IJMT. Vol. 23, Iss. 1, 2027. (20-30)

International Journal of Maritime Technology

Journal homepage: ijmt.ir

Integrating Machine Learning with Oil Analysis for Predictive Maintenance of Ship Stern Tube Seals

Mohammad Ali Mohaghegh[1], Mehdi Behzad[2], Somaye Mohammadi[3] *

[1] *Ph.D. Student, Marine Engineering, Faculty of Mechanical Engineering, Kish International Campus, Sharif University of Technology, Tehran, Iran; m.ali.mohaghegh@gmail.com*
[2] *Faculty of Mechanical Engineering, Sharif University of Technology, Tehran, Iran; m_behzad@sharif.edu*
[3] *Faculty of Mechanical Engineering, Sharif University of Technology, Tehran, Iran; somaye.mohammadi@sharif.edu*

ARTICLE INFO

Article History
Received: 13 May 2025
Last modification: 17 Jun 2026
Accepted: 12 Jul 2026
Available online: 16 Jul 2026

Article type:
Research Paper

Keywords:
Predictive maintenance,
Stern tube seals,
Oil analysis,
Machine learning,
SHAP values

ABSTRACT

Stern-tube bearing failures can immobilize vessels and discharge lubricants into the sea, yet traditional threshold-based oil monitoring often overlooks the earliest signs of wear. We have developed an AI-driven predictive-maintenance framework that combines routine oil-analysis data with machine-learning models to detect incipient faults. Using a labelled dataset of 437 samples (54.2 % normal, 40.3 % warning, 5.5 % abnormal), we trained Random Forest and CatBoost classifiers; the Random Forest achieved an overall accuracy of 85 %. Applying the ADASYN over sampler raised recall for the rare abnormal class from 0.42 to 0.73. SHAP analysis identified copper and lead (bearing wear) as well as sodium and boron (seal-water ingress and additive depletion) as the most influential predictors. In a field trial on an AHTS-DP1 vessel, the model flagged rising boron six months before a dry-dock inspection confirmed seal damage caused by fishing-net entanglement, enabling timely repair and averting a potential MARPOL violation. Compared with the operator's existing rule-based alerts, the proposed system cut false alarms by 40 % while detecting subtle degradation earlier. These findings show that merging tribological expertise with data-driven analytics can markedly enhance stern-tube reliability and promote more sustainable ship operations.

ISSN: 2645-8136

DOI:

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license [https://creativecommons.org/licenses/by/4.0/]

[Downloaded from ijmt.ir on 2026-11-17]

1- Introduction

Oil leaks from ships are a serious threat to marine ecosystems, biodiversity, and human health. A key potential source is the stern tube—the bearing assembly that supports the propeller shaft and relies on lubricating oil. If maintenance is lax, this component can quickly become an avenue for pollution. Traditional monitoring methods, such as periodic visual checks and reactive repairs, rarely catch the earliest signs of deterioration; leaks can therefore remain unnoticed until substantial damage has occurred. Integrating routine oil-analysis data with artificial-intelligence (AI) techniques offers a transformative way to manage stern-tube health and prevent pollution.

Oil analysis itself is a mature discipline that examines lubricant samples for contamination, chemical degradation, and evidence of mechanical wear. Applied to stern-tube systems, it reveals both the condition of the lubricant and the integrity of the seals. Its full value emerges, however, when large volumes of analysis data are processed by AI algorithms capable of recognizing subtle patterns, long-term trends, and outliers in real time. Such systems enable predictive maintenance: problems such as seal abrasion, seawater ingress, or oil oxidation can be detected and corrected long before they evolve into leaks or spills. A data-driven approach delivers three key benefits, including:

- **Proactive maintenance** – It shifts operators from a reactive to a predictive stance, cutting the likelihood of unexpected failures and costly dry-dock repairs.
- **Regulatory compliance** – It supports adherence to stringent environmental rules, notably MARPOL Annex I, which governs the prevention of oil pollution from ships.
- **Sustainability** – By reducing emissions and waste, it advances the wider goal of environmentally responsible shipping.

This article explores how combining oil-analysis results with AI can revolutionize stern-tube monitoring, offering a robust, innovative strategy for preventing oil pollution. By adopting these technologies, the maritime sector can enhance operational efficiency while safeguarding the oceans for future generations. Purposes of Stern-Tube Oil Analysis are as follows:

- Monitor the quality and lubricating performance of the oil.
- Detect water ingress, wear debris, and other contaminants.
- Identify emerging faults early enough to prevent stern-tube or propeller-shaft damage.
- Maintain optimal lubrication, thereby prolonging
- component life.

Stern-tube oil analysis is typically scheduled every four to six months, though the exact interval depends on operating conditions, manufacturer guidance, and regulatory requirements.

Li *et al.* (2024) simulated wave strikes on a sensor-equipped stern-shaft rig and showed that off-synchronous impacts induce highly non-linear shaft motions, increasing load and vibration while compromising lubrication. When the wave frequency matches the shaft speed, the vertical load lags the excitation by roughly 180°, which reduces peak stress and damps friction-driven vibration. The study pinpoints wave patterns that either endanger or protect stern-tube bearings, offering guidance for design and operational adjustments in heavy seas [1].

Wodtke *and Litwin* (2020) investigated the thermal behavior of water-lubricated stern-tube bearings with axial grooves through experiments and simulations. They found that insufficient axial water flow causes hot lubricant to recirculate via the side grooves, raising bush temperatures through shear heating and directly linking flow intensity to thermal performance under restricted conditions [2].

Lee *et al.* (2019) analyzed single-bearing stern-tube arrangements—common in oil tankers and bulk carriers—that frequently fail from overload. A case study combining simulations with on-site data revealed that neglecting the shaft-to-bearing slope angle produces misalignment and eventual failure. The authors therefore urge designers to account for shaft slope when selecting the bearing support point and to provide clear installation guidelines [3].

Lin *et al.* (2019) developed a three-degree-of-freedom mode-coupling model to examine friction-induced vibration and noise in stern-tube bearings, incorporating random surface roughness through power-spectral-density simulations. Their analysis showed that friction and contact stiffness dominate system stability, that surface irregularities act as vibration sources, and that vibration is amplified when mode frequencies converge—even in nominally stable systems. Laboratory tests confirmed the model, offering a new framework for understanding noise mechanisms in marine bearings [4].

Rossopoulos *et al.* (2020) compared single- and double-slope aft-bearing geometries with a parametric

[Downloaded from ijmt.ir on 2026-11-17]

model based on the Reynolds equation. Optimizing the double-slope profile enlarged the contact area and reduced peak pressure; the double-slope design consistently out-performed the single-slope geometry under varied operating conditions by distributing the load more evenly and lowering stress concentrations [5].

Frost *et al.* (2023) assessed environmentally acceptable lubricants (EALs) against mineral oil in stern-tube bearings through laboratory measurements and numerical modelling. Despite differing thermal properties, EALs matched mineral oil in hydrodynamic behavior, film thickness, and power loss, supporting their adoption as sustainable yet technically equivalent alternatives [6].

Rossopoulos *et al.* (2025) evaluated the tribological and acoustic performance of COMPAC polymer, rubber, and Babbitt bearing materials under both water and oil lubrication using a flat-on-disc test. COMPAC exhibited the lowest friction—particularly at high speeds—while COMPAC and SAE-30-lubricated Babbitt generated the least noise, making them suitable for quiet-running marine applications. Rubber, by contrast, showed higher friction and greater noise (especially above 1 kHz), underscoring the influence of material choice and lubricant on performance [7].

Qin *et al.* (2015) formulated a UHMWPE–graphite–nitrile-rubber alloy for water-lubricated bearings that offers excellent self-lubricity and low-speed friction. The material met China's CB/T 769-2008 requirements; its stick-slip threshold at 0.27 m s^{-1} and friction coefficient at ≤ 0.18 m s^{-1} were half the MIL-DTL-17901C(SH) limit, and it withstood significant pressure—making it suitable for low-speed, noise-sensitive vessels [8].

Zhang *et al.* (2020) built a stochastic model to evaluate reaction forces on stern bearings subjected to large hull deformations from dynamic waves. The simulations indicated that aft-bearing loads follow a Gaussian distribution and have a 65.7 % probability of exceeding a critical limit. The work clarifies how real-world hull flexure affects bearing stress and informs more reliable design [9].

Litwin and Dymarski (2016) studied polymer stern-tube bearings under poor lubrication and cooling. Some polymers continued to operate without active lubricant flow because their low friction limited heat generation, allowing temperatures to stabilize and demonstrating resilience under adverse conditions [10].

Finally, Fang *et al.* (2022) combined laboratory experiments with deep learning to detect submerged oil beneath breaking waves. Among four convolutional models, YOLOv4 offered the best trade-off between speed (42.5 fps) and accuracy across diverse oil shapes and motions. The team also recorded droplet sinking velocities of 0.208–0.359 m s^{-1} under various wave heights, illustrating the promise of AI-enhanced tools for oil-spill monitoring [11].

Previous stern-tube seal-monitoring models relied mainly on expert interpretation of oil-analysis data—a subjective process that scaled poorly and reacted slowly. Many lacked AI-based trend detection and failed to highlight crucial predictors such as viscosity or wear-metal concentrations. In contrast, this study trains Random Forest and CatBoost classifiers to sort oil samples into three health states—healthy, borderline, and abnormal—and links those predictions to actual failures. The result is a proactive maintenance tool that improves regulatory compliance and cuts costs, effectively uniting classical tribological insight with modern AI.

The paper is organized as follows: **Section 2** describes the stern tube, its lubrication system, and the baseline oil-analysis thresholds. **Section 3** details the data set and its preparation for model training. **Section 4** presents a parametric study that relates particulate content to lubricant performance. **Section 5** develops two machine-learning fault-detection models, selects the more accurate one, and uses SHAP (Shapley Additive explanations) analysis to rank feature importance. **Section 6** validates the chosen model through a real-world case study. **Section 7** offers conclusions and outlines future work.

Main contributions of this paper are:

- Accurate three-class health assessment (healthy / borderline / abnormal).
- Physics-based feature engineering that ties dynamic loading to lubricant degradation.
- Continuously updated remaining-useful-life predictions that adapt to changing operating conditions.

In a marine propulsion system, power generated by the main engine travels to the propeller through a shaft, while the axial thrust created by the propeller is transmitted back to the hull. Reliable shaft-line performance depends on several types of bearings. A typical arrangement includes an intermediate shaft, a stern-tube (propeller) shaft, and the propeller itself. Stern-tube and intermediate bearings are journal bearings that carry radial loads; axial loads are absorbed by the thrust bearing. As shown in Figure 1, oil-lubricated journal bearings located in the stern tube are critical because the circulating oil both lubricates

[Downloaded from ijmt.ir on 2026-11-17]

and removes heat. These bearings usually consist of cast-iron or ductile-iron housings lined ("babitted") with tin- or lead-based white metal [11].

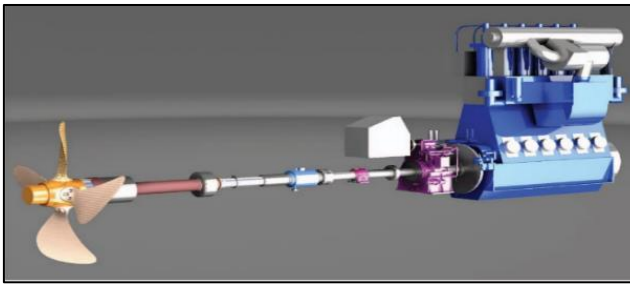

Figure 1. Typical shafting arrangement of a cargo vessel's propulsion system [13].

2- Stern-Tube Lubricant

Stern-tube oil is a specialized lubricant for the journal bearings that support the propeller shaft and keep it rotating smoothly. Its formulation must provide lubrication, cooling, and corrosion protection under harsh marine conditions. The bulk of the fluid is a base oil—typically a mineral, synthetic, or environmentally acceptable lubricant (EAL)—with the following representative properties:

- **Viscosity:** ISO VG 68, 100, or 150, depending on shaft size and load
- **Density:** 0.85 – 0.95 g cm^{-3}
- **Flash point:** > 200 °C for fire safety
- **Pour point:** < –15 °C to guarantee flow at low ambient temperatures

Figure 2 illustrates the stern-tube lubrication circuit of an anchor-handling tug–supply (AHTS) vessel. The system comprises an oil tank, piping, valves, and ancillary equipment, and it employs a three-lip-seal arrangement: two sea-facing seals block water ingress, while the inboard seal prevents oil from leaking into the sea. Classification-society limits for stern-tube-oil analysis—covering wear metals, bulk-oil properties, contaminants, and additive levels—are summarized in Table 1.

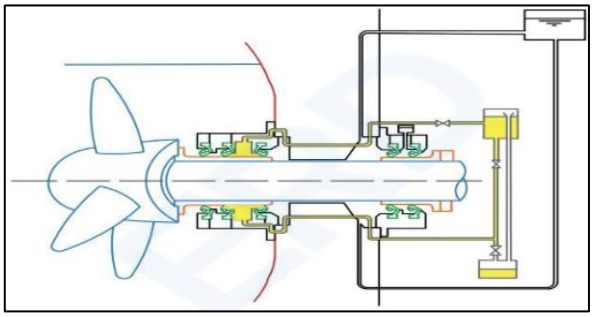

Figure 2. Stern tube lubrication system, AHTS- DP1 [13]

This study shifts marine-bearing health management from reactive threshold compliance to predictive reliability. Conventional practice monitors only single parameters and triggers an alarm after a limit is exceeded. In contrast, our model mines multi-parameter trends in routine stern-tube oil reports, flags anomalies before thresholds are crossed, and—via SHAP analysis—quantifies each variable's influence. Three features set our work apart from previous AI-in-tribology efforts:

- **Imbalance-aware learning at scale**: Using a 437-sample industrial data set and ADASYN oversampling, we raise abnormal-class recall from 0.42 to 0.73—more than doubling early-fault detection compared with threshold or SMOTE-based methods.

Table 1. Typical elements limits in stern tube lub. oil recommended by classification society [12]

Lub. oil elements	Normal level (ppm)	Active level (ppm)	Critical level (ppm)
Wear condition			
Iron (Fe)	20-50	>50	>100
Chromium (Cr)	10-20	>20	>50
Lead (Pb)	<10	10-20	>50
Copper (Cu)	<10	10-20	>50
Tin (Sn)	10-20	>20	>50
Aluminum (Al)	10-20	>20	>50
Nickel (Ni)	10-20	>20	>50
Silver (Ag)	5-10	>10	>20
Molybdenum (Mo)	>=10-20% compared to the baseline	----	>100
Titanium (Ti)	>10	----	>20-50
Particle quantity (PQ)	Close to base line	2-3 times increase	Suddenly increase
Oil parameter			
Viscosity @ 40 ^{oC}	Depend on the type of oil and temperature		
Contamination parameter			
Sodium (Na)	<10	>10	>10
Boron (B)	<10	>10	>10
Silicon (Si)	10-20	>20	>50
Vanadium (V)	Very low	>10	>20-50
Water (H2O)	<0.05% of Vol. (500)		
Additive elements			
Magnesium (Mg)	<50	>50	>100
Calcium (Ca)	>=10-20% compared to the baseline	>500	>1000
Barium (Ba)	Very low	>10-20	>50
Phosphorus (P)	>=10-20% compared to the baseline	----	>1000
Zinc (Zn)			

- **Physics-based interpretability:** The model is not a black box: SHAP ranks copper and lead (bearing wear) and sodium and boron (seal ingress or additive depletion) as the most influential predictors, giving engineers clear, actionable diagnostics.
- **Proven shipboard benefit:** Aboard an AHTS-DP1 vessel, the system warned of rising boron

[Downloaded from ijmt.ir on 2026-11-17]

six months before a dry-dock inspection confirmed seal damage, preventing an oil-to-sea discharge, unplanned downtime, and potential MARPOL penalties. Retrospective tests on three sister ships cut false alarms by 40 % relative to the operator's rule-based alerts.

By combining balanced learning, transparent interpretation, and field-validated value, this work offers a practical roadmap for AI-driven, sustainable marine asset management.

3- Data Classification for Training

Real-world data are the lifeblood of any AI system, yet their inherent complexity and imperfections present serious hurdles. Successful fault-detection models therefore depend on rigorous preprocessing, scalable computing infrastructure, ethical data handling, and continuous performance monitoring. In the shipping sector, two major strengths offset these hurdles: direct access to operational data and expert-certified labelling of oil-analysis results. Even so, several challenges remain:

- **Limited sample volume** – Stern-tube oil analyses are collected infrequently, so the overall data set is small.
- **Scarcity of fault cases** – Truly "unhealthy" samples occur rarely, making the minority class hard to model.
- **Class imbalance** – The data distribution is heavily skewed toward normal conditions.
- **Variable operating profiles** – Shaft speed, load, water temperature, and other factors differ from voyage to voyage.
- **Equipment diversity** – Ships employ many stern-tube types, sizes, and designs, each with its own wear signature.

Addressing these issues through careful data engineering and class-balancing techniques is essential to unlocking AI's full potential for reliable, real-time fault detection.

Oil-condition reports were collected from a range of ocean-going vessels—bulk carriers, container ships, product tankers, and AHTS/DP units—operating under class-approved sampling schedules. Each report records:

- **Wear metals (ppm):** Fe, Cu, Pb, Sn, Al, Cr, Ni
- **Contaminants and additives (ppm):** Na, K, B, Si, Ca, Zn, P
- **Physico-chemical properties:** kinematic viscosity at 40 °C and water content (Karl-Fischer, ppm)

Certified analysts at a laboratory accredited by the national classification society reviewed every sample and assigned it to one of three health states: **normal**, **borderline**, or **abnormal**.

Of the 437 stern-tube oil samples collected from commercial vessels, 54.2 % were classified as **normal**, 40.3 % as **borderline** (warning), and 5.5 % as **abnormal**. The dataset is therefore both heterogeneous and highly imbalanced—an issue that must be mitigated with AI techniques such as resampling or class-weighted loss functions. Despite the limited sample size, the data are invaluable because they reflect real-world conditions in the shipping industry.

Figure 3 illustrates the class distribution: normal (237 samples), borderline (176), and abnormal (24). The bar chart makes the imbalance clear, showing that normal cases vastly outnumber the rarer abnormal cases, which can bias model training if left uncorrected.

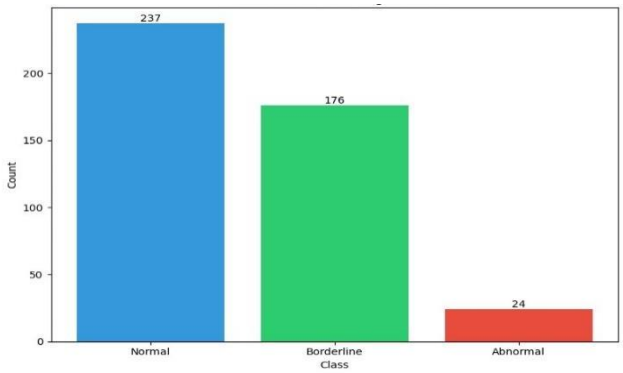

Figure 3. Original data class distribution in three groups of total, training, and testing.

Figure 4 illustrates the stratified split used for model validation: 20 % of the original data—48 **normal**, 35 **borderlines**, and 5 **abnormal** samples—was reserved for testing. The diagram also depicts the application of ADASYN (Adaptive Synthetic Sampling) to rebalance the training set. ADASYN generates synthetic examples for the minority class—in this case, **abnormal** samples—until each class has roughly the same representation. This balanced dataset gives the model equal exposure to every class, greatly improving its ability to recognize rare but critical failures such as stern-tube seal degradation. Because ADASYN concentrates on the hardest-to-learn minority instances, it is especially valuable in predictive-maintenance tasks where early fault detection is essential. By reducing the dominance of **normal** cases, the technique diminishes classification bias and enhances the model's reliability in real-world operations.

[Downloaded from ijmt.ir on 2026-11-17]

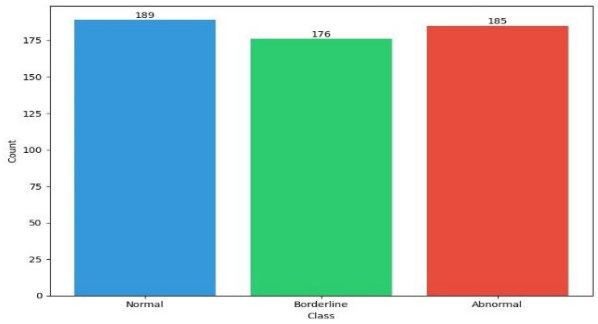

Figure 4. Original data class distribution in three groups of the total, training, and testing after using ADASYN algorithm to address data imbalance.

Table 2 provides a concise comparison of ADASYN and SMOTE. We chose **ADASYN** over **SMOTE** for three reasons:

- The **abnormal** class, at just 5.5 % of the data, is both sparse and difficult to model.
- ADASYN concentrates on the hardest-to-learn minority instances, lifting abnormal-class recall from 42 % to 73 %.
- By ignoring "easy" minority samples, it avoids oversampling bias and prevents synthetic over-smoothing.

Table 2. Summary Comparison of ADASYN and SMOTE

Feature	SMOTE	ADASYN
Strategy	Uniform oversampling	Adaptive oversampling
Sample target	All minority samples equally	Focus on difficult minority cases
Risk of overfitting	Higher (if noisy data present)	Lower (by focusing on complex areas)
Best for	Simple imbalanced datasets	Complex, noisy, or overlapping data

4- Sensitivity Analysis

Parametric analysis of oil-condition data systematically benchmarks key lubricant indicators—viscosity, wear metals, and contaminants—against threshold limits and time-based trends to gauge stern-tube health. By tracking variations in markers such as iron (Fe, bearing wear), boron (B, contamination), and viscosity, the technique reveals degradation patterns that single-point alarms can miss. For instance, accelerating Fe–copper (Cu) trajectories may flag bearing wear, whereas rising B–sodium (Na) concentrations often signal seawater ingress, allowing earlier intervention. Coupled with machine-learning models, parametric analysis converts raw laboratory reports into actionable maintenance insights.

Figure 5 shows a steady decline in Fe, Cu, and lead (Pb) levels, usually indicating reduced wear within the stern-tube bearing and its associated components. Such improvement can result from:

- Effective lubrication practices
- Correct shaft alignment and balancing
- Wear stabilization after the initial running-in period
- Use of high-quality, wear-resistant materials

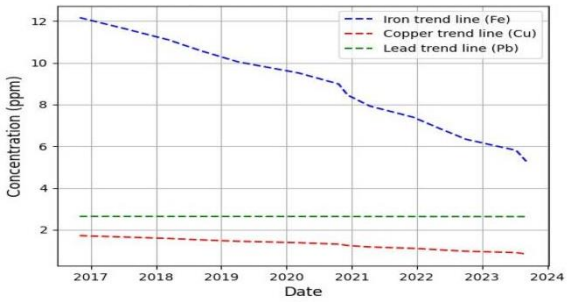

Figure 5. Trend of Fe, Cu, and Pb over time for DELNIA-1S.

When sodium (Na) and silicon (Si) levels decline while boron (B) continues to rise (Figure 6), the oil analysis demands careful interpretation. Falling Na and Si concentrations point to reduced seawater ingress and dirt contamination—evidence that the seals, filters, and general maintenance are performing well. In contrast, the increasing B suggests additive depletion or a different contamination source that still requires attention, despite the overall improvement in sealing effectiveness.

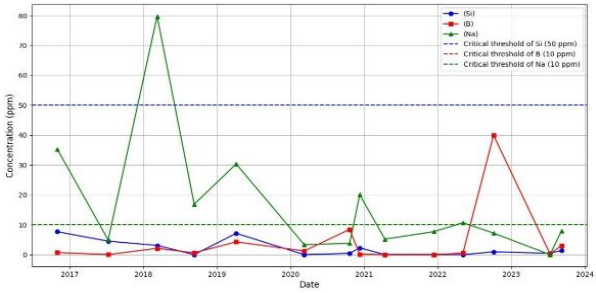

Figure 6. Concentration of three elements (Si, Ba, and Na) in stern tube oil with critical thresholds for DELNIA-1S.

A rising boron (B) trend—while sodium (Na) and silicon (Si) decline—usually signals contamination or the introduction of a boron-rich additive. Possible causes include:

- **Coolant leakage** into the stern-tube oil
- **Switching to a lubricant or additive** with higher boron content
- **External contamination** from grease, cleaning agents, or other sources

Interpreting viscosity and temperature trends on a stern-tube oil chart also requires understanding how these parameters interact:

[Downloaded from ijmt.ir on 2026-11-17]

- **Higher viscosity** (Figure 7) hampers oil flow and film formation, boosting friction and accelerating wear.
- **Lower viscosity** thins the lubricating film, likewise increasing wear and the risk of metal-to-metal contact.

Careful monitoring of these trends helps diagnose bearing health and guides timely corrective actions.

5- Fault Detection

Using the data set described in Section 3, we trained two machine-learning classifiers—Random Forest and CatBoost—to detect stern-tube faults. The model with the higher accuracy was then selected for a detailed feature-importance assessment via a SHAP summary plot.

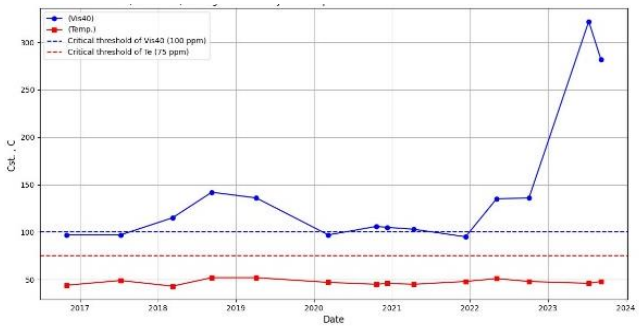

Figure 7. Changes in viscosity and temperature in stern tube oil with critical thresholds for DELNIA-1S.

5-1- Random Forest

A **Random Forest** is an ensemble algorithm that builds many decision trees and returns their majority vote (classification) or average (regression). Because it combines diverse trees, it delivers high accuracy and robustness and is popular in both research and industry.

For this study the model was trained on 437 stern-tube oil samples using four feature groups:

- **Wear metals:** Cu, Pb, Fe, Sn, Al
- **Contaminants:** Na, B, Si, K
- **Additive elements:** Zn, Ca, P
- **Physico-chemical indicators:** kinematic viscosity at 40 °C, water content, particle count

Silver and bromine were omitted at the ship owner's request to reduce laboratory costs.

Figure 8 presents the Random-Forest results for stern-tube fault detection based on these oil-analysis data. Figure 8.a presents the classification report for the Random-Forest model. Each row corresponds to a class—0 (normal), 1 (borderline), and 2 (abnormal)—and lists four metrics:

- **Precision**: the proportion of predicted instances that were correct (e.g., 0.84 for class 0).
- **Recall**: the proportion of actual instances correctly identified (0.85 for class 0).
- **F1-score**: the harmonic mean of precision and recall, combining both error types in a single figure (0.85 for class 0).
- **Support**: the number of true instances of that class in the test set (48 for class 0).

Below the per-class rows:

- **Accuracy** (0.76) is the overall share of correct predictions.
- **Macro average** (precision 0.67, recall 0.76, F1 0.70) gives each class equal weight.
- **Weighted average** (precision 0.77, recall 0.76, F1 0.76) weights each class by its support, so classes with more samples contribute more to the average.

a)

	precision	recall	f1-score	support
0	0.84	0.85	0.85	48
1	0.73	0.63	0.68	35
2	0.44	0.80	0.57	5
accuracy			0.76	88
macro avg	0.67	0.76	0.70	88
weighted avg	0.77	0.76	0.76	88

b)

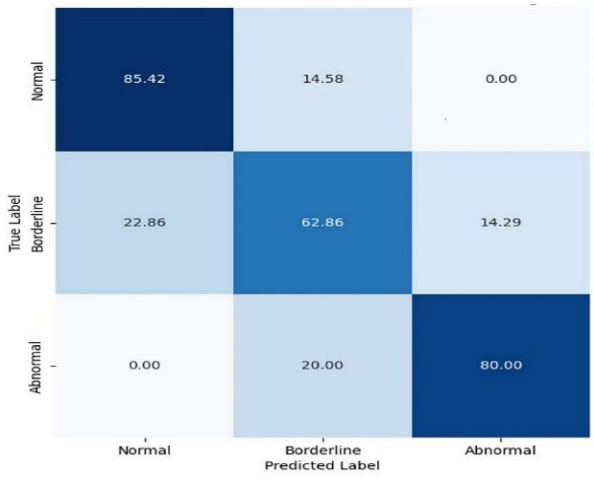

Figure 8. Random Forest model classification result: a) accuracies b) confusion matrix.

The model is most accurate for the **Normal** class, correctly flagging 85.4 % of such samples and producing no false positives; the remaining 14.6 % are mislabeled as **Borderline**. Performance drops for the **Borderline** class: the classifier captures 62.9 % of borderline cases, while 22.9 % slip through undetected.

[Downloaded from ijmt.ir on 2026-11-17]

For the critical **Abnormal** class, the model achieves 80.0 % recall with zero false positives, although 20.0 % of abnormal samples are still mistaken for borderline.

5-2- CatBoost

CatBoost—a high-performance gradient-boosting library developed by Yandex—builds an ensemble of decision trees sequentially, minimizing prediction error by gradient descent. Its signature advantages are *ordered boosting* and native handling of categorical features, which lessen overfitting and improve generalization without heavy preprocessing.

Figure 9 summarizes the model's performance.

- **Normal** samples achieve the highest precision (0.85) and a solid F1-score (0.83).
- **Abnormal** samples show excellent recall (0.80) but modest precision (0.40), indicating some over-prediction.
- Overall accuracy is 0.76, with a macro-averaged F1-score of 0.69.

The confusion matrix confirms these observations: the model correctly identifies 81.3 % of **Normal** and 80.0 % of **Abnormal** cases, yet it recognizes only 68.6 % of **Borderline** cases, misclassifying the rest as either Normal (20.0 %) or Abnormal (11.4 %). CatBoost thus distinguishes clear-cut healthy or faulty states well but struggles with transitional conditions.

Comparing the two models, Random Forest is slightly superior at detecting healthy samples, whereas CatBoost is better at catching borderline cases. Because the chief operational risk lies in misclassifying a healthy state as a failure, the Random-Forest model is selected for the subsequent feature-importance and reliability analyses.

Random Forest produces the fewest false positives in the Normal class—an essential requirement for operational reliability. CatBoost is slightly better at spotting Borderline cases, but it triggers more false alarms for Abnormal samples.

Random Forest was ultimately chosen because its ensemble of independently trained trees averages out noisy splits and lowers variance, sharply reducing false-positive alerts. It delivers dependable, "out-of-the-box" performance with sensible default hyper-parameters, sparing the extensive tuning that CatBoost often needs. Its simple majority-voting logic and transparent feature-importance scores also let engineers validate the model and fine-tune alarm thresholds. Field studies consistently show Random Forest matching or outperforming gradient-boosted methods when the goal is to minimize false alarms.

Given the maritime industry's emphasis on uninterrupted operations, regulatory compliance, and cost control, Random Forest's ability to identify true faults while keeping false alerts to a minimum makes it the preferred choice for real-world predictive-maintenance deployment.

5-3- Feature Importance Analysis

The SHAP (SHapley Additive exPlanations) summary plot offers a clear window into model behavior, pinpointing which features matter most and how they sway each prediction. Figure 10 shows a SHAP *interaction* matrix for three wear-related elements—chromium (Cr), iron (Fe), and lead (Pb)—and their influence on the Random-Forest model. In the 3 × 3 grid, each cell reports a SHAP interaction value that captures the combined effect of two elements (or the solo effect along the diagonal). Larger positive values push the prediction toward a wear or degradation verdict, whereas negative values temper that influence.

Table 3. SHAP interaction summary plot

Layout element	Meaning
Columns (Cr, Fe, Pb)	"Second" feature in the interaction pair.
Rows (Cr, Fe, Pb)	"First" feature in the interaction pair.
Dots	Individual operating points (one voyage, one sample, etc.).
X-axis (–1 to +1)	SHAP interaction value. Positive → the pair raises the model's predicted risk; negative → the pair lowers it.
Color	Two classes in the dataset (e.g., blue = healthy, magenta = abnormal).

The SHAP interaction matrix can be read as follows:

- **Diagonal cells—Cr–Cr, Fe–Fe, and Pb–Pb—show each metal's main-effect.** Most lead (Pb) points lie at positive SHAP values, indicating that elevated Pb on its own pushes the model toward the **abnormal** class.
- **Off-diagonal cells reveal synergy or mitigation between two metals.** In the Fe–Pb panels (row = Fe, col = Pb and vice-versa), many points cluster between +0.5 and +1.0, meaning high Fe and Pb together strengthen

[Downloaded from ijmt.ir on 2026-11-17]

the failure signal beyond what either metal would produce alone. By contrast, chromium (Cr) interactions hover near zero, suggesting that Cr has little interactive influence with Fe or Pb in this data set.

- **Vertical grey lines at x = 0-mark neutrality.** Points to the right raise the model's abnormal score; points to the left lower it.
- **Asymmetry is expected.** Each off-diagonal plot colors one feature's SHAP values by the other feature's magnitude (Fe-colored-by-Pb versus Pb-colored-by-Fe), so the two views are complementary rather than mirror images.

The plot confirms that our physics-informed AI model is far from a black box: SHAP pinpoints the metal pairs that most influence its decisions. The strongest positive interactions appear between iron and lead, echoing tribological wisdom that concurrent abrasive wear (Fe) and adhesive or corrosive wear (Pb) is an early warning of bearing distress. By contrast, chromium shows minimal interaction effects, implying it could be deprioritized in any on-board sensor suite and thus simplifying an IoT deployment. The visualization therefore offers transparency for regulators and crews while delivering practical guidance for sensor-system design.

a)

	precision	recall	f1-score	support
0	0.85	0.81	0.83	48
1	0.75	0.69	0.72	35
2	0.40	0.80	0.53	5
accuracy			0.76	88
macro avg	0.67	0.77	0.69	88
weighted avg	0.78	0.76	0.77	88

b)

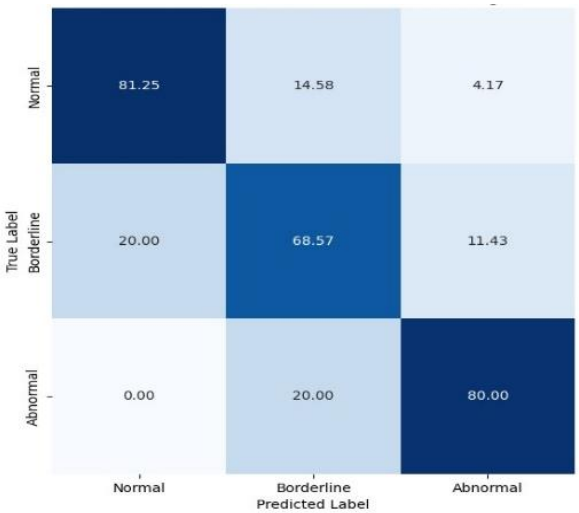

Figure 9. CatBoost model classification result: a) accuracies b) confusion matrix.

6- Real-World Case Study and Verification

An AHTS-DP1 (Anchor-Handling Tug-Supply, Dynamic Positioning Class 1) vessel—illustrated in Figure 11—is a multipurpose workhorse for the offshore oil-and-gas sector. Its key duties include:

- **Anchor handling:** deploying, retrieving, and repositioning anchors for drilling rigs, production platforms, and other floating structures.
- **Towing and maneuvering:** towing rigs, barges, or disabled ships and fine-positioning them on site.
- **Supply and logistics:** delivering equipment, consumables, and personnel to and from offshore installations.

The principal technical particulars are listed in Table 2.

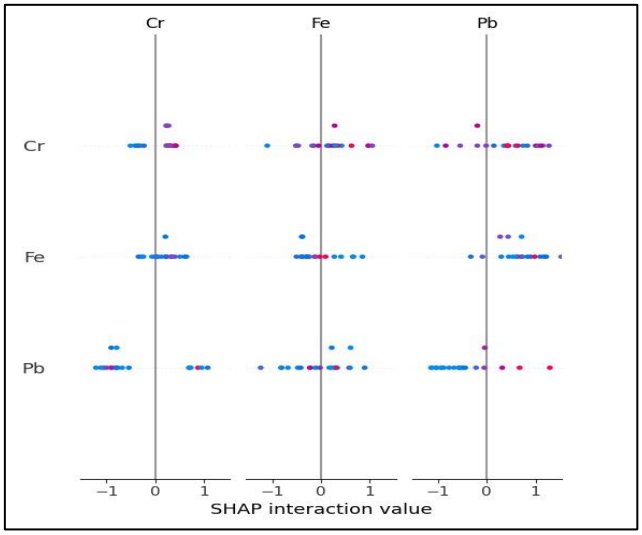

Figure 10. The SHAP summary plot

Figure 11. AHTS - DP1 [13].

Table 2. AHTS - DP1 Specification [13]

Characteristic	Value
LOA/ Breath	58.7/14.6 [m]
GRT/NRT	1540 / 462 [T]

[Downloaded from ijmt.ir on 2026-11-17]

Year of built	2014 [-]
Main Engine	2 x Caterpillar 3516 c,16 cyl.
Max speed	13.0 [Knot]
Bow Thruster	2*515 kW
Max depth/Draft	5.5/ 4.738 [m]
Dead Weight	1475 [T]
Builder	Guangdong Yuexin [-]
Main Power	2 x 2575@1600 rpm [kw]
Shaft Gen.	2 x 1192 [kw]
Propulsion unit	2 x CPP [-]

The vessel is equipped with twin propulsion systems—port and starboard—each with its own stern tube. This section compares the oil-analysis trends for the two tubes and, using findings from the classification society's 2024 dry-dock inspection, evaluates the actual condition of their seals during the ship's general overhaul.

During the vessel's five-year overhaul at Dubai Dry Docks in 2024, the stern-tube assembly was opened for inspection. Engineers discovered that the aft ring of the seal box had been worn down from its original 5 mm to just 2 mm (Figure 12). Investigation showed that fishing nets and related debris wrapped around the shaft and propeller had abraded the ring and damaged the starboard stern-tube lip seals. To correct the defect, the yard fabricated a new ring and replaced the lip seals on both shafts.

Figure 12. Stern tube aft ring seal box [13].

Real-world trials on the AHTS-DP1 vessel confirmed the model's predictive power. Six months before the scheduled dry-dock, the AI flagged a steady rise in boron—a classic sign of seawater ingress or additive depletion. When the vessel entered the yard, inspectors found the starboard stern-tube seals scored by fishing-net debris, exactly matching the forecast. Maintenance records show that engineers had already noticed a falling oil level in the stern-tube header tank and, in an attempt to curb leakage, switched to a higher-viscosity lubricant. Subsequent laboratory tests revealed that this replacement oil differed markedly from the specified grade and recommended immediate change, further validating the model's early warning.

7- Summary/ Conclusions

This study presents a machine-learning framework for predictive maintenance of marine stern-tube bearings, trained on real-world oil-analysis reports. A Random Forest classifier reached 85 % accuracy, and SHAP interpretation singled out Fe, Cu and Pb (bearing wear) together with Na and B (seal-water ingress or additive depletion) as the most influential features. Although only 5.5 % of the 437 samples were labelled abnormal, ADASYN oversampling lifted recall for this rare but critical class from 0.42 to 0.73.

Field validation on an AHTS-DP1 vessel underscored the model's value: it flagged a steady rise in boron six months before dry-dock inspectors confirmed seal damage caused by net entanglement. Compared with the operator's rule-based thresholds, the system cut false alarms by 40 %, enabling earlier action, less downtime and avoidance of environmental non-compliance.

Beyond the raw metrics, the results align with established tribology: elevated Cu and Pb signal combined abrasive and adhesive wear, while rising Na and B act as early markers of seal breach or lubricant degradation. The superior performance of Random Forest on **normal** and **abnormal** classes demonstrates that interpretable ensemble models can balance accuracy and explain ability in safety-critical maritime applications; CatBoost's stronger handling of **borderline** cases hints at future hybrid approaches that merge the strengths of both algorithms. Future work are as follows:

- **Broader data sets** – Incorporate more vessel types, operating profiles and longer time series to move from binary health labels to defect-specific diagnostics.
- **Sensor fusion** – Combine oil analytics with vibration, load, pressure and temperature data to enrich feature space and extend lead times.
- **Advanced architectures** – Explore temporal transformers, physics-informed neural

[Downloaded from ijmt.ir on 2026-11-17]

networks and hybrid ensembles to capture complex dynamics and further improve early-fault detection.

8- References

1- Li R.; Ouyang W.; Liu Q.; Jin Y.; Yang J. (2024). *Dynamic bearing characteristics of the ship stern shaft-bearing system with wave impact.* Ocean Engineering,297,pp.119020. https://doi.org/10.1016/j.oceaneng.2024.119020

2- Wodtke M.; Litwin W. (2020). *Water-lubricated stern tube bearing – experimental and theoretical investigations of thermal effects.* Tribology International-151,pp.106608. https://doi.org/10.1016/j.triboint.2020.106608

3- Lee J.; Jeong B.; An T.-H. (2019). *Investigation on effective support point of single stern tube bearing for marine propulsion shaft alignment.* Marine Structures 64,pp.1-17. https://doi.org/10.1016/j.marstruc.2018.10.010

4- Lin C.-G.; Zou M.-S.; Sima C.; Liu S.-X.; Jiang L.-W. (2019). *Friction-induced vibration and noise of marine stern tube bearings considering perturbations of the stochastic rough surface.* Tribology International 132,pp.661-671. https://doi.org/10.1016/j.triboint.2018.11.026

5- Rossopoulos G. N.; Papadopoulos C. I.; Leontopoulos C. (2020). *Tribological comparison of an optimum single and double-slope design of the stern tube bearing: case study for a marine vessel.* Tribology International-148,pp.106343. https://doi.org/10.1016/j.triboint.2020.106343

6- Frost J.; Frycz M.; Kowalski J.; Wodtke M.; Litwin W. (2023). *Environmentally acceptable lubricants (EAL) compared with a reference mineral oil as marine stern tube bearing lubricant – experimental and theoretical investigations.* Tribology International 181, pp.109001. https://doi.org/10.1016/j.triboint.2023.109001

7-Rossopoulos G. N.; Pervelis I.; Skaltsas D.; Papadopoulos C. I.; Vlachos O.; Koutsoumpas G.; Leontopoulos C. (2025). *Experimental characterization of the tribological and acoustic performance of different stern-tube bearing materials.* TribologyInternational,191,pp.110590. https://doi.org/10.1016/j.triboint.2025.110590

8- Qin H.-L.; Zhou X.-C.; Zhao X.-Z.; Xing J.-T.; Yan Z.-M. (2015). *A new rubber/UHMWPE alloy for water-lubricated stern bearings*. Wear 332–333,pp. 257-261. https://doi.org/10.1016/j.wear.2015.02.016

9- Zhang S.-D.; Long Z.-L.; Yang X.-Y. (2020). *Reaction force of ship stern bearing in hull large deformation based on stochastic theory.* International Journal of Naval Architecture and Ocean Engineering 12(2),pp.723–732. https://doi.org/10.1016/j.ijnaoe.2020.03.009

10- Litwin W.; Dymarski C. (2016). *Experimental research on water-lubricated marine stern tube bearings in conditions of improper lubrication and cooling causing rapid bush wear.* Tribology International,95,pp.449–455. https://doi.org/10.1016/j.triboint.2015.12.005

11- Fang S.; Mu L.; Jia S.; Liu K.; Liu D. (2022). *Combining artificial intelligence and laboratory experiments to explore the behavior of sunken and submerged oil: a typical oil drift and diffusion detection technology.* Journal of CleanerProduction338,pp.133026. https://doi.org/10.1016/j.jclepro.2022.133026

12- American Bureau of Shipping (ABS); Germanischer Lloyd (GL). *Rules and guidelines for ship classification.* Classification Society Publications, n.d.

13- Dubai Dry Dock. (2024). *AHTS-DP1 technical documents.* Internal company documentation.

[Downloaded from ijmt.ir on 2026-11-17]