many subsystems working together. These overlapping demands create safety challenges distinct from those on regular merchant vessels, adding to the complex and specific risks discussed in this dissertation.

1.2. Problem Statement and Research Motivation

Inspection regimes are essential for maritime safety assurance; however, the analytical foundations for inspection prioritisation were initially developed for vessel types with stable, homogeneous operational profiles. Traditional assessment tools prioritise transparency and administrative simplicity, yet they typically assume that risk factors act independently and in a

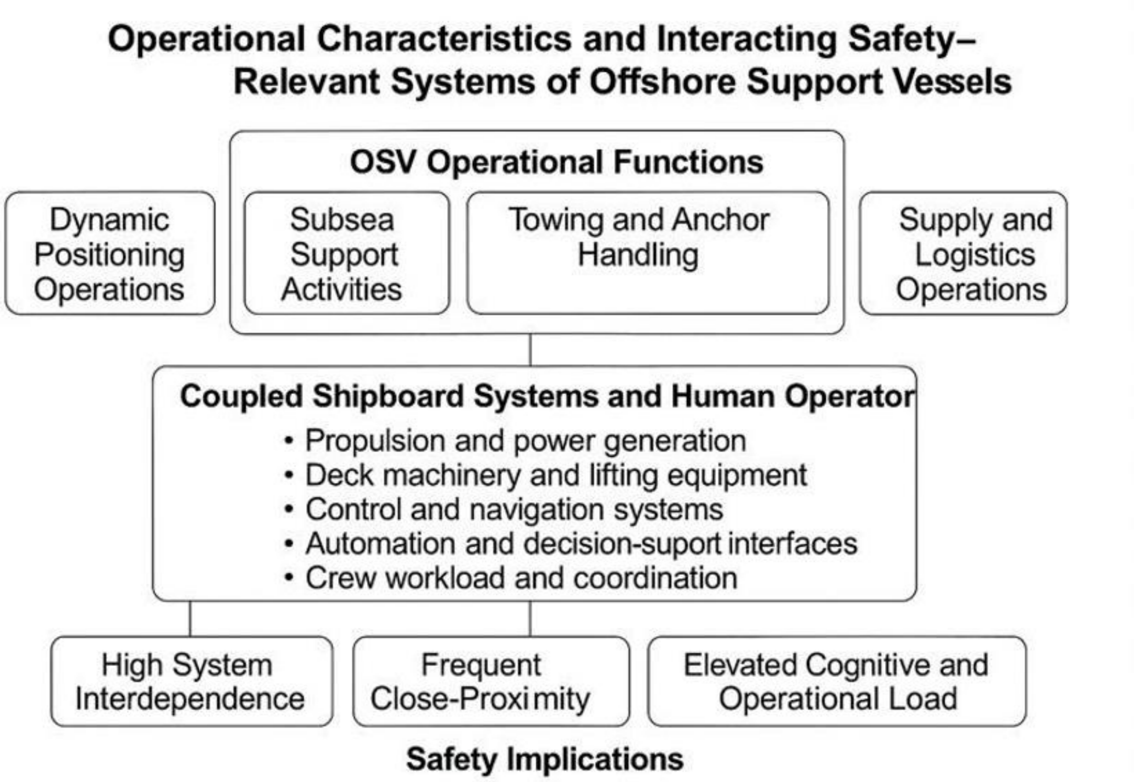

Figure 1.1. Different operational characteristics and interactions which can be found during operations carried out by OSVs, the author's illustration is based on Hetherington et al. (2006), Schroeder et al. (2015), and Lai et al. (2023)

linear manner. In contrast, safety outcomes for OSVs often result from interacting, context-dependent conditions, including simultaneous machinery loads, high task frequency, specialised equipment use, and complex human–machine interactions.

Inspection records contain patterns and operational signals that conventional assessment tools often overlook (Aven,2016). Examples include recurring clusters of deficiencies within specific OSV subtypes and correlations between operational roles and safety performance.

Machine Learning (ML) techniques can identify these structures by modelling complex, nonlinear relationships directly from data (Celik et al., 2021; Zhang et al., 2022; Lalla-Ruiz et al., 2018). However, applying ML in regulatory environments poses significant challenges. High-performing models frequently lack transparency, which complicates their acceptance in settings where inspection decisions must be explainable and defensible (Doshi-Velez & Kim, 2017; Guidotti et al., 2019).

Explainable artificial intelligence (XAI) methods, such as Shapley Additive Explanations (SHAP), provide mechanisms to interpret complex models by quantifying the contribution of individual features to model outputs (Lundberg & Lee, 2017; Xu et al., 2021). Nevertheless, interpretability alone does not address the broader challenges faced by inspection authorities. Effective inspection planning requires balancing statistical evidence, operational context, potential safety consequences, and professional judgement.

Multi-Criteria Decision-Making (MCDM) methods provide a structured approach for integrating these diverse dimensions. Techniques such as the Analytic Hierarchy Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) have been utilised in maritime risk contexts to support transparent and repeatable prioritisation (Saaty, 1980; Celik, 2009; Zavadskas et al., 2016). Preliminary stages of this doctoral research explored combining ML-based risk estimates with ranking procedures for OSVs, demonstrating the potential of linking predictive analytics with decision-structuring methods. These efforts also revealed the limitations of ad hoc integration.

Collectively, these findings reveal a disconnect between the analytical tools currently available and the integrated reasoning required in inspection environments. Addressing this gap, particularly for a vessel segment as operationally specialised as OSVs, became the central motivation for the present research. Figure 1.2 illustrates this segregation by contrasting the characteristics of available inspection data, the limitations of conventional assessment tools, and the requirements of inspection-oriented decision-making for OSVs.

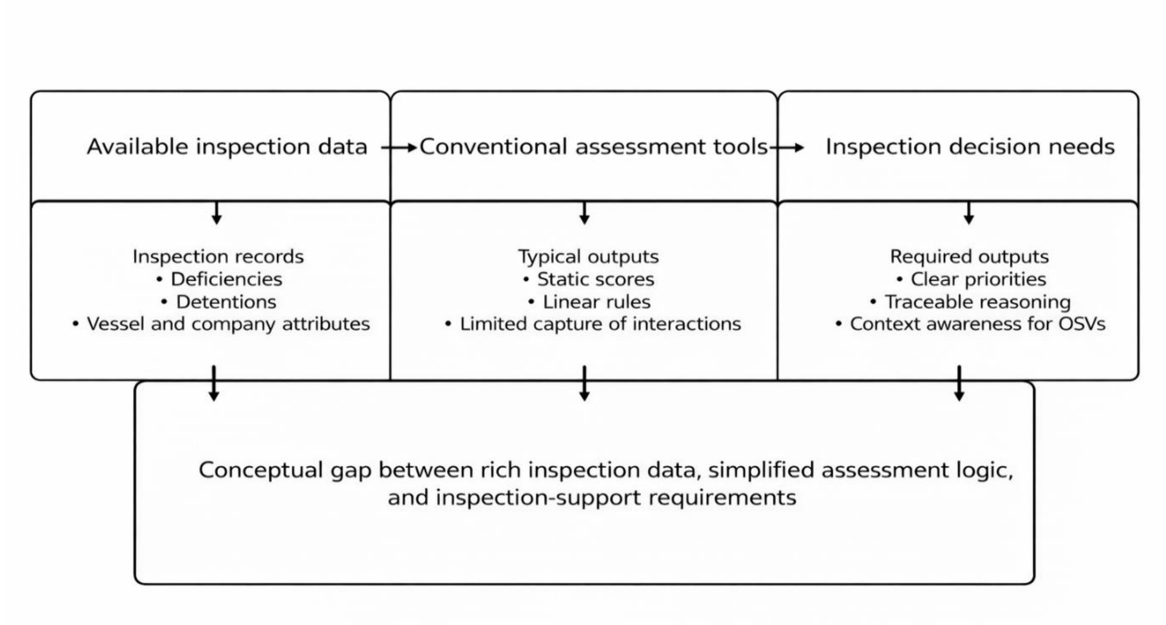

Figure 1.2. Research gap in the present methodologies in PSC inspections

1.3. Research Objectives and Hypotheses

This dissertation pursues the following objectives:

- To analyse inspection patterns of OSVs and identify risk-relevant characteristics specific to this vessel segment.
- To develop and compare ML models for predicting OSVs detention outcomes in the PSC context.
- To apply interpretable AI techniques to clarify which factors most strongly influence ML predictions.
- To incorporate predictive insights into a structured MCDM, aligning analytical outputs with inspection-priority needs.
- To evaluate the hybrid ML–XAI–MCDM framework in terms of predictive accuracy, interpretability, and practical suitability.

Research Hypotheses:

Based on the research motivation developed in this chapter, this dissertation is guided by the following hypotheses:

H1: Inspection patterns for OSV contain sufficient structure to enable the development of reliable learning-based models for predicting PSC detention outcomes.

H2: Explainable ML techniques can identify, quantify, and interpret the most influential factors associated with OSV safety performance in inspection datasets.

H3: A hybrid framework combining ML, XAI, and MCDM produces inspection-priority rankings that are more coherent, transparent, and operationally meaningful than those derived from deterministic scoring methods.

1.4. Structure of the Dissertation

The remainder of the dissertation is organised as follows. Chapter 2 reviews the relevant literature on the characteristics of OSVs; the implementation of AI techniques in scientific research on PSC and in general safety culture in the maritime industry; inspection data; ML applications; decision-analytic methods; XAI; and hybrid modelling approaches. Chapter 3 presents the research methodology, including data preparation, model development, interpretability techniques, and multi-criteria integration. Chapter 4 reports the empirical results of the predictive models and the hybrid decision-support framework. Chapter 5 discusses the research's scientific and practical contributions, while Chapter 6 outlines its limitations and directions for future work. Chapter 7 concludes by synthesising the main findings and reflecting on their implications for maritime safety oversight. The References and Appendix sections close the dissertation.

2. LITERATURE REVIEW

2.1. Maritime Safety and the Evolution of PSC Regulation

Maritime safety governance has expanded significantly over the past several decades, driven by the goal of reducing and preventing accidents in the industry, rising environmental expectations, and the globalisation of shipping. PSC emerged as a central regulatory mechanism to address uneven flag-state performance and ensure compliance with international conventions. Beginning with the Paris Memorandum of Understanding (MoU) in 1982, regional PSC agreements created harmonised inspection procedures, standard deficiency codes, and shared information systems to improve oversight efficiency and reduce the prevalence of substandard ships (Knapp & Franses, 2007; Cariou & Wolff, 2011).

At its core, PSC serves as a corrective measure within a global regulatory landscape where encouragement for strict enforcement varies among flag states. By granting coastal states the authority to inspect foreign vessels calling at their ports, PSC compensates for variations in national regulatory strength and creates a multilayered enforcement structure that improves overall safety performance (Panagakos et al., 2017). Over time, this system has substantially influenced ship design, operational practices, and the strategic behaviour of shipping companies and flag administrations.

The introduction of the EQUASIS in 2000 marked a significant turning point in maritime transparency. Established as a publicly accessible information platform, EQUASIS aggregates PSC inspection results, detention records, classification society information, and flag-state performance indicators within a single system (European Maritime Safety Agency, 2023). By providing a consistent empirical basis for cross-regional comparison, EQUASIS has enabled a wide range of safety-related analyses, including studies of vessel risk, detention patterns, and inspection-regime performance (Knapp, 2006; Cariou & Mejia, 2017; Celik et al., 2021).

In most PSC regimes, inspection prioritisation is operationalised through risk-based targeting systems that aggregate vessel-related indicators into composite risk scores to guide inspection selection (Cariou et al., 2008; Knapp & Bijwaard, 2009). The limitation lies in the deterministic nature of most PSC targeting schemes. While these schemes offer transparency and administrative simplicity, they often lack predictive accuracy, especially as vessel types become more specialised and operational environments more complex (Goerlandt & Montewka, 2015).

The global trend toward digitalisation has raised expectations for data-driven decision-making in PSC, encouraging researchers to explore analytical techniques that can enhance inspection prioritisation. Several studies have shown that ML approaches can capture nonlinear patterns and deliver stronger predictive performance of potential detentions than deterministic rules (Celik et al., 2021; Zhang et al., 2022). Studies focusing on OSVs contributed initial perspectives on detention risk modelling and highlighted the potential of combining predictive analytics with structured decision-making tools (Boko et al., 2024; Boko et al., 2025a; Boko et al., 2025b). Those findings highlighted the potential value of deeper analytical integration in inspection-oriented safety assessment.

PSC, therefore, represents both a rich empirical domain and a challenging analytical environment. Its datasets capture real-world vessel behaviour and regulatory interactions. Yet their practical use requires methods that can handle regional inconsistencies, selection biases, and the need for transparent, interpretable decision support.

Figure 2.1 illustrates the number of scientific publications employing ML and MCA methods in maritime safety research from 2010 to 2025. The horizontal axis represents the years, while the vertical axis indicates the number of studies identified annually. The figure presents two lines: one corresponding to publications utilising ML and the other to those applying MCDA.

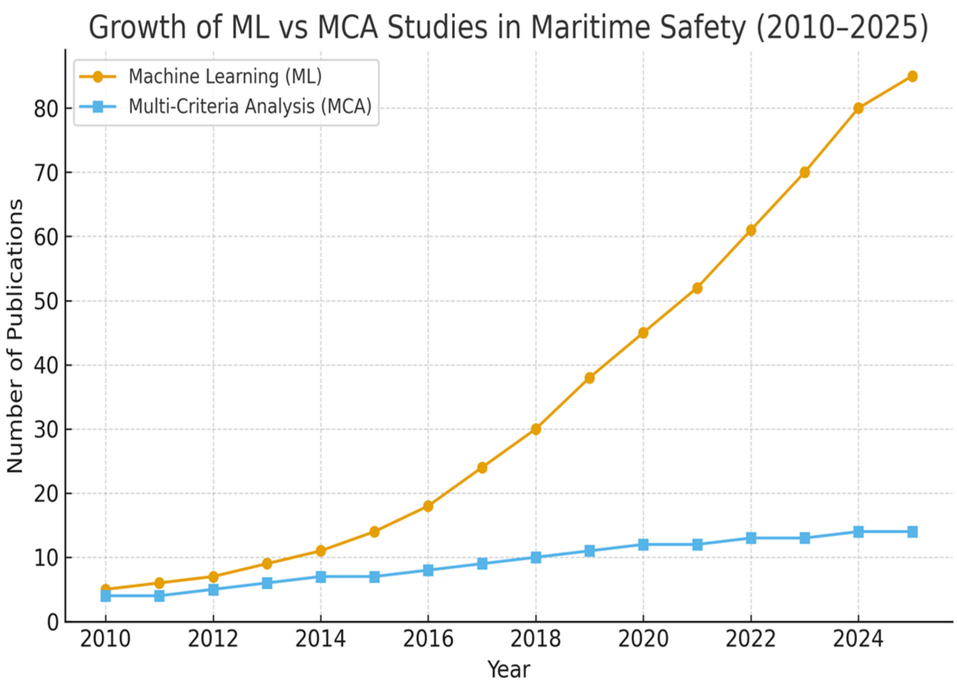

Figure 2.1. Evolution of ML and multi-criteria analysis (MCA) studies in maritime safety research (2010–2025)

This visualisation clearly demonstrates the evolving, increasing research activity in both domains – rapidly in ML - over time.

2.2. OSVs: Operational Patterns and Safety Challenges

OSVs occupy a distinctive position within the global fleet due to their specialised operational roles in offshore energy production. They serve deep-sea drilling units, undertake subsea construction, provide towage and anchor-handling support, and perform emergency and standby duties (International Maritime Organisation, 2014; Hetherington et al., 2006). The operational tempo of OSVs is intense. The technical configuration of this type of vessel further reinforces this operational complexity. Advanced propulsion systems, high-torque winches, DP control, firefighting and standby equipment, and specialised deck layouts require continuous monitoring and coordinated execution (International Maritime Organisation, 2009; Baldauf et al., 2011). Empirical studies consistently associate these technical pressures with elevated rates of deficiencies in navigation systems, propulsion machinery, fire protection arrangements, and International Safety Management (ISM)-related procedures (Celik et al., 2021; Zhang et al., 2022).

Human factors represent another critical dimension of OSV operations. Crews frequently operate under high cognitive load, managing simultaneous equipment operations, variable weather conditions, and interactions with offshore infrastructure. Research indicates that these environments increase the likelihood of procedural deviations, communication failures, and fatigue-related errors (Hetherington et al., 2006; Schroeder et al., 2015).

These human-factor challenges are evident in PSC inspection data, in which OSVs exhibit recurring deficiencies in operational readiness, navigational routines, and safety management practices.

Inspection exposure further amplifies these patterns. OSVs typically operate within regions covered by PSC memoranda of understanding (MoUs), such as the North Sea, Mediterranean, Gulf of Mexico, or Southeast Asia, and therefore undergo inspections more frequently than deep-sea merchant vessels. High inspection density increases data visibility but also intensifies selection bias. More frequent inspections result in more recorded deficiencies, regardless of whether the underlying safety performance differs from that of other ship types (Knapp, 2006; Cariou & Mejia, 2017).

Table 2.1. Thematic categorisation of academic studies related to PSC, maritime safety, ML, and MCA

Subject	No. of studies
PSC inspection in general	23
Maritime safety and PSC inspection	17
Application of MLMs in PSC	40
Application of MCDMs in PSC inspections	10
Application of combined ML and MCA methods in PSC inspection	6
MI in general	3

Despite their operational importance and distinct risk profile, OSVs remain comparatively underexamined in academic research on PSC. Many empirical PSC studies group OSVs within broad vessel categories, which obscures their specific operational challenges and safety patterns (Knapp & Franses, 2007; Celik et al., 2021). Recent analyses focusing specifically on OSVs have highlighted these limitations and shown that targeted modelling approaches reveal different risk drivers than those affecting general cargo or tanker fleets (Boko et al., 2024; Boko et al., 2025b).

These findings underscore the need for analytical tools that accurately represent the technical, operational, and organisational realities of OSV operations.

2.3. Data, PSC Records, and Analytical Approaches to Vessel Risk

PSC inspection data represents one of the most comprehensive sources available for analysing vessel safety in real operational contexts. Unlike voluntary reporting systems, PSC inspections generate independent, regulatory-verified observations of vessel condition. Studies across maritime economics, safety science, and transport regulation consistently position PSC and EQUASIS data as essential resources for examining patterns of compliance, operational failures, and structural risk across the global fleet (Cariou & Mejia, 2017; Celik et al., 2021).

As shown in Table 2.1, existing studies cluster around several recurring thematic domains, reflecting the diversity of analytical approaches within PSC-related research.

The use of PSC data for analytical modelling necessitates careful consideration of its structural characteristics. Inspection practices vary across regional MoUs, leading to significant variation in deficiency coding, inspection intensity, and reporting completeness (Knapp & Franses, 2007; Panagakos et al., 2017). For instance, the Paris MoU's historically rigid inspection procedures yield higher average deficiency rates than those of certain Asia-Pacific jurisdictions, demonstrating the influence of regional enforcement culture on recorded outcomes (Cariou & Wolff, 2011). Failure to account for these differences may lead to misinterpretation of regulatory behaviour as vessel safety performance in cross-regional comparisons.

PSC inspections are inherently non-random. Targeting regimes prioritise vessels considered higher risk based on factors such as age, flag performance, company history, and previous deficiencies (Cariou & Wolff, 2011). Empirical studies indicate that this selection process systematically biases recorded observations: vessels inspected more frequently accumulate additional deficiencies, not necessarily due to lower safety, but because of increased inspection frequency (Knapp, 2006; Cariou & Mejia, 2017). Recognising this selection bias is essential for accurate predictive modelling and interpretation of risk patterns.

Data granularity introduces a further challenge. Although deficiency codes are harmonised at the IMO level, regional variation in interpretation persists. Some MoUs record minor procedural irregularities more consistently, while others focus predominantly on technical and structural conditions. Human factors add another layer of variability: inspectors differ in training, experience, and inspection depth, which influences the consistency of recorded deficiencies (Hetherington et al., 2006; Schroeder et al., 2015). These considerations highlight the importance of treating PSC data as semi-structured regulatory information rather than a uniform dataset.

Temporal dynamics are also significant. Deficiencies accumulate throughout a vessel's operational life; however, their significance is influenced by inspection frequency, remediation actions, and evolving regulatory thresholds (Cariou & Mejia, 2017). Multiple studies report that clusters of inspections within short periods often signal unresolved operational issues, while extended intervals with clean inspections indicate stable compliance (Cariou & Mejia, 2017). Analyses of OSVs reveal similar trends, with inspection recency interacting with operational and managerial characteristics to affect both deficiency counts and detention probabilities (Boko et al., 2025b).

Although PSC data are fundamental to maritime risk analysis, they do not comprehensively capture organisational practices or cultural factors that affect safety. PSC assessments focus on observable conditions during port calls, rendering underlying management systems only partially visible. Researchers have sought to augment PSC data with company performance indicators and flag-state ratings, which serve as proxies for safety culture and the strength of regulatory oversight (Cariou & Wolff, 2011; Kontovas & Psaraftis, 2011). While these enhancements improve explanatory power, they still provide an incomplete depiction of organisational risk, especially for specialised vessel types such as OSVs.

Historically, analytical studies have primarily utilised linear statistical models such as logistic regression. These models provide interpretability and are effective at identifying broad risk factors, but they assume additive effects and are limited in their ability to capture the complex interactions that frequently influence inspection outcomes (Cariou & Mejia, 2017; Zhang et al., 2022). For example, the effect of vessel age may depend on the quality of company management, or groups of technical deficiencies may collectively increase detention risk in ways that linear models cannot readily capture. With the expansion of PSC datasets and increasing complexity of vessel operations, researchers have adopted ML techniques capable of modelling nonlinear relationships, high-dimensional feature interactions, and diverse operational patterns.

Tree-based ensemble models and neural networks (NN) have demonstrated significant improvements in predicting detentions, identifying deficiency clusters, and classifying vessel risk profiles (Breiman, 2001; Cazzulo et al., 2020; Wang et al., 2021).

The application of ML in maritime contexts has expanded to include anomaly detection, machinery condition monitoring, and navigational risk assessment (Goodfellow et al., 2016; Celik et al., 2021; Zhang et al., 2022).

In the PSC context, multiple studies have shown that ML models outperform deterministic targeting schemes by identifying subtle patterns and operational signals that traditional scoring methods overlook (Celik et al., 2021; Zhang et al., 2022).

Analyses focused on OSVs further support these results: ML-based models yield more accurate predictions of deficiency risk and detention probability, uncovering operational and technical dynamics not captured by PSC's static scoring systems (Boko et al., 2024; Boko et al., 2025b).

Despite their predictive advantages, ML models present challenges concerning interpretability and regulatory acceptance.

High-performing models frequently function as 'black boxes' (in AI terminology, a 'black box' model uses an ML algorithm to make predictions, but the reasons behind those predictions cannot be explained or traced), complicating efforts by inspectors and administrators to understand the basis of predictions (Doshi-Velez & Kim, 2017). Additionally, ML models may inadvertently learn patterns that reflect enforcement practices rather than actual vessel conditions, such as associating risk with particular MoUs due to variations in inspection rigour

(Knapp, 2006). These challenges underscore the necessity for modelling frameworks that integrate predictive accuracy with transparent explanation mechanisms.

Recent research on integrating XAI and structured multi-criteria reasoning has shown potential in addressing these challenges (Rudin, 2019; Zio & Aven, 2013; Dinis et al., 2023). Preliminary experiments indicate that connecting ML outputs with multi-criteria evaluation can facilitate more transparent decision-making in PSC contexts (Boko et al., 2025a).

Collectively, the literature demonstrates both the strengths and limitations of current analytical approaches. PSC data provides valuable insights but necessitates careful management of structural inconsistencies, selection bias, and temporal dynamics. Traditional statistical methods identify broad trends but fail to capture nonlinear interactions. ML models offer robust predictive capabilities but demand interpretability and contextualization for operational use. These challenges and opportunities underpin the methodological foundation for the hybrid framework developed in this dissertation, which seeks to integrate prediction, explanation, and structured decision-making into a unified system tailored to OSV inspection risk.

2.4. ML, XAI, and Hybrid Decision Systems in Maritime Safety

The digitalisation of maritime operations has accelerated the adoption of data-driven methods for safety assessment, anomaly detection, and operational optimisation. ML has gained prominence for its ability to model nonlinear patterns, interaction effects, and high-dimensional relationships that traditional risk models cannot capture (Breiman, 2001; Goodfellow et al., 2016). In maritime contexts, ML has been utilised for machinery failure prediction, collision avoidance, navigational anomaly detection, pollution monitoring, and PSC-related risk classification (Cazzulo et al., 2020; Wang et al., 2021; Celik et al., 2021; Zhang et al., 2022).

Tree-based ensemble models such as Random Forests (RF), Gradient Boosting, and XGBoost have been particularly influential due to their strong predictive performance and robustness to noise and missing data. These models effectively address nonlinearities and complex feature interactions, frequently outperforming traditional regression methods in safety classification tasks (Breiman, 2001; Lalla-Ruiz et al., 2018). Neural networks (NNs), including both feedforward and deep learning architectures, offer greater modelling flexibility by capturing abstract feature representations, though they entail a higher risk of overfitting and reduced interpretability (Goodfellow et al., 2016; Shukla et al., 2020).

Within the context of PSC, initial applications demonstrated that ML could enhance detention prediction and identify critical deficiency patterns. Studies employing RF and NNs on PSC datasets reported significant improvements in accuracy over deterministic targeting formulas, especially when incorporating vessel age, inspection history, technical attributes, and company performance indicators as features (Celik et al., 2021; Zhang et al., 2022). Research on OSVs further supported these results, indicating that ML models can detect operational and technical patterns that are not identifiable through traditional PSC scoring systems (Boko et al., 2024; Boko et al., 2025b). These findings underscore the value of ML as a complementary tool in PSC inspection planning.

Despite these advantages, the adoption of ML in regulatory environments is limited by concerns regarding model transparency, particularly the perception that many models function as black boxes. Regulatory authorities require interpretability and clear justification for inspection decisions. In the absence of explanations for model predictions, even highly accurate ML models may be regarded as unreliable or lacking legitimacy (Doshi-Velez & Kim, 2017). This challenge is especially pertinent in PSC, where decisions have significant implications for vessel trading rights, insurance liabilities, and operational schedules.

XAI has emerged in response to these concerns. Among the available techniques, SHAP provides a mathematically rigorous method for interpreting complex models by quantifying the contribution of individual features to predictions (Lundberg & Lee, 2017). SHAP delivers both global insights, identifying factors that generally influence predictions, and local explanations for individual vessels, enabling regulators to understand why a model classifies a specific ship as high risk. This dual capability is essential in PSC, where transparency and accountability are fundamental to decision-making.

The value of XAI in maritime applications is increasingly acknowledged. Research applying SHAP to navigational risk assessment, collision avoidance, and equipment failure detection demonstrates that interpretable ML can enhance operator trust, clarify model logic, and reveal risk patterns consistent with domain expertise (Xu et al., 2021; Guidotti et al., 2019; Roy et al., 2020). In the context of PSC inspection data, SHAP helps validate whether ML models are identifying genuine safety signals rather than patterns driven by regional inspection biases or inconsistent reporting. This validation capability is essential to ensure that ML-generated insights are aligned with safety principles and regulatory objectives (Doshi-Velez & Kim, 2017).

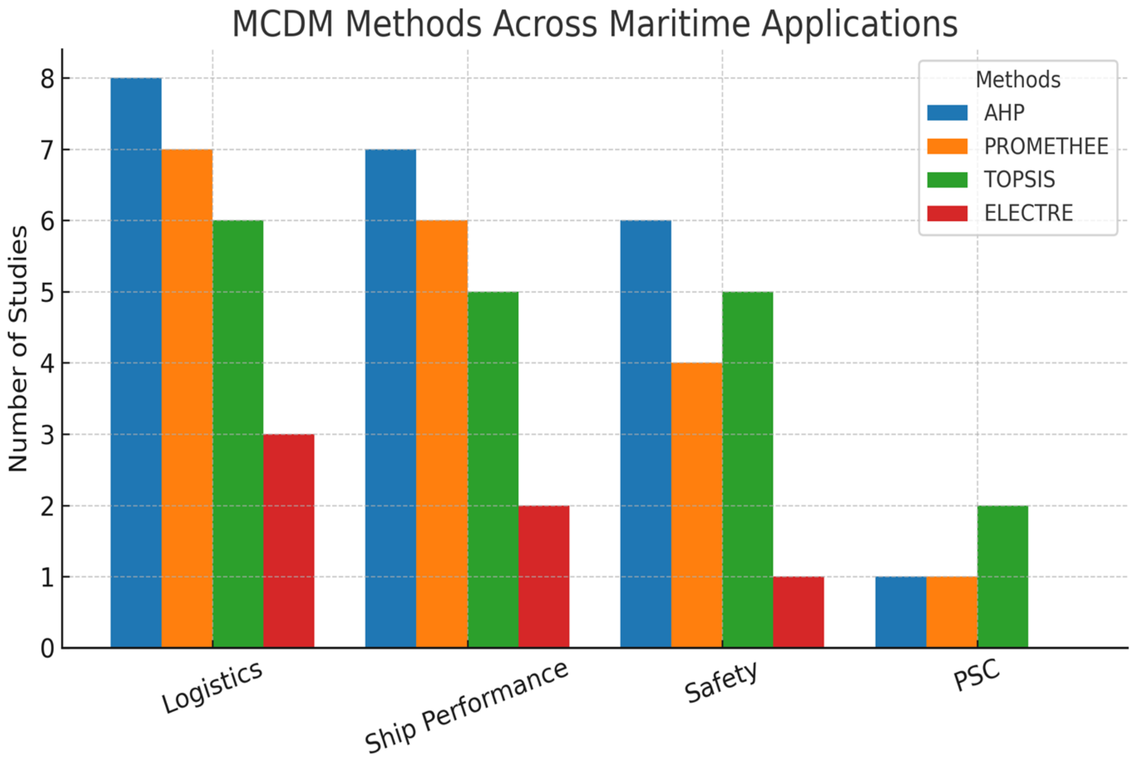

Figure 2.2. Stratification of studies according to the MCDM methods

Although ML offers predictive power and XAI provides interpretability, neither approach alone is sufficient for complex decision-making tasks such as inspection prioritisation. PSC authorities must integrate predictive insights with operational constraints, regulatory priorities, and expert judgment. MCDM methods offer a structured framework for incorporating diverse considerations into prioritisation processes. AHP and TOPSIS enable experts to evaluate multiple risk criteria, assign relative importance weights, and rank vessels based on the combined indicators (Saaty, 1980; Celik, 2009; Zavadskas et al., 2016).

To illustrate the breadth of methodological approaches in the literature, Figure 2.2 shows the distribution of different MCDM methods used across maritime application domains, such as logistics, ship performance, safety, and Port State Control (PSC). In maritime applications, MCDM has been used to assess navigational risks, port selection, ship performance, environmental impacts, and emergency response options (Dalaklis et al., 2018; Wang et al., 2014). MCDM's key advantage lies in its transparency: decision steps, weight assignments, and ranking logic are explicitly documented. This aligns closely with the needs of PSC, where inspection targeting must be defensible, consistent, and explainable to stakeholders.

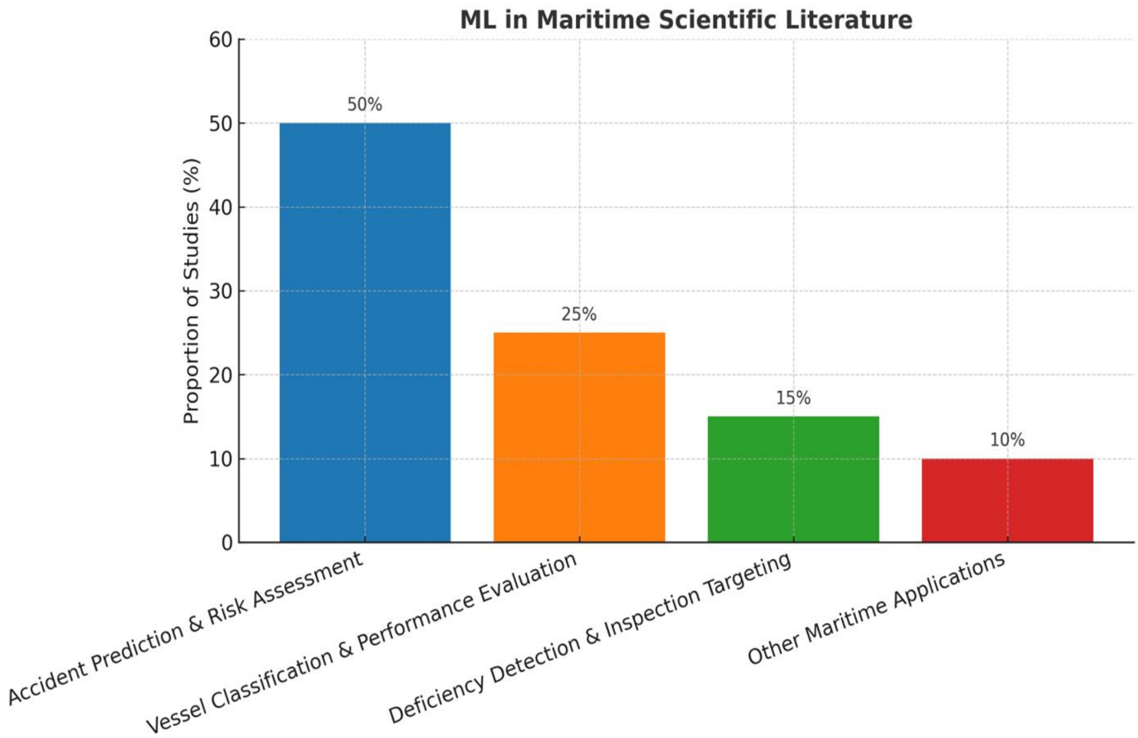

Figure 2.3. Distribution of ML applications in maritime scientific research

Figure 2.3 shows how ML is used in maritime scientific research. It organises published studies by main application areas, including accident prevention and risk assessment, vessel condition monitoring, and inspection analytics. The figure compares research activity in these areas and shows where ML studies are most common. By bringing these domains together, the diagram gives a clear picture of how widely ML is used in maritime research.

Recent study—including initial experiments in PSC contexts—suggests that integrating ML outputs into MCDM frameworks can enhance decision quality by combining predictive accuracy with expert-informed transparency (Boko et al., 2025a). Existing studies typically evaluate ML and MCDM components separately rather than integrating them into a single architecture. None provides a comprehensive methodological pipeline that links predictive modelling, model explanation, and structured decision-making into a single workflow suitable for operational PSC environments.

This absence is significant. PSC decision-making requires a balance between accuracy, interpretability, and expert judgement. ML alone cannot satisfy the transparency requirements of regulatory authorities, while MCDM alone may lack the analytical power to identify subtle risk patterns. XAI bridges the interpretability gap but does not offer prioritisation logic.

Therefore, the literature increasingly points to the value of hybrid analytical approaches, though such frameworks remain rare.

Hybrid ML–MCDM systems have been explored in other sectors—such as energy management, healthcare prioritisation, and environmental planning—with promising results (Wang et al., 2021; Zavadskas et al., 2016; Roy, 1996). These studies demonstrate that combining data-driven modelling with structured expert reasoning can yield decision tools that are both accurate and transparent. Nevertheless, their application to maritime safety, particularly PSC, remains in its infancy. For OSVs, the lack of established hybrid frameworks is even more pronounced. OSVs exhibit unique operational and safety characteristics that are poorly represented in generic PSC models.

Although ML studies have improved understanding of OSV risk patterns, and early applications of XAI have clarified model logic, the field still lacks a fully integrated methodology to transform these analytical components into a comprehensive inspection-support system. Addressing this gap constitutes the dissertation's central methodological objective (Boko et al., 2024; Boko et al., 2025a; Celik et al., 2021; Lundberg & Lee, 2017; Guidotti et al., 2019).

2.5. Synthesis of Literature, Limitations, and Research Gap

The literature reviewed across PSC regulation, OSVs operations, ML, XAI, and MCDM reveals a research landscape that is both rich and fragmented. Each field has produced meaningful advances, yet the intersections between them—where the most promising opportunities for innovation lie—remain insufficiently developed.

A recurring theme in the PSC literature is the tension between regulatory transparency and analytical complexity. Deterministic scoring systems have been widely adopted because they provide clear, straightforward inspection priorities that are justifiable (Panagakos et al., 2017). However, these systems cannot capture the nonlinear relationships and operational dynamics that studies have repeatedly documented in inspection data (Cariou & Mejia, 2017; Zhang et al., 2022). PSC authorities, therefore, operate with tools that are transparent but analytically limited.

At the same time, the body of research on applying ML to maritime safety demonstrates clear performance improvements over traditional methods. ML models have been shown to identify intricate risk patterns, detect subtle interactions, and accommodate large feature spaces that statistical models cannot (Breiman, 2001; Celik et al., 2021; Boko et al., 2024). For OSVs, ML approaches have identified operational and technical risk factors that are not adequately represented in conventional PSC frameworks (Boko et al., 2025b). These findings highlight the benefits of incorporating ML into inspection targeting.

However, the literature also recognises that ML is insufficient on its own. Regulatory environments require interpretability, and non-transparent models cannot be deployed without mechanisms enabling inspectors to understand how predictions arise (Doshi-Velez & Kim, 2017). The emergence of XAI—especially SHAP—offers promising solutions by enabling both global and vessel-level explanations of model behaviour (Lundberg & Lee, 2017; Guidotti et al., 2019). XAI techniques allow analysts to verify whether predictive models are learning safety-relevant patterns rather than inspection artefacts or regional enforcement biases. Yet, despite this progress, XAI has been applied to PSC data only sparingly, and its integration into operational decision-making frameworks remains limited.

A similar pattern appears in the MCDM literature. MCDM methods have demonstrated utility in navigational safety assessments, ship performance evaluation, and port planning (Dalaklis et al., 2018; Wang et al., 2014). In other words, MCDM has been applied to PSC, but rarely in combination with ML or XAI.

Taken together, these bodies of literature point toward a significant yet unaddressed opportunity: the development of an integrated analytical framework that links ML, transparency, and structured decision-making into a single, coherent methodology. Such a framework would allow PSC authorities to benefit from the predictive accuracy of ML, the interpretability of XAI, and the transparent prioritisation logic of MCDM.

First, PSC-specific ML research remains largely predictive, offering classification or forecasting capabilities but not structured decision-support methods. These models provide valuable insight but do not directly translate predictions into prioritisation tools suitable for regulatory use. Second, XAI has not been fully leveraged in PSC contexts. Although SHAP and similar techniques provide transparency, few studies evaluate how explainability can validate model integrity, support regulatory justification, or guide the integration of predictive results into broader decision frameworks. Third, MCDM applications in PSC are not integrated with empirical predictive modelling. Existing studies often rely solely on expert judgement or on static criteria, without utilising data-driven insights from ML models.

Finally, the literature on OSVs remains limited. OSVs are operationally and technically distinct from general cargo vessels, tankers, or bulk carriers, yet they are seldom analysed as a standalone category in PSC studies. Their inspection patterns, deficiency distributions, and operational risk factors need a tailored analysis that current PSC models cannot offer (Knapp & Franses, 2007; Boko et al., 2025b).

These limitations collectively suggest that PSC inspection targeting has not yet benefited from the full potential of modern analytical methods. The literature shows a clear need for:

- a predictive component capable of capturing nonlinear relationships.
- an interpretability layer capable of validating model logic.
- a decision-making structure capable of translating analytical insights into operational outputs.
- and a vessel-type–specific approach that accounts for OSV characteristics.

This dissertation builds on insights from previous research and directly addresses the analytical and practical needs identified in the literature. It proposes a hybrid analytical perspective that combines ML techniques, interpretability methods, and MCDM to support the assessment of OSVs safety within PSC inspection regimes. In doing so, the dissertation extends existing work in PSC analytics and outlines a methodological direction that may support future developments in data-informed inspection planning and maritime safety governance.

3. METHODOLOGY

3.1. Introduction

The preceding chapters established the operational and regulatory context for OSVs and examined the limitations of current inspection-based approaches to maritime safety assessment. Despite the increasing detail and accessibility of inspection data through regional and international information systems, the literature consistently identifies a gap between data availability and the analytical frameworks that support inspection decision-making (Knapp & Franses, 2007; Cariou & Wolff, 2011). This gap is especially pronounced in vessel segments with complex, context-dependent operational profiles.

In this context, as noted earlier, OSVs pose a distinct methodological challenge. Inspection outcomes are shaped not only by technical condition but also by operational intensity, specialised equipment, and the interactions between ship systems and human operators. Previous studies on offshore and specialised vessels reveal just how elusive these factors can be when measured with straightforward scoring rules or simple risk indicators (Hetherington et al., 2006; Schroeder et al., 2015; Lai et al., 2023).

Previous empirical research by the author investigated the application of data-driven techniques to PSC inspection data, including the use of classification models to estimate detention-related outcomes and the integration of predictive outputs into ranking procedures for OSVs (Boko et al., 2024; Boko et al., 2025b). These studies demonstrated that inspection records contain meaningful patterns related to OSV safety performance but also identified limitations when predictive results were considered in isolation, as discussed in Chapter 2. Without a clear way to interpret and prioritise results, purely predictive methods often fell short of being truly useful in a real-world regulatory setting.

The methodology adopted in this dissertation builds upon these observations. Instead of treating prediction, interpretation, and prioritisation as separate analytical tasks, the research design integrates them within a unified framework aligned with inspection-based decision-making. This integrated approach echoes the ongoing conversations in safety-critical fields, where the pursuit of analytical excellence is carefully weighed against the need for transparency, accountability, and regulatory approval (Doshi-Velez & Kim, 2017; Guidotti et al., 2019). Accordingly, this chapter examines the methodological choices underlying the proposed framework. The discussion emphasises the transformation of inspection data into analytically

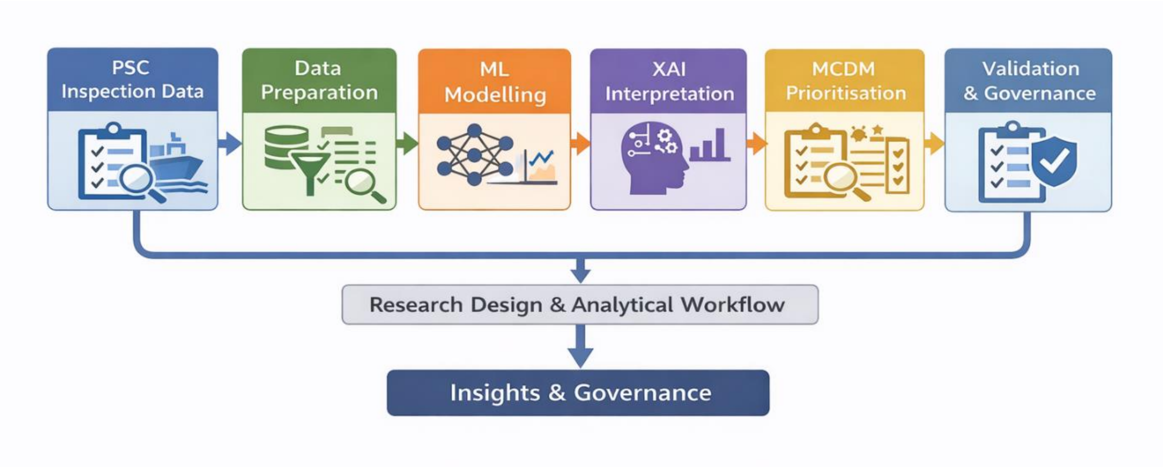

Figure 3.1. Research design and analytical workflow of the proposed hybrid framework

meaningful inputs, the development and interpretation of learning-based models, and the structuring of analytical outputs to support inspection prioritisation. By grounding these choices in both the existing literature and prior empirical research, the methodology provides a foundation for the analytical developments presented in the following sections.

3.2. Research Design and Analytical Framework

The research design adopted in this thesis is based on the premise that inspection prioritisation for OSVs represents a decision problem rather than solely a predictive or explanatory task. Although prediction and explanation are important parts of analysis, they are seen as valuable tools for helping with decision-making, not as separate goals on their own. This way of thinking directly affects how we choose, order, and combine analysis methods in the proposed framework.

The framework is intentionally modular, reflecting both analytical requirements and the regulatory environment. Inspection regimes need analytical outputs that are clear, auditable, and adaptable to changing regulatory priorities (Parasuraman et al., 2000). Monolithic modelling, whether statistical or data-driven, lacks flexibility. A modular design enables independent development, validation, and interpretation of analytical functions, while maintaining overall coherence.

Figure 3.1 presents the modular analytical framework, which demonstrates the deliberate sequencing of predictive modelling, interpretability, and decision structuring within a unified inspection-support process. The figure underscores the sequential, rather than parallel, integration of analytical stages.

In the first stage, ML models estimate detention-risk patterns using PSC inspection data. The goal is to find clear risk signals in complex data, not to make final inspection decisions. Learning-based models are used because OSV inspection outcomes depend on technical, operational, and organisational factors that interact in ways simple scoring rules or linear models cannot capture. Methods that rely solely on set indicators or expert-weighted scores are not sufficient here, as they assume factors are independent, which the evidence does not support.

The second stage uses XAI techniques to study how the model works. This step tackles a main challenge in regulatory prediction: the complexity of analysis. High accuracy is not enough for inspection authorities, who need to explain their decisions and understand what drives risk assessments. Interpretable methods help break down model outputs into the role of each feature, making it easier to see how risk estimates are made. This stage aims to make the model's logic clearer, so experts can review and understand the results, not replace their judgment.

The third stage formalises inspection prioritisation as an MCDM problem. Although transparency clarifies the generation of predictions, it does not specify how predictive information should be weighed against regulatory objectives, operational constraints, and safety considerations. MCDM methods are therefore introduced to explicitly structure this balance. With this approach, MCDM does not replace analytical modelling or act as an automated decision-making tool. Instead, it provides a straightforward and consistent way to combine predictive and interpretative results with context-based judgment to set inspection rankings.

It is important to follow these stages in the order they appear. Predictive modelling alone can be accurate but may not work well in practice if it lacks interpretability. Interpretability helps explain model behaviour, but without decision structuring, it does not clearly show how to set priorities. Decision structuring without using data-driven risk estimates relies too much on fixed assumptions or expert opinion. By combining all these stages, inspection prioritisation becomes more reliable, clear, and better suited to the real needs of PSC activities.

This integrated design allows for systematic validation at several levels. Predictive performance can be measured with numbers; interpretative consistency across models and cases can be checked; and the stability of rankings can be tested under different weighting assumptions. This kind of multi-layered validation is crucial in safety-critical fields, where credibility depends on the overall coherence and stability of the decision-support process, not just on single performance measures.

In summary, the proposed research design embodies a deliberate methodological approach. It rejects methodologies that treat prediction, explanation, or decision support as isolated solutions. Instead, it advances an integrated analytical framework tailored to the epistemic and operational demands of OSV inspection prioritisation. The primary contribution of this dissertation is the principled alignment of established techniques within a decision-theoretic framework, enabling transparent, defensible inspection decisions.

3.3. Data Sources and Data Preparation

The empirical analysis in this dissertation is based on inspection data generated through PSC activities. The selection of data sources follows the methodological principle that analytical models supporting inspection prioritisation must rely on information realistically available to inspection authorities at the time decisions are made. Accordingly, PSC inspection databases offer both a robust empirical foundation and a direct connection between analytical development and operational practice.

The primary source of inspection-related data for this study is EQUASIS. EQUASIS consolidates vessel particulars, inspection histories, deficiency records, and detention outcomes reported by multiple regional PSC regimes into a single, publicly accessible platform managed by the EMSA (EMSA, 2023). Its structure reflects the operational logic of inspection systems by integrating stable vessel attributes with event-based inspection outcomes recorded over time and across geographic regions (Knapp, 2006; Cariou & Mejia, 2017). This integration enables comparative analysis of inspection performance while maintaining the regulatory context in which the data are produced.

The use of PSC inspection data presents methodological challenges that require explicit consideration. Inspections are guided by risk-based targeting strategies that differ across MoUs and evolve rather than by random sampling (Anderson et al., 2021). As a result, inspection datasets reflect enforcement priorities, regional practices, and historical targeting logic, rather than providing an unbiased representation of fleet-wide safety conditions (Knapp & Franses, 2007; Cariou & Wolff, 2011). The dataset for this research focuses specifically on OSVs. Vessel records were filtered by ship-type classifications aligned with offshore operational roles to ensure that the analytical sample accurately represents the vessel segment under study. Variables were selected based on their relevance to inspection decision-making and their availability prior to inspection, thereby avoiding reliance on post-inspection or outcome-dependent information. This approach aligns the analytical framework with the practical informational constraints faced by PSC authorities (Celik et al., 2021; Zhang et al., 2022).

Data preparation was approached as a distinct methodological stage rather than merely a technical prerequisite for modelling. Initial processing included harmonising categorical variables, standardising vessel identifiers, and ordering inspection records chronologically. Special attention was paid to handling repeated inspections and multiple deficiencies recorded during a single inspection. Inspection outcomes were structured at the vessel–inspection level to preserve the contextual integrity of inspection events and to prevent artificial inflation of the analytical sample through replication at the deficiency level. Comparable structuring approaches have been employed in previous empirical studies of PSC inspection outcomes, including those focused on OSVs (Boko et al., 2024).

Missing data are inherent to inspection datasets compiled from multiple reporting authorities. In this study, missing values were handled using a combination of exclusion rules and imputation strategies, depending on each variable's role. Variables with systematic or structural absence were excluded to prevent bias, while limited missingness in otherwise informative variables was addressed using statistically appropriate imputation methods. The guiding principle was to preserve analytical validity and interpretability, rather than to maximise dataset size at the expense of data quality, in accordance with best practices in applied ML research (Kuhn & Johnson, 2013).

Another key aspect of data preparation was the temporal framing of inspection information. Because the study focuses on inspection prioritisation rather than retrospective classification, historical inspection records and deficiency patterns were aggregated over defined time windows to represent the information state available prior to inspection. This temporal structure enables the development of predictive models that reflect inspection planning processes and prevent the inclusion of post-inspection information in analytical inputs. Since this research focuses on inspection prioritisation rather than classifying outcomes after the fact, the analytical framework sets $t = 0$ as the time at which an inspection decision is made. All explanatory variables were built by combining historical inspection and deficiency data available before $t = 0$. This approach ensures that the model uses only information available during inspection planning. Organising the data this way avoids using post-inspection information, reduces data leakage, and follows best practices for predictive modelling in safety-critical and regulatory settings (Lipton et al., 2016; Roberts et al., 2021).

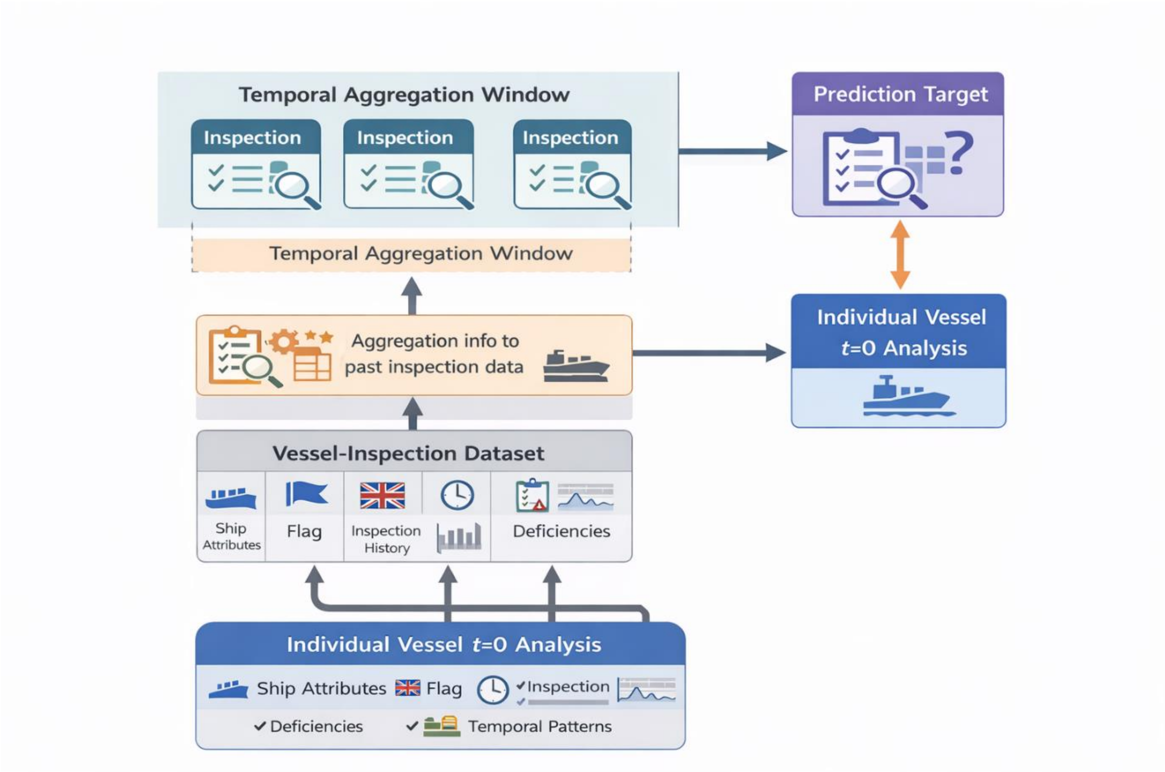

Figure 3.2. Temporal structuring and aggregation of PSC inspection data for vessel-level analysis

Similar approaches have been used in previous ML applications to PSC inspection data, including studies focused on OSVs (Boko et al., 2025b). Therefore, the unit of analysis in this research is the vessel–inspection event, with all explanatory variables constructed to represent the information state at the time of inspection (t = 0).

Figure 3.2. The following illustrates the temporal framing and aggregation logic applied to construct vessel-level inspection inputs. The figure demonstrates the aggregation of historical inspection records over a specified temporal window to generate vessel-level inputs available at the time of inspection (t = 0). This dataset integrates vessel attributes, inspection history, and deficiency patterns to facilitate predictive modelling and inspection-support analysis, while minimising the risk of information leakage.

During data preparation, a clear separation was maintained between data construction, predictive modelling, and decision structuring. Treating data preparation as a distinct methodological step enhances transparency, facilitates reproducibility, and enables subsequent analytical results to be interpreted in light of explicitly stated data assumptions. The prepared dataset serves as the empirical foundation for the modelling and decision-support methods described in the subsequent section. Grounding the analysis in inspection data representative of real-world PSC information environments establishes the conditions necessary for meaningful interpretation and evaluation of analytical outputs within the context of inspection-based maritime safety governance.

3.4. Modelling Framework and Integration Logic

The modelling strategy adopted in this dissertation is designed to extract meaningful safety-related patterns from inspection data while ensuring that analytical outputs remain interpretable and actionable within inspection-oriented decision contexts. Instead of treating prediction, interpretation, and prioritisation as separate analytical exercises, the framework integrates these components within a unified methodological logic. This integration addresses both the limitations identified in earlier PSC analytics and the practical requirements of OSV inspection planning.

Supervised ML techniques are central to the modelling framework, enabling the estimation of detention-related risk patterns from inspection data. Previous empirical studies applying ML to PSC data have shown that these models can enhance sensitivity to complex safety patterns, particularly within specialised vessel segments (Celik et al., 2021; Zhang et al., 2022). Earlier research by the author on OSVs demonstrated that similar approaches revealed structured risk patterns associated with vessel characteristics and inspection histories that were not apparent under traditional assessment schemes (Boko et al., 2024; Boko et al., 2025b).

Model selection within the framework prioritises robustness, stability, and interpretability alongside predictive performance. Instead of optimising a single model architecture, multiple learning algorithms are considered and evaluated comparatively to determine their suitability for inspection-support applications. Consequently, emphasis is placed on models that demonstrate consistent performance across validation settings and facilitate systematic interpretation.

To address the interpretability challenges inherent in ML applications, the framework incorporates XAI techniques as an integral analytical layer. Clarity is treated not as a post hoc diagnostic but as a mechanism for validating model behaviour and examining the contribution of individual variables to predicted outcomes. Techniques such as SHAP decompose model predictions to identify dominant risk drivers and interaction effects within the inspection data (Lundberg & Lee, 2017; Guidotti et al., 2019). Within PSC inspection analytics, this interpretative layer provides critical insight into whether model outputs align with established safety principles and operational expectations.

The role of understandability within the framework extends beyond transparency. By enabling systematic comparison of variable importance across models and inspection contexts, XAI supports analytical validation and facilitates discussion of model outputs in regulatory terms. This capacity for validation is essential to ensuring that ML-generated insights remain consistent with regulatory intent and can be examined within inspection decision-making processes (Doshi-Velez & Kim, 2017). Previous exploratory studies found that the absence of such interpretative mechanisms limited the practical usefulness of predictive models, particularly when results conflicted with established inspection experience (Boko et al., 2025a).

Although predictive modelling and interpretability offer valuable analytical insights, they do not, in isolation, resolve the challenge of inspection prioritisation. To translate analytical outputs into structured inspection priorities, the framework incorporates an MCDM component that synthesises quantitative and qualitative inputs within a transparent decision structure.

MCDM methods provide formal mechanisms for combining heterogeneous criteria while preserving traceability and consistency (Belton & Stewart, 2002). Within this framework, ML outputs and interpretative indicators serve as inputs to decision models that structure, rather than dictate, inspection priorities. Techniques such as AHP and TOPSIS are used to weight and aggregate criteria and to generate prioritisation rankings aligned with inspection objectives (Saaty, 1980; Celik, 2009; Zavadskas et al., 2016).

A key design principle underlying the integration logic is modularity. Predictive modelling, interpretability, and decision structuring are developed as analytically distinct components, each subject to independent validation and sensitivity analysis. Their integration follows a sequential logic that preserves interpretability at each stage and prevents conflating statistical inference with normative judgement. This modular structure enhances transparency and supports governance considerations, particularly in safety-critical regulatory environments where analytical decisions must be explainable and defensible (Kuhn & Johnson, 2013).

Collectively, the modelling framework and integration logic presented in this section implement the hybrid analytical approach introduced earlier in the chapter. By combining learning-based prediction, systematic interpretability, and structured decision support, the framework addresses limitations identified in existing PSC analytics while remaining aligned with the informational and procedural constraints of inspection regimes. The empirical performance and practical implications of this integrated approach are examined in the following chapter. Figure 3.3. summarises the overall integration logic of the proposed hybrid

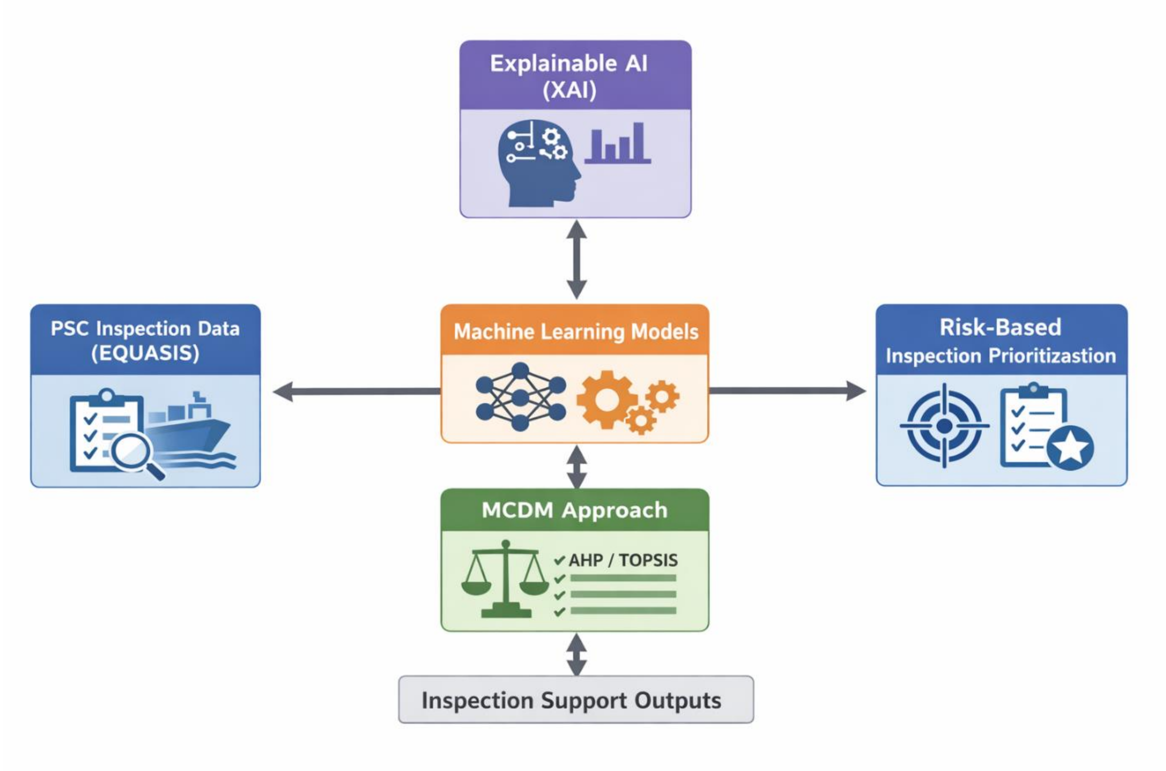

Figure 3.3. Hybrid ML – XAI – MCDM framework for inspection support

framework, using a block diagram which illustrates how to integrate PSC inspection data with ML models, XAI techniques, and MCDM methods to support the structuring of inspection priorities. The framework emphasises modularity, interpretability, and structured decision reasoning rather than automated enforcement.

3.5. Model Validation and Robustness

The reliability of analytical results in inspection-based safety assessments relies on how well the model performs and how robust, transparent, and well-governed the analysis is. In regulatory settings such as PSC, analytical tools must meet standards that go beyond mere statistical accuracy. As a result, validation is seen as a process with several layers, including reliable methods, clear interpretation, and consistent decisions.

At the statistical level, validation assesses the reliability and generalizability of predictive performance under conditions that mirror inspection prioritisation rather than retrospective classification. Because detention outcomes are imbalanced, model performance is evaluated using multiple complementary metrics. This approach prevents overreliance on a single summary measure and enables explicit analysis of trade-offs among errors, thereby reflecting the practical requirements of inspection planning (Kuhn & Johnson, 2013; Lalla-Ruiz et al., 2018).

Beyond statistical performance, robustness is evaluated through sensitivity analyses and stability checks that determine how model outputs respond to changes in data composition, parameter settings, and temporal segmentation of inspection records. Consistency in identifying risk drivers and in relative vessel rankings across different modelling assumptions is vital. In inspection environments, model instability can erode confidence and hinder practical adoption, even if average predictive performance is adequate. Previous studies have identified similar concerns when applying data-driven methods in safety-critical and regulatory domains (Goerlandt & Montewka, 2015; Celik et al., 2021).

Interpretability is a key part of the validation process. XAI techniques help explain individual predictions and check if model behaviour matches safety standards and practical knowledge from inspections. By looking at how features contribute and interact in different models and validation settings, the framework allows for systematic validation of model logic instead of relying on one-off explanations. This is especially important in regulated settings, where analytical results must align with regulatory goals and can be closely examined (Doshi-Velez & Kim, 2017; Guidotti et al., 2019).

Validation also encompasses the integrated decision-support layer. The coherence of inspection-priority rankings produced by the hybrid framework is evaluated against predictive outputs, interpretability findings, and the weighting of decision criteria. Sensitivity analyses are conducted to determine how variations in criteria weights or analytical inputs affect prioritisation outcomes. This process ensures that integrating ML outputs into MCDM does not result in disproportionate influence from any single component or unintended changes in inspection focus (Saaty, 1980; Zavadskas et al., 2016).

Governance considerations represent an additional dimension of validation. The framework is structured to support traceability of analytical decisions, enabling inspection authorities to review how data inputs, modelling assumptions, and decision criteria influence prioritisation outcomes. By maintaining a modular architecture and clear separation among predictive modelling, interpretability, and decision structuring, the framework enhances oversight and accountability. This approach aligns with broader discussions on the governance of algorithmic decision-support systems in safety-critical contexts, where transparency is critical (Montgomery, 2017; Doshi-Velez & Kim, 2017).

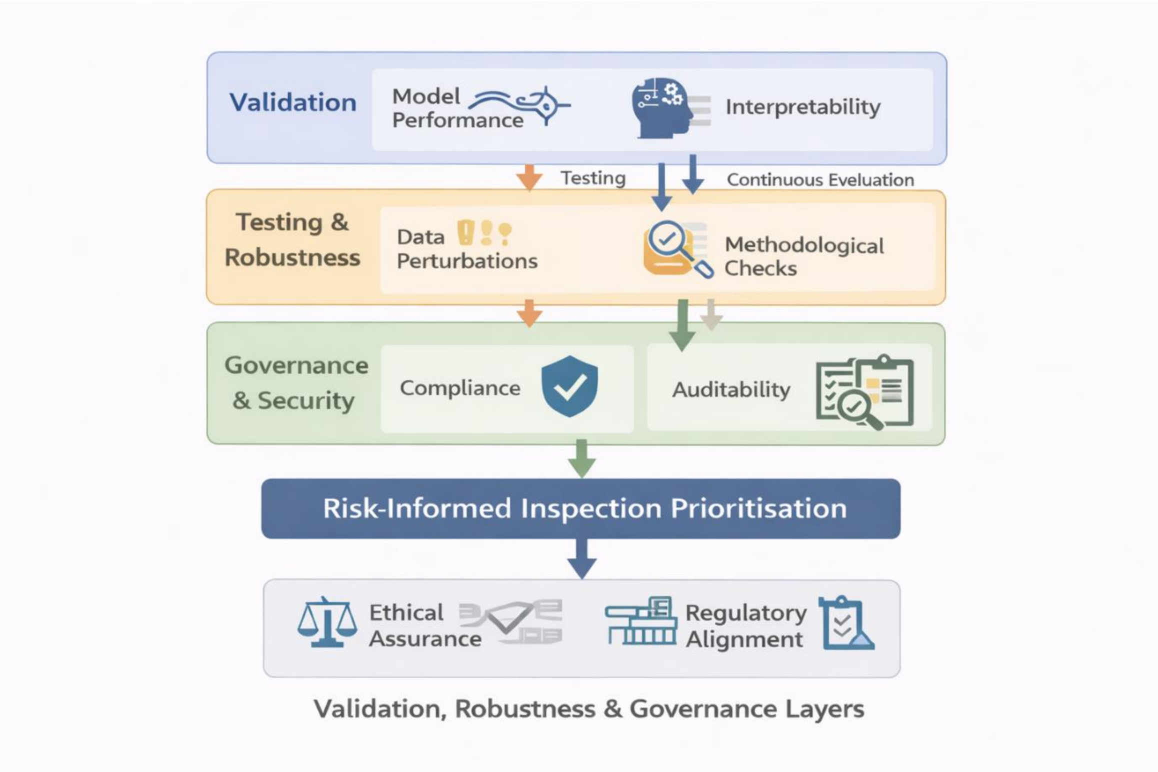

Figure 3.4. Validation, robustness, and governance layers of the proposed inspection-support framework

Reproducibility is ensured through systematic documentation of data preparation procedures, model configurations, and validation protocols. Analytical decisions are explicitly recorded to facilitate replication and independent evaluation, consistent with established scientific practice. Although the framework is not currently implemented as an operational system within a PSC authority, its methodological design anticipates future deployment requirements, including periodic model review, recalibration, and performance monitoring to accommodate evolving inspection practices and fleet characteristics.

Collectively, the validation, robustness, and governance considerations described in this section ensure that the proposed analytical framework is evaluated not only for predictive performance but also for its appropriateness within inspection-based maritime safety governance. By embedding validation across the statistical, interpretive, and decision-support layers, the methodology provides a foundation for the responsible use of data-driven tools in maritime safety oversight.

Figure 3.4. shows the multi-layer validation logic applied, illustrating how model performance, interpretability, robustness testing, and governance considerations jointly support risk-informed

inspection prioritisation while maintaining auditability, ethical assurance, and regulatory alignment.

3.6. Ethical, Legal, and Practical Considerations

The methodological design accounts for ethical, legal, and practical factors essential to the use of data-driven analytical tools in inspection-based safety governance. Instead of focusing on how well the proposed framework works in practice, these factors highlight what is needed for responsible development, interpretation, and possible use of analytical methods in regulatory settings.

Ethically, the analysis relies solely on publicly accessible inspection data. No personal data about individual seafarers or inspectors is utilised, nor is there any attempt to infer individual behaviour or responsibility. The analytical focus is maintained at the vessel and inspection system levels, aligning with the scope and intent of PSC regimes. By limiting the analysis to aggregated and anonymised records, the research addresses ethical concerns related to privacy, surveillance, and individual attribution identified in broader discussions of algorithmic decision-support systems (Doshi-Velez & Kim, 2017).

Legal considerations primarily concern the role of analytical tools in regulatory decision-making. PSC inspections are governed by international conventions and regional MoUs, which assign decision-making authority to inspectors and competent authorities. The analytical framework developed in this dissertation is not intended to supplant professional judgment or serve as an automated enforcement mechanism. Instead, it is designed as a decision-support tool that may inform inspection prioritisation while remaining subordinate to established legal mandates and procedural safeguards. This distinction is critical for maintaining compliance with regulatory principles and for preventing the delegation of legally significant decisions to opaque algorithmic processes (Guidotti et al., 2019).

Practical considerations encompass the feasibility and interpretability of analytical outputs within inspection environments. Inspection authorities face constraints related to time, resources, and institutional accountability. For analytical methods to be practically relevant, their outputs must be traceable, explainable, and compatible with existing inspection workflows. The modular structure of the proposed framework, which separates predictive modelling, interpretability, and decision structuring, supports these requirements by enabling independent examination and selective integration of each analytical component. This design addresses concerns identified in previous applications of data-driven tools in safety-critical domains, where insufficient transparency and operational compatibility have hindered practical adoption (Goerlandt & Montewka, 2015). Another practical consideration is the dynamic nature of inspection systems. Inspection priorities, targeting criteria, and reporting practices change in response to regulatory developments, fleet composition, and emerging risks. Consequently, analytical models developed from historical data require periodic review and recalibration to remain relevant. Although this dissertation does not implement an operational monitoring system, the methodological framework anticipates the necessity for lifecycle management and ongoing validation in any future application context.

Collectively, these ethical, legal, and practical considerations define the boundary conditions for the responsible use of analytical methods in PSC inspection support. By explicitly acknowledging these constraints, the dissertation positions the proposed framework as a methodological contribution rather than a prescriptive regulatory solution, thereby maintaining alignment with established principles of maritime safety governance.

4. RESULTS AND ANALYSIS

4.1. Introduction

This chapter presents the results of applying the methods described in Chapter 3. The analysis examines how well ML models perform on inspection data for OSVs. It explains and evaluates these predictions before adding them to the larger decision-support system. To meet the research goals, the results are presented clearly and in an organised manner. First, the performance of each ML model is reviewed, focusing on accuracy, robustness, and consistency in different validation tests. Instead of picking a single best model, the aim is to build a solid foundation for later interpretation and decision-making. This is especially important in regulatory work, where stable, traceable models matter more than minor performance improvements.

The results in this chapter are described and analysed, but do not discuss policy impacts or value judgments. Interpretations are limited to what the data directly supports. Broader impacts are covered in Chapter 5. This approach keeps the results clear and separate from their possible meaning for inspection work.

4.2. ML Model Results

4.2.1. Model Training and Validation Overview

The ML models were trained using the vessel–inspection event dataset described in Chapter 3. Each observation represents a single PSC inspection of an OSV, with explanatory variables capturing vessel characteristics, inspection history, and deficiency patterns available prior to the inspection event. This design maintains temporal consistency and prevents information leakage, both of which are essential for credible predictive modelling in safety-critical applications (Kuhn & Johnson, 2013; Zhang et al., 2022).

Multiple classification algorithms were implemented to address differences in model structure, learning capacity, and interpretability. The selected models include tree-based approaches and NN architectures commonly used in maritime risk prediction studies (Breiman, 2001; Celik et al., 2021; Boko et al., 2024). Hyperparameter tuning was performed on the training data using cross-validation, and the final model was evaluated on hold-out validation sets not used during training.

Due to the inherent class imbalance in detention outcomes, performance evaluation relied on multiple metrics rather than a single accuracy measure. This approach reflects best practice in maritime safety analytics, where false negatives and false positives have distinct operational implications (Lalla-Ruiz et al., 2018; Zhang et al., 2022; Saito & Rehmsmeier, 2015).

Model performance was assessed using complementary metrics that capture discrimination, classification behaviour, and probability reliability. These metrics include the area under the receiver operating characteristic curve (ROC-AUC), precision, recall, and calibration error.

4.2.2. Comparative Predictive Performance

Across all tested models, predictive performance surpassed baseline expectations established by the deterministic scoring approach. Tree-based ensemble models exhibited highly stable behaviour, consistently achieving greater discrimination between detained and non-detained inspections across validation folds. These results are consistent with previous maritime studies in which ensemble methods outperformed linear or rule-based classifiers on heterogeneous inspection data (Breiman, 2001; Boko et al., 2024).

NN models demonstrated competitive performance across several validation settings, particularly when nonlinear interactions between vessel attributes and inspection history were significant. However, their performance was more sensitive to training configuration and sample composition, leading to greater variability across validation runs. This finding underscores the importance of model robustness and reproducibility when evaluating predictive tools for regulatory applications, where consistency is paramount (Doshi-Velez & Kim, 2017).

Notably, no single model outperformed others across all evaluation criteria. Some models achieved higher discrimination metrics, while others demonstrated improved calibration or reduced variance across folds. These trade-offs informed the decision to retain multiple candidate models for subsequent interpretability analysis, rather than prematurely selecting a single predictive architecture. A comparative summary of model performance across the selected evaluation metrics is provided in Figure 4.1. below.

Figure 4.1. summarises model predictive performance across several evaluation metrics, including AUC, precision, recall, and calibration error. It highlights differences in discrimination, classification behaviour, and probability reliability among the various model architectures. Calibration error values tend to appear smaller than metrics like AUC, precision, or recall because they measure how far predicted probabilities deviate from actual outcomes,

not how well the model classifies. As a result, calibration errors usually fall within a much tighter range.

4.2.3. Calibration and Reliability of Predictions

In addition to classification performance, the calibration of predicted detention probabilities was evaluated to determine whether model outputs could be interpreted as meaningful risk indicators. Well-calibrated models generate probability estimates that closely match observed outcome frequencies, a property that is especially important in inspection prioritisation contexts (Guo et al., 2017; Zhang et al., 2022).

The analysis indicated that ensemble-based models generally demonstrated more reliable calibration than more flexible architectures. In contrast, specific high-capacity models produced overconfident predictions in regions with limited training data, underscoring the need to assess calibration alongside discrimination metrics. These findings demonstrate that predictive strength alone is insufficient for operational relevance, particularly when outputs are intended to inform downstream decision-making.

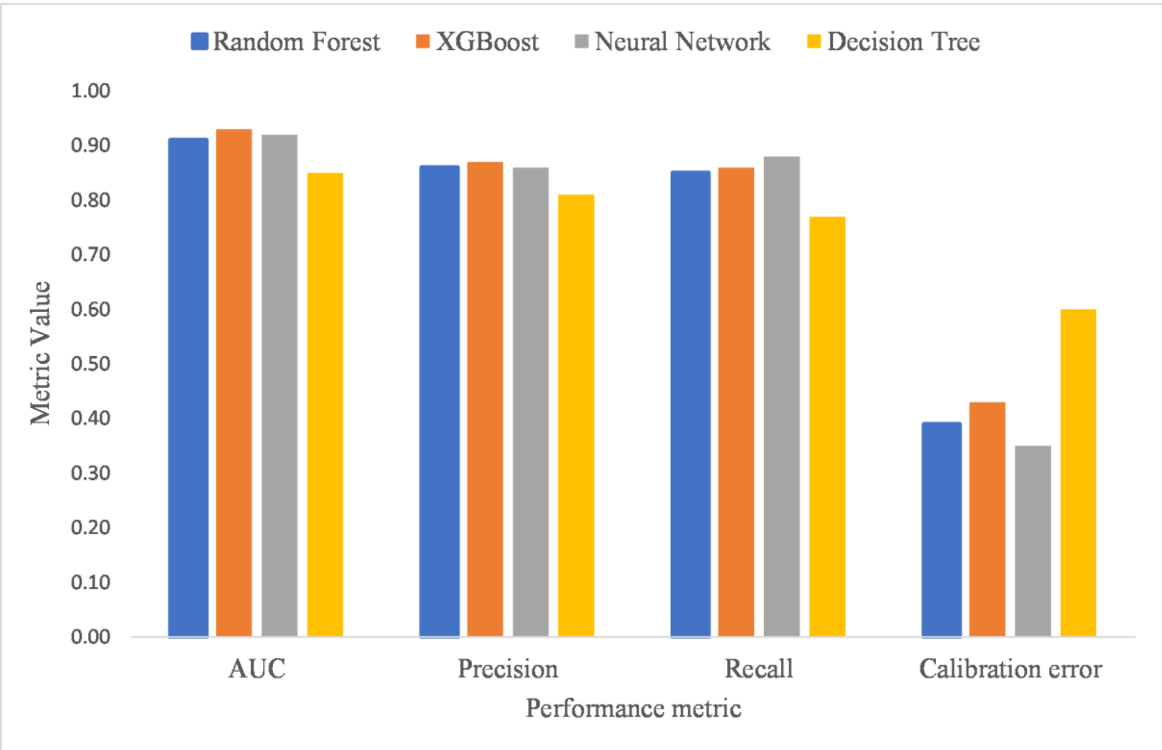

Figure 4.1. Comparative predictive performance of ML models for OSVs detention outcomes, stratified by performance metrics

4.2.4. Summary of Predictive Results

Collectively, the results demonstrate that ML models can identify meaningful patterns in PSC inspection data for OSVs. The predictive structures identified in the data support the first research hypothesis, indicating that inspection records provide sufficient information to model detention outcomes reliably for further analytical applications.

Simultaneously, the observed differences in performance stability and calibration across models reinforce the need to integrate predictive analytics with interpretability and structured decision-support mechanisms. These considerations motivate the subsequent analysis of variable influence and the integration of model outputs into the hybrid ML, XAI, and MCDM framework presented in the following sections.

4.3. Interpretation of Predictive Drivers and Risk Structure

To shed light on the patterns hidden within inspection data, XAI techniques have been used to interpret our model outputs. Rather than focusing on individual predictions, the idea was to uncover how different types of information shape the risk of detention for OSVs. Through this lens, we explored whether the model’s predictive logic aligns with real-world operational traits and established inspection routines.

Throughout all the models examined, certain variables stood out as key drivers of prediction. Inspection history—such as prior deficiencies, recent inspection frequency, and signs linked to detentions—had a powerful impact on risk scores. This highlights how safety assessments build over time, with past issues continuing to echo in future inspections. Features unique to each vessel, such as age and operational class, also shaped predictions, though their influence varied with the model’s design and training.

Notably, the contribution of individual variables was rarely independent. For instance, the predictive weight of prior deficiencies varied depending on vessel age, inspection recency, and operational role. These interaction effects were especially apparent in tree-based models, which are well-suited to capturing conditional relationships in heterogeneous datasets (Breiman, 2001; Zhang et al., 2022). Such patterns are challenging to represent using linear or additive risk formulations, which treat explanatory variables as independent contributors.

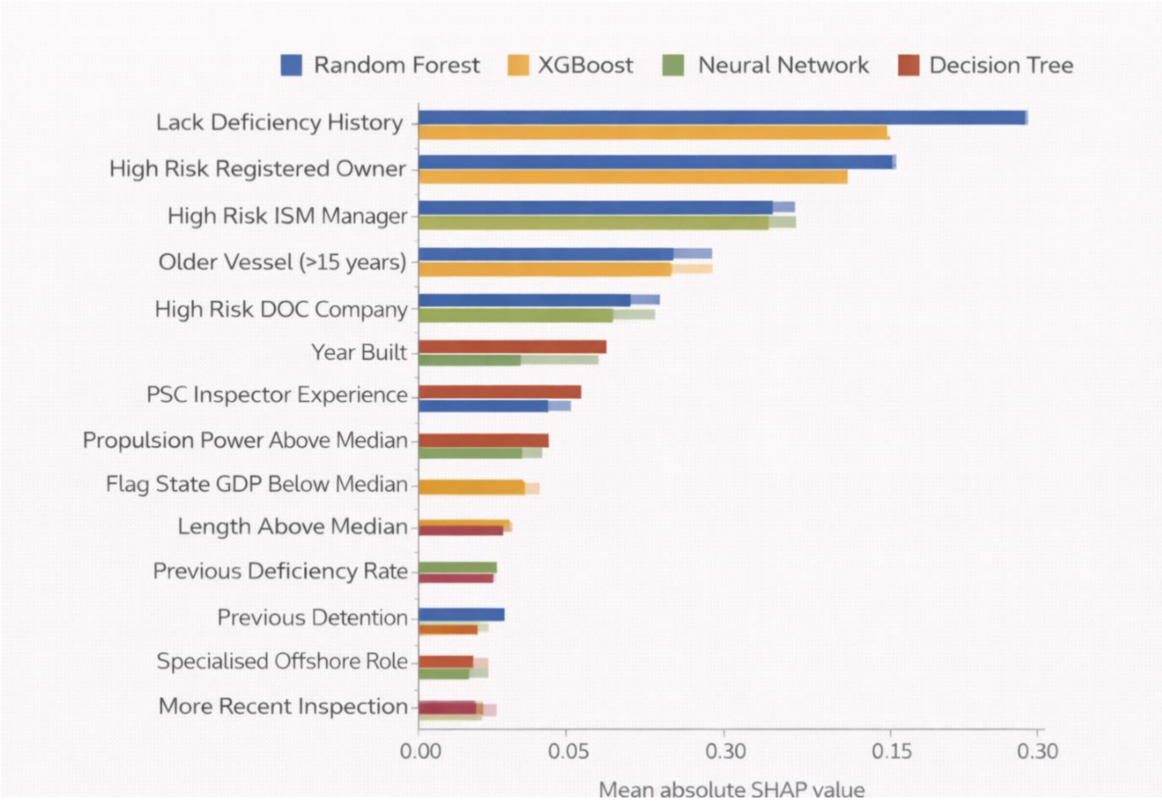

Figure 4.2. Relative importance of predictive variables for OSVs detention outcomes across ML models

Figure 4.2. presents mean absolute SHAP values for the most influential variables, illustrating the contributions of deficiency history, organisational attributes, and vessel characteristics to detention predictions across four different model architectures: RF, XGBoost, NN and Decision Tree. The interpretation analysis further identified distinctions between technical and operational signals. Although technical deficiencies and structural indicators were influential, operational and organisational variables, such as inspection history patterns and company-level attributes, frequently shaped the context in which technical issues became predictive. This finding aligns with previous maritime safety studies that emphasise the importance of organisational performance and compliance culture in addition to technical condition (Hetherington et al., 2006; Cariou & Mejia, 2017).

The influence of variables was not consistent across all models. Certain predictors were highly influential in specific configurations but less so in others, especially in higher-capacity models. This variability underscores the need for cautious interpretation. It highlights the importance of evaluating explanatory outputs alongside model robustness and calibration, rather than treating variable importance as a definitive ranking of risk factors.

In summary, the interpretability analysis confirms that the predictive models capture structured, operationally credible patterns in the inspection data. The identified drivers combine historical compliance, vessel characteristics, and inspection dynamics. These results establish a foundation for integrating predictive outputs into inspection-support reasoning without exclusive reliance on statistical representations.

4.4. Hybrid Risk Structuring and Inspection-Priority Outcomes

After evaluating predictive performance and interpretability, model outputs were integrated into the hybrid decision-support structure introduced in Chapter 3. At this stage, the analysis shifts from prediction and explanation to examining the behaviour of inspection-priority outcomes when predictive information is combined with structured decision criteria. The objective is to assess how analytically derived risk signals can be organised into coherent and transparent prioritisation outcomes.

Instead of letting ML risk scores dictate final decisions, they have been used as one piece of a larger analytical puzzle. By normalising these scores and blending them with other key criteria tied to inspection priorities and operational needs, we ensured the models' insights informed the process without letting raw probabilities drive enforcement actions.

The resulting inspection-priority rankings demonstrated high internal consistency across different model configurations. Vessels identified as higher risk by predictive models generally maintained similar relative positions when assessed through the multi-criteria structure, regardless of underlying model architecture. This indicates that the hybrid framework is less sensitive to model-specific variance than raw predictive outputs, which is a valuable property in inspection-support contexts where stability is essential (Bansal, A., Farahani, A., & Schölkopf, B., 2021).

Simultaneously, the integration process identified instances where additional decision criteria moderated predictive risk estimates. Some vessels with high predicted detention probabilities received lower priority when contextual factors were incorporated. Conversely, vessels with moderate predictive risk advanced in the prioritisation hierarchy due to consistent patterns in their inspection history. These adjustments demonstrate how structured decision reasoning can contextualise predictive signals rather than amplify them without critical assessment. Sensitivity analysis revealed that inspection-priority outcomes remained stable even when weighting assumptions changed moderately. Only under the most extreme scenarios did rankings change as anticipated, yet the core prioritisation structure proved resilient across a broad spectrum of realistic settings. This resilience shows that the hybrid framework stands on solid ground, not reliant on delicate or narrowly set parameters, which is essential for real-world trust.

Importantly, the hybrid approach did not simply rank vessels by their history of non-compliance. Instead, it produced prioritisation outcomes shaped by a thoughtful blend of predictive insights, inspection realities, and vessel features. This balanced approach mirrors the real-world logic of inspection planning, where decisions draw on a tapestry of factors rather than relying on a single factor.

In summary, the hybrid analysis demonstrates that predictive outputs can be integrated into structured inspection-support reasoning without compromising transparency or consistency. The resulting prioritisation patterns are analytically grounded, interpretable, and aligned with inspection-oriented decision contexts, thereby providing a coherent link between data-driven modelling and practical inspection planning.

4.5. Summary of Empirical Findings

This chapter presents empirical results from applying ML models and the hybrid decision-support framework to PSC inspection data for OSVs. The analysis demonstrates that inspection records contain structured information that can be utilised to model detention-related risk patterns with sufficient reliability for further analytical applications.

Interpretability analysis indicates that predictive outcomes are influenced by combinations of inspection history, vessel characteristics, and operational context, rather than by any single factor. These patterns are consistent with the operational realities of OSV activity and inspection practice. When incorporated into a structured decision framework, predictive outputs contribute to inspection-priority outcomes that are stable, transparent, and context-sensitive. The hybrid structure mitigates model-specific variability and supports coherent prioritisation without dependence on automated decision rules. The findings presented in this chapter establish an empirical foundation for evaluating the scientific and practical contributions of the proposed framework, which are addressed in the subsequent chapter.

5. DISCUSSION AND CONTRIBUTION

5.1. Purpose and Scope of the Discussion

In the previous chapter, we explored how the proposed analytical system performed on PSC inspection data for OSVs. The results reveal that ML models can uncover meaningful patterns within inspection records, while XAI techniques shed light on how predictions are formed. Together, these elements create a transparent framework that supports informed decision-making (Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G., 2019).

Rather than repeating the empirical results, this chapter examines their deeper meaning and what they reveal about inspection risk in OSVs. The discussion moves from simply describing findings to thoughtfully analysing them, considering how the trends can be put into practice, what new insights the integrated framework offers, and where its limitations lie.

Three core questions guide this section. First, what new insights arise when predictive modelling, interpretability, and decision structuring are viewed together instead of separately? Second, how does this integrated approach add value beyond traditional methods in PSC inspection analysis? Third, how can these outcomes be understood within the realities of regulatory practice and inspection governance?

5.2. Interpretation of Key Empirical Insights

The empirical findings collectively indicate that detention risk for OSVs is shaped more by the alignment of operational, technical, and inspection-related factors than by any single indicator. Across various model designs and validation settings, detention outcomes were consistently associated with combinations of conditions rather than with a single dominant variable. This observation is central to understanding why simplified scoring systems often fail to capture the complexity of risk trends in offshore operations.

For OSVs, factors such as vessel age, inspection history, and deficiency patterns do not exert uniform influence across all contexts. Their impact depends on interactions with operational roles, inspection frequency, and prior compliance trajectories. For example, recurring minor defects may be individually insignificant but can substantially elevate detention risk when combined with specific operational profiles or inspection histories. This interaction-driven structure was evident in both predictive performance and interpretability analyses.

A second insight concerns the relationship between forecast precision and operational utility. While several ML models demonstrated strong discrimination between detained and non-detained inspections, differences in calibration and stability were observed across modelling approaches. The evaluation indicates that robust inspection support depends on consistent and interpretable risk signals.

This finding reinforces a key design decision: prediction is regarded as an evidential layer rather than a prescriptive decision rule. Predictive outputs inform subsequent interpretation and prioritisation but do not directly determine inspection outcomes. In this manner, predictive modelling supports inspection judgment without constraining it.

Interpretability analysis further elucidates this role. The application of XAI techniques demonstrated that dominant predictive drivers generally align with established PSC knowledge and highlighted the variability in their influence across contexts. Notably, these explanations pertain to model behaviour. Their principal value is in providing transparency and analytical validation, rather than causal explanation. Maintaining this distinction explicitly prevents overinterpretation and ensures alignment with regulatory criteria.

5.3. Contribution of the Integrated Framework

This dissertation's main contribution is not the creation of new analytical methods, but the integration of prediction, interpretability, and decision structuring into a single inspection-support framework. Each part is already known in the literature, but bringing them together helps solve problems that none of them can address individually.

The framework overcomes these challenges by organising the analytical steps in a specific order. The prediction model identifies risk signals in complex inspection data. Interpretability helps examine how the model works and how it matches inspection knowledge. Decision structuring makes prioritisation clear and consistent. Together, these features make the framework well-suited for regulatory inspections.

The proposed configuration brings together three complementary methods in a structured way. Predictive modelling identifies patterns in inspection data that indicate a higher risk of detention. Interpretability techniques help explain how the variables interact, and MCDA translates these insights into clear inspection priorities. These methods do not work separately; instead, they build on each other in a single, unified approach. Unlike traditional PSC methods that use fixed scoring rules or separate statistics, this structure links prediction, explanation, and prioritisation to better match the real needs of inspection planning.

5.4. Significance for PSC Practice

The findings of this dissertation have several implications for inspection-based maritime safety oversight, particularly for OSVs (Hollnagel, E., 2014). These implications primarily concern decision support rather than the automation of inspection processes. At the operational level, the results indicate that inspection preparation may be enhanced by analytical tools that identify interaction-driven risk profiles. Rather than focusing solely on vessels with extreme values on individual indicators, inspection authorities could prioritise vessels whose risk arises from specific combinations of conditions. This approach aligns with current inspection practices, in which contextual judgment is frequently decisive.

At the organisational level, the framework demonstrates that predictive analytics can be integrated without compromising transparency. This integration is feasible only when predictive outputs are comprehensible and incorporated within explicit prioritisation logic. Such alignment supports internal review, fosters organisational learning, and enables informed dialogue between inspectors and administrators. From a regulatory perspective, the outcomes imply that advanced analytics can support inspection planning without displacing professional judgement. The configuration does not prescribe inspection actions; instead, it provides structured input that remains auditable.

The analysis shows that the detention risk for OSVs arises from the interactions among factors, not from any single indicator. Variables such as inspection history, previous deficiencies, and recent inspection patterns always affect predictions, but their influence varies with vessel characteristics and operating conditions. For instance, the importance of past deficiencies depends on the vessel’s age, its role, and the time since its last inspection. This means that risk signals in PSC data depend on context, not just on their own. These interactions highlight the complexity of OSV operations and explain why simple scoring systems that add up indicators often miss important details about inspection risk.

6. IMPLICATIONS

6.1. Practical and Regulatory Implications

The findings presented have direct implications for PSC authorities, inspection planners, and maritime administrations responsible for risk-based inspection strategies. The proposed framework demonstrates that inspection prioritisation can be improved by advancing beyond static scoring systems to adopt analytically structured decision support that incorporates predictive evidence, interpretability, and expert judgement.

For inspection authorities, the principal implication is methodological rather than technological (Saltelli, A. et al., 2020). The framework does not require replacing existing PSC regimes or inspection procedures. Instead, it provides a systematic mechanism for converting existing inspection data into risk signals that are both analytically robust and operationally interpretable.

This approach aligns with the practical constraints encountered by PSC organisations, where transparency, accountability, and consistency among inspectors and across regions are critical.

Specifically, the results indicate that OSVs benefit from vessel-type–specific inspection prioritisation approaches. The empirical analyses show that detention risk for OSVs is influenced by interacting technical, operational, and organisational factors that are not sufficiently addressed by generic risk matrices. Consequently, applying uniform scoring schemes across diverse vessel categories may reduce inspection effectiveness and obscure vessel-type–specific risk determinants.

From a regulatory standpoint, the framework advocates transitioning to decision-support inspections rather than automated risk ranking. Predictive outputs are provided as structured inputs to inform human decision-making, rather than as deterministic inspection directives. This distinction is essential for maintaining regulatory legitimacy and ensuring that inspection outcomes remain defensible and consistent with legal and administrative requirements.

6.2. Methodological Contributions and Transferability

While designed for OSVs, the framework also brings fresh methodological insights to maritime safety analytics. Its key innovation is the stepwise fusion of ML, XAI, and MCDM into a single, inspection-focused system. Rather than treating forecasting accuracy, interpretability, and prioritisation as rivals, the model shows how these elements can work together within a clear decision-making structure. This strategy addresses a long-standing hurdle in risk-sensitive fields, where cutting-edge analytics often struggle to gain traction due to unclear transparency or uncertain practical value.

Thanks to its modular design, the framework is highly adaptable. While this research spotlights OSVs in PSC inspections, the same analytical backbone can be tailored to other vessel types or regulatory settings, as long as the right domain variables and expert insights are included (Goerlandt, F., Khakzad, N., & Reniers, G., 2017). This flexibility positions the framework as a valuable blueprint for future studies that aim to blend data-driven modelling with regulatory decision-making.

6.3. Limitations of the Study

This dissertation has several limitations that should be acknowledged. First, this analysis uses PSC inspection data, which are not collected randomly. The choice of inspections depends on targeting rules, regional practices, and operational goals, which introduces selection bias that cannot be removed. While the modelling approach considers these factors, the results should be interpreted with reference to how the data were generated. Second, the framework is meant to help with risk prioritisation, not to explain causes. Although explainable AI methods were used to clarify the process, the relationships identified show how the model works, not the actual causes. The model is therefore meant to guide inspection planning, not to prove what causes accidents or deficiencies. Third, expert judgement is a key part of the MCDM component. This helps clarify the framework and aligns with regulatory practice, but it also introduces subjectivity. The results of the prioritisation process can vary depending on the experts, their experience, and their institutional background.

Finally, the study covers only OSVs operating within PSC regimes where data were available. While the methods can be used elsewhere, the findings should not be applied to situations with different operational or regulatory settings. These limitations define the scope of the results in this dissertation. Instead of weakening the findings, they clarify the specific situations in which the results apply.

6.4. Directions for Future Research

Overall, the results show that using a mix of existing methods rather than developing new algorithms is important for improving inspection-support analytics for complex vessel types such as OSVs. A promising next step is to apply this framework to other specialised vessel types, such as passenger ships, chemical tankers, and offshore construction units. Comparing these groups could reveal whether the risk patterns observed for OSVs also occur in other complex vessel types.

Future research could also examine dynamic risk models that account for changes in inspection results and operations over time. Using data collected over longer periods may help identify emerging risk trends and support more proactive inspection strategies.

Another area to explore is adding more data sources, like onboard sensor data, voyage details, or near-miss reports. While these are not yet part of standard PSC systems, including them could improve inspection prioritisation models if regulatory and data governance issues are addressed.

Finally, future studies could examine how inspection authorities use decision-support tools in real-world settings. Studying how inspectors respond to these tools would give helpful insights into how people and systems work together in risk-based inspections.

7. CONCLUSIONS

7.1. Summary of Investigation Aims and Approach

This research examined a key problem in today's maritime safety governance: a growing gap between the large volumes of inspection data available to PSC authorities and the analytical systems used to prioritise inspections, especially for OSVs. While inspection databases have become more detailed, traditional risk assessment tools have not kept up. They still use simple scoring methods that do not work well for vessels with complex, changing task operations and working environments.

To address this problem, the dissertation introduced and tested an integrated analytical approach that uses ML, XAI, and MCDM in a decision-support setting for inspections. This framework is not meant to automate inspections. Instead, it offers a structured way to turn inspection data into clear, transparent, and helpful prioritisation results, with human oversight at every step.

The research used real PSC inspection data for OSVs and was guided by three main hypotheses (H1–H3). These shaped the study's methods, interpretations, and decision-support contributions.

7.2. Evaluation of Research Hypotheses

Hypothesis H1

ML algorithms can recognise complex, non-linear patterns in PSC inspection data for OSVs that are not adequately captured by traditional risk assessment methods.

The empirical results presented in Chapter 4 provide substantial support for Hypothesis H1. ML models demonstrated the capacity to capture interactions among technical, operational, and organisational variables that are challenging to represent using additive or rule-based scoring systems. The observed improvements in predictive performance confirm that PSC inspection outcomes for OSVs are shaped by combinations of variables rather than isolated deficiencies.

It is important to note that the findings do not indicate that ML replaces existing inspection logic. Instead, they demonstrate that data-driven models can reveal latent risk structures within inspection records, thereby complementing established regulatory approaches. Hypothesis H1 is therefore accepted, with the qualification that predictive performance alone is insufficient for regulatory decision-making without suitable interpretative mechanisms.

Hypothesis H2

XAI techniques can improve the transparency and understandability of ML-based inspection risk models, making them suitable for regulatory use.

Hypothesis H2 is supported by the interpretive clarity analysis conducted using SHAP-based explanations. The results indicate that XAI methods provide consistent and coherent insights into model behaviour, enabling the identification of influential variables and interaction effects across OSV inspection outcomes.

The analysis confirms that XAI outputs reflect model reasoning rather than causal relationships, a distinction explicitly acknowledged in the dissertation. Nevertheless, the interpretability layer bridges a significant gap between predictive analytics and regulatory accountability. By making model outputs traceable and explainable, XAI approaches enhance confidence in data-driven assessments and support informed human decision-making. Accordingly, Hypothesis H2 is accepted. The findings demonstrate that interpretability is not merely an optional enhancement but a structural requirement for applying ML in inspection-based safety governance.

Hypothesis H3

Integrating ML outputs with MCDM methods can improve inspection prioritisation by structuring predictive information into transparent and operationally usable rankings.

The integration of predictive outputs with MCDM techniques, as discussed in Chapters 4 and 5, provides clear evidence supporting Hypothesis H3. By translating probabilistic risk estimates and interpretative knowledge into structured prioritisation scores, the framework aligns analytical outputs with the practical requirements of inspection planning.

The use of MCDM does not improve predictive validity; instead, it introduces structure, transparency, and repeatability. This distinction is essential. The value of the hybrid framework lies in its ability to connect prediction, interpretation, and prioritisation within a coherent analytical logic that adheres to regulatory constraints and maintains human oversight.

Hypothesis H3 is therefore accepted. The findings indicate that effective inspection prioritisation arises not from prediction alone, but from the deliberate structuring of analytical outputs within a decision-support framework.

7.3. Scientific Contributions

The principal scientific contribution of this dissertation lies in the development and empirical evaluation of an integrated ML–XAI–MCDM framework tailored to inspection-based maritime safety assessment. Rather than introducing new algorithms, the research advances methodological understanding by demonstrating how existing analytical tools can be combined systematically to address the specific requirements of PSC environments.

Primary contributions consist of:

- An organised strategy for modelling detention-related risk for OSVs using PSC inspection data.
- An interpretability layer that clarifies model behaviour and supports transparent regulatory reasoning.
- A decision-structuring mechanism that transforms analytical outputs into inspection-relevant prioritisation under governance constraints.
- An explicit articulation of the epistemic limits based on data inspection support, distinguishing predictive insight from causal inference and automated decision-making.

Collectively, these contributions expand the analytical resources available for maritime safety research and offer a replicable framework for inspection-oriented decision support.

7.4. Real-World Applications

From a practical perspective, the proposed framework offers inspection authorities a structured means to enhance prioritisation practices without compromising existing regulatory procedures. The framework is compatible with current PSC data infrastructures and does not necessitate changes to inspection mandates or enforcement authority. By supporting risk-informed inspection planning rather than automated selection, the framework maintains the central role of professional judgement while strengthening the analytical foundation for its application. This balance between analytical sophistication and regulatory acceptability is particularly significant for vessel segments such as OSVs, where operational complexity challenges traditional assessment logic.

7.5. Concluding Remarks

This dissertation demonstrates that inspection-based maritime safety assessment benefits from integrated analytical frameworks that combine predictive modelling, interpretability, and structured decision support. For OSVs, where safety risks result from complex interactions rather than isolated deficiencies, such integration is not only valuable but essential. By coordinating advanced analytics with regulatory logic and human review, the research advances a more transparent, accountable, and context-sensitive approach to inspection prioritisation. The framework developed in this study provides a foundation for future analytical developments in maritime safety governance while observing the institutional realities within which such systems must operate.

8. REFERENCES

Anderson, J. E., Heine, L., & Polatidis, H. (2021). Risk-based regulatory inspection: Principles, practices, and challenges. *Safety Science, 139,* 105255. https://doi.org/10.1016/j.ssci.2021.105255

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research, 253*(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023

Baldauf, M., Fischer, S., Schröder-Hinrichs, J. U., & Kataria, A. (2011). Proactive risk management in ship operations: Addressing human–system interaction. *WMU Journal of Maritime Affairs, 10*(2), 197–215. https://doi.org/10.1007/s13437-011-0016-7

Bansal, A., Farahani, A., & Schölkopf, B. (2021). Stability and robustness of machine learning models. *Foundations and Trends in Machine Learning, 14*(5), 425–540. https://doi.org/10.1561/2200000079

Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach.* Kluwer Academic Publishers.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* Springer.

Boko, Z., Skoko, I., Sanchez-Varela, Z., & Pinčetić, T. (2024). Application of advanced algorithms in Port State Control for offshore vessels using a classification tree and multi-criteria decision-making. *Journal of Marine Science and Engineering, 12*(11), 1905. https://doi.org/10.3390/jmse12111905

Boko, Z., Skoko, I., Sanchez-Varela, Z., & Milin, V. (2025a). Advancing maritime safety: A literature review on machine learning and multi-criteria analysis in PSC inspections. *Journal of Marine Science and Engineering, 13*(5), 974. https://doi.org/10.3390/jmse13050974

Boko, Z., Stanivuk, T., Radanović, N., & Skoko, I. (2025b). Machine learning-driven prediction of offshore vessel detention outcomes using Port State Control inspection data. *Journal of Marine Science and Engineering, 13*(3), 472. https://doi.org/10.3390/jmse13030472

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cariou, P., & Mejia, M. Q. (2017). Determinants of Port State Control inspection outcomes. *Maritime Policy & Management, 44*(3), 1–15. https://doi.org/10.1080/03088839.2017.1303130

Cariou, P., Mejia, M. Q., & Wolff, F.-C. (2008). Evidence on target factors used for Port State Control inspections. *Maritime Policy & Management, 35*(1), 1–17. https://doi.org/10.1080/03088830701848934

Cariou, P., & Wolff, F.-C. (2011). Ship inspection and flag performance. *Marine Resource Economics*, *26*, 71–90.

Cazzulo, A., et al. (2020). Machine learning models for ship machinery condition monitoring. *Ocean Engineering*, *216*, 108103.

Celik, M. (2009). Designing an integrated quality and safety management system for shipping operations. *Safety Science*, *47*(5), 569–577. https://doi.org/10.1016/j.ssci.2008.07.017

Celik, M., Lavasani, S. M., & Wang, J. (2021). A risk-based modelling approach for maritime safety assessment using data-driven techniques. *Ocean Engineering*, *234*, 109245. https://doi.org/10.1016/j.oceaneng.2021.109245

Dalaklis, D., et al. (2018). Multi-criteria decision methods in maritime operations. *WMU Journal of Maritime Affairs*, *17*, 231–249.

Dinis, D., Figueira, J. R., & Teixeira, A. P. (2023). A multiple-criteria approach to ship risk classification. *Socio-Economic Planning Sciences*, *90*, 101718. https://doi.org/10.1016/j.seps.2023.101718

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*.

European Maritime Safety Agency (EMSA). (2023). *The EQUASIS information system: Overview and data content.* https://www.equasis.org

Goerlandt, F., & Montewka, J. (2015). Maritime risk frameworks: A review. *Safety Science*, *76*, 245–265.

Goerlandt, F., Khakzad, N., & Reniers, G. (2017). Validity and validation of safety-related quantitative risk analysis. *Safety Science*, *99*, 127–139. https://doi.org/10.1016/j.ssci.2017.03.020

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Guidotti, R., et al. (2019). A survey of methods for explaining black-box models. *ACM Computing Surveys*, *51*(5), Article 93. https://doi.org/10.1145/3236009

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research*, *37*(4), 401–411. https://doi.org/10.1016/j.jsr.2006.04.007

Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management.* Ashgate.

International Maritime Organisation. (2009). *Guidelines for vessels with dynamic positioning systems (MSC.1/Circ.1580).* IMO Publishing.

International Maritime Organisation. (2014). *Code for the construction and equipment of mobile offshore drilling units (MODU Code).* IMO Publishing.

Knapp, S. (2006). *The effectiveness of Port State Control.* Doctoral dissertation, Erasmus University Rotterdam.

Knapp, S., & Bijwaard, G. E. (2009). The impact of inspection frequency on ship safety. *Marine Policy, 33*(3), 424–432. https://doi.org/10.1016/j.marpol.2008.09.002

Knapp, S., & Franses, P. H. (2007). Econometric analysis on the effect of Port State Control inspections on ship safety. *Marine Policy, 31*(4), 550–563. https://doi.org/10.1016/j.marpol.2006.11.004

Kuhn, M., & Johnson, K. (2013). *Applied predictive modelling.* Springer.

Lai, Y., Wang, J., & Li, K. X. (2023). Risk characteristics and inspection outcomes of offshore support vessels. *Ocean Engineering, 268*, 113366. https://doi.org/10.1016/j.oceaneng.2022.113366

Lalla-Ruiz, E., Shi, X., & Voß, S. (2018). Machine learning in maritime risk analysis: A review. *Maritime Economics & Logistics, 20*(3), 1–19. https://doi.org/10.1057/s41278-017-0094-2

Lipton, Z. C., Wang, Y.-X., & Smola, A. J. (2016). Detecting and preventing leakage in machine learning. *Journal of Machine Learning Research, 17*(1), 1–35.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

Panagakos, G., Santos, J., & Kontovas, C. (2017). Evaluating Port State Control performance. *WMU Journal of Maritime Affairs, 16*, 345–364.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A, 30*(3), 286–297. https://doi.org/10.1109/3468.844354

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *KDD Proceedings*, 1135–1144.

Roy, B. (1996). *Multicriteria methodology for decision aiding.* Kluwer.

Rudin, C. (2019). Stop explaining black-box machine learning models. *Nature Machine Intelligence, 1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Saaty, T. L. (1980). *The analytic hierarchy process.* McGraw-Hill.

Saltelli, A., et al. (2020). Five ways to ensure that models serve society. *Nature, 582*(7813), 482–484. https://doi.org/10.1038/d41586-020-01812-9

Schroeder, J., Flin, R., & Mearns, K. (2015). Human factors in Port State Control inspections. *Safety Science, 77*, 44–52. https://doi.org/10.1016/j.ssci.2015.02.004

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organisational decision-making structures in the age of AI. *California Management Review*, *61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Zavadskas, E. K., et al. (2016). Hybrid multiple-criteria decision-making methods. *Engineering Applications of Artificial Intelligence*, *47*, 1–16. https://doi.org/10.1016/j.engappai.2015.10.003

Zhang, D., Yan, X., Yang, Z., & Wall, A. (2022). Applications of machine learning in maritime safety. *Reliability Engineering & System Safety*, *217*, 108055. https://doi.org/10.1016/j.ress.2021.108055

Zio, E., & Aven, T. (2013). Uncertainties in smart-grid behaviour and modelling. *Energy Policy*, *61*, 106–115. https://doi.org/10.1016/j.enpol.2013.06.039

LIST OF FIGURES AND DIAGRAMS

LIST OF TABLES

LIST OF ABBREVIATIONS

AHP: Analytic Hierarchy Process
ANN: Artificial Neural Network
DP: Dynamic Positioning
EQUASIS: European Quality Shipping Information System
IMO: International Maritime Organisation
ISM: International Safety Management (Code)
MCDM: Multi-Criteria Decision-Making
ML: Machine Learning
MoU: Memorandum of Understanding
OSV: Offshore Support Vessel
PSC: Port State Control
RF: Random Forest
ROC: Receiver Operating Characteristic
SHAP: Shapley Additive Explanations
TOPSIS: Technique for Order Preference by Similarity to Ideal Solution
XAI: Explainable Artificial Intelligence

TERMINOLOGY AND DEFINITIONS:

Accuracy

The proportion of correctly classified inspection outcomes relative to the total number of observations. In imbalanced inspection datasets, accuracy alone may provide a misleading assessment of model performance.

Area Under the ROC Curve (AUC-ROC)

A performance metric representing the probability that a model ranks a randomly chosen high-risk vessel higher than a randomly chosen low-risk vessel across all classification thresholds.

Calibration

The degree to which predicted probabilities match observed outcome frequencies indicates the reliability of probabilistic model outputs for inspection decision support.

Decision support

Analytical assistance provided to human decision-makers to inform judgment and prioritisation without replacing professional responsibility or regulatory authority.

Explainable Artificial Intelligence (XAI)

A class of methods designed to enhance the transparency of machine learning models by providing interpretable explanations of model outputs in safety-critical and regulatory contexts.

F1-score

The harmonic mean of precision and recall provides a balanced measure of classification performance when false positives and false negatives carry different operational consequences.

False Negative (FN)

An outcome in which a vessel associated with elevated inspection risk is incorrectly classified as low risk, potentially resulting in missed inspection opportunities.

False Positive (FP)

An outcome in which a vessel is incorrectly classified as high risk, potentially leading to unnecessary allocation of inspection resources.

Imbalanced data

A dataset in which outcome classes occur with substantially different frequencies. In port state control inspection data, detentions typically represent a minority class, requiring careful selection of evaluation metrics and modelling strategies.

Inspection prioritisation

The process of ranking vessels for inspection based on structured analytical outputs, risk indicators, and regulatory criteria, without implying automated selection of inspections.

Interpretability

The extent to which the behaviour and outputs of an analytical model can be understood, explained, and meaningfully assessed by human decision-makers.

Machine Learning (ML)

A class of data-driven modelling techniques that identify patterns and relationships within data without requiring explicit rule-based specification.

Precision

The proportion of correctly identified high-risk vessels among all vessels classified as high risk by the model.

Predictive modelling

The use of statistical or machine learning techniques to estimate the likelihood of inspection outcomes based on historical data, without implying causal inference.

Recall (Sensitivity)

The proportion of correctly identified high-risk vessels among all vessels that are genuinely associated with elevated inspection risk.

Receiver Operating Characteristic (ROC) curve

A graphical representation of the trade-off between actual positive rate and false positive rate across different classification thresholds.

Selection bias (PSC context)

A structural property of port state control inspection data arising from targeted inspection strategies rather than random sampling, influencing observed inspection outcomes.

Shapley Additive Explanations (SHAP)

A model-agnostic explainability method that quantifies the contribution of individual input features to model predictions based on cooperative game theory.

True Negative (TN)

An outcome in which a low-risk vessel is correctly classified as low risk.

True Positive (TP)

An outcome in which a vessel associated with an elevated inspection risk is correctly classified as high risk.

APPENDICES

Appendix A

Paper I

Boko, Z., Skoko, I., Sanchez-Varela, Z., & Pincetic, T. (2024). Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making. Journal of Marine Science and Engineering, 12(11), 1905. https://doi.org/10.3390/jmse12111905

Journal of
Marine Science and Engineering

MDPI

Article

Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making

Zlatko Boko *, Ivica Skoko, Zaloa Sanchez-Varela and Tony Pincetic

Faculty of Maritime Studies, University of Split, 21000 Split, Croatia; iskoko@pfst.hr (I.S.); zsanchezv@pfst.hr (Z.S.-V.); tpincetic@pfst.hr (T.P.)
* Correspondence: zboko@pfst.hr

Abstract: This article examines the methods and application of classification trees and multi-criteria decision-making in the process of holding offshore vessels in port (Port State Control—PSC). This work aims to improve the efficiency and precision of the control processes in the detention of offshore vessels by using advanced analytical methods. Methodologically, a classification decision tree was used to identify the most important risk factors, enabling a faster and more accurate assessment of the possibility of detaining offshore vessels in port. Multi-criteria decision-making (MCDM) also enabled the simultaneous assessment of multiple factors, ensuring a balanced, robust, accurate, and objective approach. The research results show that the integration of these methods into the PSC process can significantly increase the safety of shipping and reduce the operating costs of offshore vessels. The application of these analytical tools can lead to a more systematic and transparent inspection process. This paper suggests further research and training of inspectors in the use of these techniques to maximize their applicability and effectiveness. Finally, this paper emphasizes the potential of classification trees and MCDM for safer and more efficient maritime transport by improving PSC procedures.

Keywords: port state control; offshore vessels; classification decision trees; multi-criteria decision-making methods

check for updates

Citation: Boko, Z.; Skoko, I.; Sanchez-Varela, Z.; Pincetic, T. Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making. *J. Mar. Sci. Eng.* **2024**, *12*, 1905. https://doi.org/10.3390/jmse12111905

Academic Editor: Xinqiang Chen

Received: 24 September 2024
Revised: 11 October 2024
Accepted: 22 October 2024
Published: 24 October 2024

1. Introduction

Port State Control (PSC) is important to the international maritime safety and pollution prevention strategy. It implements international conventions to ensure the standard operation of vessels regarding safety, security, pollution prevention, and living conditions for the crew.

PSC inspections aim to identify and rectify vessel deficiencies to reduce the risk of accidents, pollution, and other problems at sea. The inspections are carried out by qualified inspectors from national maritime administrations. They have the right to inspect the vessels, talk to the crew, and check the documentation. The inspections can be routine, targeted, or risk-based, with particular attention paid to vessels with a higher risk of defects. PSC is, therefore, a measure used by a port authority or an authorized subject (flag, agent, or recognized organization) to inspect aspects of foreign vessels calling at a country to ensure that they comply with the standards of international conventions.

PSC is carried out by international conventions adopted under the auspices of the International Maritime Organization (IMO), such as SOLAS (International Convention for the Safety of Life at Sea), MARPOL (International Convention for the Prevention of Pollution from Vessels), and MLC (Maritime Labour Convention).

The PSC aims to safeguard the living conditions of seafarers, protect human life and property, and reduce pollution of the sea, coasts, and the environment as a whole. Possible vessel problems can relate to safety, health, and pollution. Typical inspection points are vessel certificates and documentation; safety and rescue equipment; crew working conditions and their health and safety; the structural integrity of the vessel; and environmental protection standards. Suppose a vessel encounters a problem with life or property. In this case, it should not be allowed to continue its voyage unless there is a case of force majeure and there is an urgent need to resolve the situation that poses a threat to human life or the vessel. If deficiencies are found, the vessel may be detained until the problems are resolved. It should remain in position if necessary and remedial action should be taken to bring it into a safe and efficient condition for further voyages. In serious cases, the vessel may be forced to leave port or cease operations or sailing.

To improve vessel performance and crew behavior, PSC popularizes the PSC system to, directly and indirectly, influence the entire maritime industry, including vessel management, seafarer training institutions, vessel management authorities, vessel classification societies, and vessel financiers through inspections, detentions, information on inspections, frequent target lists issued under the PSC MoU, and regional training and cooperation platforms. Any poor performance of the vessel or crew may result in the vessel being detained by the port authority immediately before departure or at the next port of call.

PSC inspections are organized and coordinated at the regional level within the framework of Memorandums of Understanding (MoU), such as the Paris MoU [1], the Tokyo MoU [2], and the Caribbean MoU [3]. These agreements enable the exchange of information and common control standards. The results of PSC inspections are recorded and have an impact on the reputation of the vessel and company owner/operator. Vessels with poor PSC inspection results may have problems obtaining orders and calling at ports. Therefore, PSC is an important tool for improving safety at sea and reducing accidents and pollution in maritime transport, which helps to protect human life and the environment.

Ocean HQ [4] lists the main reasons for detaining a vessel in port. For a vessel owner or operator, port control is not a favorite topic of conversation, because if the port authority detains due to certain deficiencies, it is very likely that additional port charges will be levied. The charter price will increase, and such delays also hurt the vessel's profitability.

Each port state has the right to independently inspect all vessels calling at its ports, whereby the inspector (PSCO—Port State Control Officer) is authorized to detain the vessel if they think that it has deficiencies until the deficiencies have been rectified. The IMO sees PSC as a means of raising standards in the shipping industry, and many of the key IMO conventions contain provisions requiring governments to inspect foreign vessels calling at their ports to ensure that they comply with IMO standards.

To further promote uniformity in PSC standards, the IMO has encouraged the regionalization of countries so that countries in the same region form a PSC system. Agreements have been signed for Europe and the North Atlantic (Paris Agreement), Asia and the Pacific (Tokyo Agreement), Latin America (Viña del Mar Agreement), the Caribbean (Caribbean Agreement), West and Central Africa (Abuja Agreement), the Black Sea region (Black Sea Agreement), the Mediterranean (Mediterranean Agreement), the Indian Ocean (Indian Ocean Agreement), and the Persian Gulf (Riyadh Agreement).

PSCs often announce their review campaigns in advance and usually focus on new regulations that have recently come into force. However, vessel operators should not only focus on meeting the requirements of the new regulations or only doing what the PSC regime has announced in the campaign. Most detentions of the vessels due to PSC inspections are due to a general lack of maintenance. This means that vessels are not operating under a system of safety organization by the International Safety Management Code (ISM Code) [5], which includes relevant procedures for the vessel's maintenance and equipment.

Given the complexity and multidimensionality of the criteria used in vessel inspections, there is a need to apply advanced analysis and decision-making methods that can improve the inspection process. Classification decision trees and multi-criteria decision-making (MCDM) methods are powerful tools for analyzing the behavior of complex systems such as PSCs. On the one hand, classification decision trees allow easy decision-making based on available historical data about previous inspections. On the other hand, MCDM integrates different criteria in the process of optimal decision-making.

Therefore, the main assumptions and motivations for the research in this scientific article are as follows:

1. Improving the efficiency of PSC inspections through the application of classification trees and MCDM will significantly increase the accuracy in identifying offshore vessels that pose a higher risk or have certain deficiencies, thereby optimizing the allocation of resources and increasing the efficiency of inspections.
2. Improving maritime safety through the use of advanced analytical methods will enable more accurate prediction and detection of potential safety risks, more precise prediction of the possibilities of detention of offshore vessels due to the presence of certain irregularities, a reduction in the number of incidents, an increase in safety at sea, and a reduction in the time the vessel spends in port and thus the associated financial costs.
3. More effective inspections of offshore vessels contribute to a better understanding of potential risks, marine pollution reduction, and better marine ecosystem protection.

Through the application of modern technologies in the maritime industry, this work contributes to the dissemination of knowledge, the application of modern technologies and methods, and the promotion of innovation and improvement of existing practices. Empirical research with available data on previous PSC inspections of offshore vessels and in real conditions will provide concrete results that prove their effectiveness by suggesting concrete improvements for this complex process. In this way, this article will contribute to the development of PSC practice, enable a better understanding and application of advanced decision-making methods in the maritime industry, and improve safety on a global scale.

This paper is divided into six chapters. After the introduction, the second chapter gives an overview of previous research on PSC and the application of various mathematical, statistical, and other approaches, including machine learning (ML) applications using classification trees and MCDM to optimize the decision-making process in PSC. The third chapter briefly explains the three-stage methodological approach, which consists of 1. the application of the classification tree "rpart", 2. the MCDM TOPSIS method, and 3. the integration of the results of "rpart" and TOPSIS.

The fourth chapter describes the data source and structure in detail and explains the three-step approach to solving the problem: 1. applying the classification decision tree "rpart"; 2. applying the TOPSIS method; and 3. combining the results of both algorithms. The fifth chapter documents, reviews, and evaluates the results obtained, while the last chapter is dedicated to discussion. At the end, you will find a conclusion, a list of biographies, and a list of abbreviations used.

2. Theoretical Background of the Integration of Machine Learning Algorithms and Methods in PSC

Although the basic international legal framework for PSC was established by various IMO conventions, including SOLAS, MARPOL, and MLC, PSC has evolved over the years into an important instrument for these conventions. As PSC is a key element of global maritime safety and environmental protection, research to date shows considerable scientific interest from researchers contributing in various ways by investigating and demonstrating the positive impact of PSC inspections on vessel safety, seafarers' working conditions, and environmental protection, all to reduce the number of maritime accidents and incidents, improve overall safety, and reduce pollution.

The Maritime Anti-Corruption Network Report [6] promotes some of the practices that have emerged from the challenges facing the maritime industry. This is because the interactions of PSCOs during inspections are sometimes very complex. The report therefore clarifies the discretionary rights and powers of PSCOs, which on the other hand make it difficult for seafarers to plan the vessel's sailing in port or to comply with the conditions, as well as obtain a clear interpretation of the rules, which should be harmonized with those applicable internationally.

In dissertation [7], the researcher highlights the challenges associated with ensuring sufficient resources and capacity to carry out PSC inspections. PSC is recognized as an internationally accepted safety enforcement program that detects and corrects deficiencies, promotes best practices, supports compliance, and collects and evaluates safety data. However, the author points out that all interested parties in this process play a key role, namely the shipping companies, the seafarers, and even PSC itself.

Study [8] analyzes the risk factors of vessel detention and identifies the most important factors of maritime safety and environmental protection. Using a dataset collected over six years and comprising a total review of 178,153 records from 2010 to 2015, the authors develop a Bayesian network (BN) model to analyze the influential inspection factors that lead to vessel detention.

The analysis includes data related to flag state, vessel type, recognized organization inspection body, and the vessel's age. The results of this study guide PSC officials and vessel owners in identifying critical areas to improve maritime safety. The results promote environmental sustainability and help to create a cleaner environment by suggesting a similar methodological approach for future PSC inspections.

Recognizing that PSC is an important measure to improve maritime safety and reduce the number of shipping accidents, study [9] points out that some MoUs have started to implement a new inspection regime (NIR), but its effectiveness has yet to be investigated. Therefore, the study reviews the NIR effectiveness and the degree of improvement in vessel safety. As in the previously cited study, the authors use the BN model to determine the relationship between the NIR, vessel defects, detentions, and marine casualties and to examine the impact of the NIR on maritime safety. Study [10] analyzes the factors responsible for detaining vessels under PSC using the Grey Rational Analysis (GRA) model with increased entropy weighting to understand how different factors influence the decision to detain a vessel. Entropy-weighted GRA is an advanced technique for solving multi-criteria decision problems combining elements of GRA and entropy weighting. GRA is a part of the grey system theory, which Julong Deng developed [11]. The empirical analysis conducted is based on data on vessel detention in the Asia–Pacific region collected under the Tokyo MoU over the past ten years. Study [12] uses derived data from the vessel arrival reporting system to identify a specific type of inspection vessel using a combination of the GRA model and the entropy weight method (EWM). In this case study, different types of arriving vessels at five selected ports in Malaysia were analyzed to determine the possible outcome of vessel inspections. Based on 100,623 records of vessel arrivals collected over five years (2015–2019), the arriving vessels were identified, analyzed, evaluated, and assessed.

In study [13], the authors use a concentrated inspection campaign (CIC) derived from the PSC Addendum, i.e., a defined individual series of inspections of vessels for non-compliance carried out in three consecutive months at the end of each calendar year. The study uses GRA, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), and the MCDM method to analyze 71,376 records with 496 possible deficiency codes for 21 vessel types according to the Paris MoU, to improve the inspection pattern. This study combines the three-sigma rule to identify elements of vessel inspection that are unlikely to be met, based on the decisions of Member States' Port Facility Security Officers (PFSOs). In study [14], the authors analyze PSC implemented by the port authority, i.e., the system for inspecting foreign vessels that enter the port and are detained because they do not meet the expected standards. The study analyzes the detection of critical deficiencies that influence the decision to detain the vessel by combining the cloud model, Criteria Interaction through the Inter-Criteria Correlation (CRITIC) method, and perspective theory. The cloud model is used to solve problems related to the possible uncertainties of the PSC inspection results, while the CRITIC method is used to determine the weighting values of defects and to describe the dependencies between different defects. In study [15], the authors use a variant of the Deck of Cards Method (DCM), which enables a subjective evaluation of decision-making by building an MCDA (multi-criteria decision analysis) model. The elements presented in the paper, from the criteria framework to the criteria scales, from the criteria value functions to the criteria weights, reflect the subjective judgment. Although the DCM was proposed by Simos [16], later improved by Figueira and Roy [17] and extended to other contexts by Corrente et al. [18], to determine the weighting of criteria in the originally conceived downloaded MCDA methods ELECTRE (Elimination and Choice Expressing Reality) and PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation), the DCM is used in this study to formalize the construction of the system MCDA model. However, the proposed model shows that the values lead to classifications that require additional input of data and resources by the PSC authorities, which is not always applicable.

Therefore, the authors of study [19] used data obtained through vessel inspections in the national ports of PSC to assess their compliance with radio traffic safety regulations by applying binary logistic regression. This study analyses the relationship between the severity of maritime communication deficiencies and some characteristics of the inspected vessels. To this end, the study uses 23,725 PSC inspection data and proposes several classification criteria to better assess the risk associated with maritime distress communication. This study uses regression to analyze inspection deficiencies associated with skill deficiencies among Global Maritime Distress and Safety System (GMDSS) operators.

Other studies [20] used and proposed other machine learning algorithms applicable in the shipping industry to evaluate the effectiveness of different machine learning algorithms, but did not analyze the details of each algorithm. The results of the cited study show that different machine learning algorithms can predict the braking performance of a 9500 TEU container vessel better than deviations based on a specific formula, resulting in a lower mean square error (MSE).

The implementation and evaluation of the classification tree model in PSC have been carried out in a series of case studies with empirical data that have been successfully implemented in different ports.

Paper [21] analyzes the spatio-temporal distribution of pollution caused by ships using large datasets from the Automatic Identification System (AIS). The authors analyze pollution patterns in the vicinity of ports and their impact on the marine environment. Using sophisticated analytical methods, the study emphasizes the importance of understanding these patterns to make informed decisions about the management of marine resources and the protection of the environment. The study also emphasizes the need for better monitoring and regulation of shipping traffic to reduce the negative impact on the ecosystem of ports.

Study [22] covers various machine learning methods, including classification decision trees, in the context of PSC. The research applies machine learning techniques, decision trees, and connectivity analyses to investigate the relationship between vessel selection in the pre-inspection phase and vessel defects in the inspection phase.

Research [23] incorporates unknown variables into a practical decision problem by predicting unknown variables using auxiliary data. The authors use the classical approach of model development by machine learning to generate point estimates in a deterministic decision technique. ML regression models are used for the prediction. The authors note that the quantitative objective is discrete and that the properties are analogous to the categorical objectives in the classification problems. Considering that it is much easier to describe and estimate the distribution of categorical variables than that of quantitative variables, this study innovatively proposes the use of random forest (RF) models in combination with regression and classification features to generate distributions of discrete quantitative targets. A review of previous research found that numerous studies have investigated the factors that influence the detention of vessels to suggest improvements to the effectiveness of PSC inspections. However, most investigations and studies rarely consider the three following aspects simultaneously: 1. the conditions under which offshore vessels are detained; 2. the application of mathematical methods and machine learning techniques, such as classification decision trees in combination with MCDM methods, to simplify the process of making decisions about the possibility of detaining the vessel; and 3. the application of methods that can predict whether or not a vessel with certain defects will be detained on the values of variables and the results of previous inspections. The aim of this study is therefore to create a systematic, robust, and accurate framework that combines classification trees and MCDM methods to identify the most important deficiencies in the inspection of offshore vessels and to predict whether a vessel will be detained based on these deficiencies and other variable factors.

The theoretical framework for analyzing the detention of vessels in ports comprises the theoretical knowledge of the importance of this process for the efficient management of maritime traffic. Vessel detention in ports directly impacts the business and economic performance of both vessels and ports. Additionally, detention can result in economic losses, logistical delays, and customer dissatisfaction. Therefore, it is crucial to identify the key factors for vessel detention to improve the overall efficiency of maritime transport.

The detention times of vessels have a direct impact on the utilization of port capacity, the optimal flow of cargo, and the speed of cargo handling. Understanding the importance of vessel detention in ports is crucial for the competitiveness of ports and the overall efficiency of maritime transport. Detention of vessels can be related to various deficiencies, such as technical problems, lack of human resources, administrative difficulties, or legal inadequacies. Technical deficiencies include inadequacies in equipment or infrastructure, while operational deficiencies may result from a lack of coordination of logistics processes. Legal and regulatory gaps can make it difficult to complete administrative procedures and customs formalities quickly.

Technical deficiency identification [24] is key to understanding the problem of vessel detention in ports. Operational deficiencies [25] refer to deficiencies in logistical processes or operational procedures that can slow down the detention of vessels in ports. Legal and regulatory deficiencies [26] refer to deficiencies in regulations, laws, or standards that may affect the detention of vessels in ports. The identification of these deficits is crucial for the development of strategies and measures to improve maritime transport as a whole.

Offshore vessels play an important role in various sectors, particularly in the oil and gas industry, and in other areas such as renewable energy, underwater research, rescue operations, and transport. The category of offshore vessels includes 1. drilling rigs; 2. support vessels, i.e., platform supply vessels (PSVs) and anchor handling supply vessels (AHTS); 4. construction and maintenance vessels; 5. tankers that can be used as floating production, storage, and offloading (FPSO) vessels; 6. research vessels; 7. renewable energy vessels; and 8. rescue vessels [27].

In this article, the deficiencies of offshore vessels are categorized by the practice and available provisions of SOLAS-ISM. Paris/Tokyo MoU [28] is only one of the determinants for classifying the "seriousness" of offshore vessel detention. The main determinants of the Tokyo MoU "deficiency" findings used in this paper are listed in Table 1 [29].

Table 1. Primary items of deficiency codes in Tokyo MoU.

Code	Description	Code	Description
01000	Certificates and Documentation	10000	Safety of Navigation
02000	Structural Conditions	11000	Life-Saving Appliances
03000	Water/Weathertight Conditions	12000	Dangerous Goods
04000	Emergency Systems	13000	Propulsion and Auxiliary Machinery
05000	Radio Communications	14000	Pollution Prevention
06000	Cargo Operations Including Equipment	15000	ISM (International Safety Management)
07000	Fire Safety	16000	ISPS (International Safety for Port and Vessels)
08000	Alarms	99000	Other
09000	Working and Living Conditions	18000	Labor Conditions

Classification decision trees are popular machine learning algorithms used for classification and prediction. There are several types of classification and regression decision tree algorithms, but some of the best-known and most commonly used are as follows:

- "rpart" (Recursive Partitioning and Regression Trees) [30] is used for classification and regression. CART (Classification and Regression Trees) [31] is a method for creating classification and regression trees for forecasting and modeling. CHAIN (Classification Hierarchy Analysis) [32,33] is not a standard method like "rpart" or CART.

In this study, the authors opted for a combination of the "rpart" classification decision tree and the TOPSIS multi-criteria decision method.

MCDM is a decision-making process in which several factors or criteria must be considered. The characteristics of this approach include analyzing several alternatives, quantifying the criteria, ranking the options, and finding the optimal solution. This process can be complex, as several parameters need to be simultaneously evaluated, and various methods are used to facilitate decision-making in such situations. The review of previous research presented in the second chapter of this article has shown that MCDM is increasingly present in the nautical industry, especially in the processes related to keeping vessels in port. Many studies have focused on method development and techniques that can improve the efficiency of decision-making in this context.

The benefits of applying MCDM to the detention of offshore vessels in port include the possibility of better decision-making, more efficient use of resources, and optimization of performance. However, challenges include the complexity of the criteria analysis process, the need for model validation, and adaptation to changing conditions. It is important to carefully balance the benefits and challenges to achieve optimal management of this process.

Some of the best-known methods of MCDM are as follows:

- AHP [34] is a method for structuring complex decisions in a hierarchical form. Criteria and alternatives are compared in pairs, and the result is a weighting of the criteria and alternatives that enables a decision to be made.
- TOPSIS [35] is a method based on the concept that the best alternative is the one that is closest to the ideal solution and furthest from the worst solution. The distance of each alternative to the ideal and anti-ideal solution is calculated.
- ELECTRE [36] is a method that uses the concept of dominance to compare alternatives. It considers the weighting of the criteria and the dominant relationships between the alternatives. PROMETHEE [37] is a method based on the comparison of pairs of alternatives and the expression of preferences for each alternative. The result represents the ranking of alternatives according to overall preferences. SAW (Simple Additive Weighting) [38] is a method in which each alternative is assigned an overall value resulting from the sum of the weighted scores for each criterion. The alternatives are ranked based on these totals.

3. Methodological Approach of Integration of Classification Trees and MCDM Methods in PSC

By integrating classification models and MCDM into the port management process of offshore vessels, it is possible to combine the advantages of both approaches for more efficient decision-making. Classification models provide a structure for data analysis, while MCDM incorporates different parameters and priorities into the decision-making process. This integration improves the quality of decisions and optimizes processes to keep vessels in port.

There are various methods for integrating classification models and MCDM into the port laytime process. The combination of methods aims to combine the advantages of classification models and MCDM for a more efficient and accurate decision-making process.

The "rpart" classification tree algorithm and the TOPSIS MCDM method are used in this study.

The algorithm "rpart" (Revised Partitioning Algorithm) [39] uses a decision tree approach for classification and regression. The steps of the procedure are as follows:

1. Selection of the best partitioning:
   - The algorithm starts with the entire data training set as the root.
   - For each variable, the best partitioning is calculated, i.e., the point that divides the data into two parts with the largest difference in target values. Criteria such as the Gini index or entropy (information gain) are used for classification [40].
2. Division of the data: The data are divided into two parts according to the best split. This process is repeated recursively for each new node.
3. Creating leaves: The splitting process continues until each node fulfills a stop criterion (e.g., minimum number of samples in a leaf, maximum tree depth, or if further splitting does not bring any significant improvement).
4. Assignment of classes to nodes: The leaves are classified according to the majority class of the data or according to the mean for regression tasks.
5. Pruning of neighborhoods: To reduce the complexity of the tree and avoid overfitting, pruning can be performed. This is carried out using methods such as "cost-complexity pruning" [41].

TOPSIS is a method for organizing alternatives according to several criteria. The steps of the TOPSIS method include the following:

1. Creating a decision matrix—creating a matrix that contains the scores for all alternatives according to all criteria.
2. Normalizing the matrix "r_{ij}" (total dataset: number of alternatives (i) $\times$ number of variables (j)) by eliminating the unit of measurement, e.g., by vector normalization:

$$r_{ij} \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (1)$$

   where x_{ij} is the value of alternative i according to criterion j.
3. Creation of a weighted, normalized matrix and use of weighting criteria:

$$\vartheta_{ij} = \omega_j \cdot d_{ij} \quad (2)$$

   where ω_j is the weight of criterion j.
4. Identification of ideal and anti-ideal points: The ideal solution A^* (best possible) and the anti-ideal solution A^- (worst possible) are determined as follows:

$$A^* = \vartheta 1*,\ \vartheta 2*, \ldots, \vartheta n* \quad (3)$$

   and

$$A^- = \{\vartheta_1^-, \vartheta_2^-, \ldots, \vartheta_n^-\} \quad (4)$$

   where ϑ_1^* and ϑ_1^- are the maximum and minimum values for criterion j.
5. Calculating the distance from the ideal Equation (5) and anti-ideal Equation (6) point: The distance of each alternative from the ideal and anti-ideal solution is calculated using the Euclidean distance:

$$D_i^* = \sqrt{\sum_{j=0}^{n} \left(\vartheta_{ij} - \vartheta_j^*\right)^2} \quad (5)$$

$$D_i^- = \sqrt{\sum_{j=0}^{n} \left(\vartheta_{ij} - \vartheta_j^-\right)^2} \quad (6)$$

6. Calculation of the relative approximation to the ideal solution Equation (7): The relative proximity of the individual alternatives to the ideal solution is calculated as follows:

$$C_i^* = \frac{D_i^-}{D_i^* + D_i^-} \quad (7)$$

where C_i^* is between 0 and 1. Higher values indicate a better alternative.
7. The alternatives are categorized by ranking the values of C_i^* from the highest to the lowest.

4. Research Methodology

The methodological approach pursued in this paper comprises several steps shown by a block diagram (Figure 1).

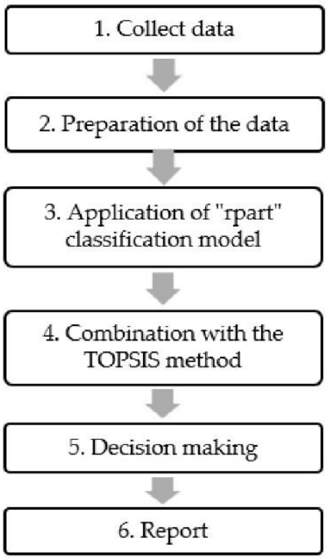

Figure 1. Blog diagram of the methodology used.

1. Collect vessel data:
   - Input: vessel data (e.g., vessel type, inspection history, equipment condition, and regulatory compliance);
   - Description: Collect all relevant information about the vessel to be used for the analysis.
2. Preparation of the data
   - Input: Raw data about the vessel;
   - Processes: Data cleansing, normalization, and transformation;
   - Output: Clear and formatted data, ready for analysis.
3. Application of the "rpart" classification model:
   - Input: Prepared data;

- Processes:
  Training: use the data to train the "rpart" model.
  Prediction: Use the trained model to predict whether the vessel needs to be detained.
- Output: Predict whether the vessel should be detained or not (e.g., "Yes" or "No").

4. Combination with the TOPSIS method:
   - Input: Predictions from the "rpart" model as well as additional criteria (e.g., safety and environmental impact).
   - Operations:
     Normalization and weighting of criteria;
     Calculation of ideal and anti-ideal solutions;
     Calculation of the distance to the ideal and anti-ideal solution;
     Ranking of alternatives.
   - Output: Ranking of the alternatives and final result (e.g., recommendation to keep or not to keep the ship).
5. Making the decision:
   - Input: Ranking results of the TOPSIS method;
   - Work steps: Analyze the results and make a decision on whether to keep the vessel;
   - Output: Decision (e.g., "Detain the vessel" or "'Don't detain the vessel").
6. Report:
   - Input: The decision made based on the analysis;
   - Processes: Creation of reports for PSC (Port State Control) with an explanation of the decision;
   - Output: Report for PSC.

4.1. Dataset Description

The input dataset is represented by the matrix $X = [x_{ij}]$, $i = 1, \ldots m; j = 1, 2 \ldots, n$, where m is the number of observed instances (alternative or inspection of offshore vessels) and n is the number of observed variables.

The source of the data used in this empirical study on the detention and control of offshore vessels is EQUASIS [42]. The input dataset consists of data on 405 inspections and 27 variable values. Since one of the variables is the name of the vessel, this variable is ignored in the following to avoid possible legal issues.

The target variable in the formation of the decision tree is "Detention" and it takes the values "Yes" or "No". If the offshore vessel is detained, this specification has the value "Yes" for a specific dataset describing the inspection, otherwise "No". The total number of recorded inspections that make up the input dataset is 405, of which 40 offshore vessels were detained and 365 were not.

Tables 2–9 describe all variables that define their type, range, or mean value, and the number of occurrences grouped according to the values of the variable "Detention".

Standardization or normalization of data is not required when most of the data are on the same scale (the case in this research). The criteria are in the same range, so no additional scaling is required. Furthermore, algorithms such as decision trees, random forests, or Naive Bayes do not rely on distances between the data and therefore do not require standardization or normalization. These algorithms work on the principle of decision rules, so differences in the scaling do not affect their accuracy. The TOPSIS method, on the other hand, requires normalization of the data in the first step of the approach.

Table 2. Variable number (Var No), variable name, representation and description, type, and possible values or range.

Var No	Variable_Name	Representation and Description	Type	Possible Values/ Range
1	Detention	Vessel Detention	Character	Yes/No
2	InitialInspection	Initial Inspection	Integer	[0, 1]
3	MoreDetailedInspection	More Detailed Inspection	Integer	[0, 1]
4	FollowUpInspection	Follow-Up Inspection	Integer	[0, 1]
5	StandardInspection	Standard Inspection	Integer	[0, 1]
6	NumberOfInspections	Number Of Inspections	Integer	[0, 1]
7	NumberOfDeficiency	Number Of Deficiency	Integer	[0, 38]
8	ISM	ISM Deficiency	Integer	[0, 31]
9	MARPOL	MARPOL Deficiency	Integer	[0, 3]
10	CertificateDocumentation	Certificate and Documentation Deficiency	Integer	[0, 4]
11	PropulsionAuxiliaryMachinery	Propulsion Auxiliary Machinery Deficiency	Integer	[0, 10]
12	SafetyOfNavigation	Safety Of Navigation Deficiency	Integer	[0, 2]
13	RadioCommunications	Radio Communications Deficiency	Integer	[0, 5]
14	EmergencySystems	Emergency Systems Deficiency	Integer	[0, 3]
15	FireSafety	Fire Safety Deficiency	Integer	[0, 4]
16	MLC	MLC Deficiency	Integer	[0, 5]
17	Alarms	Alarms Deficiency	Integer	[0, 1]
18	ISPS	ISPS Deficiency	Integer	[0, 5]
19	OtherTypeofDeficiencies	Other Types of Deficiencies	Integer	[0, 1]
20	WaterWeatherTightConditions	Water/Weathertight Condition Deficiency	Integer	[0, 6]
21	LifeSavingAppliances	Life-Saving Appliances Deficiency	Integer	[0, 2]
22	GT	GT—Gross tonnage of the offshore vessel	Integer	[2012, 80106]
23	CountryOfInspection (COI)	Country Of Inspection of the offshore vessel	Character	1 of 36 Countries
24	Flag	Offshore vessel's flag	Character	1 of 30 Flags
25	MoU	Memorandum of Understanding	Character	1 of 10 MoUs
26	YOB	Year of offshore vessel built	Character	1 of 26 YOBs

Table 3. Variable name and count of instances grouped by the variable Detention = "No" and Detention = " Yes".

Variable Name	All Dataset			Detention No			Detention Yes		
	Total	=0	=1	Total	=0	=1	Total	=0	=1
InitialInspection	405	192	213	365	172	193	40	20	20
FollowUpInspection	405	399	6	365	359	6	40	40	0
StandardInspection	405	399	6	365	362	3	40	37	3
ISPS	405	398	7	365	360	5	40	38	2

Table 4. Variable name, and count of instances grouped by the variable Detention = "No" and Detention = " Yes".

No	Variable_Name	All Dataset				Detention No				Detention Yes			
		Total	=0	=1	>1	Total	=0	=1	>1	Total	=0	=1	>1
1	MoreDetailedInspection	405	238	165	2	363	214	149	2	40	24	16	0
2	NumberOfInspections	405	0	16	389	365	0	15	350	40	0	1	39
3	NumberOfDeficiency	405	6	84	315	365	6	80	279	40	0	4	36
4	ISM	405	319	68	18	365	290	64	11	40	29	4	7
5	MARPOL	405	319	68	18	365	290	64	11	40	29	4	7
6	CertificateDocumentation	405	189	114	102	365	177	105	83	40	12	9	19
7	PropulsionAuxiliaryMachinery	405	381	19	5	365	346	18	1	40	35	1	4
8	SafetyOfNavigation	405	267	100	38	365	253	87	25	40	14	13	13
9	RadioCommunications	405	349	47	9	365	322	37	6	40	27	10	3
10	EmergencySystems	405	354	43	8	365	323	38	4	40	31	5	4
11	FireSafety	405	251	109	45	365	235	98	32	40	16	11	13
12	MLC	405	328	50	27	365	305	42	18	40	23	8	9
13	Alarms	405	391	11	3	365	357	7	1	40	34	4	2
14	OtherTypeofDeficiencies	405	306	79	20	365	282	65	18	40	24	14	2
15	WaterWeatherTightConditions	405	375	25	5	365	342	20	3	40	33	5	2
16	LifeSavingAppliances	405	306	67	32	365	280	60	25	40	26	7	7

Table 5. "GT"—gross tonnage of the offshore vessel; its minimum, mean, and maximum values are grouped by Detention = "No" and Detention = " Yes".

GT	Dataset			Detention No			Detention Yes		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
	2012	4939.3	80106	2012	5032.8	80106	2012	4086.6	6776

Table 6. Country Of Inspection (COI) (36 different occurrences), count of instances grouped by the variable Detention = "No" and Detention = " Yes".

No	COI	All Dataset			Detention No			Detention Yes		
		Total	=0	=1	Total	=0	=1	Total	=0	=1
1	Australia	405	360	45	365	328	37	40	32	8
2	Brazil	405	404	1	365	364	1	40	40	0
3	Bulgaria	405	402	3	365	362	3	40	40	0
4	China	405	393	12	365	355	10	40	38	2
5	Colombia	405	403	2	365	363	2	40	40	0
6	Cyprus	405	372	33	365	333	32	40	39	1
7	Denmark	405	387	18	365	347	18	40	40	0
8	Egypt	405	399	6	365	360	5	40	39	1
9	France	405	403	2	365	363	2	40	40	0
10	Germany	405	400	5	365	361	4	40	39	1
11	Ghana	405	402	3	365	362	3	40	40	0
12	Greece	405	399	6	365	361	4	40	38	2
13	Hong Kong	405	403	2	365	363	2	40	40	0
14	India	405	403	2	365	363	2	40	40	0
15	Israel	405	398	7	365	360	5	40	38	2
16	Italy	405	402	3	365	362	3	40	40	0
17	Japan	405	403	2	365	363	2	40	40	0
18	Malaysia	405	403	2	365	363	2	40	40	0
19	Malta	405	387	18	365	347	18	40	40	0
20	Morocco	405	404	1	365	364	1	40	40	0
21	Netherlands	405	389	16	365	350	15	40	39	1
22	New Zealand	405	403	2	365	365	0	40	38	2
23	Nigeria	405	399	6	365	363	2	40	36	4
24	Norway	405	387	18	365	352	13	40	35	5
25	Poland	405	404	1	365	364	1	40	40	0
26	Romania	405	403	2	365	363	2	40	40	0
27	Russia	405	399	6	365	359	6	40	40	0
28	Singapore	405	381	24	365	344	21	40	37	3
29	South Africa	405	403	2	365	363	2	40	40	0
30	Spain	405	389	16	365	349	16	40	40	0
31	Sweden	405	403	2	365	363	2	40	40	0
32	Thailand	405	403	2	365	363	2	40	40	0
33	Trinidad and Tobago	405	404	1	365	364	1	40	40	0
34	Tunisia	405	398	7	365	359	6	40	39	1
35	United Kingdom	405	286	119	365	249	116	40	37	3
36	United States of America	405	397	8	365	361	4	40	36	4

Table 7. Flag—offshore vessel's flag (30 different appearances); count of instances grouped by the variable Detention = "No" and Detention = " Yes".

No	Flag	All Dataset			Detention No			Detention Yes		
		Total	=0	=1	Total	=0	=1	Total	=0	=1
1	Azerbaijan	405	401	4	365	361	4	40	40	0
2	Bahamas	405	385	20	365	347	18	40	38	2
3	Belgium	405	394	11	365	355	10	40	39	1
4	Belize	405	403	2	365	365	0	40	38	2
5	Brazil	405	375	30	365	337	28	40	38	2
6	Canada	405	384	21	365	346	19	40	38	2
7	Cyprus	405	393	12	365	353	12	40	40	0
8	Denmark	405	387	18	365	347	18	40	40	0
9	Egypt	405	404	1	365	364	1	40	40	0
10	France	405	392	13	365	352	13	40	40	0
11	Germany	405	404	1	365	365	0	40	39	1
12	Gibraltar	405	385	20	365	345	20	40	40	0
13	Greece	405	384	21	365	344	21	40	40	0
14	Liberia	405	395	10	365	360	5	40	35	5
15	Luxembourg	405	375	30	365	336	29	40	39	1
16	Malaysia	405	400	5	365	362	3	40	38	2
17	Malta	405	386	19	365	346	19	40	40	0
18	Marshall Islands	405	401	4	365	363	2	40	38	2
19	Mexico	405	392	13	365	354	11	40	38	2
20	Nigeria	405	402	3	365	363	2	40	39	1
21	Norway	405	358	47	365	319	46	40	39	1
22	Panama	405	401	4	365	362	3	40	39	1
23	Russia	405	397	8	365	358	7	40	39	1
24	Singapore	405	400	5	365	361	4	40	39	1
25	St Vincent and Grenadines	405	396	9	365	359	6	40	37	3
26	Tuvalu	405	398	7	365	358	7	40	40	0
27	United Arab Emirates	405	402	3	365	362	3	40	40	0
28	United Kingdom	405	399	6	365	361	4	40	38	2
29	United States of America	405	403	2	365	364	1	40	39	1
30	Vanuatu	405	349	56	365	316	49	40	33	7

Table 8. Memorandum (MoU) variable (10 different appearances), count of instances grouped by the variable Detention = "No" and Detention = " Yes".

No	MoU	All Dataset			Detention No			Detention Yes		
		Total	=0	=1	Total	=0	=1	Total	=0	=1
1	Abuja	405	395	10	365	359	6	40	36	4
2	Black	405	404	1	365	364	1	40	40	0
3	Black Sea	405	402	3	365	362	3	40	40	0
4	Caribbean	405	404	1	365	364	1	40	40	0
5	Indian Ocean	405	379	26	365	343	22	40	36	4
6	Mediterranean	405	358	47	365	323	42	40	35	5
7	Paris	405	171	234	365	343	22	40	31	9
8	Tokyo	405	333	72	365	304	61	40	29	11
9	US Coastguard	405	397	8	365	361	4	40	36	4
10	Vina Del Mar	405	402	3	365	362	3	40	40	0

Table 9. YOB—Year of offshore vessel built (26 different appearances), count of instances grouped by the variable Detention = "No" and Detention = " Yes".

No	YOB	All Dataset			Detention No			Detention Yes		
		Total	=0	=1	Total	=0	=1	Total	=0	=1
1	1992	405	404	1	365	364	1	40	40	0
2	1997	405	404	1	365	364	1	40	40	0
3	1998	405	396	9	365	356	9	40	40	0
4	1999	405	387	18	365	347	18	40	40	0
5	2001	405	401	4	365	361	4	40	40	0
6	2002	405	388	17	365	351	14	40	37	3
7	2003	405	398	7	365	360	5	40	38	2
8	2005	405	389	16	365	349	16	40	40	0
9	2006	405	380	25	365	343	22	40	37	3
10	2007	405	388	17	365	348	17	40	40	0
11	2008	405	395	10	365	356	9	40	39	1
12	2009	405	389	16	365	349	16	40	40	0
13	2010	405	372	33	365	335	30	40	37	3
14	2011	405	366	39	365	330	35	40	36	4
15	2012	405	348	57	365	320	45	40	28	12
16	2013	405	366	39	365	332	33	40	34	6
17	2014	405	377	28	365	337	28	40	40	0
18	2015	405	383	22	365	346	19	40	37	3
19	2016	405	380	25	365	341	24	40	39	1
20	2017	405	399	6	365	359	6	40	40	0
21	2018	405	402	3	365	362	3	40	40	0
22	2019	405	403	2	365	363	2	40	40	0
23	2020	405	403	2	365	365	0	40	38	2
24	2021	405	404	1	365	364	1	40	40	0
25	2022	405	400	5	365	360	5	40	40	0
26	2023	405	403	2	365	363	2	40	40	0

4.2. Implementation Methodology

For the application of the decision tree algorithm "rpart" and the MCDM TOPSIS method (which only uses numerical values of the letter variables "Detention", "Country of detention", "Flag", and "MoU" and the numerical value "YOB"), these are converted into so-called "dummy" variables [43]. In this way, both algorithms can use the same data structure and group the occurrence of each value by the variable "Detention", which can have the value "Yes" or "No". Dummy variables are a useful tool to translate categorical data into a form that can be used by machine learning algorithms. This is necessary because most machine learning algorithms are designed for numerical inputs. Furthermore, the use of dummy variables avoids the problem of introducing a wrong order or ordinality between the categories, which could happen if the categories were converted directly to integers. Each new variable can have the values 0 or 1, with 1 indicating the presence of a particular category.

When applying the TOPSIS method, the data from all 405 alternatives and 125 variables were used. Based on the frequency of occurrence for the entire set of input data, grouped by the variable "Detention", the criteria and the target (impact) of each criterion are defined. When using the TOPSIS method, the text variable "Detention" is also converted into a numerical value, as the TOPSIS method can only process numerical values. For this reason, there is one more variable (125) in the TOPSIS method than in the creation of a classified decision tree.

The list of all variables (125) used in the TOPSIS method with the assigned weighting values and "impact" can be found in Table 10. The weighting values of each criterion and the "target" effect (min = "-"; max = "+") of each of the 125 variables in the TOPSIS method are determined based on the number of occurrences in the input dataset. Larger values are assigned to variables that appear more than once in the input dataset when the vessel is detained, i.e., when “Detention” = Yes.

Table 10. The list of all variables, assigned weight (W), and impact (I) in the TOPSIS method.

Variable Name	W	I	Variable Name	W	I	Variable Name	W	I	Variable Name	W	I
InitialInspection	20	-	COI_Greece	5	+	Flag Denmark	1	+	MoU Vina Del Mar	1	+
MoreDetailedInspection	16	-	COI_HongKong	1	+	Flag Egypt	1	+	YOB_1992	1	+
Follow-UpInspection	1	-	COI_India	1	+	Flag France	1	+	YOB_1997	1	+
StandardInspection	3	-	COI_Israel	5	+	Flag Germany	2	+	YOB_1998	1	+
NumberOfinspections	39	-	COI_Italy	1	+	Flag Gibraltar	1	+	YOB_1999	1	+
NumberOfDeficiency	36	-	COI_Japan	1	+	Flag Greece	1	+	YOB_2001	1	+
ISM	7	-	COI_Malaysia	1	+	Flag Liberia	10	+	YOB_2002	2	+
MARPOL	7	-	COI_Malta	1	+	Flag Luxembourg	2	+	YOB_2003	3	+
CertificateDocumentation	19	-	COI_Morocco	1	+	Flag Malaysia	3	+	YOB_2005	4	+
PropulsionAuxiliary Machinery	4	-	COI_Netherlands	3	+	Flag Malta	1	+	YOB_2006	5	+
SafetyOfNavigation	13	-	COI_NewZealand	5	+	Flag Marshall Islands	3	+	YOB_2007	6	+
RadioCommunications	3	-	COI_Nigeria	10	+	Flag Mexico	3	+	YOB_2008	7	+
EmergencySystems	4	-	COI_Norway	1	+	Flag Nigeria	2	+	YOB_2009	8	+
FireSafety	13	-	COI_Poland	1	+	Flag Norway	2	+	YOB_2010	9	+
MLC	9	-	COI_Romania	1	+	Flag Panama	1	+	YOB_2011	11	+
Alarms	2	-	COI_Russia	1	+	Flag Russia	1	+	YOB_2012	12	+
ISPS	2	-	COI_Singapore	1	+	Flag Singapore	1	+	YOB_2013	6	+
OtherTypeOfDeficiencies	2	-	COI_SouthAfrica	1	+	Flag_St Vincent and Grenadines	1	+	YOB_2014	1	+
WaterWeatherTight Conditions	2	-	COI_Spain	1	+	Flag Tuvalu	1	+	YOB_2015	3	+
LifeSavingAppliances	7	-	COI_Sweden	1	+	Flag_United Arab Emirates	0.1	+	YOB_2016	1	+
GT	3	-	COI_Thailand	1	+	Flag_UnitedKingdom	1	+	YOB_2017	1	+
COI_Australia	25	+	COI_Trinidadand Tobago	1	+	Flag_UnitedStatesof America	1	+	YOB_2018	1	+
COI_Brazil	1	+	COI_Tunisia	3	+	Flag Vanuatu	1	+	YOB_2019	1	+
COI_Bulgaria	1	+	COI_UnitedKingdom	5	+	MoU Abuja	4	+	YOB_2020	5	+
COI_ChinaPeopless Republic	5	+	COI_UnitedStatesof America	10	+	MoU_Black	1	+	YOB_2021	1	+
COI_Colombia	1	+	Flag_Azerbaijan	1	+	MoU_Black Sea	1	+	YOB_2022	1	+
COI_Cyprus	3	+	Flag_Bahamas	3	+	MoU_Caribbean	1	+	YOB_2023	1	+
COI_Denmark	1	+	Flag_Belgium	2	+	MoU_Indian Ocean	4	+	Detention_0	10	-
COI_Egypt	3	+	Flag_Belize	3	+	MoU_Mediterranean	1	+	Detention_1	100	+
COI_France	1	+	Flag_Brazil	3	+	MoU_Paris	12	+			
COI_Germany	3	+	Flag_Canada	3	+	MoU_Tokyo	11	+			
COI_Ghana	1	+	Flag_Cyprus	1	+	MoU_US CoastGuard	4	+			

The implementation of the classification tree and MCDM TOPSIS methods in this work was carried out using several R packages [44,45].

The complexity of algorithms that combine “rpart” and the TOPSIS method can be analyzed separately for each algorithm and then integrated to obtain the overall complexity of the combined approach.

“rpart” is an implementation of the classification decision tree algorithm. The complexity of this algorithm depends on several factors: Number of instances (N): total number of samples in the training set. Number of attributes (M): number of attributes or features in the data. Tree depth (D): the depth of the final tree.

For most implementations, the complexity of the “rpart” algorithm can be roughly estimated as $O(N * M * \log N)$, where 1. the algorithm traverses all data instances (N) and all attributes (M) to compute the optimal splitting points; 2. the logarithmic factor results from the process of building the tree, where the splitting of the data on each branch can be considered as a binary search.

The TOPSIS method involves several steps, each of which has its complexity: 1. normalization of the data matrix: if there are m alternatives and n criteria, the normalization has a complexity of $O(m * n)$. 2. The weighting of the normalized matrix: this is also $O(m * n)$, since each element of the matrix is multiplied by the corresponding weight. 3. Calculating the distance to the ideal and anti-ideal solution: this involves summing the differences for each alternative, which has a complexity of $O(m * n)$. 4. Ranking the alternatives: the complexity of the ranking is $O(m * log\ m)$.

The overall complexity of the TOPSIS method can be approximated as $O(m * n)$, since all steps are linear concerning the number of alternatives and criteria.

The combined complexity depends on the order in which these methods are applied and on their mutual interactions. If "rpart" is first applied to identify relevant alternatives and then TOPSIS to rank them, the overall complexity can be a combination of the individual complexities: "rpart" complexity: $O(N * M * log\ N)$. TOPSIS complexity: $O(m * n)$, where m is the number of alternatives that come from the "rpart" analysis.

Assuming that m is approximately N (the number of alternatives remaining after the "rpart" analysis), the combined complexity can be $O(N\backslash*M\backslash*logN) + O(N\backslash*n)$.

In practice, this combined complexity can be higher, especially if M and n are large or if multiple iterations are required between these steps. Ultimately, the importance of these factors depends on the specific implementation and data size, but this provides a framework for considering the overall complexity of combining "rpart" and TOPSIS methods.

5. Results

In this case study, the input dataset (405 instances of inspection of the offshore vessels $\times$ 125 variables) is randomly divided into a training and testing dataset in a ratio of 70:30. The test dataset accounts for 30% of the total data and is used to evaluate the model after it has been trained. This joint step is performed for several important reasons:

1. It makes it possible to evaluate the model's performance using data that the model did not see during training. This gives a better picture of how the model generalizes to new, unknown data.
2. Avoidance of overfitting. Since the model is only trained on one dataset, there is a risk that it will overfit the specifics of that dataset. The test dataset is used to check whether the model generalizes well or whether it has only learned the specific features of the training dataset.
3. Validation of model selection, as different models and their parameters can be compared using a test dataset. This way, the best model can be selected and its hyperparameters optimized. Hyperparameters are parameters that are used to control the process of machine learning models. They are preset and do not change during model training. In contrast to model parameters that are learned from the data, hyperparameters are usually set before training and influence model performance.
4. Data partitioning enables the evaluation of various model performance metrics (such as Accuracy, Precision, Recall, etc.) against the test dataset. These metrics help to understand how effective the model is in real-life situations.

Sometimes a third dataset, called the validation dataset, is also used to tune the model's hyperparameters and select a better model during training. The dataset is used for the final evaluation. The training dataset for the "rpart" classification decision tree therefore consists of 283 instances and the test dataset of 122 (possible alternatives). Both datasets have 123 independent variables, while the variable "Detention" is a target variable that can have a value of 0 (no) and 1 (yes).

The textual and graphical (Figure 2) expression of the "rpart" decision tree model in R is interpreted as follows.

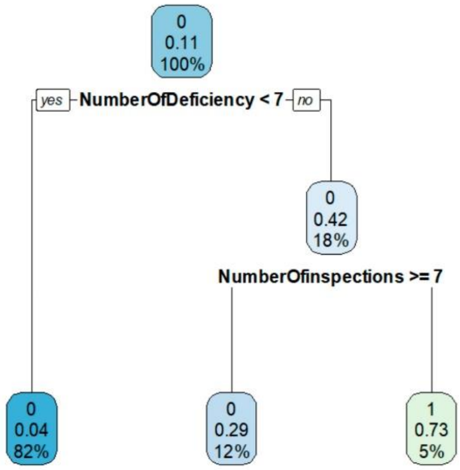

Figure 2. Generated decision tree classification model for the training dataset.

Textual representation of the model result:

The expression presented in Table 11 describes how the decision tree uses the Number of Deficiency and the Number of Inspections to classify data into two classes (Detention) (0—No; 1—Yes), and how decisions are made based on these attributes.

Table 11. Decision tree results interpretation.

n = 283 node), split, n, loss, yval, (yprob) * denotes terminal node 1) root 283 31 0 (0.89045936 0.10954064) 2) NumberOfDeficiency < 6.5 233 10 0 (0.95708155 0.04291845) * 3) NumberOfDeficiency >= 6.5 50 21 0 (0.58000000 0.42000000) 6) NumberOfinspections >= 6.5 35 10 0 (0.71428571 0.28571429) * 7) NumberOfinspections < 6.5 15 4 1 (0.26666667 0.73333333) *	1): Identification of the node (root node). root: The root node of the tree. 283: Number of instances in the root node (entire training dataset). 31: Number of incorrectly classified examples if all examples are classified as class 0. 0: Class prediction (highest probability) at the root node. (0.89045936 0.10954064): Probabilities for each class (class 0: 89.05%, class 1: 10.95%).	2): Identification of the node. Number of Deficiency < 6.5: Rule for division (if the Number of Deficiency is less than 6.5, go to this node). 233: Number of instances in this node. 10: Number of misclassified examples if all are classified as class 0. 0: Class prediction (class 0). (0.95708155 0.04291845): Probabilities for each class (class 0: 95.71%, class 1: 4.29%). *: Marking that the node is terminal (leaves).
3): Node identification. Number Of Deficiency >= 6.5: Rule for the subdivision (if the Number Of Deficiency is greater than or equal to 6.5, go to this node). 50: Number of instances in this node. 21: Number of misclassified examples if all are classified as class 0. 0: Class prediction (class 0). (0.58000000 0.42000000): Probabilities for each class (class 0: 58%, class 1: 42%).	6): Identification of the node. Number of Inspections >= 6.5: Splitting rule (if Number Of Deficiency is greater than or equal to 6.5, go to this node). 35: Number of instances in this node. 10: Number of misclassified examples if all are classified as class 0. 0: Class prediction (class 0). (0.71428571 0.28571429): Probabilities for each class (class 0: 71.43%, class 1: 28.57%). *: Marking that the node is terminal (leaves).	7): Identification of the node. Number of Inspections < 6.5: Splitting rule (if the Number of Inspections is less than 6.5, go to this node). 15: Number of instances in this node. 4: Number of misclassified examples if all are classified as class 1. 1: Class prediction (class 1). (0.26666667 0.73333333): Probabilities for each class (class 0: 26.67%, class 1: 73.33%). *: Marking that the node is an end node (leaves).

node: Identification of the node in the tree.
split: The rule used to split the data in this node.

n: Number of instances in this node.

loss: The number of instances that would be misclassified if all instances in a node were classified according to the majority class.

yval: Class prediction in this node.

(yprob): Probability for each class in this node.

indicates an end node: Indication that the node is an end node (leaf) and does not subdivide further.

The reason for using a threshold of 7 for the "Number Of Deficiency" and the "Number of Inspections" is to round off and simplify the interpretation of the results. In practice, the threshold value of 7 is used as an integer value, which is intuitive and easier to interpret when making decisions. However, in the analyses, the threshold of 6.5 is used as a more precise, numerical threshold derived from statistical models and data analyses.

The difference between the thresholds in these parts of the text reflects different aspects of data interpretation—one is more a function of rounding to simplify interpretation, the other for the precision of analysis.

The model created this way with the training dataset was tested with the test dataset. The result of the prediction of a possible outcome is shown in a confusion matrix.

A confusion matrix [46] is an instrument for evaluating the performance of a classification model. It helps to understand how the model categorizes the examples and makes it possible to identify the types of errors that the model makes.

- Actual 0: Actual examples of class 0.
- Actual 1: Actual examples of class 1.
- Predicted 0: Examples predicted by the model as class 0.
- Predicted 1: Examples predicted by the model as class 1.

Elements of the confusion matrix interpretation (Table 12):

- True positives (TPs): number of true positive examples correctly categorized as positive by the model (4 (Actual 1, Predicted 1));
- True negatives (TNs): Number of true negative examples correctly categorized as negative by the model (110 (Actual 0, Predicted 0));
- False positives (FPs): The number of true negative examples the model incorrectly categorized as positive (3 (Actual 0, Predicted 1));
- False negative (FN): The number of true positive examples the model incorrectly categorized as negative (5 (Actual 1, Predicted 0)).

Table 12. Confusion matrix.

	Predicted 0	**Predicted 1**
Actual 0	110	3
Actual 1	5	4

This means the following: for 110: out of 113 real negative examples (class 0), the model correctly categorized 110 as negative; 3: out of 113 real negative examples (class 0), the model categorized 3 as positive; out of 9 real positive examples (class 1), the model correctly categorized 4 as positive; 5: out of 9 real positive examples (class 1), model 5 incorrectly categorized as negative.

The metric derived from the confusion matrix is the proportion of correctly classified examples from the classification decision tree ("rpart") for the test dataset, which is calculated according to Equation (8)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{110 + 4}{110 + 4 + 3 + 5} \approx 0.934 \quad (8)$$

Precision is defined by Equation (9) and has a value of 0.57143, Specificity is defined by Equation (10) and has a value of 0.97345, while Recall is defined by Equation (11) and has a value of 0.4444. *F*1 *Score* is defined by Equation (12) and has a value of 0.5.

$$Precision = \frac{TP}{TP + FP} = \frac{4}{4 + 3} \approx 0.571 \quad (9)$$

$$Specificity = \frac{TN}{TN + FP} = \frac{110}{110 + 3} \approx 0.973 \quad (10)$$

$$Recall = \frac{TP}{TP + FN} = \frac{4}{5 + 5} \approx 0.444 \quad (11)$$

$$F1\ Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} = 0.5 \quad (12)$$

The *F*1 *Score*—Equation (12)—is a performance measure of the classification model that represents the harmonic mean between *Precision* and *Recall*. It is used when there is an imbalance between the classes and when both *Precision* and *Recall* are important. The *F1 Score* is useful to achieve a balance between *Precision* and *Recall* and not just optimize one of these parameters. So, an F1 Score with a value of 0.5 indicates a trade-off between *Precision* and *Recall*. Since the *Precision* has a value of 57.14%, the predictions labeled as positive by the model are positive. The model correctly identified 44.44% of the actual positive instances. If the *F*1 *Score* is low (as in this case), this indicates a significant number of false positive or false negative predictions or even both. Depending on the application context, such performance can sometimes be unacceptable, especially if the consequences of prediction errors are severe.

The reason for this may be that only 10% of the total dataset contains truly positive values, namely 40 (for Detention = 1) and 365 (for Detention = 0).

The *F*1 *Score*, which has a value of 0.5, shows that both the *Precision* and the *Recall* are quite low and that the "rpart" model is not particularly efficient. For this reason, the MCDM method TOPSIS is used in the next phase.

Of course, these metrics allow a comprehensive analysis of the performance of the "rpart" classification model and help to evaluate how effective the model is in solving a given classification problem with a classification decision tree.

Based on the assigned weights (W) and impact (I) for each of the variables in the total set of alternatives (with 405 instances × 125 variables) (Table 10), the TOPSIS method ranked all 40 in the top 40 with distance = 1 in the order shown in Table 13. This shows that the weighting of the criteria and the values of the impact vector of the input dataset considerations are correct. Table 13 thus shows the 40 best alternatives identified using the TOPSIS method. The columns are No Alt (alternatives); Detention_1; and Score and Rank, sorted from best to worst.

Note: In the TOPSIS method, the best-rated alternative is the one with the highest score, but in this empirical study the one with the highest probability of detention, i.e., Detention = 1, was rated as such.

For the trained classification decision tree model "rpart", a prediction of the probability of detention (Detention Probability) is also created for all examples (405) of the input dataset. The probability values are then added as a new column to the original dataset (column Detention Probability "rpart", Table 14). In this way, information on the probability of the offshore vessel being detained is obtained (Column D = 1, Table 14) for each example in the dataset. These data are useful and are used in the further analysis.

Table 13. The first 40 best classification alternatives (No Alt) of the TOPSIS method.

No Alt	Detention_1	Score	Rank	No Alt	Detention_1	Score	Rank
404	1	0.64004	1	390	1	0.60817	21
403	1	0.63895	2	405	1	0.60740	22
397	1	0.63747	3	388	1	0.60662	23
396	1	0.63503	4	283	1	0.60567	24
152	1	0.63494	5	387	1	0.60561	25
51	1	0.63361	6	348	1	0.60485	26
381	1	0.63266	7	347	1	0.60379	27
395	1	0.62939	8	391	1	0.60315	28
380	1	0.62630	9	402	1	0.60272	29
394	1	0.62623	10	171	1	0.60120	30
125	1	0.62549	11	384	1	0.60103	31
93	1	0.62532	12	393	1	0.59639	32
130	1	0.62456	13	382	1	0.58679	33
321	1	0.62314	14	355	1	0.57963	34
271	1	0.62196	15	383	1	0.57542	35
401	1	0.62081	16	400	1	0.56435	36
389	1	0.61294	17	324	1	0.56393	37
385	1	0.61245	18	323	1	0.56292	38
398	1	0.61153	19	392	1	0.54063	39
399	1	0.60908	20	386	1	0.50631	40

Table 14. The first 40 best classification alternatives by combining the "rpart" algorithm and the TOPSIS method.

No	No Alt	D = 1	Detention Probability (rpart)	TOPSIS Score	Combined Score	No	No Alt	D = 1	Detention Probability (rpart)	TOPSIS Score	Combined Score
1	A397	1	0.733	0.637	0.685	21	A343	0	0.733	0.358	0.545
2	A396	1	0.733	0.635	0.684	22	A248	0	0.733	0.349	0.541
3	A93	1	0.733	0.625	0.679	23	A385	1	0.286	0.612	0.449
4	A271	1	0.733	0.622	0.678	24	A399	1	0.286	0.609	0.447
5	A389	1	0.733	0.613	0.673	25	A388	1	0.286	0.607	0.446
6	A398	1	0.733	0.612	0.672	26	A387	1	0.286	0.606	0.446
7	A390	1	0.733	0.608	0.671	27	A402	1	0.286	0.603	0.444
8	A405	1	0.733	0.607	0.670	28	A171	1	0.286	0.601	0.443
9	A348	1	0.733	0.605	0.669	29	A384	1	0.286	0.601	0.443
10	A347	1	0.733	0.604	0.669	30	A393	1	0.286	0.596	0.441
11	A391	1	0.733	0.603	0.668	31	A355	1	0.286	0.580	0.433
12	A382	1	0.733	0.587	0.660	32	A324	1	0.286	0.564	0.425
13	A383	1	0.733	0.575	0.654	33	A323	1	0.286	0.563	0.424
14	A400	1	0.733	0.564	0.649	34	A386	1	0.286	0.506	0.396
15	A392	1	0.733	0.541	0.637	35	A404	1	0.043	0.640	0.341
16	A49	0	0.733	0.391	0.562	36	A403	1	0.043	0.639	0.341
17	A139	0	0.733	0.387	0.560	37	A152	1	0.043	0.635	0.339
18	A260	0	0.733	0.377	0.555	38	A51	1	0.043	0.634	0.338
19	A94	0	0.733	0.377	0.555	39	A381	1	0.043	0.633	0.338
20	A346	0	0.733	0.375	0.554	40	A395	1	0.043	0.629	0.336

The combination of the "rpart" (Detention Probability) and TOPSIS (TOPSIS Score) results is carried out by calculating the weighted sum of the Combined Score (columns 6 and 12—Table 14) for each example in the dataset.

The new column "Combined Score" in the dataset thus represents the weighted sum of "Detention Probability" (probability of detention from "rpart") of and "TOPSIS Score" (the result of the TOPSIS method) and is calculated according to Equation (13), where and represents the *i*-th alternative of the entire dataset. Neither of these two values ("Detention

Probability" and "TOPSIS Score") is favored, as each component is assigned a weight of 0.5 (50%).

$$CombinedScore[\mathrm{i}] = 0.5 \ * DetentionProbability[\mathrm{i}] + 0.5 * TopsisScore[\mathrm{i}] \quad (13)$$

The final results (Table 14) show that of the 40 alternatives (Alt) with the highest adherence probability, 33 (or 82.5%) belong to the group in which the variable Detention has the value (D = 1 or Yes), while 7 have the value (D = 0 or No), or 17.5%. This means that the results of the 40 most likely detention vessels identified by the combination of "rpart" and TOPSIS match 82.5% of cases. (See Table 14)

This means that although "rpart" is used in the entire dataset (as can be seen in Figure 3, which shows that there are 45 checked vessels in the entire dataset, of which 40 are Not-detained vessels (N) and 5 are Detained vessels (D)) it is possible that one of the 5 detained vessels is not identified as "Detained" by the "rpart" algorithm.

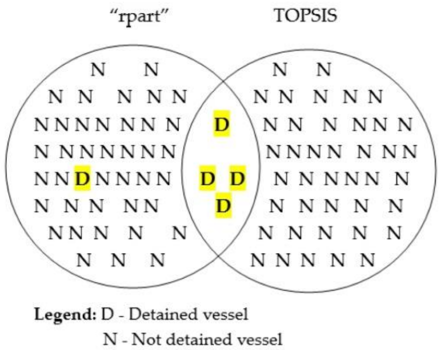

Figure 3. Results interpretations.

The TOPSIS method, on the other hand, individualizes all 4 detained vessels in the total set of 45 vessels (of which 41are Non-detained vessels (N) and 4 are Detained vessels (D)).

However, by combining these two methods, it is possible to determine with greater certainty which vessels are more likely to be detained.

6. Discussion

A total of 405 instances (alternatives) and 125 variables were used for the empirical study in this article. When applying the "rpart" classification decision tree, the data were randomly divided into a training dataset (283 × 124) and a test dataset (122 × 124) in a ratio of 70:30. The number of independent variables is 123, while the dependent variable is "Detention", which can be classified as "Yes" and "No". For both the training and test datasets, when applying the "rpart" classification tree, all categorical variables ("Country of Inspection", "Flag", "Memorandum", and "YOB") were converted into numerical variables to show the number of occurrences of each variable in the input dataset for the entire dataset, but also grouped by the variable "Detention".

When applying the MCDM method TOPSIS, data from all 405 alternatives and 125 variables were used. Based on the frequency of occurrence for the entire set of input data, the criteria and the target (impact) of each criterion were defined. When using the TOPSIS method, the text variable "Detention" was also converted to a numeric value, as the TOPSIS method can only process numeric values. For this reason, there is an additional variable (125) in the TOPSIS method.

Data validation can be crucial when using the TOPSIS method to ensure the accuracy and reliability of the results. Several methods were used to validate the data in the TOPSIS method as follows:

- All data were checked. The input dataset does not contain any missing values and outliers (data outside the possible ranges that can significantly affect the results). If this were the case, the data would be deleted before application.
- Data consistency was checked, i.e., all values of the variables are in the same format and make sense in the context of the problem to be analyzed. The data were normalized so that all values were in the same range (i.e., between 0 and 1). This is an important element of validation as the TOPSIS method works with distances between points, so different ranges can distort the results. The weighting values were checked for all variables and all reflect the actual priorities.
- A basic sensitivity analysis was performed by changing the weighting values of the criteria and analyzing how this change affected the ranking of the solutions.
- A stability check was performed by analyzing how small changes in the input data affected the results of large changes in the ranking.

Additionally, the TOPSIS method yielded all 40 vessels detained. This essentially confirms that partial (minor) changes to the input dataset do not significantly alter the results, which is also a cross-validation step. Cross-validation was applied to reduce the risk of over-learning the model and improve its robustness, especially in real-world applications with high data variability.

The classification decision tree model "rpart" correctly identified 44.44% of the true positive cases (actually detained vessels). This is justified by the small number of actual vessels detained (Detention = 1 is 40 or 10%) in the total number of alternatives (405).

However, the MCDM TOPSIS method, which is based on the fixed weight values of the criteria and the impact (min or max) of each criterion, provided excellent results and ranked all positive instances (Detention = 1) among the 40 most trustworthy, i.e., the best.

The combined results of "rpart" with TOPSIS MCDM show that of the 40 most likely alternatives (Alt) for retaining offshore vessels, 33 (or 82.5%) belong to the group that would likely be detained.

The results confirmed that by applying a combination of classification decision trees and MCDM, it is possible to improve the decision-making process when offshore vessels are detained in port.

Finally, analyzing the complexity of algorithms [47] helps to understand the performance and resource requirements, which is crucial for efficient application in different scenarios and with different data sizes. The overall complexity of the algorithm includes the calculation of the temporal complexity, i.e., how much time is required to execute the algorithm, and the spatial complexity, i.e., how much computer memory and other relevant resources are required to execute the algorithm. Both are listed in Table 15.

Table 15. The complexity of algorithms.

	rpart	**TOPSIS**
Time complexity	$O(n \times m) \log m$	$O(n \times m)$
Spatial complexity	$O(m \times n)$	$O(m \times n)$

In the "rpart" algorithm, m stands for the number of instances and n for the number of variables, and in TOPSIS, m stands for the number of instances and n for the number of criteria.

The analysis revealed several important advantages and challenges in implementing this approach. By implementing classification trees in software tools, it is possible to automate part of the PSC process for keeping offshore vessels in port, ensure greater consistency in decision-making, and reduce the possibility of human error.

The classification trees enabled the identification of key factors that influence the decision to detain offshore vessels. This helps inspectors to focus on critical aspects during the inspection, increasing efficiency and reducing the time needed for decision-making. The decision tree generated in the R package provided a visually understandable representation of the decisions, facilitating their interpretation and application in practice, so that inspectors (but also vessels and other interested parties) can quickly, easily, and clearly understand the logic behind the decision and adapt their inspection procedures accordingly.

On the other hand, MCDM has proven to be a comprehensive, flexible approach and a powerful decision support tool. Based on historical data (qualitative and quantitative) on the detention of offshore vessels, MCDM enabled an assessment and prediction of the possibility of future detention of vessels, resulting in a more balanced and objective approach. The proposed approach has also been shown to apply to the specific needs and priorities of different ports and inspection teams, allowing for minor and/or major changes in the assignment of weighting values to variables and their targets according to the respective safety and operational requirements. Therefore, the tools of the MCDM methodology provide structural support in the decision-making process and help inspectors make informed and transparent decisions, reduce subjectivity, and increase confidence in the PSC process.

The main drawbacks of this approach relate to the following important elements. It is necessary to focus on the potential challenges in practical application, including 1. training of inspectors should consider the resources and time required to train inspectors in the application of machine learning and MCDM methods and the strategies to overcome these challenges; 2. technical requirements should be analyzed through an analysis of technical barriers and 3. organizational barriers point to the importance of changing the management and culture within organizations to enable the successful adoption of new technologies. Furthermore, recommendations should be made for future research related to analyzing all these aspects and developing practical guidelines and case studies that would facilitate implementation in different contexts.

Although classification trees and MCDM offer significant benefits, their implementation requires initial investment in software tools and training of inspectors. Ensuring adequate training and support is critical to the success of these methods. On the other hand, the success of these analytical methods depends on the quality and reliability of the data. Therefore, deficiencies in data collection or irregularities in recording can hurt the accuracy and usefulness of the results. In addition, changes to existing PSC procedures may be met with resistance from inspectors and other involved parties. It is important to communicate the benefits appropriately and ensure the involvement of all relevant stakeholders in the change process.

7. Conclusions

In this article, the application of a combination of machine learning algorithms, i.e., classification trees and MCDM, in the process of holding offshore vessels in port, better known as PSC, was investigated. The analysis showed that these methods can significantly improve the efficiency and accuracy of the inspection process and detention of vessels. Classification trees are a powerful tool for identifying key risk factors, enabling inspectors to make faster and more accurate detention decisions.

The use of MCDM has further improved the decision-making process as multiple factors can be considered simultaneously, ensuring a balanced approach that minimizes subjectivity. The combination of these methods has enabled the creation of a more systematic, transparent, reliable, and robust PSC framework that can lead to increased safety of navigation and reduced operational costs.

Based on the results presented, further exploration and integration of machine learning techniques into existing PSC procedures is recommended. The use of advanced analytical tools can significantly contribute to improving safety standards and operational efficiency in the maritime sector. In addition, training inspectors on these tools and methods can further increase their applicability and effectiveness.

Finally, this paper points to the significant potential of improving the PSC process of modern analytical machine learning methods, which can contribute to safer and more efficient maritime transport.

The research conducted has several potential limitations related to the quality and scope of the data, the complexity and robustness of the model, methodological limitations, practical implementation, validation and generalization, and technical performance. If the data used to train and test the model are not sufficiently representative or contain errors, the results may be unreliable. If the data come from a limited number of sources or geographical areas, the models may not be generalizable to all situations or regions. Classification trees can be prone to overfitting, where the model responds too accurately to the specifics of the training rather than to general patterns. Algorithms such as "rpart" can reach their limits in terms of the complexity of the data and the interpretability of the results. If the research is still in the development phase, there may not be sufficient evidence of the stability and reliability of the results. The weighting of the criteria can be subjective and depends on the accuracy of the weightings. The implementation of advanced algorithms may require significant technical resources and staff training, which can be a barrier for organizations with limited budgets. There is a possibility that inspectors and other users may be unwilling to accept new methods due to resistance to change or lack of training. If a model is not tested on independent datasets, one does not always have insight into its ability to generalize the results across different scenarios. The results may relate to a specific context or period and may not apply to all types of offshore vessels or different regions.

Understanding and recognizing these limitations will certainly help to formulate more realistic conclusions and recommendations for this work and identify areas for future research and improvement.

Author Contributions: Conceptualization, Z.B. and I.S.; methodology, Z.B., I.S. and Z.S.-V.; software, T.P.; validation, Z.B., I.S. and Z.S.-V.; formal analysis, Z.B.; investigation, Z.B. and T.P.; resources, Z.B. and T.P.; data curation, Z.B.; writing—original draft preparation, Z.B. and I.S.; writing—review and editing, I.S.; visualization, T.P.; supervision, I.S. and Z.S.-V.; project administration, T.P.; funding acquisition, Z.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AHP	Analytical Hierarchy Process
AHTS	Anchor Handling Tug Supply
AI	Artificial intelligence
AIS	Automatic Identification System
BN	Bayesian network
CART	Classification and regression trees
CHAID	Chi-Square automatic interaction detector
CIC	Concentrated inspection campaign
COI	Country of Inspection
CRITIC	Criteria interaction through the inter-criteria correlation
DCM	Deck of Cards Method
DRS	Decent Reasoning System
ELECTRE	Elimination Et Choix Traduisant la REalité (Elimination and Choice Expressing Reality)
EWM	Entropy weight method
FN	False negative
FP	False positive

FPSO	Floating production storage and offloading
GRT	Gross register tones
GT	Gross tonnage
IMO	International Maritime Organization
ISM	International Safety Management
ISPS	International Safety for Port and Vessels
k-NN	k-Nearest Neighbors
MACN	Maritime Anti-Corruption Network
MARPOL	International Convention for the Prevention of Pollution from Vessels
MAUT	Multi-Attribute Utility Theory
MCDA	Multi-criteria decision analysis
MCDM	Multi-criteria decision-making
MDI	More detailed inspection
MLC	Maritime Labour Convention
MoU	Paris Memorandum of Understanding
MSE	Mean squared error
NIR	New inspection regime
PFSO	Port Facility Security Officer
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
PSC	Port State Control
PSCO	Port State Control Officer
PSV	Platform Supply Vessel
rpart	Revised Partitioning Algorithm
SAW	Simple Additive Weighting
SOLAS	International Convention for the Safety of Life at Sea
STCW	International Convention on Standards of Training, Certification, and Watchkeeping for Seafarers
TEU	Twenty-foot equivalent unit
TN	True negative
TOPSIS	Technique for Order Preference by Similarity to an Ideal Solution
TP	True positive
Var No	Variable number
YOB	Year of built

References

1. Paris MoUamen. Available online: https://parismou.org/system/files/2023-06/Paris%20MoU%20including%2045th%20amendment.pdf (accessed on 17 June 2024).
2. Tokyo MoUdetentionlist. Available online: https://www.tokyo-mou.org/inspections_detentions/detention_list.php (accessed on 21 June 2024).
3. Caribbean MoU. Available online: https://www.caribbeanmou.org (accessed on 21 June 2024).
4. Oceanshq. Available online: https://www.oceanshq.com/articles/top-reasons-for-port-state-control-detentions/ (accessed on 29 May 2024).
5. ISM Code. Available online: https://www.imo.org/en/ourwork/humanelement/pages/ISMCode.aspx (accessed on 25 May 2024).
6. MACN. Available online: https://macn.dk/wp-content/uploads/2023/01/strengthening-integrity-transparency-during-port-state-control-inspections-2022.pdf (accessed on 20 June 2024).
7. Lyu, P. Support a Safety Learning Culture Learning Culture in Port State Control Regime, Digital Repository of the World Maritime. Master's Thesis, World Maritime University, Malmö, Sweden, 28 October 2023.
8. Chuah, L.F.; Mokhtar, K.; Bakar, A.A.; Othman, M.R.; Osman, N.H.; Bokhari, A.; Mubashir, M.; Abdullah, M.A.; Hasan, M. Marine environment and maritime safety assessment using Port State Control database. *Chemosphere* **2022**, *304*, 135245. [CrossRef] [PubMed]
9. Zheng, L. The Effectiveness of New Inspection Regime on Port State Control Inspection. *Open J. Soc. Sci.* **2020**, *8*, 440–446. [CrossRef]
10. Jihong, C.; Shuanghong, Z.; Lang, X.; Zheng, W.; Yijie, F.; Tianxiao, Z. Identification of key factors of vessel detention under Port State Control. *Mar. Policy* **2019**, *102*, 21–27. [CrossRef]
11. Deng, J.L. Control problems of grey systems. *Syst. Control. Lett.* **1982**, *1*, 288–294. [CrossRef]
12. Mohd Tarmizi, O.; Li, T.; Yuli, C.; Noorul Shaiful Fitri, A.R. Empirical analysis on port state control inspection for foreign-registered vessels in Malaysian ports. *Asian J. Vessel. Logist.* **2021**, *37*, 127–139. [CrossRef]

13. Lai, C.Y.; Liu, C.-P.; Huang, K.-M. Optimization of the Concentrated Inspection Campaign Model to Strengthen Port State Control. *J. Mar. Sci. Eng.* **2023**, *11*, 1166. [CrossRef]
14. Zhu, J.H.; Yang, Q.; Jiang, J. Identifying crucial deficiency categories influencing vessel detention: A method of combining cloud model and prospect theory. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108949. [CrossRef]
15. Dinis, D.; Figueira, J.; Teixeira, A.P. A multiple criteria approach for vessel risk classification: An alternative to the Paris MoU Vessel Risk Profile. *Socio-Econ. Plan. Sci.* **2023**, *90*, 101718. [CrossRef]
16. Simos, J. L'évaluation Environnementale: Un Processus Cognitif négocié. Ph.D. Thesis, Ecole Polytechnique Federal de Lausanne (EPFL), Lausanne, Switzerland, 1989. [CrossRef]
17. Figueira, J.; Roy, B. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *Eur. J. Oper. Res.* **2002**, *139*, 317–326. [CrossRef]
18. Corrente, S.; Figueira, J.; Greco, S. Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *Eur. J. Oper. Res.* **2021**, *291*, 738–756. [CrossRef]
19. Rey-Charlo, R.E.; Cueto, J.L.; Piniella, F. Analyzing Port State Control Data to Explore Future Improvements to GMDSS Training. *J. Mar. Sci. Eng.* **2023**, *11*, 2379. [CrossRef]
20. Gómez Ruiz, M.Á.; de Almeida, I.M.; Pérez Fernández, R. Application of Machine Learning Techniques to the Maritime Industry. *J. Mar. Sci. Eng.* **2023**, *11*, 1820. [CrossRef]
21. Chen, X.; Dou, S.; Song, T.; Wu, H.; Sun, Y.; Xian, J. Spatial-Temporal Ship Pollution Distribution Exploitation and Harbor Environmental Impact Analysis via Large-Scale AIS Data. *J. Mar. Sci. Eng.* **2024**, *12*, 960. [CrossRef]
22. Chung, W.H.; Hung, Y.Y.; Yuan, C.C. Application of Machine Learning to Maritime Safety. In *ISMI*; IEEE: Piscataway, NJ, USA, 2022. Available online: https://www.conf.tw/site/userdata/1449/ISMI_paper/ISMI2022_paper_5987.pdf (accessed on 10 June 2024).
23. Yan, R.; Wang, S.; Zhen, L.; Jiang, S. Classification and regression in prescriptive analytics: Development of hybrid models and an example of vessel inspection by port state control. *Comput. Oper. Res.* **2024**, *163*, 106517. [CrossRef]
24. Jacobsen, K.; Notteboom, T. Technical challenges in port operations: A review of the current state of the art. *J. Shipp. Port Res.* **2017**, *3*, 45–59.
25. Richardson, C.; Wilmsmeier, G. Operational inefficiencies affecting vessel stay at ports: Insights from a global survey. *Transp. Res. Part E Log. Transp. Rev.* **2016**, *88*, 31–45.
26. Djukic, T.; Tadic-Djincic, J. Legal aspects of vessel detention in port: Case study of Port of Bar. *Inter. J. Marit. Law* **2018**, *24*, 187–203.
27. Offshore Vessels. Available online: https://www.marineinsight.com/types-of-ships/what-are-offshore-vessels/ (accessed on 20 June 2024).
28. Paris MoU. Available online: https://parismou.org/PMoU-Procedures/Library/memorandum (accessed on 23 May 2024).
29. Tokyo MOU Deficiency Codes. Available online: https://www.tokyo-mou.org/doc/Tokyo%20MOU%20deficiency%20codes%20(December%202019).pdf (accessed on 21 June 2024).
30. Cran.rpart. Available online: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf (accessed on 20 June 2024).
31. Lewis, R. An Introduction to Classification and Regression Tree (CART) Analysis. In Proceedings of the 2000 Annual Meeting of the Society for Academic Emergency Medicine, San Francisco, CA, USA, 22–25 May 2020.
32. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier Chains: A Review and Perspective. *J. Artif. Intell. Res.* **2021**, *70*, 683–718. [CrossRef]
33. Freitas, A.; De Carvalho, A. A Tutorial on Hierarchical Classification with Applications in Bioinformatics. *Resear. Trends Data Min. Techn. Applic.* **2007**, 175–208. [CrossRef]
34. Siekelova, A.; Podhorska, I.; Imppola, J. Analytic Hierarchy Process in Multiple-Criteria Decision-Making: A Model Example. *SHS Web Conf.* **2021**, *90*, 01019. [CrossRef]
35. Borawska, A. Multiple-Criteria Decision Analysis Using Topsis Method for Interval Data in Research into the Level of Information Society Development. *Folia Oeconomica Stetin.* **2014**, *13*, 63–76. [CrossRef]
36. Taherdoost, H.; Madanchian, M.A. Comprehensive Overview of the ELECTRE Method in Multi Criteria Decision-Making. *J. Manag. Sci. Eng. Res.* **2023**, *6*, 5–16. [CrossRef]
37. Taherdoost, H. Using PROMETHEE Method for Multi-Criteria Decision Making: Applications and Procedures. *Iris J. Econ. Busi. Manag.* **2023**, *1*, 45–67. [CrossRef]
38. Dyer, J.S. Multiattribute Utility Theory (MAUT). In *Multiple Criteria Decision Analysis, International Series in Operations Research & Management Science*; Greco, S., Ehrgott, M., Figueira, J., Eds.; Springer: New York, NY, USA, 2016; Volume 233. [CrossRef]
39. Rpart. Available online: https://cran.r-project.org/web/packages/rpart/rpart.pdf (accessed on 25 June 2024).
40. Biró, T.S.; Néda, Z. Gintropy: Gini Index Based Generalization of Entropy. *Entropy* **2020**, *22*, 879. [CrossRef]
41. Pruning in, R. Available online: https://stat.ethz.ch/R-manual/R-devel/library/rpart/html/prune.rpart.html (accessed on 20 June 2024).
42. Equasis.org. Available online: https://www.equasis.org/EquasisWeb/authen/HomePage?fs=HomePage (accessed on 23 May 2024).
43. Dummy_cols in R. Available online: https://www.rdocumentation.org/packages/fastDummies/versions/1.7.3/topics/dummy_cols (accessed on 25 June 2024).

44. R Project. Available online: https://www.r-project.org (accessed on 10 June 2024).
45. TOPSIS in R. Available online: https://www.rdocumentation.org/packages/MCDA/versions/0.1.0/topics/TOPSIS (accessed on 20 June 2024).
46. Bari, A.M.; Jamil, A.H.M.S.; Mamtaz, M.; Monirujjaman, K.M.; Aljahdali, S.; Kaur, M.; Singh, P.; Masud, M.A. Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. *J. Healthc. Eng.* **2021**, *2021*, 9917919. [CrossRef] [PubMed] [PubMed Central]
47. Krakauer, D.C. Unifying complexity science and machine learning. *Front. Complex Syst.* **2023**, *1*, 1235202. [CrossRef]

Appendix B

Paper II

Boko, Z., Skoko, I., Sanchez Varela, Z., & Milin, V. (2025). Advancing Maritime Safety: A Literature Review on Machine Learning and Multi-Criteria Analysis in PSC Inspections. Journal of Marine Science and Engineering, 13(5), 974. https://doi.org/10.3390/jmse13050974

Journal of
Marine Science and Engineering

MDPI

Review

Advancing Maritime Safety: A Literature Review on Machine Learning and Multi-Criteria Analysis in PSC Inspections

Zlatko Boko *, **Ivica Skoko**, **Zaloa Sanchez Varela** and **Vice Milin**

Faculty of Maritime Studies, University of Split, 21000 Split, Croatia; iskoko@pfst.hr (I.S.); zsanchezv@pfst.hr (Z.S.V.); vmilin@pfst.hr (V.M.)
* Correspondence: zboko@pfst.hr

Abstract: This literature review provides a structured quantitative analysis of existing research on the application of machine learning models (MLMs) and multi-criteria decision-making methods (MCDM) in the context of port state control (PSC). The aim of the study is to capture current research trends, identify thematic priorities, and demonstrate how these analytical tools have been used to support decision-making and risk assessment in the maritime domain. Rather than evaluating the effectiveness of individual models, the study focuses on the distribution and frequency of their use and provides insights into the development of methodological approaches in this area. Although several studies suggest that the integration of MLMs and MCDM techniques can improve the objectivity and efficiency of PSC inspections, this report does not provide a comparative assessment of their performance. Instead, it lays the groundwork for future qualitative studies that will assess the practical benefits and challenges of such integration. The findings suggest a fragmented but growing research interest in data-driven approaches to PSC and highlight the potential of advanced analytics to support maritime safety and regulatory compliance.

Keywords: literature review; machine learning models; multi-criteria analysis; port state control; maritime safety

check for updates

Academic Editor: Xinqiang Chen

Received: 31 March 2025
Revised: 15 May 2025
Accepted: 16 May 2025
Published: 17 May 2025

Citation: Boko, Z.; Skoko, I.; Sanchez Varela, Z.; Milin, V. Advancing Maritime Safety: A Literature Review on Machine Learning and Multi-Criteria Analysis in PSC Inspections. *J. Mar. Sci. Eng.* **2025**, *13*, 974. https://doi.org/10.3390/jmse13050974

1. Introduction

This paper analyzes the integration of machine learning (ML) models and multi-criteria analysis (MCA) in assessing vessel safety risks within the context of port state control (PSC) inspections, applying a systematic literature review (SLR).

The focus of this study is on a systematic literature review on the application of advanced ML and MCA methods to improve evaluation, forecasting, and planning processes in the context of PSC. The approach is primarily quantitative, with the intention of collecting and processing relevant literature documenting trends, scope, and specific applications of these methods in the maritime sector. The aim is not to critically evaluate each individual study or to assess the merits and limitations of particular techniques, but rather to systematically collect, categorize, and analyze existing work in order to assess the scope, dynamics, and focus areas of current research.

This provides an insight into the extent to which ML and MCA techniques have already been integrated into inspection risk assessment, resource planning in shipping companies, and the development of strategic regulatory approaches within international frameworks such as Memoranda of Understanding (MoU). While the practical relevance of these methods for improving maritime safety is obvious, the main contribution of this study is to take stock of existing knowledge and identify potential research gaps.

The study is part of a broader research project (PhD), with the next phase focusing on collecting newer and more comprehensive data sources to support further investigations while strengthening the application of advanced MLMs and MCA techniques in the context of PSC.

Maritime safety is a key component of the global transport system, and PSC [1,2] is an integral part of the international strategy aimed at reducing the risk of marine accidents. This system operates within the framework of the International Maritime Organization (IMO). It relies on regional PSC agreements, such as the Paris MoU and the Tokyo MoU, which facilitate the exchange of information and best practices among signatory states. PSC inspections play a crucial role in international maritime oversight, aiming to ensure safety and the protection of life at sea, as well as the preservation of the environment from potential maritime accidents [3–6].

The main objective of PSC inspections is to identify and eliminate safety risks and potential violations that could jeopardize the safety of the vessel, the crew, and the environment. Based on binding international conventions such as SOLAS (Safety of Life at Sea), MARPOL (Marine Pollution) and STCW (Standards of Training, Certification and Watchkeeping), PSC inspectors are authorized to inspect vessels in port, regardless of the flag under which the vessel sails, to check compliance with international standards. During these inspections, which may encompass various aspects of the vessel's operation, inspectors assess the vessel's condition, including its structural components, safety equipment, environmental protection systems, and the qualifications of its crew. In the event of non-compliance or irregularities, inspectors may decide to detain the vessel in port, which can have serious consequences for both the vessel owners and its operations.

Modern risk assessment methods for PSC vessel inspections extend far beyond traditional approaches, which primarily rely on statistical analyses and expert judgment. These are often limited by the subjective judgements of the inspectors and, from the shipowner's perspective, reduce the ability to assess potential violations or the likelihood of the vessel being detained. The use of MLMs [7,8] enables the automated processing of large data sets [9], pattern recognition, and a more accurate determination of safety risks. At the same time, MCA provides a systematic framework for decision-making by considering multiple relevant factors simultaneously, which improves the interpretation of the results generated by MLMs.

The application of these methods enables a more precise prediction of inspection results. It is therefore extremely useful for both inspectors, who need to formulate requirements by legal standards, and shipowners, who want to minimize costs and optimize the time vessels spend in port. By utilizing advanced technologies, including artificial intelligence (AI), machine learning (ML), and algorithms, it is possible to optimize procedures for identifying potential risks based on historical data on violations and deficiencies on vessels. This approach not only improves the efficiency of inspections but also helps reduce the number of accidents caused by technical defects [10–13].

In this context, current research [14–19] focuses on the development of sophisticated MLMs that enable automatic pattern recognition in data from previous inspections and a more reliable prediction of future violations and safety risks. The main challenges in the field of PSC inspections and vessel safety are to increase the accuracy of predictions and improve the robustness of these models to ensure their applicability in real operational conditions.

The effectiveness of different ML and multi-criteria decision-making (MCDM) methods in PSC depends on the specific requirements of the inspection process. Methods such as decision trees are highly interpretable and easy for inspectors to understand, but often have lower accuracy in more complex scenarios. In contrast, methods such as Random

Forests (RFs) and Support Vector Machines (SVMs) offer higher predictive performance and accuracy, but their complexity reduces transparency and makes the results more difficult to interpret. Bayesian network (BN) algorithms are fast and simple and are therefore suitable for initial analyses, but are limited by the assumption of feature independence. On the other hand, MCDM methods such as the Analytic Hierarchy Process (AHP) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) allow a structured ranking of vessels based on risk and other factors, which support informed decision-making, although they rely on subjective judgments that may influence the final outcome. In practice, an optimal approach often requires the combination of interpretable and powerful methods to find a balance between the reliability of the model and the confidence of the inspectors.

The combination of MLMs and MCA opens up the possibility of improving the efficiency of PSC inspections and increasing maritime safety [20–23].

The aim of this paper is therefore to conduct a SLR for the period of the last ten years (from 2015 to 2025) to analyze existing approaches to the integration of MLMs and MCA in the assessment of vessel safety risks in the context of PSC reviews, to identify essential methods and research gaps, and to define possible directions for future research.

The paper is organized in such a way that the introductory section presents the basic concepts of the research, focusing on the importance of applying MLMs and MCA in assessing safety risks within the context of PSC inspections. The *Materials and Methods* chapter outlines the framework of the research methodology, including the search and analysis strategy for relevant literature, the criteria for selecting sources, and the data used for analysis. The *Results* and *Discussion* chapters present the analysis results, along with a discussion of the key findings and challenges encountered in implementing MLMs and MCA methods within the context of PSC inspections. Finally, the *Conclusions* chapter summarizes the main findings of the study. It formulates recommendations for future research and improvement of the methodological approach to safety risk assessment through the application of modern technological solutions and their implementation in the field of PSC inspections.

1.1. *Maritime Safety and PSC Inspection*

In the section "*Maritime Safety and PSC Inspection*", the studies analyzed highlight the most important aspects of maritime safety and the role of PSC inspections in maintaining this safety. The main findings of these studies relate to the importance of PSC inspections, the impact of the COVID-19 pandemic in this area, the most commonly identified deficiencies in PSC inspections, human factors as key elements of maritime safety, and the researchers' emphasis on the need for technological improvements and better regulations.

The data collected during PSC inspections form the basis for improving safety in the maritime industry and enable the development of strategies to improve inspection processes and predict potential risks on vessels [24,25].

Considering that PSC inspections are one of the most important mechanisms for ensuring maritime safety and reducing operational risks, their role in identifying violations on vessels is crucial, particularly in contributing to the reduction of marine casualties and environmental incidents.

Studies [26–29] confirm that PSC inspections play a key role in the implementation of international regulations such as SOLAS and MARPOL and ensure a high level of safety worldwide. Their effectiveness in reducing the risk of accidents through preventive measures and penalizing vessels that do not meet safety standards is particularly emphasized.

The COVID-19 pandemic has had a significant impact on PSC inspections, making them more challenging to conduct and reducing the scope of inspection oversight. Studies [30–35] show that the restrictions introduced during the pandemic have led to fewer inspections, difficulties in physically monitoring vessels, and adjustments to regulatory procedures to ensure uninterrupted shipping traffic. Due to the limited access of inspectors to vessels, the number of safety violations increased, particularly in areas such as technical maintenance and crew training. These problems also highlighted the need to digitize inspection procedures to ensure their continuous application even in exceptional circumstances.

The data analysis of PSC inspections reveals that the most common deficiencies and violations on vessels are related to technical aspects, safety procedures, and human factors. During Paris MOU inspections, problems related to safety equipment, navigation systems, and vessel maintenance are most frequently identified. Additionally, inadequate crew training and non-compliance with safety procedures often contribute to serious incidents at sea.

Research indicates that vessels with a history of repeated violations are inspected more frequently, underscoring the importance of consistently applying inspection protocols. The human factor is a crucial aspect of maritime safety, as most accidents at sea are attributed to human error. Inadequate training, crew fatigue, and poor management of safety procedures can significantly affect the safety of shipping. A particular problem is the lack of awareness of the importance of compliance with safety regulations, which emphasizes the need for continuous training and education of seafarers.

Current research [36–39] points to the need for technological improvements and better regulation in the area of PSC controls. The automation of data analysis, the use of digital technologies in vessel monitoring, and the improvement of international co-operation between port states can significantly increase the efficiency of inspection procedures. The introduction of digitalized inspection systems would certainly enable faster and more accurate identification of safety deficiencies. At the same time, improvements in regulations would contribute to more consistent enforcement of international safety standards.

The results of the previously cited studies and research confirm that PSC inspections play a crucial role in improving maritime safety, with particular emphasis on the importance of the human factor, the impact of global crises, and the need to modernize and optimize inspection methods to increase their efficiency and ensure a higher level of safety in the shipping industry [40].

In the context of PSC, it is important to understand that although the core process often involves physical vessel inspections, digitization can significantly enhance and support various aspects of the process. Physical inspection of vessels is still important, but digital tools and technologies can improve the efficiency and accuracy of the inspection process.

For example, digitization can enable faster access to data through electronic records of previous inspections, vessel tracking, and compliance data, and streamline data collection. In addition, advanced technologies such as remote monitoring, sensors, and AI-based predictive analytics can help inspectors identify potential issues before they become critical, increasing safety and reducing the risk of accidents.

While digitalization cannot completely replace physical inspections, it can improve the process by allowing inspectors to better analyze data, make informed decisions faster, and gain easier access to important information. This approach enables a more efficient and safer inspection process, contributing to the overall improvement of the system.

1.2. Application of MLMs in PSC

PSC inspections are an essential mechanism for monitoring maritime traffic, allowing ports to ensure that vessels comply with international safety, environmental, and labor standards. Traditional inspection methods rely on manual inspections, the inspector's experience, and predefined lists of criteria. However, the growing amount of data generated from previous inspections, digital registers, and sensor systems opens up the possibility of applying MLMs aimed at optimizing inspection procedures, reducing operational risks, and improving predictive analysis in the identification of high-risk vessels [41–49].

The application of MLMs in the context of PSC helps recognize patterns in large datasets, reduces the need for manual intervention, and increases the efficiency of control mechanisms.

MLMs [50–52] can be divided into two main categories:

1. Supervised learning methods (SLMs) use labeled data to train models that enable the prediction or classification of new data based on previous experience.
2. Unsupervised learning methods (ULMs) do not require labeled data and are applied to detect hidden patterns and structures in the data, which helps discover new relationships or clusters among the data.

The most commonly used SLMs are as follows:

Linear regression is a statistical model used to predict continuous values based on the linear relationship between variables. In the context of PSC, linear regression is used to analyze quantitative variables that influence the frequency of incidents or inspection outcomes, such as vessel age, vessel size, and previous safety violations. Although this approach can be helpful in analyzing trends, its use is limited to situations where there is a linear relationship between the predictors and the target variable [19,53–56].

Logistic regression is a binary classification method that predicts the probability of an event occurring, i.e., it provides a yes/no answer (e.g., whether a vessel needs to be inspected or not). In PSC reviews, logistic regression helps to assess vessels with a higher probability of a safety risk based on historical accident data and non-compliance or other risk factors [16].

Decision trees are classification and regression techniques based on hierarchical data decomposition. They work by successively branching nodes based on selection criteria, with decisions being binary or multi-class depending on the algorithm used and the nature of the problem. In the context of PSC inspections, these methods enable a systematic analysis of relevant variables, such as vessel type and age, frequency and condition of previous inspections, among others, allowing for an objective assessment of risk and determination of the need for detailed inspections. Their interpretability and transparency make them an effective tool for making informed regulatory decisions [57,58].

SVMs belong to the group of ML and are based on models that are used for classification or regression by defining an optimal hyperplane that separates the data classes. There are several variants and extensions of SVMs, such as the Linear SVM, which is used when the data are linearly separable; non-linear SVM, which uses kernel functions to work with non-linearly separable data; Support Vector Regression (SVR), a version of SVM adapted for regression problems; and one-class SVM, which is used for anomaly and outlier detection. In the context of PSC inspections, an SVM enables the accurate classification of vessels based on their likelihood of meeting safety standards and identifies those that require priority inspection oversight. A key advantage of this method is its ability to efficiently analyze high-dimensional and non-linearly separable data, which is crucial for processing complex inspection criteria [49,59].

RF [53,60] is an ensemble method based on a set of decision trees that enables collective decision making to improve prediction accuracy and reduce the risk of overfitting. In the context of PSC inspections, RF is used to predict irregularities and risks on vessels, relying on a wide range of variables, including vessel type, inspection history, and other relevant factors. Its main advantage lies in its robustness and its ability to analyze complex and heterogeneous data. AdaBoost [61] is also an ensemble MLM that combines several weak classifiers (e.g., decision trees) into a more potent model. This technique uses labeled data to train the model and iteratively improves its performance by assigning a higher weight to samples that are difficult to classify. Applying these methods enhances predictions by sequentially learning and weighting the samples, resulting in improved precision in data analysis.

The most commonly used ULMs in the context of PSC inspections are as follows:

The K-Means algorithm is a clustering method that groups data into a predefined number of clusters based on their similarity to one another. In the context of PSC inspections, K-Means is used to classify vessels based on common characteristics, such as vessel type, age, or inspection history. This enables the more efficient allocation of inspection resources and helps identify hidden patterns among vessels that might otherwise go unnoticed using traditional analytical approaches [62].

Cluster analysis represents a group of MLMs and statistical analyses that allow segmentation of data into homogeneous groups based on their similarities. These methods are crucial for uncovering hidden patterns in large datasets, enabling informed decisions in various areas, including maritime safety and inspection processes. In the context of PSC inspections, cluster analysis is used to identify unsafe or high-risk groups of vessels based on their specific characteristics, such as the frequency of incidents, technical deficiencies, compliance with international standards, and frequency of environmental violations. By applying cluster analysis methods, it is not only possible to recognize potentially risky segments of the fleet but also to identify patterns that may indicate systemic failures in the maintenance and monitoring of vessels. Additionally, cluster analysis enables the identification of long-term trends in the maritime industry, which can significantly enhance inspection planning strategies. Based on the analysis results, it is possible to optimize the distribution of inspection resources and focus on the groups of vessels that pose the most significant risk to safety and environmental protection. By integrating cluster analysis with other machine learning methods, such as RF and AdaBoost, the accuracy of predictions can be further improved, enabling a proactive approach to maritime risk management [63].

Artificial neural networks (ANNs) [64–67] and deep learning [68,69] are highly developed models that imitate the human brain. ANNs can recognize patterns in complex and unstructured data. In the context of PSC inspections, ANNs can be utilized to identify complex relationships between various variables that impact the safety of vessels, including weather conditions, vessel type, previous violations or incidents, and others. The advantage of ANNs lies in their ability to learn and adapt to changes in the data, making them very effective in predicting risks and accidents. ANNs and their advanced variants, such as Convolutional Neural Networks (CNNs) [70,71] and Recurrent Neural Networks (RNNs) [72], are increasingly being used to analyze large data sets generated during PSC inspections.

BNs and Bayesian models (BMs) can be applied to both supervised and unsupervised learning, depending on their usage and the presence of labels in the data. When BNs are used for classifications or regressions where the data are already labeled (with output variables), they are categorized as supervised methods. For example, when they are used to predict the probability of a particular outcome based on known characteristics, such as predicting the risk of a shipping accident based on historical data. The application of BNs falls under unsupervised learning when they are used to uncover hidden structures in data without prior knowledge of the initial variables. In addition, BNs are often used in semi-supervised learning (SSL) when there is a combination of labeled and unlabeled data.

The application of BNs and BMs to PSC inspections [38,73–82] represents one of the most advanced solutions for probabilistic modeling and decision-making under uncertainty. These methods enable the formalization of complex dependencies between variables and provide a robust framework for analyzing dynamic systems with incomplete and uncertain data. Their application to PSC inspections contributes to a more accurate risk assessment, improvement of inspection strategies, and optimization of resource allocation, thus increasing the efficiency and effectiveness of marine surveillance. The cited theoretical and empirical studies confirm that BNs facilitate a more in-depth analysis of the factors that influence the outcomes of inspections, including the likelihood of detention, identification of high-risk vessels, and assessment of compliance. By using probabilistic models, researchers are developing sophisticated methods for predicting security threats and optimizing inspection procedures, reducing uncertainty in the decision-making process. Some papers focus on improving vessel liability prediction algorithms, while others investigate the integration of BNs with MCA to achieve more accurate modeling of risk scenarios and increase regulatory efficiency. The common goal of these studies is to improve maritime safety through the use of advanced MLMs. BNs are utilized as key tools for building intelligent risk assessment systems and facilitating the development of sophisticated, data-driven inspection strategies. Their application not only increases the efficiency of PSC procedures but also allows for the more precise identification of potential threats and the making of optimal regulatory decisions. The synergy of these studies reflects the goal of developing predictive models that improve preventive measures and enhance global vessel safety.

The application of MLMs to PSC inspections represents a significant advance in improving the efficiency, accuracy, and predictive analysis of vessel safety and regulatory compliance. Monitored methods enable accurate vessel classification based on historical inspection data, optimizing the identification of high-risk vessels and increasing the efficiency of inspection processes. Non-monitored techniques, on the other hand, uncover hidden patterns in the data, enable proactive detection of potential irregularities, and optimize the allocation of inspection resources. Integrating these methods into PSC inspections not only improves the detection of violations but also significantly contributes to increasing vessel safety, reducing operational risks, and minimizing the likelihood of accidents at sea through more accurate and informed regulatory decisions.

1.3. Application of MCDMs in PSC Inspections

MCA and MCDM methods [83,84] serve as basic approaches for optimizing inspection procedures in the context of PSC inspections. Their application facilitates a structured and quantitatively sound evaluation of alternative solutions, considering multiple, often conflicting criteria, and provides a methodologically rigorous framework for decision-making in complex regulatory and operational environments. Unlike MLMs, which rely on algorithmically generated patterns derived from data, MCA is based on explicit decision modeling that enables transparent analysis of the determinants influencing inspection outcomes.

In the context of PSC inspections, the implementation of MCA involves defining relevant evaluation criteria, selecting and analyzing alternatives based on predefined parameters, determining the relative weighting of the criteria through expert judgment or empirical data, and integrating the results using advanced aggregation methods. This analytical approach enables the formal quantification of safety risks, the objectification of inspection decisions, and the optimization of resource allocation, which ultimately increases the efficiency of regulatory oversight and reduces uncertainty in assessing safety risks [85].

In the risk assessment of vessels, MCA enables the integration of various factors, including technical specifications, operational parameters, inspection history, flag state, detention frequency, and compliance with international conventions.

The best-known methods used in this context include the AHP, the TOPSIS, PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations), and ELECTRE (Elimination and Choice Expressing Reality). Each of these methods offers different advantages in the multi-criteria optimization process.

The AHP method [86] facilitates the hierarchical decomposition of decision problems and uses pairwise comparison analysis to quantify the relative importance of criteria.

The TOPSIS method [76,87] ranks alternatives based on their proximity to ideal and anti-ideal solutions, ensuring the selection of an option that optimally balances competing criteria.

The PROMETHEE [88] method incorporates preference functions that enable a more adaptive and dynamic evaluation of alternatives, aligning with the decision-makers' subjective preferences.

The ELECTRE [89] method employs a process of eliminating dominant alternatives, supporting decision-making in highly uncertain environments with competing criteria. This is particularly beneficial for inspection procedures that require a comprehensive approach to safety risk assessment.

Additionally, the synergistic integration of MCDM with probabilistic models, such as Bayesian networks (BNs), facilitates the formal consideration of uncertainties and dependencies between key inspection parameters. This approach enhances the accuracy of risk prediction, optimizes inspection strategies, and contributes to the overall improvement of maritime safety [8,26,33,40,76].

The application of MCA methods in the context of PSC inspections is a key element of the modern approach to safety risk assessment and regulatory oversight in the maritime sector. Its application enables the systematic analysis and quantification of risk factors, refines inspection methodologies, and improves the efficient allocation of inspection resources, ultimately promoting a strategically oriented and science-based decision-making model for maritime safety.

1.4. Application of Combined ML and MCA Methods in the Context of PSC Inspections

In today's PSC inspection practice, there is a growing need for more efficient and accurate methods to analyze safety risks on vessels. Traditional approaches that rely on human judgment and simple algorithmic techniques are being gradually replaced by advanced, integrated, and novel methods known as "hybrid" or "combined" methods. These methods enable the efficient processing of large amounts of data and decision-making based on complex analysis. In this context, the combination of MLMs and MCA represents a significant step towards improving the accuracy and efficiency of safety assessments for vessels subject to PSC review.

The combination of MLMs and MCA in PSC inspections improves the assessment of vessel safety risks. DL methods, such as stacked autoencoders [90], have proven to be highly effective in extracting relevant features from high-dimensional inspection data. Furthermore, algorithms such as XGBoost, which are based on gradient boosting methods, are characterized by iterative model refinement in risk classification and prediction. In addition to these methods, the application of reinforcement learning (RL) techniques enables adaptive decision making in inspections by iteratively learning optimal strategies from previous results [58,91,92].

The application of the Gale–Shapley algorithm [46] plays a crucial role in optimizing the allocation of inspection resources. This algorithm facilitates the allocation of inspection capacity based on a multi-criteria evaluation of vessels, ensuring a balance between inspection efficiency and accuracy. Consequently, it optimizes the allocation of inspectors to the vessels that pose the highest safety risk. In addition, principal component analysis (PCA) [58] helps to reduce the dimensionality of the data and identify the most critical risk factors, allowing further study and a more accurate assessment of safety parameters.

The integration of MLMs with MCA techniques such as AHP and TOPSIS enables decision-making based on the simultaneous assessment of multiple aspects of vessel safety. These methods enable the incorporation of expert judgment and quantitative data, ensuring objectivity in the evaluation process [76,87].

The combined application of MLMs and MCA significantly improves the ability to identify high-risk vessels, optimizes inspection processes, and contributes to a more efficient allocation of inspector capacity. This approach enables more accurate risk assessment and better resource allocation, thereby increasing the overall efficiency of inspections and maritime safety.

In summary, the integration of MLMs and MCA, along with other advanced statistical and mathematical methods in PSC inspections, not only enables a more accurate risk assessment but also optimizes the allocation of inspection resources [93]. This ensures that capacity is targeted to those vessels that pose the most significant risk to maritime safety. This approach enhances the accuracy and speed of decision-making while also providing greater transparency and accountability in the inspection process, ultimately contributing to improved global maritime safety.

Table 1 provides a brief quantitative summary of the academic studies, organized by their thematic focus on PSC and associated methods. The largest group of studies (40) addresses the use of MLMs in PSC, reflecting a strong research focus on improving inspection processes through advanced technologies, data analytics, and algorithm-based decision-making.

Table 1. Distribution of references according to thematic focus in connection with PSC inspections.

Subject	References
PSC inspection in general	[1–23]
Maritime safety and PSC inspection	[24–40]
Application of MLMs in PSC	[16,19,38,41–49,52–82]
Application of MCDMs in PSC Inspections	[8,26,33,40,76,83–89]
Application of combined ML and MCA methods in the context of PSC Inspections	[46,58,76,90–93]
ML in general	[50,51,68]

The second most represented category includes papers addressing the general aspects of PSC inspections (23 sources), highlighting the continued academic focus on understanding the regulatory framework, operational structure, and core functions of the PSC system.

Maritime safety in the context of PSC inspections is represented by 17 sources, confirming the crucial role of PSCs in ensuring the safety of life, property, and the marine environment.

The use of MCDM methods in the context of PSCs is documented in 10 sources, while combined approaches involving both ML and MCDM techniques appear in six references. This reflects a growing, albeit still ongoing, trend towards integrated analytical models in this area.

The least represented group is "ML in general" with only three sources, suggesting that general ML concepts in this paper are explored through their specific application to PSC rather than as a standalone research area.

Overall, Table 1 illustrates a clear shift towards digitalization and data-driven methods in PSC, with a focus on automating decision making, improving safety procedures, and reducing subjectivity in vessel condition assessment. This thematic division shows an emerging research trend towards integrated, data-driven approaches for the evaluation and optimization of PSC performance. The focus is on analytical decision support systems that aim to improve the objectivity and efficiency of inspection processes across the maritime sector.

Given the complexity and scope of the topic, this work represents the first stage of a broader research initiative. The present work focuses on mapping the research landscape and identifying key trends and patterns based on frequency, distribution, and thematic grouping. A subsequent qualitative analysis is already planned, which will examine the specific methods, data sets, and results of the identified studies in detail.

However, some initial qualitative reflections on the performance and characteristics of the methods used have already been made—particularly in the papers cited as references [58,67], which provide methodological insights into selected approaches. We believe that this multi-layered methodology allows for a more structured and meaningful synthesis of the literature, starting with a macro-level overview and moving to a more fine-grained analysis in the next phase.

2. Materials and Methods

Data processing in this study was performed using R-4.5.0 packages for Windows [94], which provide a robust framework for efficient data analysis and manipulation. R, a versatile tool for statistical analysis and data visualization, offers a comprehensive selection of packages that support various analysis techniques, from basic statistical tests to advanced modeling approaches. Key packages, such as dplyr (1.1.4), ggplot2 (3.5.1), and tidyr (1.3.1), have been utilised to process, clean, and visualise the data, enabling insightful interpretation and comprehensive analysis of the results.

2.1. Bibliography Data Description

In this study, the input dataset comprises 93 sources stored in an Excel spreadsheet, which contains the collected and relevant literature on PSC inspections, MLMs, and MCA. Only relevant sources were selected, organized, and systematized. The table is sorted alphabetically by the last name of the first author and contains the following key information: Column 1—Source ID; Column 2—Full source citation (Author 1, A.B.; Author 2, C.D. Title of the article. Abbreviated name of the journal, Year, volume, pages); Column 3—List of authors; Column 4—Title of the article; Column 5—Keywords; Column 6—Year of publication; Column 7—Abbreviated name of the journal/conference; Column 8—Full name of the journal/conference; Columns 9–12—Indexing data in relevant databases (Web of Science, Scopus, Google Scholar, ScienceDirect), where 1 means that the source is indexed and 0 otherwise; Column 13—Total number of authors of the publication. This structured breakdown enables efficient filtering and categorization of the sources according to the criteria of the PICOC method, ensures a comprehensive analysis, and serves as a basis for further discussion and interpretation of the results.

Keywords were generated to identify relevant terms and phrases related to the research topic. Terms were selected that most accurately reflect the content of the sources, thereby improving findability in search engines and databases. The selected keywords were strategically integrated into the title, summary, and paragraphs of the text to maximize the article's visibility and accessibility to the target audience. Abbreviations were introduced for the keywords in the table, while all synonyms were standardized to ensure that each term appeared only once. For example, vessel detention is abbreviated as vd, while artificial neural network (ANN) is shortened to "nn", which also applies to other terms. This approach enhances searchability, ensures consistency, and facilitates the efficient retrieval of relevant terms, thereby making the study more accessible to researchers and experts in the field.

Table 2 shows the frequency distribution of the different types of sources in the collected literature. It categorizes the various types of sources (e.g., journals, conference papers, books, dissertations) and indicates how often each type appears in the dataset.

Table 2. Frequency of different source types.

Journal or Source Type	Frequency
Journal	86
Proceedings	2
Book or chapter in book	2
PhD Dissertation	1
Master of Science thesis	1
Web link	1

2.2. Literature Review Methodology

There are various methods for conducting literature reviews. The most commonly used methods include the systematic literature review (SLR), the narrative literature review (NLR), meta-analysis, scoping review (SR), critical review (CR), and the population, intervention, comparison, outcome, and context (PICOC) method. Each of these methods has specific applications and advantages. However, given the aim and purpose of this study, the SLR method with the application of PICOC is considered the most appropriate. The SLR method enables a rigorous, systematic, and objective analysis of existing research, which is crucial for a topic that integrates MLMs and MCA in the context of PSC security risk analysis. This approach enables the selection of relevant sources based on predefined criteria, reducing subjectivity and ensuring the accuracy of the analysis [95,96].

The combination of the SLR method with PICOC enables a precise formulation of parameters for literature selection, thus increasing the precision and relevance of the analysis. The PICOC method is essential for formulating the research question and filtering out relevant literature sources through five key components:

- (P)opulation refers to all studies, papers, dissertations, and articles related to PSC inspections.
- (I)ntervention focuses on the application of methods such as MLMs and MCA in analyzing PSC inspections.
- (C)omparison analyzes different approaches to the application of MLMs and MCA in the field of PSC inspections and compares them with other methods of safety risk analysis.
- (O)utcome identifies and evaluates safety risks on vessels and the improvement of existing methods and models in the context of PSC inspections.
- (C)ontext refers to the application of these methods in the field of maritime inspections, with a particular focus on vessel safety research and the implementation of the relevant technologies.

By employing the SLR method in conjunction with PICOC, this paper offers a comprehensive analysis of methodological approaches, identifies research gaps, and explores the connections between studies in this area. This approach enables the systematic integration of MLMs and MCA methods within the context of PSC inspections, which is crucial for improving existing models and developing new strategies in this area.

Table 3 presents a comprehensive breakdown of the steps involved in the systematic research process. The initial phase focuses on formulating a well-defined research question using the PICOC methodology, with an emphasis on PSC inspections and the application of MLMs and MCA. This is followed by the establishment of rigorous selection criteria, which clearly define inclusion and exclusion parameters based on the research scope, methodology, and publication years.

Table 3. Step activity description.

1. DEFINITION OF THE RESEARCH QUESTION	Formulate a clear and concise research question using the PICOC method. Focus on the safety risks associated with PSC controls, including MLMs and MCA.
2. FORMULATION OF CRITERIA FOR STUDY SELECTION	Use the PICOC framework to define specific criteria for selecting relevant studies. Define the inclusion and exclusion criteria for appropriate studies, based on factors such as research area, methodology, and publication year.
3. LITERATURE SEARCH	Systematic search in relevant databases, including Web of Science, Scopus, Google Scholar, and ScienceDirect. Use of specific keywords and phrases related to MLMs, MCA, and PSC inspections.
4. STUDY SELECTION	Application of the previously defined inclusion and exclusion criteria to select only relevant studies. This step involves reviewing abstracts, keywords, and conclusions.
5. ASSESSMENT OF THE QUALITY OF THE STUDIES	Assessment of the quality of the selected studies based on methodological criteria.
6. DATA EXTRACTION	Systematic extraction of key data from the selected studies, including methodology, results, variables analyzed, and applications of MLMs and MCA in relation to PSC inspections.
7. DATA ANALYSIS AND SYNTHESIS	Analyzing and synthesizing the data from all studies to identify common themes, methodological approaches, and research gaps. Integrate findings into a unified framework that links MLMs, MCA, and PSC inspections.
8. WRITING THE REPORT	Produce a report summarizing the key findings of the literature review. This report will include a critical review of methodological approaches, identification of research gaps, and suggestions for future areas of research.
9. DISCUSSION AND CONCLUSIONS	Discussion on the importance of integrating MLMs and MCA in analyzing PSC inspections and the benefits and challenges of using these methods. Suggestions for further development and improvement of research in this area.

The next stage involves conducting a systematic search across key academic databases, including Web of Science, Scopus, Google Scholar, and ScienceDirect, utilizing carefully selected keywords and phrases related to MLM, MCA, and PSC inspections. Once the studies are selected, their quality is critically evaluated against predefined methodological criteria. Subsequently, essential data are extracted from the chosen studies, including research methodology, key findings, analyzed variables, and the implementation of MLMs and MCA within the context of PSC inspections.

The extracted dataset is then systematically analyzed and synthesized to identify prevailing thematic patterns, methodological trends, and existing research gaps. Finally, a structured report is compiled, offering a synthesized overview of key insights from the systematic literature review, a critical evaluation of methodological approaches, identification of unresolved research gaps, and recommendations for future investigations in this domain.

The methodological steps (Table 3) provide a structured and comprehensive application of the SLR methodology, using PICOC for the precise formulation of the research question and criteria. This approach ensures objectivity, rigor, and thoroughness in the literature review, which is essential for integrating different approaches in assessing safety risks during PSC inspections.

3. Results

Table 4 illustrates the distribution of the number of authors per paper in the collected sources. Papers with three and four authors are the most common (24), suggesting that research in this area tends to be conducted in smaller teams, allowing for effective collaboration and coverage of different areas of expertise. This distribution suggests that research in this area often requires the collaboration of multiple researchers, but without the need for excessively large teams. Single-authored articles (11) are relatively rare, and articles with five or more authors are even rarer, suggesting that a larger number of authors is not always necessary to conduct high-quality research in this area.

Table 4. Distribution of authors per paper in the dataset.

Number_of_Authors	Frequency
3	24
4	24
2	17
1	11
5	10
6	3
7	2
9	2
Total	93

The focus was on identifying sources dealing with PSC and the application of MCA and MLMs in the context of PSC risk assessment and decision-making. The combined search terms included terms such as PSC, PSC inspection, vessel detention, ML, effectiveness, risk, safety, BN, and the like, with the term "PSC" necessarily included. These terms can also be seen in Figure 1, which helps to explain the frequency of certain terms in the visualization.

The top 10 keywords (Figure 1) represent the most frequently used terms in the publications, listed in order of occurrence. The most frequently used keywords are PSC, with 83 occurrences, and VD (vessel detention), with 31. ML (machine learning) also occurs 13 times. Other keywords such as effectiveness, safety, and risk appear 13 times each, BN (Bayesian network) 12 times, prediction 11 times, MCA (multi-criteria analysis) 10 times, and NN (neural network) 9 times. This visualisation gives an overview of the predominant topics in the analyzed publications and reflects the most important concepts investigated in this area.

The distribution of publications by database (Figure 2) shows the number of publications in four well-known academic databases. Google Scholar contains the most significant number of publications, with 89, followed by Scopus with 88 and Web of Science with 87. The lowest number of publications is found in the ScienceDirect database, with 58. This diagram illustrates the distribution of publications across various databases, providing a better understanding of research coverage and availability within the academic community.

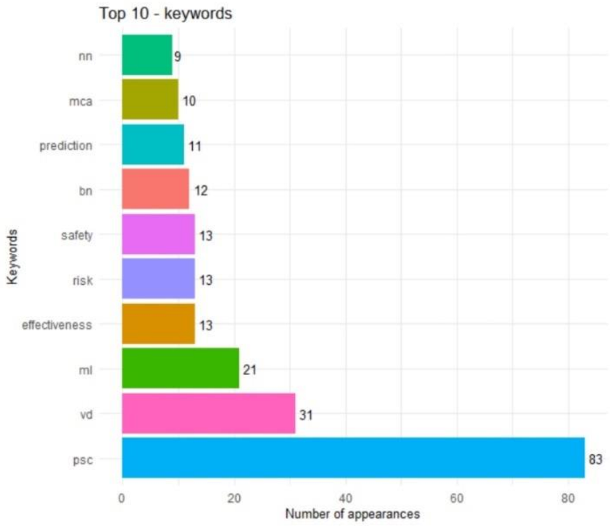

Figure 1. Top 10 keywords.

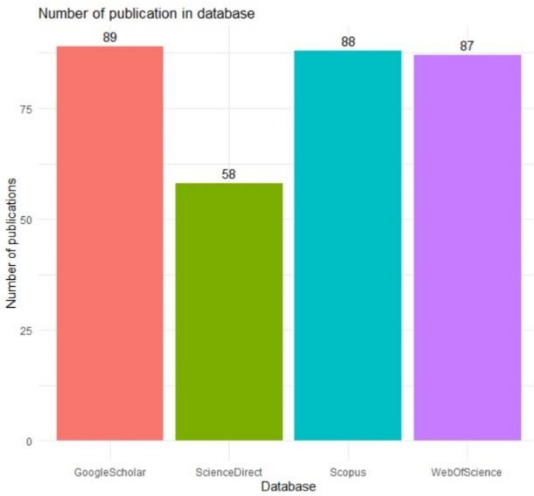

Figure 2. Number of publications per database.

The frequency of authors with two or more publications (Figure 3) shows the number of publications by individuals who have at least two works in the dataset. The leading author is Wang, S. et al. with 12 publications, closely followed by Yan, R. et al. with 11.

Authors Yang, Zh. and Yang, Za., with seven, and Wan, C. et al., Yin, J. et al., and Yu, Q. et al. have each published five papers.

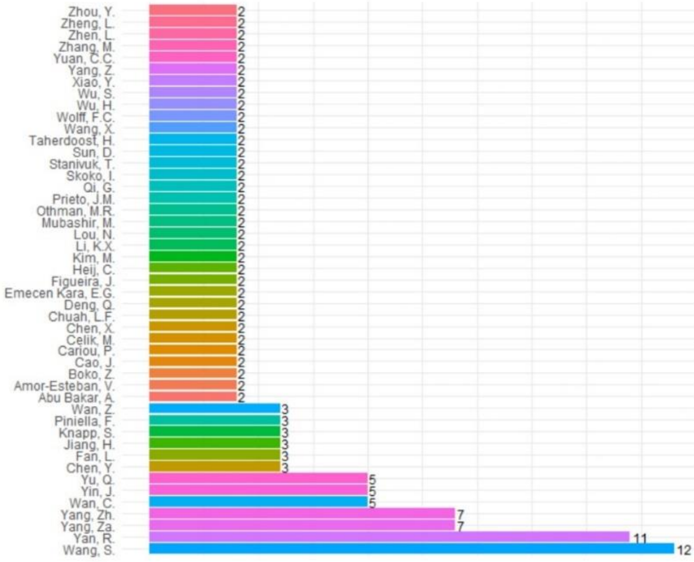

Figure 3. The frequency of authors with two or more papers.

The distribution of publications by year (Figure 4) and the distribution of publications by year and source (Figure 5) illustrate the trends in publication output over time. The highest number of publications was recorded in 2023 with a total of 19 publications, while the lowest number was observed in 2017 with only 1 publication. These visualizations provide a clear overview of the fluctuations in the number of publications over the years, allowing for the identification of periods with the highest and lowest publication activity.

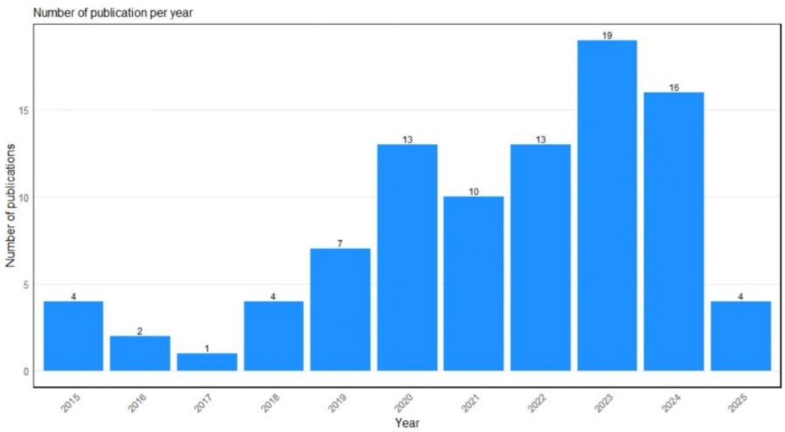

Figure 4. Number of publications per year.

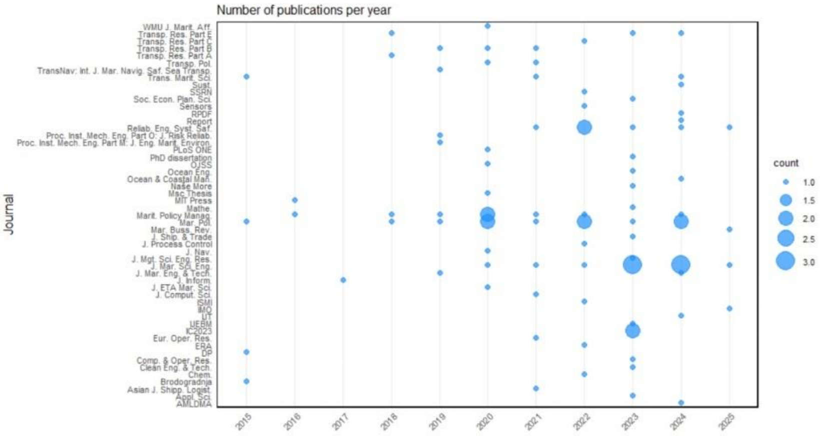

Figure 5. Number of publications per year and per source.

Figure 5 further refines this analysis by plotting the years on the x-axis and the abbreviated titles of the different publication types (such as journals, dissertations, conference proceedings, and similar sources) on the y-axis. The size of the circles in this graph represents the number of publications: larger circles indicate years with a higher number of publications. In comparison, smaller circles highlight years with fewer publications. This visualization allows for a nuanced analysis of publication trends over time and shows the years with the most and least publications.

Figure 6 focuses on the distribution of publications across specific journals and proceedings. The x-axis shows the number of publications per year, while the y-axis lists the different journals and proceedings in which these papers were published. For example, *Marine Policy* contains 11 publications, the *Journal of Marine Science and Engineering* 10, *Marine Policy and Management* 8, and *Reliability Engineering and Systems* 6. This visualization provides insight into the authors' preferences in choosing publication venues and highlights the most essential and relevant journals and proceedings for specific years.

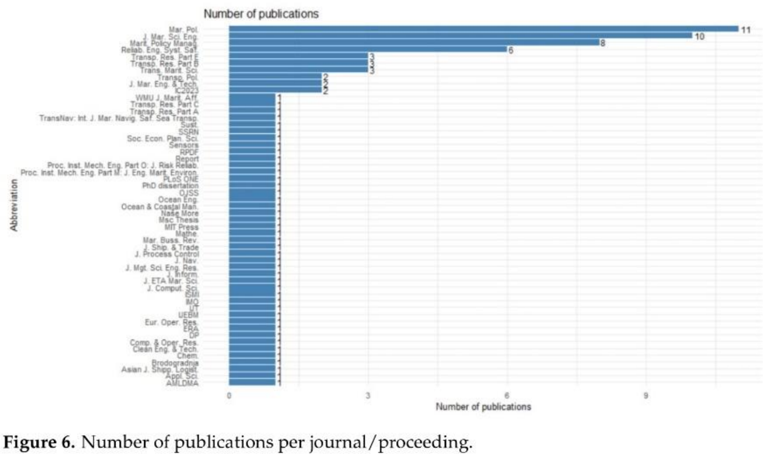

Figure 6. Number of publications per journal/proceeding.

The author collaboration network with cluster visualization (Figure 7) depicts the scientific collaboration structure among authors based on their shared publications. It comprises 41 distinct clusters, where nodes represent individual authors and 435 edges reflect their co-authorship relationships. The clusters are visually distinguished by different colors, facilitating the identification of research groups with stronger collaboration. This type of analysis offers valuable insights into the patterns of collaboration and highlights key contributors within the scientific community.

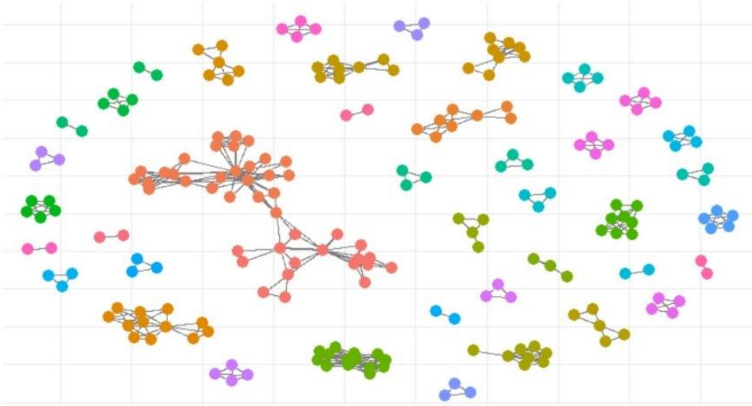

Figure 7. Author collaboration network with cluster visualization.

The author collaboration network with a focus on the surnames of first authors (Figure 8) provides a more granular view of scientific collaboration, concentrating on the primary researchers behind the publications. In this network, nodes represent the first authors, while the 426 connections illustrate their co-authorship links. Structural analysis of this network reveals the distribution of authorship and the contributions of individual researchers, thereby allowing the identification of dominant collaboration patterns within the dataset under examination.

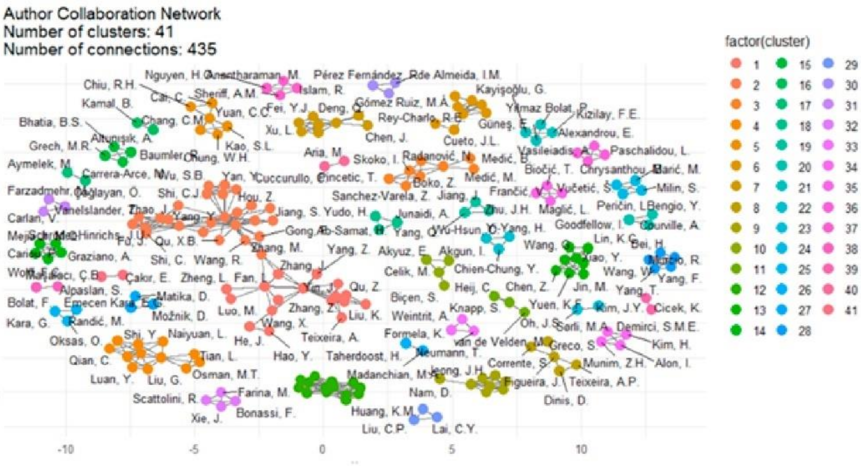

Figure 8. Author collaboration network with first author surnames.

4. Discussion

The data analysis in this study was performed using the R package, which enables efficient data analysis, manipulation, and visualization. R has proven its exceptional capabilities as a tool for statistical analysis, offering a wide range of available packages that support various methodological approaches, from basic statistical tests to advanced modeling techniques. This tool enabled the extraction of relevant results that were subsequently analyzed as part of the ongoing research.

The analysis of the data revealed several vital conclusions. Most of the papers in the collected sources involve three or four authors, suggesting that smaller research teams predominantly conduct research in this area. These teams tend to foster effective collaboration and involve a wide range of expertise. Single-authored papers are relatively rare, while publications with five or more authors are even rarer. This distribution suggests that a large number of authors is not necessarily a prerequisite for high-quality research results in this field.

Table 5 provides a structured comparative overview of selected ML and MCDM methods in terms of their applicability to PSC inspections. Each method is evaluated based on its main advantages, limitations, and overall effectiveness in supporting risk assessment, decision-making, and PSC compliance. The analysis aims to highlight the suitability of each approach for operational implementation in maritime safety inspections.

Table 5. Comparative analysis of methods in the context of PSC inspections.

Method	Advantages	Limitations	Effectiveness in PSC Context
Decision Trees	High interpretability; easy to apply	Lower accuracy in complex scenarios	Suitable for training and quick decision-making, but limited in complex cases
RF	High accuracy; robust to noisy data	Low transparency; complex structure	Very effective for risk detection, but requires additional explanation of results
SVM	Excellent performance on complex datasets	Difficult to interpret for non-experts	Accurate, but less applicable without expert support in operational settings
BN	Fast and simple; suitable for small or text-based data	Assumes feature independence, which is often unrealistic	Useful for preliminary classification and filtering
AHP	Structured multi-criteria decision-making	Subjectivity in determining weighting factors	Effective when decision priorities are clearly defined
TOPSIS	Clear rankings; focuses on proximity to the ideal solution	Limited flexibility for unstructured or dynamic data	Useful for ranking vessels based on risk level

The thematic focus of the publications analyzed is made clear by the use of key terms. The most frequently occurring key terms, such as PSC (83 occurrences) and VD (31 occurrences), indicate important research directions, while terms such as ML and efficiency appear in fewer publications. This distribution of key terms likely reflects the current challenges and technological advancements in the maritime industry, as well as the increasing use of advanced methods to solve complex problems.

The distribution of publications across different databases reveals that most papers are indexed in the Google Scholar database (89 entries). At the same time, Scopus and Web of Science contain nearly identical numbers of publications (88 and 87, respectively). This indicates a wide availability of research results in various academic sources, which enhances the visibility and accessibility of relevant information within the scientific community.

The trend in the number of publications shows a significant increase in 2023, when the majority of papers were published. This increase could be attributed to increased research activities, the initiation of new projects, and a growing interest in the topic. An analysis of publication trends over time provides valuable insights into the periodic fluctuations in research intensity, which may indicate progress and developments within the field.

The authorship network, comprising 41 clusters and 435 links between authors, provides a detailed insight into the structure of scientific collaboration. This network exhibits a clear concentration of cooperation, a crucial factor for progress in the field. Additionally, identifying the dominant research teams and their interactions offers valuable insights into collaboration practices and key players within the scientific community. Such information is crucial for understanding the dynamics of research collaborations and is likely to play a significant role in shaping future research efforts.

In the coming years, significant progress is expected in the application of MLMs to the risk analysis of vessels, particularly in the context of PSC. A systematic review of the available literature reveals that the application of MLMs can provide significant benefits in enhancing the accuracy, efficiency, and speed of vessel security risk assessments. While these techniques promise to improve on existing methods, several critical challenges still need to be overcome to realize their potential fully.

One of the main problems is the availability and quality of data, as information on vessel accidents and inspection results is often not standardized or publicly available in sufficient quantity. This lack of high-quality data is an obstacle to the application of ML techniques, as these methods require large amounts of accurate and up-to-date information. Without adequate data, it is challenging to develop reliable models that accurately predict safety risks, which limits the application of machine learning (ML) in the context of PSC inspections.

Although MCA facilitates systematic decision-making in complex situations, its integration with MLMs has not been sufficiently explored, particularly in the context of PSC inspections. The synergy between these methodological approaches promises more accurate and efficient assessments, but requires further research to fully realize its benefits in the context of vessel safety analysis.

Another challenge that needs to be addressed concerns the interpretation of the results arising from the combination of MLMs and MCA methods. Many MLMs operate as "black boxes", i.e., it is not always possible to clearly explain the reasons for categorizing a vessel as high-risk. This lack of transparency can lead to uncertainty among inspectors and weaken confidence in automated decisions. The literature recognizes the need to develop more transparent and interpretable models that enable inspectors to better understand the decisions made and facilitate the verification and validation of these decisions in practice.

To overcome these obstacles, further research is needed in the coming years in the areas of data collection, standardization, and development of new methodological approaches to integrate MLMs and MCA techniques.

In addition, the development of transparent and interpretable models that provide clear explanations for decisions is crucial to increase confidence in these technologies and their application in vessel surveys. The transparency of the models not only improves the understanding of the decision-making process but also helps to increase accountability and accuracy in the application of the review criteria.

Future research directions should therefore focus on further optimizing the methods of data collection and validation to increase the accuracy of bibliometric analyses. Particular attention should be paid to enhancing MLMs for assessing security risks, thereby increasing the efficiency and accuracy of testing procedures.

Although the results of this study provide valuable insights into the research trends and key players in this field, the challenges related to data quality and standardization remain a significant obstacle to the further development and implementation of advanced analytical approaches in vessel safety verification.

5. Conclusions

To improve the interpretability of machine learning models in the context of PSC, it is critical to apply techniques that improve model transparency and facilitate understanding by inspectors and regulators. Models with inherent interpretability, such as decision trees and rule-based systems, provide explicit and understandable decision rules that align well with operational requirements and allow inspectors to follow the decision process step-by-step. This transparency is critical to gaining the trust and acceptance of practitioners who rely on clear explanations to justify inspection results.

For more complex MLMs, which often work as a "black box", post-hoc explanation methods are used to explain the behavior of the model without compromising the predictive performance. Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) generate interpretable explanations by quantifying the contribution of individual input features to individual predictions. These methods allow inspectors to understand which factors influenced the model's decision, increasing transparency and facilitating compliance.

In addition, improving interpretability not only promotes inspector confidence but also helps to identify potential biases or errors in the model, contributing to continuous improvement and safer vessel operations. Therefore, incorporating interpretability techniques is critical to the effective use of machine learning tools in PSC inspections to balance predictive accuracy with the practical need for understandable and justifiable decisions.

This systematic literature review examined the integration of MLMs and MCA within the framework of PSC inspections. By applying the SLR and PICOC methodologies, the study identified and synthesized key contributions, methodological trends, and research gaps in the existing literature.

Findings indicate that MLMs, particularly in the form of predictive models, have strong potential to enhance risk identification and prioritization in PSC inspections. These models allow for more accurate forecasting of vessel detention probabilities and improved assessment of safety-related factors. When combined with MCA, they support structured and transparent decision-making by enabling the evaluation of multiple, often conflicting, risk indicators.

Despite these opportunities, the review also highlighted critical challenges. One of the main limitations is the scarcity of empirical studies that apply MLM and MCA to real-world PSC data. Moreover, methodological heterogeneity and lack of standardization in current approaches hinder broader applicability and reproducibility.

Future research should focus on developing advanced MLM techniques—such as DL and ANNs—and on improving the operational integration of MCA in PSC environments. Emphasis should also be placed on collaborative efforts with maritime authorities to validate these models in practice, ensuring they contribute meaningfully to the improvement of maritime safety and regulatory compliance.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ANN	Artificial Neural Network
BM	Bayesian Model
BN	Bayesian Network
CR	Critical Review
DL	Deep Learning
ELECTRE	Elimination Et Choix Traduisant La Realité (Elimination And Choice Translating Reality)
IMO	International Maritime Organisation
k-NN	K-Nearest Neighbors
LIME	Local Interpretable Model-agnostic Explanations
MARPOL	International Convention For The Prevention Of Pollution From Ships
MCA	Multi-Criteria Analysis
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MLM	Machine Learning Model
MoU	Memorandum Of Understanding
NLR	Narrative Literature Review
NN	Neural Network
PCA	Principal Component Analysis
PICOC	Population, Intervention, Comparison, Outcome, And Context.
PROMETHEE	Preference Ranking Organization Method For Enrichment Evaluation
PSC	Port State Control
SHAP	SHapley Additive exPlanations
SLM	Supervised Learning Method
SLR	Systematic Literature Review
SOLAS	International Convention For The Safety Of Life At Sea
SR	Scoping Review
SSL	Semi-Supervised Learning
STCW	International Convention On Standards Of Training, Certification, And Watchkeeping For Seafarers
SVM	Support Vector Machine
SVR	Support Vector Regression
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
ULM	Unsupervised Learning Method

References

1. International Maritime Organization. Port State Control. Available online: https://www.imo.org/en/ourwork/msas/pages/portstatecontrol.aspx (accessed on 11 January 2025).
2. Bhatia, B.S.; Carrera-Arce, M.; Baumler, R.; Grech, M.R. Seafarers vs. Port State Control: Decoding Work/rest Compliance Data Disparity. *Mar. Policy* **2024**, *163*, 106105. [CrossRef]
3. Bolat, F.; Alpaslan, S. Cumulative analysis of Port State Control based on Paris MoU inspections. *Trans. Marit. Sci.* **2021**, *10*, 224–246. Available online: https://hrcak.srce.hr/258064 (accessed on 15 May 2025). [CrossRef]
4. Liu, K.; Yu, Q.; Yang, Z.; Wan, C.; Yang, Z. BN-based port state control inspection for Paris MoU: New risk factors and probability training using big data. *Reliab. Eng. Syst. Saf.* **2022**, *224*, 108530. [CrossRef]
5. Chuah, L.F.; Mokhtar, K.; Abu Bakar, A.; Othman, M.R.; Osman, N.H.; Bokhari, A.; Mubashir, M.; Abdullah, M.A.; Hasan, M. Marine environment and maritime safety assessment using Port State Control database. *Chemosphere* **2022**, *304*, 135245. [CrossRef] [PubMed]
6. Biočić, T.; Frančić, V.; Hasanspahić, N.; Maglić, L. The analysis of the deficiencies resulting from Paris MoU PSC inspections. *Naše More* **2023**, *70*, 228–238. [CrossRef]

7. Wu-Hsun, C.; Yi-Yang, H.; Chien-Chung, Y. Application of Machine Learning to Maritime Safety. In Proceedings of the International Symposium on Maritime Innovation (ISMI), Singapore, 17–21 October 2022; IEEE: Piscataway, NJ, USA, 2022. Available online: https://www.conf.tw/site/userdata/1449/ISMI_paper/ISMI2022_paper_5987.pdf (accessed on 15 May 2025).
8. Dinis, D.; Figueira, J.; Teixeira, A.P. A multiple criteria approach for ship risk classification: An alternative to the Paris MoU Ship Risk Profile. *Soc. Econ. Plan. Sci.* **2023**, *90*, 101718. [CrossRef]
9. Tsou, M.C. Big data analysis of port state control ship detention database. *J. Mar. Eng. Technol.* **2019**, *18*, 113–121. [CrossRef]
10. Yang, Z.; Yang, Z.; Yin, J.; Qu, Z. A risk-based game model for rational inspections in port state control. *Transp. Res. Part E* **2018**, *118*, 477–495. [CrossRef]
11. Zheng, L. The Effectiveness of New Inspection Regime on Port State Control Inspection. *Open J. Soc. Sci.* **2020**, *8*, 440–446. [CrossRef]
12. Heij, C.; Knapp, S. Shipping inspections, detentions, and incidents: An empirical analysis of risk dimensions. *Marit. Policy Manag.* **2019**, *46*, 866–883. Available online: https://www.researchgate.net/publication/334747907_Shipping_inspections_detentions_and_incidents_an_empirical_analysis_of_risk_dimensions (accessed on 15 May 2025). [CrossRef]
13. Şanlıer, Ş. Analysis of port state control inspection data: The Black Sea Region. *Mar. Pol.* **2020**, *112*, 103757. [CrossRef]
14. Cariou, P.; Wolff, F.C. Identifying substandard vessels through port state control inspections: A new methodology for concentrated inspection campaigns. *Mar. Pol.* **2015**, *60*, 27–39. Available online: https://ideas.repec.org/a/eee/marpol/v60y2015icp27-39.html (accessed on 15 May 2025). [CrossRef]
15. Chen, J.H.; Zhang, S.H.; Xu, L.; Wan, Z.; Fei, Y.J.; Zheng, T.X. Identification of key factors of ship detention under port state control. *Mar. Pol.* **2019**, *102*, 21–27. [CrossRef]
16. Zhu, J.H.; Yang, Q.; Jiang, J. Identifying crucial deficiency categories influencing vessel detention: A method of combining cloud model and prospect theory. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108949. [CrossRef]
17. Rey-Charlo, R.E.; Cueto, J.L.; Piniella, F. Analyzing Port State Control Data to Explore Future Improvements to GMDSS Training. *J. Mar. Sci. Eng.* **2023**, *11*, 2379. [CrossRef]
18. Lyu, P. Support a Safety Learning Culture in Port State Control Regime. Ph.D. Dissertation, World Maritime University, Malmö, Sweden, 28 October 2023. Available online: https://commons.wmu.se/cgi/viewcontent.cgi?article=3290&context=all_dissertations (accessed on 15 May 2025).
19. Munim, Z.H.; Sørli, M.A.; Kim, H.; Alon, I. Predicting maritime accident risk using Automated Machine Learning. *Reliab. Eng. Syst. Saf.* **2024**, *248*, 110148. [CrossRef]
20. Randić, M.; Matika, D.; Možnik, D. SWOT analysis of deficiencies on ship components identified by port state control inspections with the aim to improve the safety of maritime navigation. *Brodogradnja* **2015**, *66*, 61–72.
21. Stanivuk, T.; Medić, B.; Medić, M. Statistical review of the annual report on the performance of maritime safety inspection in Croatia. *Trans. Marit. Sci.* **2015**, *4*, 41–51. [CrossRef]
22. Formela, K.; Weintrit, A.; Neumann, T. Overview of definitions of maritime safety, safety at sea, navigational safety and safety in general. *TransNav* **2019**, *13*, 285–290. [CrossRef]
23. Emecen Kara, E.G.; Oksas, O.; Kara, G. The similarity analysis of Port State Control regimes based on the performance of flag states. *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.* **2019**, *234*, 147509021987426. [CrossRef]
24. Graziano, A.; Cariou, P.; Wolff, F.C.; Mejia, M. Port State Control Inspections in the European Union: Do Inspector's Number and Background Matter? *Mar. Pol.* **2018**, *88*, 230–241. [CrossRef]
25. Almorza, D.; Prieto, J.M.; Amor-Esteban, V.; Piniella, F. Port state control inspections under the Paris Memorandum of understanding and their contribution to maritime safety: Additional risk classifications and indicators using multivariate techniques. *J. Mar. Sci. Eng.* **2024**, *12*, 533. [CrossRef]
26. Yuan, C.C.; Chiu, R.H.; Cai, C. Important factors influencing the implementation of independent Port State Control regimes. *J. Mar. Sci. Eng.* **2020**, *8*, 641. [CrossRef]
27. Xiao, Y.; Wang, G.; Lin, K.C.; Qi, G.; Li, K.X. The effectiveness of the New Inspection Regime for Port State Control: Application of the Tokyo MoU. *Mar. Pol.* **2020**, *115*, 103857. [CrossRef]
28. Yang, Z.; Yang, Z.; Teixeira, A. Comparative analysis of the impact of new inspection regime on port state control inspection. *Transp. Policy* **2020**, *92*, 65–80. [CrossRef]
29. Xiao, Y.; Qi, G.; Jin, M.; Yuen, K.F.; Chen, Z.; Li, K.X. Efficiency of port state control inspection regimes: A comparative study. *Transp. Policy* **2021**, *106*, 165–172. [CrossRef]
30. Henry, D.C. Shipping and COVID-19: Protecting seafarers as frontline workers. *WMU J. Marit. Aff.* **2020**, *19*, 279–293. [CrossRef]
31. Nam, D.; Kim, M. Implication of COVID-19 outbreak on ship survey and certification. *Mar. Pol.* **2021**, *131*, 104615. [CrossRef]
32. Maşalacı, Ç.B.; Çakır, E. Differences in deficiency types causing ship detentions at the Black Sea Region during the COVID-19 pandemic and pre-pandemic. *Mar. Pol.* **2023**, *151*, 105553. [CrossRef]

33. Prieto, J.M.; Amor-Esteban, V.; Almorza-Gomar, D.; Turias, I.; Piniella, F. Application of Multivariate Statistical Techniques as an Indicator of Variability of the Effects of COVID-19 on the Paris Memorandum of Understanding on Port State Control. *Mathematics* **2023**, *11*, 3188. [CrossRef]
34. Peričin, L.; Vučetić, Š.; Barić, M.; Milin, S. Port State Control Inspections During the COVID Pandemic-Case Study: Republic of Croatia. *Trans. Marit. Sci.* **2024**, *13*. [CrossRef]
35. Deng, Q.; Wan, Z.; Chen, J. Factors influencing ship detentions: Pre-and post-pandemic analysis. *Ocean. Coast. Manag.* **2024**, *257*, 107344. [CrossRef]
36. Heij, C.; Knapp, S. Predictive power of inspection outcomes for future shipping accidents-an empirical appraisal with special attention for human factor aspects. *Marit. Policy Manag.* **2018**, *45*, 604–621. [CrossRef]
37. Fu, J.; Chen, X.; Wu, S.; Shi, C.; Wu, H.; Zhao, J.; Xiong, P. Mining ship deficiency correlations from historical port state control (PSC) inspection data. *PLoS ONE* **2020**, *15*, e0229211. Available online: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229211 (accessed on 15 May 2025). [CrossRef]
38. Fan, L.; Zhang, M.; Yin, J.; Zhang, J. Impacts of dynamic inspection records on port state control efficiency using Bayesian network analysis. *Reliab. Eng. Syst. Saf.* **2022**, *228*, 108753. [CrossRef]
39. Gómez Ruiz, M.Á.; de Almeida, I.M.; Pérez Fernández, R. Application of Machine Learning Techniques to the Maritime Industry. *J. Mar. Sci. Eng.* **2023**, *11*, 1820. [CrossRef]
40. Lai, C.Y.; Liu, C.P.; Huang, K.M. Optimization of the Concentrated Inspection Campaign Model to Strengthen Port State Control. *J. Mar. Sci. Eng.* **2023**, *11*, 1166. [CrossRef]
41. Yan, R. Development of Machine Learning and Data Mining Models for Port State Control Inspection. Master's Thesis, Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, 2020. Available online: https://theses.lib.polyu.edu.hk/bitstream/200/10643/3/5082.pdf (accessed on 3 March 2025).
42. Chung, W.H.; Kao, S.L.; Chang, C.M.; Yuan, C.C. Association rule learning to improve deficiency inspection in port state control. *Marit. Policy Manag.* **2020**, *47*, 332–351. [CrossRef]
43. Yan, R.; Wang, S.; Peng, C. Ship selection in port state control: Status and perspectives. *Marit. Policy Manag.* **2021**, *48*, 600–615. [CrossRef]
44. Chen, Y.; Lou, N.; Liu, G.; Luan, Y.; Jiang, H. Risk analysis of ship detention defects based on association rules. *Mar. Policy* **2022**, *142*, 105123. [CrossRef]
45. Qian, C.; Lou, N.; Jiang, H.; Zhou, Y.; Shi, Y. Ship detention prediction based on ensemble learning method. In Proceedings of the 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 4–6 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. Available online: https://www.researchgate.net/publication/373961303_Ship_Detention_Prediction_Based_on_Ensemble_Learning_Method (accessed on 15 May 2025).
46. Farzadmehr, M.; Carlan, V.; Vanelslander, T. Contemporary challenges and AI solutions in port operations: Applying Gale–Shapley algorithm to find best matches. *J. Ship. Trade* **2023**, *8*, 27. [CrossRef]
47. Demirci, S.M.E.; Cicek, K. Intelligent ship inspection analytics: Ship deficiency data mining for port state control. *Ocean Eng.* **2023**, *278*, 114232. [CrossRef]
48. Junaidi, A.; Yudo, H.; Ab-Samat, H. Identification of data analysis methods and focus trends in Port State Control inspections: A comprehensive literature review. *Int. J. Technol.* **2024**, *15*, 179–194. Available online: https://ijtech.eng.ui.ac.id/article/view/6492/ (accessed on 15 May 2025). [CrossRef]
49. Sheriff, A.M.; Anantharaman, M.; Islam, R.; Nguyen, H.O. An in-depth analysis of port state control inspections: A bibliometric analysis and systematic review. *Marit. Bus. Rev.* **2025**, *10*, 2454754. [CrossRef]
50. Nielsen, M. *Neural Networks and Deep Learning*; Determination Press: Online, 2015. Available online: http://neuralnetworksanddeeplearning.com/ (accessed on 15 May 2025).
51. Aria, M.; Cuccurullo, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Inform.* **2017**, *11*, 959–975. [CrossRef]
52. IET. Ship inspection by port state control. In *Applications of Machine Learning and Data Analytics Models in Maritime Transportation*; IET: Stevenage, UK, 2024; Chapter 2. [CrossRef]
53. Yan, R.; Wang, S. A semi-"smart predict then optimize" (semi-spo) method for efficient ship inspection. *Transp. Res. Part B* **2020**, *142*, 100–125. [CrossRef]
54. Yan, R.; Wang, S.; Cao, J.; Sun, D. Shipping domain knowledge informed prediction and optimization in Port State Control. *Transp. Res. Part B* **2021**, *149*, 52–78. [CrossRef]
55. Yan, R.; Wang, S.; Zhen, L.; Jiang, S. Classification and regression in prescriptive analytics: Development of hybrid models and an example of ship inspection by port state control. *Comput. Oper. Res.* **2023**, *163*, 106517. [CrossRef]
56. Deng, Q.; Wan, Z. Long-Term or Short-Term? Prediction of Ship Detention Duration Based on Machine Learning. *J. Mar. Sci. Eng.* **2024**, *12*, 2220. [CrossRef]

57. Wang, S.; Yan, R.; Qu, X.B. Development of a non-parametric classifier: Effective identification, algorithm, and applications in port state control for maritime transportation. *Transp. Res. Part B* **2019**, *128*, 129–157. [CrossRef]
58. Boko, Z.; Skoko, I.; Sanchez-Varela, Z.; Pincetic, T. Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making. *J. Mar. Sci. Eng.* **2024**, *12*, 1905. [CrossRef]
59. Wu, S.B.; Chen, X.Q.; Shi, C.J.; Fu, J.J.; Yan, Y.; Wang, S.Z. Ship detention prediction via feature selection scheme and support vector machine (SVM). *Marit. Policy Manag.* **2022**, *49*, 140–153. [CrossRef]
60. Knapp, S.; van de Velden, M. Predicting Inspection Outcomes and Evaluating Port State Control Targeting Using Random Forests. *Report* **2024**. [CrossRef]
61. Bei, H.; Yang, F.; Wang, W.; Yang, T.; Murcio, R. Maritime safety through multi-source data fusion: An AdaBoost-based approach for predictive ship detention by port state control. *Mar. Pol. Manag.* **2024**, *51*, 123–145. [CrossRef]
62. Hou, Z.; Yan, R.; Wang, S. On the K-Means Clustering Model for Performance Enhancement of Port State Control. *J. Mar. Sci. Eng.* **2022**, *10*, 1608. [CrossRef]
63. Osman, M.T.; Tian, L.; Chen, Y.; Abdul Rahman, N.S.F. Empirical analysis on port state control inspection for foreign-registered ships in Malaysian ports. *Asian J. Shipp. Logist.* **2021**, *37*, 127–139. [CrossRef]
64. Akyuz, E.; Akgun, I.; Celik, M. A fuzzy failure mode and effects approach to analyse concentrated inspection campaigns on board ships. *Marit. Policy Manag.* **2016**, *43*, 887–908. [CrossRef]
65. Yan, R.; Wu, S.; Jin, Y.; Cao, J.; Wang, S. Efficient and explainable ship selection planning in port state control. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103924. Available online: https://www.researchgate.net/publication/364647513_Efficient_and_explainable_ship_selection_planning_in_port_state_control (accessed on 15 May 2025). [CrossRef]
66. Vasileiadis, A.; Alexandrou, E.; Paschalidou, L.; Chrysanthou, M. Artificial neural network and its applications. *ResearchPDF* **2024**, *1*, 1–15. Available online: https://www.researchgate.net/publication/380122805_Artificial_Neural_Network_and_Its_Applications (accessed on 12 January 2025).
67. Boko, Z.; Stanivuk, T.; Radanović, N.; Skoko, I. Machine Learning-Driven Prediction of Offshore Vessel Detention: The Role of Neural Networks in Port State Control. *J. Mar. Sci. Eng.* **2025**, *13*, 472. [CrossRef]
68. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
69. Biçen, S.; Celik, M. A Deep Learning Approach to Analyse Ship Inspection Reports via Natural Language Processing Integrated with Artificial Neural Network. *J. Mar. Eng. Technol.* **2024**, *23*, 291–304. [CrossRef]
70. Chen, X.; Yang, Y.; Wang, S.; Wu, H. Ship Type Recognition via a Coarse-to-Fine Cascaded Convolution Neural Network. *J. Nav.* **2020**, *73*, 813–832. [CrossRef]
71. Lim, J.J.; Kim, D.W.; Hong, W.H.; Kim, M.; Lee, D.H.; Kim, S.Y.; Jeong, J.H. Application of Convolutional Neural Network (CNN) to Recognize Ship Structures. *Sensors* **2022**, *22*, 3824. [CrossRef]
72. Bonassi, F.; Farina, M.; Xie, J.; Scattolini, R. On recurrent neural networks for learning-based control: Recent results and ideas for future developments. *J. Process Control* **2022**, *114*, 92–104. [CrossRef]
73. Yang, Z.; Yang, Z.; Yin, J. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. *Transp. Res. Part A* **2018**, *110*, 38–56. [CrossRef]
74. Fan, L.; Zhang, Z.; Yin, J.; Wang, X. The efficiency improvement of port state control based on ship accident Bayesian networks. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2019**, *233*, 71–83. [CrossRef]
75. Fan, L.; Zheng, L.; Luo, M. Effectiveness of port state control inspection using Bayesian network modelling. *Marit. Policy Manag.* **2020**, *49*, 261–278. [CrossRef]
76. Yang, Z.; Wan, C.P.; Yang, Z.; Yu, Q. Using Bayesian network-based TOPSIS to aid dynamic port state control detention risk control decision. *Reliab. Eng. Syst. Saf.* **2021**, *213*, 107784. [CrossRef]
77. Yang, Z.; Yu, Q.; Yang, Z.; Wan, C. A Machine Learning-Based Bayesian Model for the Development of Highly-Effective Port State Control Inspection Policy. *SSRN* **2022**. Available online: https://ssrn.com/abstract=4255426 (accessed on 15 May 2025).
78. Chuah, L.F.; Mohd Rof'ie, N.R.; Mohd Salleh, N.H.; Abu Bakar, A.; Oloruntobi, O.; Othman, M.R.; Mohamed Fazlee, U.S.; Mubashir, M.; Asif, S. Analyzing the influencing factors of Port State Control for a cleaner environment via Bayesian network model. *Clean. Eng. Technol.* **2023**, *14*, 100636. [CrossRef]
79. Jiang, H.; Chen, Y.; Naiyuan, L.; Zhou, Y. Dynamic Evolution Analysis of Ship's Port State Control Inspection Based on Bayesian Networks. In Proceedings of the 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 4–6 August 2023. [CrossRef]
80. Yang, Z.; Wan, C.; Yu, Q.; Yin, J.; Yang, Z. A machine learning-based Bayesian model for predicting the duration of ship detention in PSC inspection. *Transp. Res. Part E* **2023**, *180*, 103331. [CrossRef]
81. Kamal, B.; Altunışık, A. A data-driven Bayes approach for investigating International Safety Management Code-sourced detention of ships in Port State Controls. *Mar. Pol.* **2024**, *169*, 106346. [CrossRef]
82. Yang, Z.; Yu, Q.; Yang, Z.; Wan, C. A data-driven Bayesian model for evaluating the duration of detention of ships in PSC inspections. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *181*, 103371. [CrossRef]

83. Corrente, S.; Figueira, J.; Greco, S. Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *Eur. Oper. Res.* **2021**, *291*, 738–756. [CrossRef]
84. Çağlayan, Ö.; Aymelek, M. An Integrated Multi-Criteria Decision Support Model for Sustainable Ship Queuing Policy Application via Vessel Traffic Service (VTS). *Sustainability* **2024**, *16*, 4615. [CrossRef]
85. Yan, R.; Wang, S. Ship detention prediction using anomaly detection in port state control: Model and explanation. *ERA* **2022**, *30*, 3679–3691. [CrossRef]
86. Yilmaz Bolat, P.; Kayişoğlu, G.; Güneş, E.; Kizilay, F.E. Weighting Key Factors for Port Congestion by AHP Method. *J. Eta Marit. Sci.* **2020**, *8*, 252–273. [CrossRef]
87. Emecen Kara, E.G. Determination of maritime safety performance of flag states based on the Port State Control inspections using TOPSIS. *Mar. Pol.* **2022**, *143*, 105156. [CrossRef]
88. Taherdoost, H. Using PROMETHEE Method for Multi-Criteria Decision Making-Applications and Procedures. *Iris J. Econ. Bus. Manag.* **2023**, *1*. [CrossRef]
89. Taherdoost, H.; Madanchian, M. A Comprehensive Overview of the ELECTRE Method in Multi Criteria Decision-Making. *J. Mgt. Sci. Eng. Res.* **2023**, *6*, 5–16. [CrossRef]
90. Kim, J.Y.; Oh, J.S. A pilot study of stacked autoencoders for ship mode classification. *Appl. Sci.* **2023**, *13*, 5491. [CrossRef]
91. He, J.; Hao, Y.; Wang, X. An Interpretable Aid Decision-Making Model for Flag State Control Ship Detention Based on SMOTE and XGBoost. *J. Mar. Sci. Eng.* **2021**, *9*, 156. [CrossRef]
92. Yan, R.; Wang, S.; Peng, C.; Sun, D. An Artificial Intelligence Model Considering Data Imbalance for Ship Selection in Port State Control Based on Detention Probabilities. *J. Comput. Sci.* **2021**, *48*, 101257. [CrossRef]
93. Wang, R.; Zhang, M.; Gong, F.; Wang, S.; Yan, R. Improving port state control through a transfer learning-enhanced XGBoost model. *Reliab. Eng. Syst. Saf.* **2025**, *253*, 110558. [CrossRef]
94. R Project for Statistical Computing. Available online: https://www.r-project.org/about.html (accessed on 25 March 2025).
95. Carrera-Rivera, A.; Ochoa, W.; Larrinaga, F.; Lasa, G. How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX* **2022**, *9*, 101895. [CrossRef]
96. Mulyanto, A.; Hartati, S.; Wardoyo, R. Systematic Literature Review of Text Feature Extraction. In Proceedings of the Seventh International Congress on Information and Communication Technology, London, UK, 21–24 February 2022.

Appendix C

Paper III

Boko, Z., Stanivuk, T., Radanović, N., & Skoko, I. (2025). Machine Learning-Driven Prediction of Offshore Vessel Detention: The Role of Neural Networks in Port State Control. Journal of Marine Science and Engineering, 13(3), 472. https://doi.org/10.3390/jmse13030472

Journal of
Marine Science and Engineering

MDPI

Article

Machine Learning-Driven Prediction of Offshore Vessel Detention: The Role of Neural Networks in Port State Control

Zlatko Boko *, **Tatjana Stanivuk**, **Nenad Radanović** and **Ivica Skoko**

Faculty of Maritime Studies, University of Split, 21000 Split, Croatia; tstanivu@pfst.hr (T.S.); neno.radanovic@yahoo.com (N.R.); iskoko@pfst.hr (I.S.)
* Correspondence: zboko@pfst.hr

Abstract: This study investigates the application of different neural network (NN) models in assessing the risk of the detention of offshore vessels during port state control (PSC) inspections. The focus is on the use of different NN models ("nnet", "mlp", "neuralnet", "rsnns") to identify the main risk factors based on historical data on vessels and their inspections. The main objective of this research is to improve maritime safety and the efficiency of inspection procedures by applying techniques that can more accurately predict the probability of detention of the offshore vessels. These models make it possible to analyse complex patterns in the data, such as the relationships between the country of inspection, flag, memorandum, age, tonnage and previous deficiencies, and the risk of detention. Understanding these patterns is crucial for inspection teams' proactive action as it helps direct resources to potentially high-risk vessels. Implementing these models into PSC processes helps to optimise resource allocation, reduce unnecessary costs, and increase the reliability of decision-making processes. NN models significantly help in recognising non-linear patterns and provide high accuracy in risk prediction. The study also includes a comparative analysis of the elements that determine the accuracy, sensitivity, and other performance aspects of the models to determine the most appropriate approach for practical implementation. The results emphasise the importance of applying artificial intelligence (AI) in various aspects of modern maritime safety management. This research opens up new opportunities for the development of intelligent support systems that not only increase safety but also improve the efficiency of inspection processes on a global scale.

Keywords: neural networks; risk assessment; offshore vessel detention; port state control (PSC); inspection procedures efficiency; artificial intelligence (AI)

check for updates

Academic Editors: Baiqiao Chen, Sheng Xu and Ersee Uzunoglu

Received: 10 February 2025
Revised: 26 February 2025
Accepted: 26 February 2025
Published: 28 February 2025

Citation: Boko, Z.; Stanivuk, T.; Radanović, N.; Skoko, I. Machine Learning-Driven Prediction of Offshore Vessel Detention: The Role of Neural Networks in Port State Control. *J. Mar. Sci. Eng.* **2025**, *13*, 472. https://doi.org/10.3390/jmse13030472

1. Introduction

Risk assessment in the shipping industry is a key challenge for ensuring shipping safety, with a particular focus on vessel inspections under the port state control (PSC) system. The PSC system aims to enforce compliance with international safety and environmental standards by inspecting vessels entering ports. Optimising this process requires precision and efficiency in predicting the likelihood of vessels being detained during inspections, which is crucial in improving safety and reducing operational costs.

Traditional risk analysis methods in this context rely on statistical techniques and expertise based on experience. However, with the increasing complexity of data and the growing number of factors influencing the outcome of inspections, the need for advanced approaches is becoming increasingly apparent. In this regard, developing and applying artificial intelligence (AI) methods, particularly neural network (NN) models, represent a significant step towards more accurate and efficient risk assessments.

This research investigates the application of different NN models to predict the probability of detention of offshore vessels during PSC inspections. In particular, it focuses on the use of NN models such as "nnet", "mlp" (multilayer perceptron), "neuralnet", and "rsnns" (R package for self–organising maps and neural networks), which are widely known for their ability to analyse complex and non-linear relationships in large datasets. Using these models, the study aims to identify key risk factors that can help anticipate potential inspection problems, such as vessel characteristics, previous inspection reports, flags, number and type of deficiencies, vessel age, and other relevant factors.

The application of these models not only enables more accurate predictions, but also better resource optimisation within the inspection teams. Through the use of methods such as "nnet", "mlp", "neuralnet", and "rsnns", critical patterns and interactions between variables can be identified that may be overlooked by conventional methods. This makes the goal of increasing maritime safety and reducing operating costs achievable and contributes significantly to the efficiency of the PSC system.

This study provides valuable insights into how advanced machine learning (ML) methods, especially NNs, can be used to analyse risk in the maritime industry. Their application can optimise decision-making, improve safety, and increase the efficiency of inspections globally.

The structure of this research study is designed to provide a systematic and clear presentation of the research process, results, and conclusions while maintaining a high scientific and professional standard. The paper is divided into seven main sections. The introductory section defines the research problem, sets out the research objectives, and emphasises the topic's relevance to current challenges in the maritime industry. This section emphasises the need to improve the efficiency and accuracy of risk assessment for the detention of vessels under PSC.

The second section reviews previous research and existing studies on applying ML, particularly NNs, in maritime safety management. It provides a detailed literature review and analyses the main contributions, shortcomings, and open questions that are a starting point for further research.

The third section analyses the significance and importance of PSC inspections, focusing on their role in improving maritime safety. This section also analyses the concept and structure of NNs, explains their operating mechanisms, and examines their different types and specific applications for predicting complex patterns.

The fourth section describes the methodological approach used in this study. In this section, the stages of data collection, processing, and analysis are systematically presented, the evaluation measures used are explained in detail, and the technical specifications of the NN models used, including details of design, training, and performance evaluation, are outlined.

The fifth section analyses the obtained results and highlights the main research findings. The results are interpreted in terms of accuracy, sensitivity, specificity, and other relevant metrics used to evaluate model performance.

The sixth section provides a discussion contextualising the research findings within the existing literature, highlighting their implications and limitations and making recommendations for future research. The seventh section concludes the paper by summarising the key findings and outlining potential directions for further advancements in the application of advanced machine learning models in maritime security.

This structured approach enables a coherent and comprehensive examination of the research problem. It provides the reader with a clear understanding of all relevant aspects of the study and its contribution to science and practice.

2. Recent Research on the Application of Machine Learning Methods in Port State Control Inspections

Maritime safety is a critical component of global logistics and international trade, and its importance continues to grow, given the increasing density of marine traffic and the ever-more demanding requirements for compliance with international maritime standards. In this context, vessel inspections carried out by naval authorities as part of PSC have become essential to ensure safety, prevent accidents, and protect the environment. PSC inspections make it possible to identify deficiencies in vessel operations, comply with applicable regulations, and ensure that foreign-flagged vessels fulfil international safety and environmental standards.

However, given the dynamic nature of maritime transport, the challenges associated with risk assessment and identifying potential hazards require introducing advanced technological solutions that can significantly improve the efficiency of inspections. In this regard, NN models represent a promising approach to improving risk assessment processes, as they allow for the analysis of large datasets and the identification of hidden patterns that are not immediately recognisable using traditional analysis methods. As a fundamental component of AI, NNs can predict events based on complex relationships between variables, making them a valuable tool for predicting the likelihood of vessels being detained in ports.

The advantage of using NNs in this context is their ability to process data from multiple sources, including information on previous inspections, vessel characteristics, weather conditions, and the specifics of ports and maritime authorities. Such modelling can enable more accurate and faster decision-making regarding the need for further inspections, vessel detention, or other corrective actions, ultimately increasing safety and reducing the risk of accidents.

The following section examines recent research examining the relationship between the applications of various ML techniques, in particular NN models, and PSC inspections. The analysis includes recent studies that focus on implementing advanced ML models, deep learning (DL), and NNs to improve risk assessment accuracy and optimise prediction related to vessel detention probability. These studies aim to contribute to the efficiency, safety, and transparency of the inspection process in the global maritime sector.

The research [1] investigated the application of NNs in traffic safety modelling. The research aimed to develop an NN-based model capable of effectively analysing and predicting safety risks in different traffic scenarios. The authors demonstrated that NNs offer advantages over traditional statistical methods, especially in non-linear relationships between variables that influence safety. The experimental results highlighted the ability of NNs to detect patterns in complex datasets, making them suitable for analysing security incidents and identifying key risk factors. The research also emphasised the flexibility of NNs in processing different types of data and their applicability to a wide range of traffic safety problems. The authors concluded that such models could significantly improve decision-making processes and implement preventive measures in transport systems. The study results suggest that NNs are a valuable tool for modelling safety risks and supporting the development of strategies to improve safety in transport systems. While the research results are promising in detecting structures in maritime transport, it is essential to note that this paper is not explicitly focused on PSC inspections, particularly PSC inspections of offshore vessels.

Researchers [2–4] analyse the application of Bayesian networks (BNs) in the context of PSC inspections but differ in their specific objectives and analytical approaches. All three use BNs to model different aspects of PSC inspections and emphasise their ability to use probabilistic reasoning and decision-making under uncertainty. The main objective of these investigations is to increase the efficiency of PSC inspections by improving the prediction and analysis of factors influencing vessel detention. In addition, all three researchers emphasise the importance of PSC inspections for improving maritime safety and reducing operational risks. However, paper [2] focuses on evaluating the effectiveness of PSC inspections through BN modelling, analysing the probabilities of detecting violations and their causes. The main contribution of this paper is to show how a BN can improve decision-making through probabilistic analyses of inspection data. Secondly, paper [3] integrates the BN with machine learning (ML) methods the duration of vessel detention following a PSC inspection. In contrast to the first paper, which primarily analyses the results of the inspections, this second paper focuses on the factors that influence the duration of detention, allowing for more accurate planning and resource allocation. Paper [4] examines the dynamic evolution of PSC inspections by analysing the changes in inspection patterns over time using the BN. The paper differs from the previous two in that it does not consider PSC inspections as a static process but analyses how inspection strategies and the detection of violations evolve under the influence of various factors. Nevertheless, all three papers mentioned contribute to understanding how BNs can improve PSC inspections by analysing inspection efficiency, predicting the duration of detention of vessels, or observing changes in inspection strategies over time.

The papers [5,6] investigate the application of convolutional neural networks (CNNs) in the maritime sector, focusing on recognising vessel types and identifying vessel structures. Both papers use CNN architectures to improve automated recognition processes and emphasise the benefits of DL in feature extraction and classification in complex maritime environments. Although they share a common methodological basis, these papers differ in their specific objectives and approaches. In contrast to research focusing on PSC inspections, paper [5] does not analyse the regulatory aspects of maritime safety but is primarily concerned with the automatic detection of vessels and their structural components. Paper [5] proposes a cascaded CNN model with coarse to delicate processing stages for recognising vessel types to improve classification accuracy through progressively refined feature extraction. The main contribution of this paper lies in its hierarchical approach, which enables more accurate discrimination between vessel types by integrating global and detailed feature representations. This method improves recognition performance in complex maritime scenarios where vessels often exhibit significant class differences. Paper [6] applies CNNs for vessel structure recognition, focusing on identifying specific vessel components rather than classifying entire vessels. This paper emphasises the potential of DL for structural analysis, which could play a crucial role in automated inspection, maintenance, and safety monitoring. In contrast to the first paper [6], primarily concerned with the differentiation of vessel types, this work investigates the recognition and classification of vessel structures based on visual input, thus contributing to the broader field of vessel integrity assessment.

The articles [7,8] investigate the application of ML techniques in the context of PSC to improve the processes of prediction and optimisation of vessel detentions in ports. Still, they differ in their methodological approaches and the ML techniques applied. Article [7] utilises anomaly detection as the primary technique, which enables the identification of unusual patterns in vessel data and operations. Anomaly detection helps recognise potential risks that are not immediately obvious and may indicate vessels with a high-security risk, thereby contributing to more accurate predictions of vessel detentions. Paper [8] employs K-means clustering as a methodology for grouping vessels based on similarities in performance characteristics. K-means clustering allows for segmenting vessels into groups, thus facilitating better analysis of vessel behaviour patterns and optimising resources in the inspection process. This technique helps identify vessels that pose a greater risk, enabling more efficient resource allocation for inspections and thereby increasing the overall efficiency of the PSC systems; paper [7] contributes to developing an anomaly detection model that can predict the accuracy and recognition of unexpected but potentially dangerous events leading to vessel detentions. This approach enables the rapid detection of accidents or irregularities that traditional methods may overlook. In contrast, article [8] optimises inspection strategies by utilising clustering techniques, allowing for selective and targeted vessel inspections based on their specific characteristics, thus reducing inspector workload and focusing on potentially high-risk segments of vessels. While both works address similar themes within PSC, the differences lie in the applied ML techniques: anomaly detection in the first work facilitates the identification of unusual and high-risk situations. In contrast, K-means clustering optimises the distribution of inspection resources, thereby improving the system's overall efficiency.

The articles [9–12] deal with the application of ML in the maritime industry, with a particular focus on PSC. Although all papers deal with the improvement of PSC processes by ML, they use different techniques and approaches. The commonality of all papers lies in the shared goal of improving vessel inspection processes and optimising resources through ML techniques. All the papers cited take a data-driven approach, analysing large datasets to identify patterns that enable better assessment of vessel safety risks. The paper [9] uses a wide range of ML techniques to analyse data from the maritime industry, but with a broader scope that includes other sectors of the maritime industry, while the other papers focus exclusively on PSC. In [10], the authors apply data mining techniques to analyse vessel deficiencies within PSC and identify patterns that may indicate a higher risk of vessel detention. Paper [11] provides a broader frame for applying ML in the maritime industry, but without specific application to individual ML models. At the same time, in [12] the authors focus on analysing vessel inspection reports using natural language processing (NLP) in combination with DL and artificial neural networks (ANN), which enables the analysis of unstructured data. The contribution of [9] lies in its holistic approach to the application of ML in the maritime industry, which provides a broader perspective on how different techniques can improve the safety and efficiency of PSC processes. The authors in [10] make a specific contribution through the data mining of defects, which enables better prioritisation of inspections and prediction of vessels with a higher risk of detention. The research in [11] contributes to the development of a theoretical framework for the application of ML in PSC and shows how different models can improve the inspection process and safety in vessel operations. The authors in [12] integrate NLP with ANN to analyse unstructured text data from inspection reports, enabling a deeper understanding of inspection results and more accurate risk assessments.

The authors of article [13] investigate the application of advanced algorithms in PSC, specifically for offshore vessels, using classification trees and multi-criteria decision-making (MCDM). Their research aims to develop a methodology that enables more efficient data analysis on offshore vessels while identifying the key risk factors that influence inspection decisions and the possible detention of vessels. Classification trees allow precise categorisation of vessels based on risk, while MCDM techniques facilitate the weighting of the different criteria in decision-making. That article's contribution lies in optimising the vessel inspection process within PSC by applying sophisticated algorithms that can improve the accuracy of risk prediction and resource optimisation.

The detention of offshore vessels in ports under PSC can significantly negatively impact the operational efficiency of vessel operators and all stakeholders, including vessel owners, crew members, and regulators. Although these vessels typically do not carry cargo, their detention due to non-compliance with regulatory standards or the need for additional inspections can lead to significant delays in the completion of maritime tasks, a reduction in the time a vessel can spend at sea, and the incurrence of additional operational costs.

This detention can further affect the operator's competitiveness in the market and have a negative impact on its reputation. In addition, such deployments can increase the vessel's maintenance costs and lead to complications with safety protocols and compliance with international laws.

To improve operational efficiency and mitigate the negative effects of vessel detention, this paper explores the application of AI-based models, specifically NNs. These models have the potential to predict better and identify the key factors that can lead to vessel detention and improve the accuracy of predictions related to future risks, such as non-compliance with regulations or unplanned inspections. Using NNs makes it possible to analyse complex patterns in large datasets automatically. This enables more accurate decision-making, faster identification of critical factors, and optimisation of strategies to reduce costs, speed up the inspection process, and improve the overall competitiveness of vessel operators. Therefore, this approach is crucial to enhancing predictive capabilities in fleet safety management and operational efficiency.

In contrast to previous studies, which are relatively scarce and rely predominantly on basic ML methods without incorporating NN architectures for analysing ship inspection data in the context of PSC, this research introduces their application. However, given the specific requirements of the task, we utilise simpler NN models that provide an optimal trade-off between accuracy, computational efficiency, and practical applicability. This paper's unique contribution lies in examining the optimisation of the PSC process from multiple perspectives, including inspectors, vessel owners, crew members, and other stakeholders involved in the process.

The implementation of NNs, unlike traditional data analysis approaches, utilises advanced architectures to predict inspection durations and outcomes with a high degree of accuracy and efficiency. The diversity of models is reflected in the paper of four different NN models, each thoroughly evaluated based on relevant performance metrics. The processing of offshore vessel inspection data based on historical records enables the identification of patterns related to past deficiencies, flag states, and other risk factors that significantly influence vessel detention decisions.

The comparative analysis of the models provides a detailed overview of the performance of four different NN models. Parameters such as precision, accuracy, sensitivity, specificity, and AUC (Area Under Curve) are carefully analysed, and the results are presented graphically to illustrate the performance of each model. This analysis aims to find the optimal model to integrate large and heterogeneous datasets effectively, enabling accurate and efficient decision-making in PSC inspections.

Another research focus is implementing a system for automatic risk detection and preventive management. Such a system helps reduce human error, increase the efficiency of inspection procedures, and optimise the safety of offshore vessels. This paper's comprehensive application of advanced NN techniques represents a significant advance in the PSC inspection data analysis field, focusing on improving process safety and efficiency.

The ultimate goal is to optimise the inspection process, reduce the costs associated with vessel detention, and innovate by applying advanced technologies to achieve better resource allocation, more accurate risk assessment, and fewer procedural delays.

3. The Application of Neural Networks in Optimising Port State Control Inspection Process

3.1. PSC

PSC [14] plays a crucial role in the international oversight of maritime safety and makes it possible to ensure the safety and protection of life at sea and the environment from potentially harmful accidents at sea. Based on binding international conventions such as SOLAS (Safety of Life at Sea), MARPOL (Marine Pollution) and STCW (Standards of Training, Certification and Watchkeeping), PSC enables inspectors to inspect vessels in port, regardless of which flag the ship flies, to check compliance with international standards.

The main objective of PSC inspections is to identify and mitigate safety risks and potential non-compliance that could jeopardise the safety of the vessel, its crew, and the environment. These inspections cover various operational aspects of the vessel, with inspectors assessing the condition of structural components, safety equipment, environmental protection systems, and crew qualifications. In cases where irregularities are found, the authorities may decide to detain the vessel in port, significantly impacting shipowners and the vessel's operation.

PSC is a global process and part of the international strategy to reduce the risk of accidents in maritime transport. It is linked to regional PSC agreements (e.g., Paris Memorandum of Understanding (MoU) [15], Tokyo MoU [16]) that facilitate the exchange of information and best practices between signatory countries. The data collected during PSC inspections form the basis for improving safety in the shipping industry and enables the development of strategies to improve inspection processes and predict potential risks on vessels.

Modern approaches utilise new technologies, including AI and ML methods, to improve the efficiency of PSC inspections. This includes optimising methods to identify potential risks based on historical data on vessel violations and deficiencies. This approach enables a better assessment of the likelihood of a vessel having defects or other malfunctions. This improves the efficiency of inspections and reduces the number of accidents caused by technical problems on vessels.

In this context, modern researchers are focusing on developing ML-based models that can help automatically recognise patterns in data from previous inspections and predict future violations or risks. Increasing the accuracy of predictions and improving the robustness of the models are key challenges in the field of PSC.

3.2. Neural Networks: Fundamentals and Application in PSC Data Analysis

NNs are advanced ML models inspired by the functions of the human brain. They are used for pattern recognition and learning relationships between input data and the target variable [17,18].

This paper uses NNs to analyse large and diverse datasets in the context of PSC data. The aim is to predict vessel detention in harbours, assess the level of risk, and monitor compliance with international standards.

The essential components of NNs are neurons organised in layers. The NN layers include an input layer, one or more hidden layers and an output layer. The neurons in each layer are connected to the neurons in the neighbouring layers by weight coefficients representing the strength of the connections between the neurons. Each neuron receives input values, multiplies them by the weighting coefficients, adds a bias term, and applies an activation function, such as a Rectified Linear Unit (ReLU) or the sigmoid function. These activation functions allow the model to learn non-linear relationships within the data.

During training, the NN applies the backpropagation algorithm to adjust the weights and minimise the difference between the predictions and the actual values. Through this optimisation process, the NN can learn patterns from the data and generalise them to unseen examples [19].

In NNs, the weighting values are determined randomly or according to predefined rules and then adjusted iteratively during training. The model calculates the output value by combining input values with weighting coefficients. The bias term serves as an additional component with which the function's output can be shifted and is normally initialised with zero. After the first iteration, the model calculates the error by comparing the prediction with the output value. Based on this error, the backpropagation algorithm updates the weights. The new weighting values enable more accurate predictions in the subsequent iterations.

When training an NN, historical data adjust the model's parameters. The learning process optimises the weights and distortions in the network by minimising the error between the predictions and the actual values. The backpropagation algorithm and optimisers, such as the Adam algorithm [20–22], are used to achieve optimal parameter values. Once training is complete, the model is validated and tested on separate datasets to assess its generalisation ability on unseen data.

The application of NNs in analysing PSC inspection data offers significant advantages. These models can recognise latent patterns and relationships not immediately apparent in the raw data, such as correlations between geographical regions, vessel flags, and risk levels. They can also predict how likely vessels are to be detained, enabling more efficient resource allocation and increasing safety in the harbour.

However, using NNs presents some challenges. These models require large amounts of high-quality training data, and their complexity makes interpretability difficult. Despite these challenges, implementing NN models can significantly improve the PSC's inspection procedures, facilitate data-driven decisions, and improve compliance with international standards.

In this paper, the NN models "nnet", "mlp", "neuralnet", and "rsnns" were implemented using appropriate packages in the R [23] programming language. The "nnet" model uses basic functionalities to implement multilayer perceptrons with standard optimisation algorithms. The "mlp" package enables the efficient application of multilayer perceptrons with additional options for regularisation and hyperparameter optimisation. The "neuralnet" model offers flexibility in designing and training complex NNs with various activation functions. Finally, the "rsnns" package facilitates the implementation of advanced recurrent NNs designed explicitly for dynamic systems. The following sections provide a brief overview of how the individual models work.

3.2.1. Nnet

The package "nnet" [24] in the programming language R represents the basic implementation of simple NNs, primarily focused on classification and regression problems [25]. It allows the creation of multilayer perceptron networks ("mlp") with a single hidden layer that uses an activation function similar to the logarithmic function. The network is trained using the backpropagation algorithm, which optimises the weights to minimise the error in the network. This package is characterised by its ease of use and quick application for simple models. However, it is limited in optimisation and hyperparameter tuning compared to more advanced tools. Due to its simplicity, "nnet" is often used for educational purposes and less complex experiments.

3.2.2. Mlp

The multilayer perceptron (mlp) [26] is one of the best-known types of NNs in the field of DL. It consists of an input layer, one or more hidden layers, and an output layer, where all neurons between the layers are connected by weights that are optimised during the training process. "Mlp" uses activation functions such as sigmoid, tanh, or ReLU, which enable the learning of non-linear functions. ReLU, for example, is one of the most commonly used activation functions in deep neural networks (DNNs), especially in the context of the working layers of networks. The ReLU function transforms input values according to a simple rule: all negative numbers are set to zero, while other (positive) values remain unchanged.

Mathematically, the ReLU function is defined as $f(x) = \max(0, x)$, where x is the input value that the function receives. This simple form allows the network to learn quickly while avoiding the saturation problem that can occur with other activation functions, such as sigmoid or tanh. ReLU enables faster network training and better generalisation by reducing the vanishing gradient problem often occurring with deeper networks.

One of the most essential features of "mlp" is the ability to adjust the number of layers and neurons, contributing to the model's high flexibility. "Mlp" is used for problems where the data are complex and non-linear, such as pattern recognition, classification, and time series. This type of NN is advantageous in situations where traditional linear models are insufficient to achieve the desired results.

3.2.3. Neuralnet

The package "neuralnet" [27] in R is an alternative to the package "nnet". It enables the implementation of NNs with multiple hidden layers, which makes it possible to work with deeper architectures. This package is characterised by greater flexibility than "nnet" as it allows the user to select the number of hidden layers and the number of neurons in each layer, enabling the creation of more complex models. In addition, the package supports various activation functions, including log-sigmoid, tanh, and ReLU, and provides optimisation via the backpropagation algorithm, which is the basic method for training networks. One of the advantages of this package is that it allows detailed tuning of hyperparameters, such as the learning rate and the number of iterations. Although "neuralnet" offers more flexibility compared to more advanced tools such as "keras" or "tensorflow", its application is somewhat slower.

3.2.4. Rsnns

The package "rsnns" [28] is one of the most comprehensive tools for implementing NNs in the R environment. This package offers a wide range of possibilities for working with different types of NNs, including "mlp", radial basis function (RBF) networks, and self-organising maps (SOMs). In addition to standard techniques, "rsnns" enables advanced optimisation algorithms such as stochastic optimisation, adaptive training algorithms, error analysis, and result visualisation. The package offers a comprehensive set of functionalities for implementing complex models and is therefore suitable for research projects that require precise control over network architectures and optimisation. Compared to more straightforward packages, "rsnns" offers more power in analysing and modelling complex tasks in pattern recognition, prediction, and data analysis.

RBF is a function that is often used in NNs, especially in layers with radial basis function networks (RBFNs). The RBF is a non-linear activation function used in many applications such as classification, regression, and interpolation. The main feature of the RBF is its dependence on the distance between the input and the centres in the network. Mathematically, the RBF is usually represented as $\phi(x) = exp\left(-\frac{\|x-c\|^2}{2\sigma^2}\right)$, where x is the input vector, c is the centre of the function, $\|x - c\|$ is the Euclidean distance between the input and the centre, and σ is a parameter to control the width of the function.

RBFNs use this function to transform data into a higher dimensional space so that the models can better capture non-linear patterns. RBFNs are particularly popular in the context of interpolation and for classification tasks in situations where the data are complex and linear models are challenging to apply. In practice, the RBF is used in NN layers that contain radial basis functions as activation functions so that the models can handle complex problems efficiently. The advantages of RBFs include their ability to generalise new data well and their efficiency in pattern recognition in non-linear spaces.

Each of the NNs packages and techniques used in this research offers specific advantages and capabilities depending on the problem's complexity and the researcher's or engineer's particular needs. While "nnet" and "neuralnet" enable the rapid implementation of basic models, "rsnns" offers advanced features useful for more complex projects. "Mlp" is considered the basic model for many deep learning tasks and is the standard technique for prediction and classification in many areas. The choice of the package depends on the specific requirements, the size of the data, the complexity of the model, and the available computing resources.

4. Methodology

The research in this paper uses a methodology consisting of several key steps, including the description, preparation, and normalisation of the data, the splitting of the data into training and test datasets, the training of four different NN models on the same training dataset, identifying the key input variables of each NN model in the training dataset, testing or predicting on the test dataset, graphing the trained NNs, analysing the results, comparing the predictive performance of the obtained NN results, and evaluating and discussing the results.

After data collection, the data are subjected to detailed processing, which includes cleaning and preparation for further analysis. This process includes normalising numerical values, coding categorical data using methods such as one-hot coding, and correcting missing values by imputation or eliminating unreliable datasets [29]. After these transformations, the data are fully organised, cleaned, and free of not-available (NA) values, which ensures its consistency and reliability for further analysis and modelling.

The input dataset is represented by the matrix $\{D_{ij}\}, i = 1, \ldots m; j = 1, 2 \ldots, n$, where m is the number of observed instances (offshore vessel inspections) and n is the number of observed variables.

At the beginning of the paper, before normalisation, a graphical representation of the correlation between all input variables was displayed, highlighting only correlations greater than ± 0.3. A correlation greater than 0 indicates a positive correlation, meaning that the variables are associated in the same direction, while a correlation less than 0 indicates a negative correlation, where the variables are associated in the opposite direction.

The normalisation of data [30] is crucial to preparing data for applying machine learning (ML) methods for several reasons. Firstly, different variables in the dataset may have different units of measurement, which can lead to one variable dominating another during model training. Normalisation ensures that all variables have the same units of measurement and that the contribution of each variable to the model is balanced. Another reason is that many ML algorithms assume that the data are in a particular range and will not perform well if the data have different scales. Normalisation can also help to speed up the convergence process during model training. In this case, a normalisation technique called "centring" and "scaling" was used, where each variable is first centred so that its mean becomes zero and then scaled to have a standard deviation of 1. This technique ensures that all variables have the same influence on the NN model, which contributes to better model efficiency and accuracy during training.

After data preparation, the paper applies four different NN models: "nnet", "mlp", "neuralnet", and "rsnns". These models are used to predict the probability of vessel detention, with particular attention paid to optimising the parameters of each model to achieve the highest possible accuracy and robustness.

Each NN model is presented graphically to explain its structure based on the training dataset and its performance. Finally, a comparative plot is provided for all four NN models regarding key metrics, including accuracy, precision, sensitivity, F1 score, and others.

The discussion analyses the results obtained using the four NN models and compares their key performance metrics. The discussion focuses on processing and describing each model's advantages and weaknesses in predicting vessel detections during PSC inspections, emphasising the factors affecting model performance. This section also addresses each NN model's computational and time complexity, which is critical for evaluating the model's efficiency under real-world conditions. Considering the modelling specifics, special attention has been paid to the computational cost of training and testing the model and the time required to implement each model.

Finally, the conclusion gives an overview of the paper's limitations and potential areas for improvement, such as using other techniques, improving the results, or optimising the existing models.

The R programming language, which specialises in statistical analysis and data processing, was used to implement the NN models. Relevant R libraries for NNs were used, briefly described in the previous chapter. Table 1 provides an overview of the methodology used.

Table 1. Methodology overview.

Phase	Description
Data description	The dataset is represented as a matrix $\{D_{ij}\}, i = 1, \ldots m; j = 1, 2 \ldots, n$ where m is the number of observed instances (offshore vessel inspections), and n is the number of observed variables.
Correlation analysis	The graphical representation of the correlations between all numerical variables in the input dataset shows correlations greater than ± 0.3. Positive correlations (greater than 0) indicate variables related in the same direction, while negative correlations (less than 0) indicate variables related in the opposite direction.
Data preparation and normalisation	The data contain different types of variables, such as categorical and numerical variables. Categorical variables (e.g., country of inspection, vessel flag) are converted into factor representations using one-hot coding or embedding techniques. Numerical variables (e.g., number of defects, vessel tonnage) are normalised to ensure uniformity and easy processing in NNs.
Normalisation method	Normalisation is achieved by "centring" and "scaling", whereby each variable is first centred so that its mean is zero and then scaled so that its standard deviation is one. This ensures that all variables have the same influence on the NN model.
Data splitting	The input dataset (421) is split into training (337) and test sets (84) to train and evaluate the models.
Model training	Four different NNs models are applied: "nnet", "mlp", "neuralnet", and "rsnns". These models are trained on the same training dataset to predict the probability of a vessel detention, focusing on optimising model parameters to achieve higher precision and robustness.
Feature importance	The most important input variables for each model are identified within the training dataset.
Testing and prediction	The models are tested, and predictions are made on the test dataset to assess their generalisation performance.
Model evaluation	A graphical representation of each NN model is provided, showing their structures and performance metrics, including accuracy, precision, sensitivity, F1 score, etc.
Results analysis	The results are analysed, focusing on each model's advantages and disadvantages in predicting vessel detections during PSC inspections.
Discussion and comparison	The discussion includes a comparative analysis of the models, focusing on their strengths, weaknesses, and factors affecting performance. In addition, the computational and time requirements during the training and testing phases are discussed.
Conclusion	The conclusion highlights the paper's limitations and possible improvements through other techniques or the optimisation of existing models.

4.1. Dataset Description

The input dataset is represented as a matrix $\{D_{ij}\}$ consisting of data from 421 inspections, each described by 26 variable values relating to the country of inspection, memorandum, flag, and other data such as inspection status, number, and type of defects and operational factors that may affect the vessel's safety.

The PSC input data used in this paper are available on the Equasis website [31] and relates to inspections carried out by the competent authorities in ports to verify vessels' compliance with international safety, environmental, and technical standards. The recorded inspection data are essential for assessing vessel safety and reducing the risk of accidents at sea. The website contains detailed information on the number of inspections carried out throughout a vessel's life cycle, including the results of these inspections, which may indicate irregularities or deficiencies in the vessel's equipment and maintenance. In cases where serious problems are identified, a vessel may be detained in port until the irregularities are rectified or the required safety standards are met. In addition, the website provides insight into the sanctions that can be imposed for non-compliance, including a possible ban on further voyages or mandatory repairs. Equasis also provides an analysis of inspections by vessel flag and port state, enabling a global assessment of the frequency and effectiveness of PSC measures. This is a valuable tool for researching trends and assessing safety factors in international maritime transport.

The input data refer exclusively to offshore vessels built in 1997 or later with a gross tonnage (GT) between 3000 and 13,000. The observation period extends from 2005 to the present.

The target variable "Detention" can have a numerical value of 0 or 1. If a vessel is detained, "Detention" takes the value 1; otherwise, it takes the value 0. The total number of recorded inspections in the input dataset is 421, with 56 vessels detained (13.30%).

Three of the 26 input variables in the dataset are text variables (CountryOfInsp, Memo, and Flag), which were converted to factors so that they could be used in learning models as NNs can only process numerical values.

During data preparation, the input dataset was randomly split into a training and a test dataset, with a ratio of 80% (337) for training and 20% (84) for testing. This split allows for efficient validation, i.e., testing the model on unseen data, which increases the accuracy of the predictions.

Tables 2–4 describe all variables, define their type, the cut-off or mean values, and the frequency of their occurrence.

Table 2. Target variable "Detention".

	Dataset Number	**Detention_No**	**Detention_Yes**
Detention	421	365	56

The variables "CountryOfInsp" (36 options), "Flag" (32 options), and "Memo" (12 options) are categorical variables. Table 3 shows a tabular representation of the 10 most frequent values for these three variables in the input dataset (421).

The general descriptive statistics for the remaining 22 numerical independent variables can be found in Table 4.

The descriptive statistics of the numerical variables grouped according to the outcome variable "Detention" are presented in Table 5.

Table 3. Tabular representation of the 10 most frequent values for three categorical variables (CountryOfInsp, Memor, and Flag) in the input dataset (421).

No	CountryOfInsp	N	Memor	N	Flag	N
1	United Kingdom	119	Paris	236	Vanuatu	57
2	Australia	50	Tokyo	79	Norway	50
3	Cyprus	35	Mediterranean	49	Brazil	31
4	Singapore	24	Indian Ocean	20	Luxembourg	30
5	Norway	20	Abuja	11	Canada	22
6	Denmark	18	US Coast Guard	8	Greece	21
7	Malta	18	Indian	7	Bahamas	20
8	Spain	17	Black Sea	3	Gibraltar	20
9	Netherlands	16	Paris MoU	3	Malta	19
10	People's Republic of China	13	VinaDelMar	3	Denmark	18

Table 4. Descriptive statistics for numeric variables in the input dataset.

No	Variable	Variable Description	Data Type	Min	Max	Mean	Median	Unique Values
1	YOB	Year of Ship Built	numeric	1992	2023	2010.10	2011	27
2	InitialInsp	Initial Inspection	numeric	0	1	0.52	1	11
3	DetInsp	More Detailed Inspection	numeric	0	1	0.40	0	9
4	FollowUpInsp	Follow Up Inspection	numeric	0	1	0.03	0	5
5	StandInsp	Standard Inspection	numeric	0	1	0.02	0	2
6	NoInsp	Number Of Inspections	numeric	1	38	11.40	10	24
7	NoDef	Number Of Deficiency	numeric	0	31	4.17	3	21
8	ISM	ISM Deficiency	numeric	0	3	0.18	0	4
9	MARPOL	MARPOL Deficiency	numeric	0	4	0.28	0	5
10	CertDoc	Certificate Documentation Deficiency	numeric	0	10	1.05	1	8
11	PAMach	Propulsion Auxiliary Machinery Deficiency	numeric	0	2	0.08	0	3
12	SafNav	Safety Of Navigation Deficiency	numeric	0	5	0.53	0	6
13	RadioCom	Radio Communications Deficiency	numeric	0	3	0.18	0	4
14	EmergSyst	Emergency Systems Deficiency	numeric	0	4	0.16	0	5
15	FireSafety	Fire Safety Deficiency	numeric	0	5	0.56	0	6
16	MLC	MLC Deficiency	numeric	0	5	0.28	0	6
17	Alarms	Alarms Deficiency	numeric	0	5	0.06	0	4
18	ISPS	ISPS Deficiency	numeric	0	2	0.02	0	3
19	OTDef	Other Types of Deficiencies	numeric	0	11	0.34	0	6
20	WWTightCond	Water Weather Tight Conditions Deficiency	numeric	0	2	0.09	0	3
21	LifeApl	Life-Saving Appliances Deficiency	numeric	0	5	0.35	0	6
22	GT	Gross Tonnage	numeric	2012	80106	4908.71	3832	100

Figure 1 was created based on the correlation, i.e., the relationships between the different variables. The value in each cell of the matrix with dimensions $(n \times n)$, where n stands for the model's total number of variables (26) and represents the correlation coefficient between two variables. A correlation coefficient close to 1 indicates a strong positive correlation, while a coefficient close to -1 indicates a strong negative correlation. A coefficient close to 0 indicates a weak or no linear relationship between the variables. In Figure 1, correlation values greater than ± 0.3 are shown graphically and labelled as significant. Figure 1 shows a positive correlation between "NoDef" and "ISM", with a moderate positive correlation of 0.50. "NoDef" and "MARPOL" also show a positive correlation of 0.44, indicating a moderately positive relationship between these variables. "NoDef" and "CertDoc" show a relatively strong positive correlation of 0.56, which means that as one variable increases, the other also increases. "PAMach" and "SafNav" show a moderate positive correlation of 0.65, which means that higher values for "PAMach" generally go hand in hand with higher values for "SafNav". "SafNav" and "RadioCom" show a moderate positive correlation of 0.41. "SafNav" and "NoDef" also show a moderate positive correlation of 0.65.

Table 5. Descriptive statistics of the numerical variables, grouped by the target variable "Detention".

Detention	0				1			
Variable	**Min**	**Max**	**Mean**	**Median**	**Min**	**Max**	**Mean**	**Median**
YOB	1992	2023	2009.95	2011	1999	2023	2011.02	2012
InitialInsp	0	1	0.53	1	0	1	0.48	0
DetInsp	0	1	0.41	0	0	1	0.32	0
FollowUpInsp	0	1	0.02	0	0	1	0.14	0
StandInsp	0	1	0.01	0	0	1	0.07	0
NoInsp	1	38	11.65	10	1	26	9.75	9
NoDef	0	15	3.45	3	1	31	8.82	8.5
ISM	0	3	0.12	0	0	2	0.54	0
MARPOL	0	4	0.25	0	0	3	0.52	0
CertDoc	0	6	0.93	1	0	10	1.80	1
PAMach	0	2	0.05	0	0	2	0.23	0
SafNav	0	4	0.41	0	0	5	1.36	1
RadioCom	0	3	0.14	0	0	3	0.45	0
EmergSyst	0	4	0.13	0	0	3	0.32	0
FireSafety	0	5	0.48	0	0	5	1.11	1
MLC	0	5	0.22	0	0	4	0.66	0
Alarms	0	4	0.03	0	0	5	0.27	0
ISPS	0	1	0.01	0	0	2	0.09	0
OTDef	0	6	0.30	0	0	11	0.63	0
WWTightCond	0	2	0.07	0	0	2	0.18	0
LifeApl	0	4	0.31	0	0	5	0.66	0
GT	2012	80106	5032.78	3888	2012	6776	4100.05	3776.5

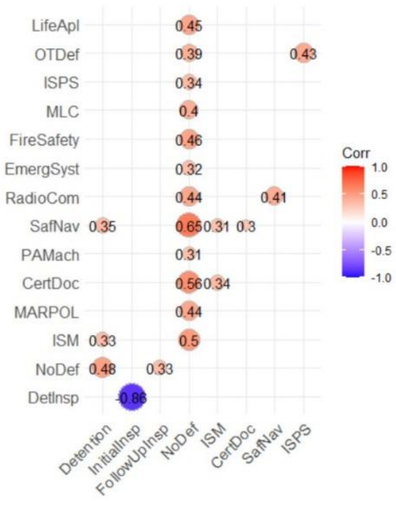

Figure 1. Correlation matrix (with significant correlations above ±0.3).

The variables "InitialInsp" and "DetInsp" have a strong negative correlation of −0.86. When "InitialInsp" increases, "DetInsp" decreases, and vice versa.

4.2. Metrics

Several key metrics are used to evaluate and compare the predictive performance of the four NN models. Based on the confusion matrix, precision, accuracy, sensitivity, F1 score, and other statistical measures are calculated.

In the confusion matrix (Table 6), the values of True Positives (TPs), False Positives (FPs), True Negatives (TNs), and False Negatives (FNs) are arranged according to the actual (true) and predicted classes. In the binary classification, which was applied to all four NN models in this paper, the confusion matrix has a 2×2 structure.

Table 6. Confusion matrix form.

	Predicted 0	**Predicted 1**
Actual 0	TN	FP
Actual 1	FN	TP

In the confusion matrix, TP represents the number of correctly classified positive cases, i.e., the cases in which the actual value was 1, and the model predicted 1. FPs denote the number of negative cases incorrectly classified as positive, i.e., the cases in which the actual value was 0 but the model predicted 1. TNs denote the number of correctly classified negative cases, i.e., the cases in which the actual value was 0, and the model predicted 0. FNs denote the number of positive cases incorrectly classified as negative, i.e., cases where the actual value was 1 but the model predicted 0.

The most important metrics for the performance evaluation of NN models are shown in Table 7.

Table 7. Metrics.

Metric	**Description**	**Formula**
Accuracy (ACC)	Represents the overall percentage of correct predictions relative to all predicted instances.	$ACC = \frac{TP+TN}{TP+TN+FP+FN}.$
95% Confidence Interval (CI)	Represents the confidence interval for accuracy.	$SE = \sqrt{\frac{Accuracy \cdot (1-Accuracy)}{TP+TN+FP+FN}}$
No Information Rate (NIR)	Indicates the accuracy of a model in predicting the most frequent class.	$NIR = max\left(\frac{TP+FN}{TP+TN+FP+FN}, \frac{FP+TN}{TP+TN+FP+FN}\right)$
The p-value (ACC > NIR)	Tests whether the model's accuracy is significantly higher than the NIR value.	Obtained by testing the difference between the model accuracy and the NIR value.
Kappa	A measure of agreement that accounts for the agreement occurring by chance.	$Kappa = \frac{P_0-P_e}{1-P_e}$
P_e (Expected Agreement)	Represents the percentage of correct predictions expected by chance, taking into account the class distribution in the dataset.	$P_e = \left(\frac{TP+FN}{TP+TN+FP+FN}\right) \times \left(\frac{TP+FP}{TP+TN+FP+FN}\right) \times \left(\frac{TN+FP}{TP+TN+FP+FN}\right) \times \left(\frac{TN+FN}{TP+TN+FP+FN}\right).$
McNemar Test p-value	Evaluates whether there is a significant difference in the classification errors for positive and negative samples.	$P = \frac{(b-c)^2}{b+c}$
Sensitivity (Recall or TPR)	Measures how well the model identifies positive instances.	$Sensitivity = \frac{TP}{TP+FN}.$
Specificity (True Negative Rate (TNR))	Measures how well the model identifies negative instances	$Specificity = \frac{TN}{TN+FP}.$
Positive Prediction Value (PPV)	Indicates how accurate the positive instances (TP) are.	$PPV = \frac{TP}{TP+FP}.$
Negative Prediction Value (NPV)	Indicates how accurate the negative instances (TN) are.	$NPV = \frac{TN}{TN+FN}.$
Detention Rate	Measures the success of the model in recognising positive instances.	$DetentionRate = \frac{TP}{TP+TN+FP+FN}.$
Balanced Accuracy	The average of sensitivity and specificity, useful for unbalanced datasets.	$BalancedAccuracy = \frac{Sensitivity+Specificity}{2}$
F1	The harmonic mean of precision and sensitivity.	$F1 = 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity}.$
FPR	The ratio of FPs to the total number of TNs instances. Indicates how often the model incorrectly classifies negative cases as positive.	$FPR = \frac{FP}{FP+TN} = 1 - TNR$
AUC	Represents the area under the ROC curve, quantifying the model's ability to distinguish between positive and negative instances.	$AUC = \int_0^1 TPR(t)dFPR(t)$

The values listed and explained in Table 7 are essential metrics for evaluating classification models, especially when misclassification or neglect of positive samples is undesirable.

All four NN models “nnet”, “mlp”, “neuralnet”, and “rsnns” were trained and tested on the same dataset. Based on the predictions of each model and the values obtained from the confusion matrix (TN, FP, FN, and TP), the corresponding performance indicators were calculated and presented. This comparison aims to determine which model shows the most effective approach in recognising the positive class, especially when dealing with unbalanced data.

A dataset is considered unbalanced when the frequency of one class significantly exceeds that of the others, which is the case for the present dataset as it contains only 56 (13.3%) positive instances. Consequently, the metrics described above provide a comprehensive framework for analysing the performance of classification models and can be used to interpret the results derived from the confusion matrix.

4.3. Implementation

The methodology in this paper was implemented using R 4.4.2 [2], the latest stable version of the R statistical programming language, released on 31 October 2024. This version offers numerous improvements over previous versions, including improvements in features, packages, installation processes, compatibility, and development and documentation tools. R 4.4.2 offers advanced optimisation features for code execution and supports advanced features that improve statistical analysis and modelling efficiency and accuracy.

The following R packages and their functions were used to implement the four NN models in this methodology:

1. The “Nnet” package [24] provides a basic implementation of feedforward NNs with one or more hidden layers. It uses the “nnet()” function for training models, including logistic regression, classification, and regression models. Due to its simplicity and efficiency, this package is beneficial for basic classification and regression tasks.
2. The “Mlp” package [26] implements deep neural networks (DNNs) for classification and regression problems. It offers additional functionalities, such as optimising hyperparameters and training multilayer perceptron models (MLP). The “mlp” package is often used to model complex relationships between input and output data in various applications.
3. The “Neuralnet” package [27] supports the implementation of complex feedforward NNs with multiple hidden layers and is suitable for regression and classification problems. It provides functions for optimising weights and evaluating model performance, which makes it very useful for analysing high-dimensional data.
4. The Rsnns package [28] implements various types of NNs, including “mlp” and radial basis function (RBF) networks. It also facilitates model evaluation and optimisation using multiple techniques. It is particularly useful for advanced ML and predictive modelling applications and for processing large datasets.

Using these NN models in R enables precise data analysis, model training, and parameter optimisation, which is crucial for AI research applications.

5. Results

The results of the NN model “nnet” (Figure 2) show the network’s training parameters, including the weights between the layers and the bias values. Like the others, this NN model was trained using functions of the R programming language.

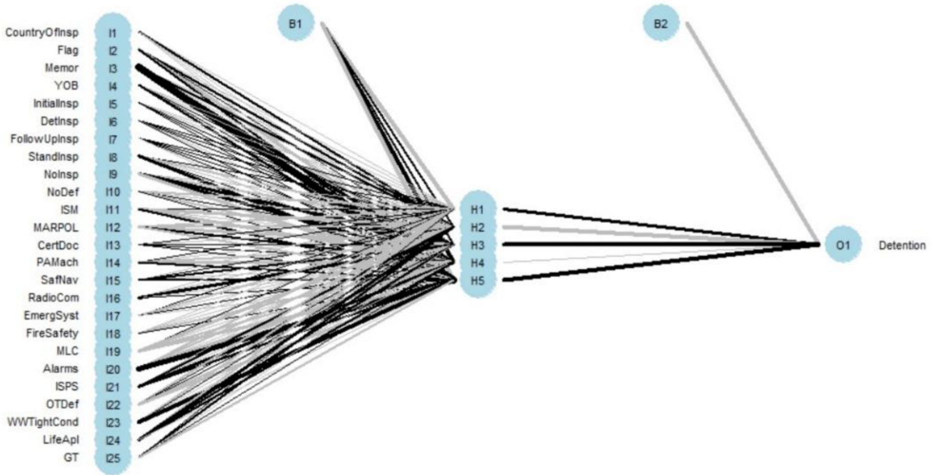

Figure 2. Graphical representation of the neural network structure (nnet).

The NN model "nnet" created for the training dataset consists of three layers (input, hidden, and output). The input layer comprises 25 variables the model uses as input data. The hidden layer contains five neurons that help model complex non-linear relationships between inputs and outputs. The weight matrix between the input layer and the hidden layer (hidden_weights) has 25 rows and 5 columns (the first eight are shown in Table 8), i.e., each of the 25 input neurons is connected to each of the five hidden neurons (V1–V5). Each weight in the matrix determines how strongly a particular input neuron influences a particular hidden neuron. For example, the first input neuron weights −20.596 for the first hidden neuron, which means a strong negative influence. Similarly, the eighth input neuron weights 31.749 for the fifth hidden neuron, which is a strong positive influence. Large weights (either in the positive or negative direction) indicate that the corresponding input significantly influences the activation of the hidden neuron. Smaller weights indicate a weaker influence.

Table 8. The weight matrix between the input and hidden layer is in the "nnet" model.

Neuron	V1	V2	V3	V4	V5
1	−20.596	3.077	12.024	−4.306	−5.624
2	−0.407	8.617	5.900	9.350	−1.044
3	−0.215	9.402	9.120	−1.448	7.716
4	−1.036	6.248	−5.041	16.136	0.302
5	−5.222	5.551	5.004	−2.012	−21.093
6	−2.628	−1.050	−0.388	−12.726	−1.450
7	4.087	−0.423	0.318	11.415	5.714
8	1.992	1.500	0.157	2.456	31.749
...	...	...	...	...	...

The weight matrix (vector) between the hidden and output layers (Table 9) with dimensions (5 × 1) illustrates how each of the five hidden neurons contributes to the final output of the network. The output layer consists of a single output neuron that is orientated to the prediction task performed. Since hidden neuron 3 weighs 16.590, it has a significant positive impact on the output. Similarly, hidden neuron 2 has a weight of −11.743, which lowers the output value. Combining positive and negative weights allows the network to find optimal relationships between the hidden layers and the output.

Table 9. The weight matrix between the hidden and output layers is in the "nnet" model.

Hidden_Output	W_i
V1	9.969
V2	−11.743
V3	16.590
V4	−1.079
V5	−10.577

The bias values for the hidden layer are listed in Table 10. The vector hidden_biases with the dimensions (5 × 1) contains the bias values for the five hidden neurons. The bias is an additional constant that allows the NN to match the outputs of the hidden neurons better, regardless of the input weights. For example, the bias (B1) for the first hidden neuron is −16.236, which means that this neuron will only be activated if the combined input exceeds this value. Similarly, the bias for the fourth hidden neuron is 16.910, indicating a shift in activation of this neuron towards higher input values. This means that the bias works similarly to the intercept in linear models and helps the network to adapt optimally to the data.

Table 10. Hidden layer bias values.

Hidden_Biases
−16.236
13.008
−17.539
16.910
−4.501

The bias for the output neuron (B2) with the dimension (1 × 1) is 17.454 and acts as an additional parameter that controls the activation level of the output neuron. This means that even if all hidden neurons output zero, the output neuron will have an initial value of 17.454 before the activation function is applied.

The output of the NN is defined by Equation (1):

$$Y = f\left(\sum_{i=1}^{n} W_i H_i + B\right) \tag{1}$$

where

- W_i is the weight between the hidden layer and the output layer;
- H_i is the output of the hidden neuron;
- B is the bias (in this case, 17.454);
- f is the activation function, and the bias shifts the sum before activation.

These elements are important as they enable the model to approximate the data better. Without the bias, the NN would be limited to models that pass through the coordinate origin (0,0). The bias also improves the network's ability to learn non-linear relationships, allowing the model to fit the transfer function flexibly. Finally, this method accelerates convergence during training, as the bias enables the network to adapt to optimal values more quickly.

In this case, the positive bias value of 17.454 indicates that the model generally predicts higher output values unless weights and activation functions neutralise them.

Figure 2 illustrates that the NN model "nnet" is designed to process 25 input variables, transform them through five hidden neurons and produce a single output. The weights and bias values structure shows that the network can model complex relationships within the data.

Further improvement of the model would require experimenting with increasing or decreasing the number of hidden neurons (e.g., testing the network with 10 or 3 hidden neurons) and applying weight regularisation to prevent overfitting. This analysis clearly shows that the NN has successfully learnt certain patterns in the data. However, for the final evaluation of its performance, additional analyses (e.g., accuracy tests) would be beneficial.

Figure 3 shows the ranking of the 10 most influential variables and illustrates the relative importance of each variable in the "nnet" model. According to the results, the most crucial variable is "Alarms" with a relative importance of 0.0683, which indicates the most substantial influence on the model's prediction. The following most essential variables are "LifeApl" (0.0679) and "NoInsp" (0.0662), which also significantly influence the results of the model, albeit to a lesser extent. Other variables such as "MARPOL", "Memor", and "PAMach" play a lesser but still significant role. This visualisation improves the interpretability of the model, facilitates the identification of key factors, and directs further research to the most influential variables.

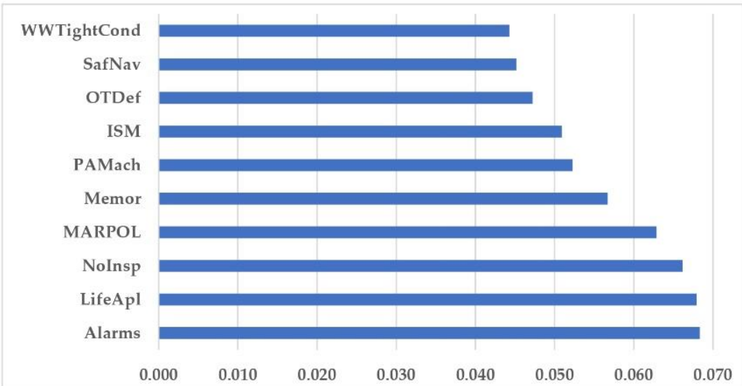

Figure 3. Top 10 most important variables in the "nnet" model.

The evaluation results of the NN model "nnet" for the test dataset (84 instances) are shown in the confusion matrix (Table 11).

Table 11. Confusion matrix of "nnet" model for the test dataset.

	Predicted 0	**Predicted 1**
Actual 0	63	5
Actual 1	9	7

The results of the NN model "mlp" show the performance of a multilayer perceptron ("mlp") trained on a dataset with 337 samples and 25 predictors for a binary classification problem (classes "0" and "1"). The model was validated with a 10-fold cross-validation, increasing the reliability of the performance estimation (Figure 4).

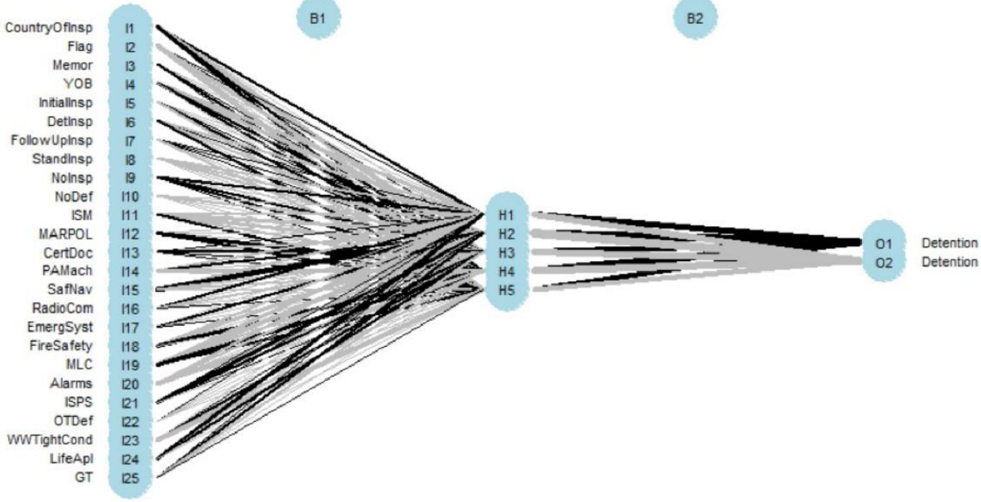

Figure 4. Graphical representation of the neural network structure ("mlp").

The model was tested with three sizes of hidden layers (size = 1, 3, and 5) and evaluated for accuracy. The best results were obtained with five neurons, as this configuration provided the highest accuracy (91.7%), shown graphically in Figure 4.

Figure 5 illustrates the relative importance of the 10 most influential variables in the "mlp" model, while the confusion matrix for the "mlp" model for the test dataset has the following values: TN = 70, TP = 7, FP = 5 and FN = 2 (Table 12).

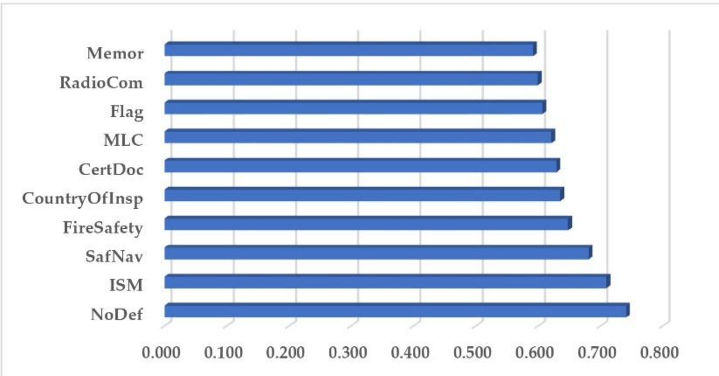

Figure 5. Top 10 most important variables in the "mlp" model.

Table 12. Confusion matrix of "mlp" model for the test dataset.

	Predicted 0	Predicted 1
Actual 0	70	5
Actual 1	2	7

Figure 6 illustrates the NN model "neuralnet", which consists of 25 input variables, a single hidden layer with five neural units, and an output layer with two neural units. The model architecture determines how the network transforms the input data to optimise the classification task. The input layer consists of 25 variables that represent key factors for analysing the detention of vessels in ports. These include attributes such as the country of inspection (CountryOfInsp), the flag of the vessel (Flag), the year of construction (YOB), the number of previous inspections (NoInsp), defects (NoDef), and various safety and regulatory parameters.

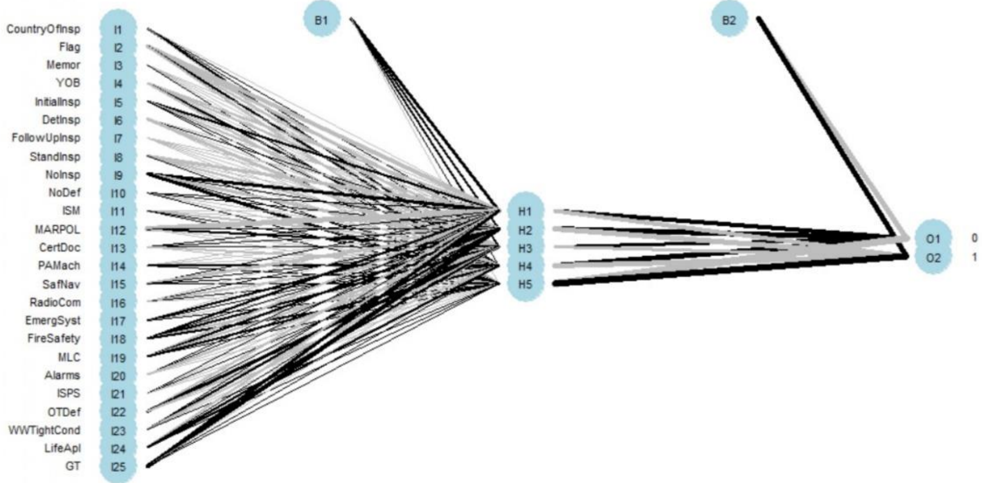

Figure 6. Graphical representation of the neural network structure ("neuralnet").

The hidden layer consists of five neural units (1layhid1 to 1layhid5) that extract and combine significant patterns from the input data. Each neuron in the hidden layer receives weighted input signals (B1), processes them through activation functions, and passes the outputs with the weighted output signal (B2) to the next layer. The weighting coefficients of the model determine the strength and significance of the connections between the individual input variables and the hidden neurons.

The output layer consists of two neural units ("0" and "1") that represent a binary classification, i.e., the prediction of whether a vessel will be kept (1) or released (0). Each output unit receives input signals from all five hidden neurons, whose values are combined using the corresponding weights and then transformed by an activation function to obtain the final probability prediction.

The model was trained with 328 iterations (steps) and achieved a threshold error of 0.009, indicating high model convergence. The model error is 2.082, which suggests the network's overall accuracy.

This NN structure enables the detection of complex non-linear relationships between the variables and optimises the decision-making process regarding the detention of vessels based on historical inspection data.

In addition to the 25 input variables in the model, there is an axis intercept, an additional component in analysing the variables' importance. The intercept is a crucial component in models such as the NN as it defines the base prediction of the model before any input variables are applied. In other words, the intercept reflects the baseline output value, which is used as an output reference for the predictions regardless of the values of the input variables.

In the context of the NN, the intercept is a parameter that allows the network to "shift" or "adjust" its output value relative to a predefined reference point. This is particularly important as it allows the model to be better generalised and adapted to different data patterns. Without the intercept, the model would be forced to make predictions across all network layers that are fixed at zero, significantly limiting its flexibility and accuracy. Due to this fundamental role, the intercept is often treated as a separate variable when evaluating the importance of variables, as it affects the initial accuracy of the model. Thus, although the model may have 25 input variables, the intercept is often included in the analysis of variable importance, which explains the presence of 26 variables in the variable importance scores.

The importance of the top 10 input variables in the "neuralnet" model is shown in Figure 7, while the confusion matrix of the "neuralnet" model can be seen in Table 13.

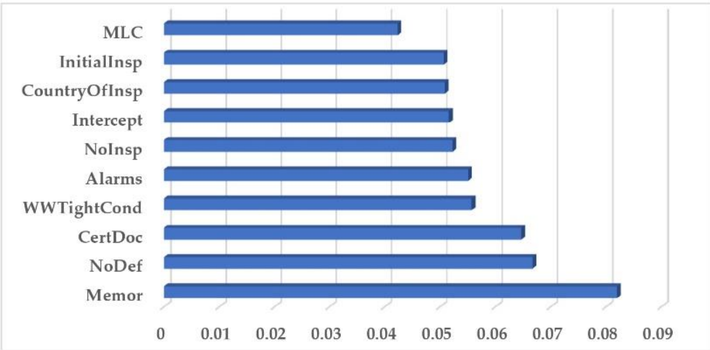

Figure 7. Top 10 most important variables in the "neuralnet" model.

Table 13. Confusion matrix of "neuralnet" model for the test dataset.

	Predicted 0	Predicted 1
Actual 0	68	5
Actual 1	4	7

Figure 8 shows the structure of the NN model "rsnns", which consists of 25 input units (representing different vessel and inspection features), five hidden units (using the same activation function as the input units), and two output units (one using a logistic activation function and the other using the same activation as the hidden units).

This network uses standard backpropagation for training and a topological order for updating the units. The connections between the units are defined by weights that are updated during the training process.

The meaning of the 10 most important variables in the "rsnns" model is shown in Figure 9, while the confusion matrix for the "neuralnet" model is shown in Table 14.

Table 14. Confusion matrix of "rsnns" model for the test dataset.

	Predicted 0	Predicted 1
Actual 0	69	2
Actual 1	6	7

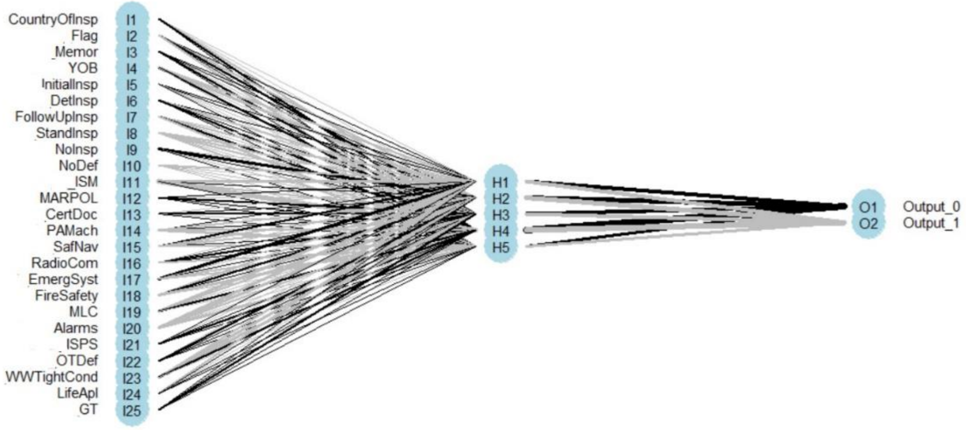

Figure 8. Graphical representation of the neural network structure ("rsnns").

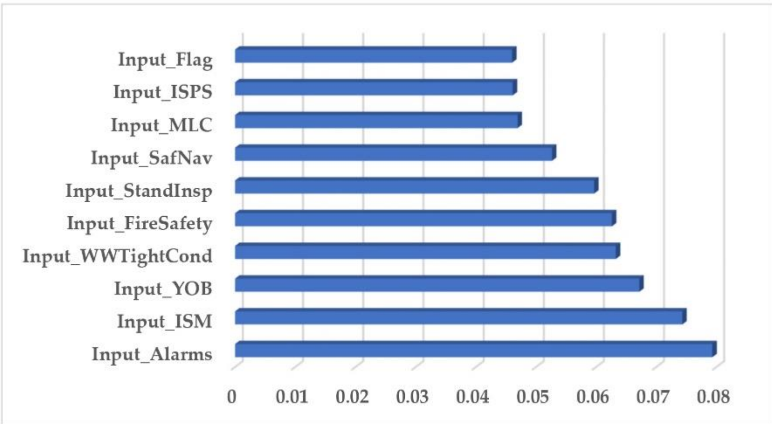

Figure 9. Top 10 most important variables in the "rsnns" model.

The NN model "nnet" consists of three layers: an input layer with 25 variables, a hidden layer with five neurons, and an output layer. The weight matrices of the model between the layers determine the influence of the individual inputs and hidden neurons on the final output. Bias values for the hidden neurons adjust the activation thresholds, improving the model's ability to capture complex non-linear relationships. The output layer, guided by these weights and biases, predicts values based on the data, with the biases helping the network to adapt more quickly during training. The performance of the model was evaluated using accuracy metrics and a confusion matrix and showed reliable prediction for binary classification tasks. Further improvements could include adjusting the number of hidden neurons and adjusting the weights. The importance of the model's variables was also analysed to identify the most important factors influencing the predictions. Similar NN models such as "mlp", "neuralnet", and "rsnns" were compared, each with different configurations and training parameters, which showed effective prediction and analysis of the importance of the variables.

Although all models ("nnet", "mlp", "neuralnet", "rsnns") were trained and tested on the same dataset, the differences in the results can be explained by variations in the optimisation techniques, the model parameters, and their ability to adapt to the specifics of the dataset.

Model complexity: Each of these models uses different architectures and methods to adjust the weights and activation functions. For example, "nnet" and "neuralnet" models may use different approaches in defining the number of layers and neurons, which affects their ability to adapt appropriately to the data.

Optimisation methods: Differences in optimisation methods (such as weight adjustment algorithms, learning rates, and regularisation) can lead to differences in the models' ability to minimise the loss function. These factors can significantly affect the accuracy of the model in the test dataset.

Model complexity: Some models, such as "mlp" or "rsnns", may use deeper or more complex architectures that are better able to model non-linearities in the data, while simpler models may have limited ability to capture all features of the dataset.

Although all models have been trained with the same data, their different architectures and optimisation approaches can lead to differences in performance. Understanding how each model processes the data and how their complexity and optimisation techniques are aligned can give you deeper insights into these results.

6. Discussion

The analysis of the confusion matrices for all four NN models "nnet", "mlp", "neuralnet", and "rsnns" is shown in Table 15. The results provide a deeper insight into the performance of the models based on the distribution of correct and incorrect classifications.

Table 15. Models' performance metrics with confusion matrix elements.

Model	nnet	mlp	neuralnet	rsnns
TN	63	70	68	69
FP	5	5	5	2
FN	9	2	4	6
TP	7	7	7	7

The models show similar values in terms of the number of TNs, with "mlp" (70) and "rsnns" (69) performing slightly better than "nnet" (63) and "neuralnet" (68). This indicates that "mlp" and "rsnns" are slightly more accurate in identifying negative cases. For FPs, all models except "rsnns" have the same value (5), while "rsnns" achieves the lowest number (2), which may be advantageous in contexts where minimising FPs is crucial.

The number of FN instances shows clear differences between the models, with "mlp" achieving the best result with only two FN instances, while "nnet" has the highest number (9), which reduces its sensitivity and increases the probability of missing TP instances. All models have the same number of TP instances (7), meaning they recognise equally positive instances.

From these results, it can be concluded that the "mlp" model is the most efficient, as it achieves the lowest number of FN instances at a relatively low FN rate. The "rsnns" model is more accurate in avoiding FP classifications, but it misses more positive instances, which can be critical in contexts where identifying positive instances is crucial. The "nnet" model shows the weakest performance as it has the highest number of FN classifications, which reduces its reliability and limits its practical usability.

The comparative analysis of the models "nnet", "mlp", "neuralnet", and "rsnns" (Table 16) reveals significant performance differences based on various metrics.

Table 16. Comparative analysis of the performance of the models "nnet", "mlp", "neuralnet", and "rsnns".

Model	nnet	mlp	neuralnet	rsnns
Accuracy	0.833	0.917	0.893	0.905
CI (Lower)	0.754	0.858	0.827	0.842
CI (Upper)	0.913	0.976	0.959	0.968
NIR	0.810	0.893	0.869	0.845
p-Value—NIR	0.050	0.050	0.050	0.050
Kappa	0.830	0.916	0.891	0.904
Pe	0.721	0.781	0.764	0.771
McNemar-*p*-Value	1.143	1.286	0.111	2.000
Sensitivity	0.438	0.778	0.636	0.538
Specificity	0.926	0.933	0.932	0.972
Precision	0.583	0.583	0.583	0.778
NPV	0.875	0.972	0.944	0.920
Detection Rate	0.083	0.083	0.083	0.083
Balanced Accuracy	0.682	0.856	0.784	0.755
F1 Score	0.500	0.667	0.609	0.636
FPR	0.074	0.067	0.068	0.028
AUC	0.729	0.778	0.764	0.849

Regarding accuracy, the "mlp" model stands out with a value of 0.917, while "rsnns" is close at 0.905, indicating superior classification ability compared to "nnet" (0.833) and "neuralnet" (0.893). This trend is further supported by the Kappa coefficient, where "mlp" (0.916) and "rsnns" (0.904) show better agreement with actual data compared to the other models. The AUC confirms this conclusion, with the "rsnns" model achieving the highest AUC value (0.849), suggesting its superior ability to differentiate between positive and negative cases. In contrast, the "nnet" model exhibits the lowest AUC value (0.729), indicating its weaker ability to separate classes.

In terms of sensitivity, the "mlp" model performs the best with a value of 0.778, significantly higher than "nnet" (0.438), meaning that "mlp" is more effective at identifying positive cases. Specificity is high across all models, with "rsnns" achieving the highest specificity (0.972), indicating superior efficiency in recognising negative cases.

Precision is the same for the "nnet", "mlp", and "neuralnet" models (all at 0.583), while "rsnns" achieves a better precision of 0.778, suggesting its capability to avoid false positives. NPV is also higher in the "mlp" (0.972) and "neuralnet" (0.944) models, suggesting that these models are more reliable in predicting negative outcomes.

The balanced accuracy metric, which considers both sensitivity and specificity, shows that the "mlp" model is the most balanced with a value of 0.856, while "neuralnet" (0.784), "rsnns" (0.755), and "nnet" (0.682) are comparatively lower, further confirming the superiority of "mlp" in maintaining a balance between recognising positive and negative cases.

Finally, the F1 score, which reflects the balance between precision and sensitivity, is highest for the "mlp" model (0.667), with "rsnns" following closely at 0.636, indicating that it is also effective in balancing these metrics.

In conclusion, the "mlp" model stands out as the most balanced and accurate model, with superior sensitivity, balanced accuracy, and overall performance. The "rsnns" model excels in specificity and AUC values, making it the most efficient in distinguishing between positive and negative cases. While each model has its strengths and weaknesses, the choice depends on the specific priorities of the task.

Figure 10 shows the ROC graphical representation of the performance of the binary classification models ("nnet", "mlp", "neuralnet", and "rsnns") at different decision thresholds. The primary purpose of the ROC curve is to analyse the model's ability to distinguish between positive and negative instances. This metric is particularly useful when there is an imbalance between the classes, as it allows the evaluation of the trade-off between the correct detection of positive instances and false alarms.

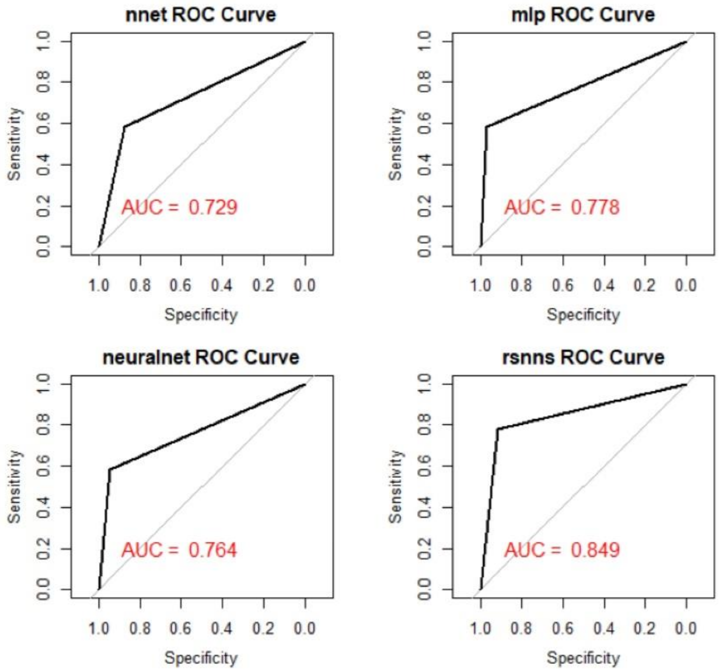

Figure 10. Comparative ROC curve plot for models: “nnet”, “mlp”, “neuralnet”, and “rsnns”.

An important measure derived from the ROC curve is the AUC, which quantifies the model’s overall performance. An AUC value of 0.5 indicates a model with no predictive power, while values between 0.5 and 0.7 indicate a weak or moderate model, values between 0.7 and 0.9 indicate a good model, and values above 0.9 indicate an exceptionally accurate model.

The model “rsnns” achieves the highest AUC value (0.849), indicating its superior ability to distinguish positive and negative cases. It is followed by “mlp” with an AUC value of 0.778, indicating a solid but slightly weaker precision than “rsnns”. The “neuralnet” model achieves an AUC value of 0.764, close to the value of “mlp” but indicates a slightly weaker performance in class separation. “Nnet” has the lowest AUC value (0.729), indicating the lowest efficiency in class separation and a greater tendency to overlap between positive and negative instances. From these results, it can be concluded that the “rsnns” model has the best overall classification ability, while “nnet” shows the weakest discriminatory power among the models analysed.

Relatively low AUC values can be explained by a significant imbalance in the data, with a low percentage of truly detained vessels in the dataset (about 13.3%). When the positive class (detained vessels) represents a tiny proportion of the total dataset, the model tends to favour the negative class to maximise accuracy, which has a negative impact on sensitivity. This data distribution makes it more difficult for the models to recognise the rare detention cases, directly reducing the AUC value.

Based on all metrics presented, the "mlp" model can be considered the most efficient model due to its superior accuracy, balance between sensitivity and specificity, and overall reliability as measured by the kappa coefficient and AUC value.

Figure 11 shows that the "mlp" model has the best accuracy and the best F1 score and is therefore the most balanced model in terms of precision and sensitivity. The model "rsnns" has the best specificity and precision, indicating that it is the most accurate in identifying negative cases. The "nnet" model has a significantly lower sensitivity, suggesting that it misses many positive cases. The Kappa values confirm that all models are sound in terms of fit to the actual data, but "mlp" stands out as the best model in all key metrics.

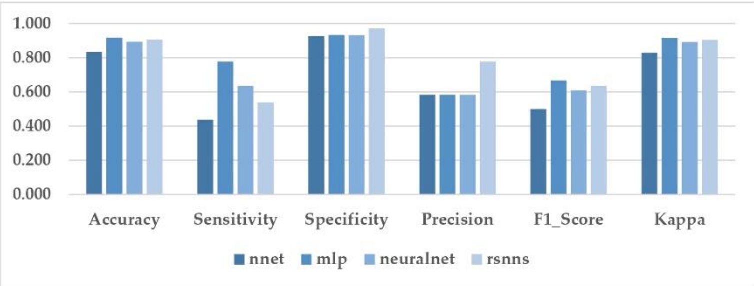

Figure 11. Main performance metrics for models: "nnet", "mlp", "neuralnet", and "rsnns".

The complexity of algorithms for NNs depends on the number of layers, the number of neurons per layer, the number of epochs, and the optimisation method. The package "nnet" in R allows the implementation of simple NNs for classification and regression tasks based on gradient descent for training. Its time complexity is $O\left(L \cdot N^2 \cdot T\right)$, while the required memory is $O(L \cdot N)$. These networks are limited by their simple architecture.

"Mlp" is the most commonly used NN type, consisting of an input layer, one or more hidden layers, and an output layer. Training is performed using the backpropagation algorithm, which has a time complexity of $O\left(L \cdot N^2 \cdot T\right)$ and a memory complexity of $O(L \cdot N)$. This makes "mlp" flexible, but computationally intensive for complex problems.

The "neuralnet" package facilitates the implementation of feedforward networks with one or more hidden layers, which rely on gradient descent for training. Its time complexity remains $O\left(L \cdot N^2 \cdot T\right)$ and the memory complexity is $O(L \cdot N)$. This package is particularly suitable for visualisation and simpler architectures.

The package "rsnns" offers advanced options for implementing different types of networks, including "mlp" and radial-based networks. It uses optimisation techniques such as "quickprop" and "rprop", which can reduce the time complexity, although the fundamental estimation remains the same $O\left(L \cdot N^2 \cdot T\right)$. The memory complexity is $O(L \cdot N)$. However, this package includes better memory management features, improving performance for more extensive networks.

Therefore, all four NN models (Table 17) have similar computational complexity but differ in flexibility, optimisation capabilities, and scalability. More extensive models require significant computational resources while choosing a suitable algorithm and optimisation method can improve performance in practical applications.

Table 17. Time and storage complexity of models: "nnet", "mlp", "neuralnet", and "rsnns".

	Time Complexity	Storage Complexity
nnet	$O(L \cdot N^2 \cdot T)$	$O(L \cdot N)$
mlp	$O(L \cdot N^2 \cdot T)$	$O(L \cdot N)$
neuralnet	$O(L \cdot N^2 \cdot T)$	$O(L \cdot N)$
rsnns	$O(L \cdot N^2 \cdot T)$	$O(L \cdot N)$

The selection of relatively simple ANN models in this study was justified by a combination of computational constraints, domain-specific requirements, and the structure of the dataset. While the DNN and more complex architectures have demonstrated remarkable capabilities in various domains, their advantages must be weighed against practical considerations such as interpretability, training complexity, and resource requirements. In the context of PSC, where regulators need to make transparent and explainable decisions, simpler models offer a more practical solution.

More comprehensive ANN architectures, including deep learning models with multiple hidden layers, are usually characterised by high-dimensional feature spaces or unstructured data formats. However, the dataset used in this study consists of structured tabular data, where conventional ANN models often perform comparably to more complex architectures while being more interpretable. Furthermore, the implementation of highly complex models would require significantly higher computing power, making them less feasible within the operational constraints of the maritime regulatory framework.

While small language models and other advanced deep learning techniques have revolutionised many fields, their application in PSC is still largely unexplored. As their main strength lies in the processing of unstructured text data, their direct integration into structured maritime datasets poses a challenge. However, future research could explore hybrid approaches that combine classical ANN architectures with advanced representation learning techniques to evaluate their potential to improve prediction accuracy while maintaining transparency and efficiency of decision-making.

When implementing NN models, it is important to realise that these models can require significant computational power, especially when working with large datasets. We have explored various strategies to optimise resources and reduce computational costs. In addition, access to high-quality, up-to-date data is essential for the effective application of these models. In our work, we have emphasised the importance of continuous data collection and the development of robust databases that facilitate the training of models with real data.

Another consideration is the regulatory framework, which can limit the implementation of models. Regulatory requirements often demand transparency and explainability, which is a particular challenge when applying deep neural networks in complex environments. In addition, challenges can arise in real-world applications, as the need for fast data processing and real-time model integration can complicate their use. These issues require the integration of additional tools aimed at optimising performance and reducing latency, which is crucial for effective real-time prediction.

7. Conclusions

The application of NN models to predict the detention of vessels is an advanced and innovative approach in the context of PSC. This approach enables more efficient targeting of inspections of vessels with a high risk of non-compliance with international safety and environmental standards. Compared to alternative methods, the NN approach offers better adaptability, higher prediction accuracy and a better ability to model complex, non-linear relationships.

One of the main limitations of using NNs in PSC inspections is the quality and reliability of the data. The data collected from various sources, such as previous inspections, vessel characteristics and other technical parameters, may be incomplete or inaccurate, affecting the model's performance. In addition, the heterogeneity of data from different systems and sources can make it difficult to integrate and analyse.

Social and regulatory complexity also poses a challenge. NN-based models must be harmonised with the legal framework and ethical standards for maritime safety. The use of automated decision-making systems in the context of inspections must be carefully balanced with human oversight to avoid the possibility of errors in the process. One of the limitations of this research relates to the need to train inspectors and other relevant stakeholders in using NN modelling in PSC inspections.

This requires significant investment in training and adapting existing workflows to ensure the efficient and accurate application of these models in real-world conditions. In addition, the limited interpretability of NN models can pose a challenge when explaining or justifying the decisions of inspectors or regulators.

While NNs can achieve high performance in pattern recognition, their "black box" nature can make it challenging to understand the rationale behind certain predictions, which is crucial for safety-related decisions. Therefore, future research needs to focus on improving the interpretability of models through model explanation methods (e.g., LIME (Local Interpretable Model-agnostic Explanations) or SHAP (Shapley Additive Explanations)) that enable a better understanding of modelling decisions. In addition, research into new techniques for dealing with imperfect and unbalanced data is crucial for improving prediction accuracy.

Further research into developing hybrid models that combine NNs with traditional statistical methods can help reduce the risk of error and increase the system's robustness. In addition, models need to be developed that allow for continuous learning and adaptation to new data and regulatory changes, which could significantly improve the long-term applicability of these systems in the context of PSC inspections.

Author Contributions: Conceptualization, Z.B.; methodology, T.S.; validation, I.S. and Z.B.; formal analysis, N.R.; investigation, Z.B. and N.R.; resources, N.R. and T.S.; data curation, Z.B. and N.R.; writing—original draft preparation Z.B.; writing—review and editing, I.S. and Z.B.; visualisation, Z.B. and I.S.; supervision, I.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in Equasis. Available online: https://www.equasis.org/EquasisWeb (accessed on 11 January 2025).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under Curve
BMU	Best Matching Unit

BN	Bayesian Network
CI	Confidence Interval
DL	Deep Learning
DNN	Deep Neural Network
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GT	Gross Tonnage
IMO	International Maritime Organization
LIME	Local Interpretable Model-agnostic Explanations
LOF	Local Outlier Factor
MARPOL	Marine Pollution
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MoU	Memorandum of Understanding
Mlp	Multilayer Perceptron
NA	Not Available
NIR	No Information Rate
NLP	Natural Language Processing
NN	Neural Network
Pe	Expected Agreement
Po	Observed Agreement
PSC	Port State Control
RBF	Radial Basic Function
RBFN	Radial Basis Function Network
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
Rsnns	R Support Vector Machines and Neural Networks
SE	Standard Error
SGD	Stochastic Gradient Descent
SHAP	Shapley Additive Explanations
SOLAS	Safety of Life at Sea
SOM	Self-Organising Map
STCW	Standards of Training, Certification and Watchkeeping
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

References

1. Hashemi, R.R.; Le Blanc, L.A.; Rucks, C.T.; Shearry, A. A neural network for transportation safety modeling. *Expert Syst. Appl.* **1995**, *9*, 247–256. [CrossRef]
2. Fan, L.; Zheng, L.; Luo, M. Effectiveness of port state control inspection using Bayesian network modelling. *Marit. Policy Manag.* **2020**, *49*, 261–278. [CrossRef]
3. Yang, Z.; Wan, C.; Yu, Q.; Yin, J.; Yang, Z. A machine learning-based Bayesian model for predicting the duration of ship detention in PSC inspection. *Transp. Res. Part E Logist. Transp. Rev.* **2023**, *180*, 103331. [CrossRef]
4. Jiang, H.; Chen, Y.; Naiyuan, L.; Zhou, Y. Dynamic Evolution Analysis of Ship's Port State Control Inspection Based on Bayesian Networks. In Proceedings of the 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi'an, China, 4–6 August 2023; p. 10243671. [CrossRef]
5. Chen, X.; Yang, Y.; Wang, S.; Wu, H. Ship Type Recognition via a Coarse-to-Fine Cascaded Convolution Neural Network. *J. Navig.* **2020**, *73*, 813–832. [CrossRef]
6. Lim, J.-J.; Kim, D.-W.; Hong, W.-H.; Kim, M.; Lee, D.-H.; Kim, S.-Y.; Jeong, J.-H. Application of Convolutional Neural Network (CNN) to Recognize Ship Structures. *Sensors* **2022**, *22*, 3824. [CrossRef] [PubMed]

7. Yan, R.; Wang, S. Ship detention prediction using anomaly detection in port state control: Model and explanation. *Electron. Res. Arch.* **2022**, *30*, 3679–3691. [CrossRef]
8. Hou, Z.; Yan, R.; Wang, S. On the K-Means Clustering Model for Performance Enhancement of Port State Control. *J. Mar. Sci. Eng.* **2022**, *10*, 1608. [CrossRef]
9. Gómez Ruiz, M.Á.; de Almeida, I.M.; Pérez Fernández, R. Application of Machine Learning Techniques to the Maritime Industry. *J. Mar. Sci. Eng.* **2023**, *11*, 1820. [CrossRef]
10. Demirci, S.M.E.; Cicek, K. Intelligent ship inspection analytics: Ship deficiency data mining for port state control. *Ocean Eng.* **2023**, *278*, 114232. [CrossRef]
11. IET. Ship inspection by port state control. In *Applications of Machine Learning and Data Analytics Models in Maritime Transportation*; IET: London, UK, 2024; Chapter 2. [CrossRef]
12. Biçen, S.; Celik, M. A Deep Learning Approach to Analyse Ship Inspection Reports via Natural Language Processing Integrated with Artificial Neural Network. *J. Mar. Eng. Technol.* **2024**, *23*, 291–304. [CrossRef]
13. Boko, Z.; Skoko, I.; Sanchez-Varela, Z.; Pincetic, T. Application of Advanced Algorithms in Port State Control for Offshore Vessels Using a Classification Tree and Multi-Criteria Decision-Making. *J. Mar. Sci. Eng.* **2024**, *12*, 1905. [CrossRef]
14. IMO-PSC. Available online: https://www.imo.org/en/ourwork/msas/pages/portstatecontrol.aspx (accessed on 11 January 2025).
15. Paris MoU. Available online: https://parismou.org/PMoU-Procedures/Library/memorandum (accessed on 14 January 2025).
16. Tokyo MoU. Available online: https://www.tokyo-mou.org (accessed on 15 January 2025).
17. Sima, L.; Bucukovski, J.; Carlson, E.; Yien, N.L. Advanced Computing and Related Applications Leveraging Brain-inspired Spiking Neural Networks. *arXiv* **2023**, arXiv:2309.04426.
18. Vasileiadis, A.; Alexandrou, E.; Paschalidou, L.; Chrysanthou, M. Artificial neural network and its applications. *ResearchPDF* **2024**, *1*, 1–15. Available online: https://www.researchgate.net/publication/380122805_Artificial_Neural_Network_and_Its_Applications (accessed on 12 January 2025).
19. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
20. Zhang, J.; Gouza, F.B. GADAM: Genetic-Evolutionary ADAM for Deep Neural Network Optimization. *arXiv* **2018**, arXiv:1805.07500.
21. Kobayashi, Y.; Iiduka, H. Conjugate-gradient-based Adam for stochastic optimization and its application to deep learning. *arXiv* **2020**, arXiv:2003.00231.
22. Xie, X.; Zhou, P.; Li, H.; Lin, Z.; Yan, S. Adan: Adaptive Nesterov Momentum Algorithm for Faster Optimizing Deep Models. *arXiv* **2022**, arXiv:2208.06677. [CrossRef] [PubMed]
23. CRAN. Available online: https://cran.r-project.org/bin/windows/base/ (accessed on 7 January 2025).
24. Nnet. Feed-Forward Neural Networks and Multinomial Log-Linear Models. Available online: https://cran.r-project.org/web/packages/nnet/index.html (accessed on 17 December 2024).
25. DataCamp. Available online: https://campus.datacamp.com/courses/supervised-learning-in-r-regression/tree-based-methods?ex=7 (accessed on 21 January 2025).
26. Mlp. Available online: https://www.rdocumentation.org/packages/RSNNS/versions/0.4-17/topics/mlp (accessed on 8 January 2025).
27. Neuralnet. Available online: https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf (accessed on 11 January 2025).
28. Rsnns. Available online: https://cran.r-project.org/web/packages/RSNNS/index.html (accessed on 8 January 2025).
29. One-Hot-Encoding. Available online: https://www.geeksforgeeks.org/ml-one-hot-encoding/ (accessed on 12 January 2025).
30. Nielsen, M. *Neural Networks and Deep Learning*; Determination Press, 2015; Available online: http://neuralnetworksanddeeplearning.com/ (accessed on 25 February 2025).
31. Equasis. Available online: https://www.equasis.org/EquasisWeb (accessed on 11 January 2025).

BIOGRAPHY

Zlatko Boko serves as a Teaching Assistant at the Faculty of Maritime Studies, University of Split, where he is currently pursuing a PhD. He has more than twenty-five years of professional experience in the offshore maritime industry, having served as Master and Deck Officer on various offshore support vessels, including platform supply vessels and anchor-handling ships.

His research interests include maritime safety analytics and port state control inspection systems, with a focus on the application of machine learning, explainable artificial intelligence, and multi-criteria decision-making methods. He has participated in several international projects related to offshore operations and the development of the blue economy.

He is the author and co-author of scientific papers published in peer-reviewed journals and conference proceedings. His doctoral thesis addresses predictive modelling and risk-based approaches to offshore maritime safety and inspection systems.