TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Syed Muhammad Hamza Bacha 246076IVSM

Hydromast Fault Detection

Master's Thesis

Supervisor: Asko Ristolainen
PhD
Co-Supervisor: Laura Piho
PhD

Tallinn 2026

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Syed Muhammad Hamza Bacha 246076IVSM

Hüdromasti vea tuvastamine

Magistritöö

Juhendaja: Asko Ristolainen

PhD

Kaasjuhendaja: Laura Piho

PhD

Tallinn 2026

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis and that this thesis has not been presented for examination or submitted for defense anywhere else. All used materials, references to the literature, and work of others have been cited.

Author: Syed Muhammad Hamza Bacha

11.05.2026

Abstract

Continuous monitoring of rivers and coastal areas is essential for understanding water flow dynamics, yet environmental sensors deployed for extended periods face degradation from debris accumulation, biological fouling, and mechanical wear. The Hydromast, a bio-inspired water flow monitoring device developed at the Center for Biorobotics, Tallinn University of Technology, measures water velocity and direction through the deflection of a mast detected by a Hall effect sensor. During long-term deployment, the device is susceptible to two fault types: debris and bio-fouling, which obstruct the mast and alter its position, and membrane fatigue, which gradually degrades the silicone membrane and permanently shifts the sensor baseline. This thesis develops a three-stage fault detection and classification pipeline to identify these faults automatically from Hall sensor data. In the first stage, 100 statistical features are extracted from 10-second measurement windows and used to train a Random Forest classifier for binary fault detection, achieving 95.74% accuracy and 100% recall on chronologically split test data. In the second stage, physics-based rules examine Hall sensor displacement patterns and variability changes to classify detected faults as debris/bio-fouling. In the third stage, Spearman rank correlation analysis applied to daily Hall Z means identifies the monotonic progressive drift characteristic of membrane fatigue. The Random Forest classifier was trained on real world data from the Pohjaka site and evaluated on a held-out test set from the same site. The pipeline was then validated across three independent scenarios: controlled laboratory experiments with known fault conditions, and two field deployments at Särevere and Arbavere. Laboratory tests confirmed 0% false positives under clean conditions and 95.6–100% detection across all fault types. At Arbavere, three co-located devices exhibited different fault patterns during the same period, with two devices showing progressive irreversible Z decline classified as membrane fatigue and one showing intermittent reversible fluctuations classified as debris/bio-fouling, demonstrating the pipeline's ability to differentiate fault types under real environmental conditions. The pipeline requires no site-specific retraining, needing only a baseline measurement from a known-clean period, making it practical for automated monitoring of

Hydromast deployments.

The thesis is in English and contains 38 pages of text, 6 chapters, 11 figures, 6 tables.

Annotatsioon

Hüdromasti vea tuvastamine

Jõgede ja rannikualade pidev seire on veevoolu dünaamika mõistmiseks hädavajalik, kuid pikema aja jooksul kasutusel olevad keskkonnasensorid kannatavad prahi kogunemise, bioloogilise saastumise ja mehaanilise kulumise tõttu. Tallinna Tehnikaülikooli biorobootika keskuses välja töötatud bioloogiast inspireeritud veevoolu seireseade Hüdromast mõõdab vee kiirust ja suunda painduva masti kaudu, mida registreerib Halli andur. Pikaajalise kasutamise käigus on seade vastuvõtlik kahele peamisele riketüübile: prahi ja bioloogilise saastumise suhtes, mis takistavad masti tööd ja muudavad selle asendit, ning membraani väsimuse suhtes, mis kahjustab järk-järgult silikoonmembraani ja nihutab anduri baasjoont püsivalt. Käesolevas magistritöös arendatakse välja kolmeastmeline rikke tuvastamise ja klassifitseerimise protsess, et tuvastada need rikked automaatselt. Esimeses etapis arvutatakse 10-sekundilistest mõõteakendest 100 statistilist tunnust, mida kasutatakse Random Foresti klassifikaatori treenimiseks binaarsete rikete tuvastamiseks, saavutades kronoloogiliselt jagatud testandmetel 95,74% täpsuse ja 100% saagise. Teises etapis analüüsitakse füüsikaseadustel põhinevate reeglite abil Halli anduri nihkumismustreid ja muutuste kõikumist, et liigitada tuvastatud rikked prahi või bioloogilise saastumisena. Kolmandas etapis tuvastatakse membraani väsimusele iseloomulik monotoonne progressiivne nihkumine, kasutades Halli anduri Z-väärtuste päevaste keskmiste suhtes Spearmani järjestuskorrelatsiooni analüüsi. Süsteemi valideeriti neljas stsenaariumis: mudel treeniti Põhjaka väliskatsete andmetel ja kontrolliti nii laborikatsetes teadaolevate riketega, kui ka kahe sõltumatu välitesti käigus Säreveres ja Arbaveres. Laboratoorsed katsed kinnitasid 0% valepositiivseid tulemusi puhtates tingimustes ja 95,6–100% tuvastamist kõigi riketüüpide puhul. Arbaveres näitasid kolm samas kohas asuvat seadet sama perioodi jooksul erinevaid riketüüpe: kaks seadet näitasid progressiivset pöördumatut Z-väärtuse langust, mis klassifitseeriti membraani väsimuseks, ja üks näitas vahelduvaid pöördumatuid kõikumisi, mis klassifitseeriti prahi/bioloogilise saastumisena, demonstreerides väljatöötatud mudeli võimet eristada riketüüpe reaalsetes tingimustes. Mudel ei vaja kohaspetsiifilist ümberõpet,

vaid ainult baasmõõtmist teadaolevalt puhtast perioodist, mis muudab selle praktiliseks Hüdromasti paigaldiste automatiseeritud veatuvastuseks.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 38 leheküljel, 6 peatükki, 11 joonist, 6 tabelit.

List of abbreviations and terms

ADV	Akustiline Doppleri kiirusmõõtur (*Acoustic Doppler Velocimeter*)
AE	Autoenkooder (*Autoencoder*)
CNN	Konvolutsiooniline närvivõrk (*Convolutional Neural Network*)
IMU	Inertsiaalne mõõteseade (*Inertial Measurement Unit*)
IoT	Asjade internet (*Internet of Things*)
LSTM	Pika lühimälu närvivõrk (*Long Short-Term Memory*)
RMS	Ruutfunktsiooni keskmine (*Root Mean Square*)

Table of contents

List of figures

List of tables

1 Introduction

1.1 Motivation

The continuous monitoring of rivers and streams has become essential as climate change and human interference in the natural environment continue to affect rivers, seas, and oceans around the world. River flow patterns have been affected by anthropogenic climate change globally, with observed trends in low, mean, and high flows linked to human-induced climate change [1]. It is important to understand water flow dynamics to predict floods easily, conduct an ecological research, and understand water resource management, given that river discharge is a key factor in sustaining regional water resources and ecological stability [2].

Bio-inspired sensors have been developed for continuous monitoring of the environment that imitate natural sensing mechanisms found in marine life. The Hydromast, developed at the Taltech Center for Biorobotics, is one such device inspired by the lateral line organ of fish [3]. Fish use neuromasts to detect water movements, pressure changes, and vibrations in their environment, allowing them to locate prey and avoid predators [4]. The device converts this biological principle into an artificial sensor capable of measuring water velocity and flow direction. The Hydromast device, despite its remarkable benefits for environmental monitoring, faces several hindrances if stationed for longer periods in natural settings. The accumulation of silt, leaves, branches, and sediments at the bed of rivers and streams obstructs the sensors and affects their precise measurement. Furthermore, the growth of undesirable organisms, algae, bacteria and others on submerged surfaces, known as biofouling, can affect measurement accuracy, sometimes within a week, thereby gradually reducing sensor functionality [5]. In addition, prolonged exposure to environmental conditions can lead to wear and tear of moving parts, resulting in gradual performance degradation over time. These challenges are particularly concerning due to their steady onset, which complicates the differentiation between genuine environmental changes and sensor degradation without direct physical examination.

The deployment of environmental sensors in distant areas presents challenges for regular maintenance, making continuous visits difficult and costly. Physical checks conducted weekly or monthly increase the risk that defective sensors produce incorrect, unnoticed data. Thus, this creates a significant deficiency in data quality assurance, which can lead to inaccurate analysis of environmental conditions. Therefore, there is a clear need for an automated fault-detection system capable of detecting device malfunctions and notifying maintenance teams before data integrity is significantly compromised.

Machine learning approaches have made tremendous progress in anomaly detection and predictive maintenance in various domains. By deeply studying patterns in sensor data, machine learning models can learn to differentiate between normal operating conditions and faults or anomalies [6]. Predictive maintenance driven by machine learning enables proactive decision-making by learning complex degradation patterns and predicting potential failures from historical and streaming sensor data. Recent studies have shown that machine learning methods such as Random Forests, Neural Networks, and Support Vector Machines can achieve high accuracy in detecting anomalies in environmental sensor systems [7]. However, the application of these approaches to bio-inspired flow measuring devices like the Hydromast remains largely unexplored.

This thesis addresses the challenge of automatic fault detection for the Hydromast by developing a machine learning-based system capable of identifying sensor faults from raw sensor data. The system uses data from the device's 3D Hall effect sensor to detect physical changes caused by debris accumulation or membrane fatigue. By providing early warning of developing problems, this system aims to improve the reliability of Hydromast deployments and reduce maintenance costs through targeted interventions rather than routine inspection visits. The proposed fault detection system is validated using both controlled laboratory experiments, where fault conditions are precisely introduced, and real-world deployment data collected from Hydromast devices in Estonian rivers.

1.2 Problem Statement

Hydromast is a low-cost environmental monitoring device that is designed to measure the water flow speed, direction and water column height. It has been used to monitor rivers [8], coastal areas [9], and ports [10]. The device component costs are approximately 1,000

euros [8], significantly lower than traditional equipment (roughly 1/10th of a commercial ADV). This allows to increase the number of measurement points and increase spatial resolution of flow monitoring.

However, devices deployed in those monitoring stations encounter harsh conditions and severe wear and tear which ultimately affects the sensor data. As with any device, faults may arise during prolonged deployment. The flexible sensing unit is prone to wear over time. Especially when high, out of range flow speeds are continuously present. Plants, debris, and sometimes small organisms accumulate on device surfaces [8], all of which compromise measurement accuracy, as illustrated by the three primary fault types shown in Figure 1.

(a) Membrane fatigue (b) Debris accumulation (c) Biofouling

Figure 1. Three primary fault types affecting Hydromast device during long-term deployment: (a) silicone membrane wear from repeated hydrodynamic stress cycles; (b) debris accumulation including leaves and organic matter in river environments; (c) marine biofouling with barnacles and organisms observed at Varna Port, Bulgaria.

Currently, fault detection relies on physical inspection or looking at data afterwards and identifying anomalies. However, manual inspection of each sensor is time-consuming, especially when multiple sensors are deployed in remote and sparse sites. Additionally, accurate measurements are already compromised for days or weeks when detection relies solely on historical data analysis. Hence, for efficient monitoring, sophisticated automated methods are needed for practical long-term applications.

1.3 Research Questions

This research addresses the challenge of monitoring device condition by developing a system that automatically detects when faults occur in a Hydromast. The system will use signal processing and machine learning methods to identify patterns of faulty behavior.

The focus of this research work will be on the following two research questions:

RQ1: Which statistical features can be used to identify the faults in recorded measurements?

RQ2: How do the statistical features distinguish between membrane fatigue and debris or bio-fouling?

Development of such a fault detection will assist the researchers in identification of non-usable data reducing the time and efforts. Additionally, this will increase the understanding of weather the device has encountered extreme fatigue under higher flow rates such as in floods and heavy rains.

2 Background

This chapter provides the technical context for the Hydromast device and the challenges associated with long-term deployment in aquatic environments. It discusses the main categories of water monitoring sensors currently in use, the specific degradation mechanisms that impact the Hydromast, and the opportunities offered by machine learning for automated fault detection.

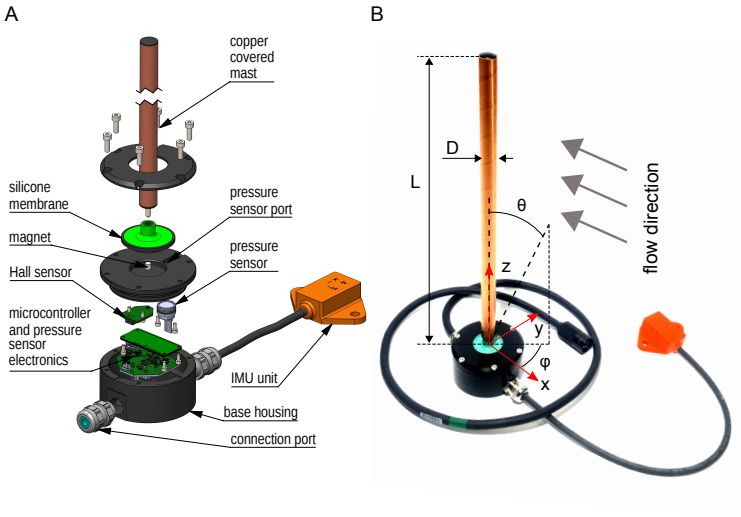


Figure 2. Exploded view of the Hydromast showing key components: copper-covered mast, silicone membrane, magnet, Hall effect sensor, pressure sensor, microcontroller electronics, and base housing. Water flow causes the mast to deflect, moving the magnet relative to the Hall sensor, enabling flow velocity measurement. (Adapted from https://ieeexplore.ieee.org/document/10411837)

2.1 Water Monitoring Sensors

Scientists require the measurement of water flow in rivers and coastal zones to facilitate hydrological modeling, ecosystem management, and the evaluation of climate impacts. Their investigations encompass the dynamics of water movement, sediment transport, flood events, and the influence of climate change on aquatic systems. While conventional sensors yield precise measurements, they are subject to considerable deployment limitations [8]. These instruments are often costly and require human operators for each measurement interval. Furthermore, long-term deployment is rendered impractical by elevated equipment

expenses and the inherent risks of damage or loss within aquatic settings.

Examining specific technologies, Acoustic Doppler velocimeters (ADV), while capable of measuring mean velocity within 1% accuracy, exhibit inherent noise variance consisting of electronic circuitry limitations and flow-related Doppler broadening, with measurement errors ranging from ± 0.95 to ± 3.0 mm s^{-1} depending on velocity range settings [11]. Furthermore, ADV velocity outputs in turbulent flows represent a combination of Doppler noise, signal aliasing, velocity fluctuations, and installation vibrations, requiring extensive post-processing to obtain reliable turbulence measurements [12]. Mechanical propeller current meters, historically prevalent in hydrological measurements, present problems during high discharge conditions and require direct physical contact with the water body, making measurements costly, time-consuming, and labor-intensive [13]. For submerged instrumentation across all sensor types, biofouling represents the single most significant factor affecting operation, maintenance, and data quality, capable of disrupting measurement quality within less than one week of deployment [5]. This biological accumulation increases the cost of ownership to prohibitive levels for maintaining operational sensor networks and can affect sensing characteristics, service life, sensitivity, and data reliability [14]. These widespread limitations across conventional flow measurement technologies underscore the need for robust, low-cost alternatives capable of extended unattended deployment in natural aquatic environments.

The Hydromast, developed at Tallinn University of Technology, offers a budget-friendly alternative for existing commercial sensors [15]. Unlike systems that rely on acoustic measurement techniques, the Hydromast employs a bio-inspired mechanical transduction method. This device features a vertical cylindrical mast affixed to an elastomer base. Water flow causes the mast to deflect laterally; this deflection is then measured by an integrated Hall effect sensor. Consequently, flow velocity is determined from the measured deflection angle [15, 16]. The system allows for the deployment of numerous devices at a fraction of the cost associated with traditional ADV systems. Hydromast has been utilized in Estonian rivers for continuous, long-term monitoring [8], and its affordability allows for extended deployment, even in the face of environmental exposure risks. However, like all submerged instrumentation, Hydromast remains susceptible to biofouling and debris accumulation, which can compromise measurement accuracy over time [5, 8]. This

limitation motivates the development of automated fault detection systems capable of identifying sensor degradation without requiring physical inspection.

2.2 Issues with Monitoring Sensors

The Hydromast's performance is primarily compromised by three degradation mechanisms. Initially, the elastic membrane experiences mechanical fatigue due to persistent hydrodynamic loading [8]. Material cracking and structural failure result from repeated stress cycles, especially during periods of elevated flow. Consequently, this degradation negatively impacts deflection response characteristics, thereby introducing systematic measurement inaccuracies.

Furthermore, the accumulation of debris emerges as suspended organic matter, which includes leaves, branches, and aquatic vegetation, which becomes entrained around the device structure. This accumulated material subsequently modifies the local flow field, thereby causing measurements to diverge from ambient conditions, even when the devices are operating correctly.

Thirdly, biofouling arises from the growth of submerged surfaces by microorganisms. In freshwater ecosystems, algal growth is observed, whereas barnacles and other macrofouling organisms adhere to surfaces in marine environments. Furthermore, bacterial biofilms modify device geometry and surface properties, thereby impacting the precision of measurements.

Current methods for detecting faults have significant limitations. Physically checking each device requires going to the site, which is time-consuming and logistically difficult, especially for devices in remote or marine environments. Alternatively, looking at past data to find problems only works after a long time of collecting faulty data. Neither of these methods is practical for quickly finding faults in large networks [8]. For example, a network of twenty devices spread across several river sites can't be inspected manually every week. As a result, delayed fault detection leads to long periods of unreliable data.

2.3 Bridging Machine Learning with Data Processing

Machine learning techniques offer promising avenues for automating fault detection within sensor systems. These methodologies facilitate the computational analysis of sensor data patterns that signify operational irregularities. Under typical operating conditions, Hydromast presents specific data signatures. Accumulation of debris results in discernible pattern deviations, whereas membrane degradation produces distinct fault signatures. Consequently, machine learning algorithms can be trained to identify these fault-specific patterns.

Automated fault detection systems present numerous operational benefits. Initially, early fault detection allows for predictive maintenance scheduling, thereby mitigating the risk of complete sensor failure. Second, the capacity for real-time anomaly detection allows for immediate evaluation of data quality, thereby precluding the integration of flawed measurements into subsequent analyses. Third, automated monitoring enhances operational efficiency by allocating maintenance resources solely to sensors that necessitate intervention, as opposed to adhering to predetermined inspection schedules. Ultimately, these systems facilitate the management of extensive sensor networks, a task that would be unfeasible through manual monitoring.

Time-series data produced by Hydromast are particularly amenable to machine learning analysis [16]. The continuous acquisition of measurements generates detailed records of sensor responses to fluctuating environmental conditions. Machine learning classification algorithms can then analyze these temporal patterns to differentiate between normal operational behavior and various fault conditions.

3 Previous Studies

3.1 Related work on Environmental Sensors

Sensor fault detection has been extensively studied in industrial monitoring contexts, where equipment failures may compromise production efficiency or safety [6]. Research has examined fault detection in marine and offshore sensor systems [17], identifying unique challenges including limited physical accessibility for maintenance, harsh environmental conditions with dynamic hydrodynamic loading, and requirements for spatially distributed coverage. However, these studies primarily address sensor technologies such as acoustic and pressure-based systems with different failure modes than bio-inspired mechanical sensors like Hydromast.

Studies of offshore platform monitoring have demonstrated that sensor placement and network density significantly affect fault detection capabilities. Furthermore, the elevated costs of offshore maintenance operations relative to terrestrial deployments emphasize the importance of early fault detection for cost-effective system operation.

Recent work by Potharaju et al. [7] demonstrated a two-stage machine learning approach for environmental sensor anomaly detection, combining Isolation Forest for unsupervised anomaly labeling with supervised classification using Random Forest, Neural Networks, and AdaBoost, achieving 99.93% accuracy. This validates the effectiveness of integrating unsupervised and supervised learning for sensor fault detection in scenarios lacking pre-labeled training data. However, their approach generated labels entirely from unsupervised methods, whereas the present study leverages expert domain knowledge to establish ground truth labels, enabling validation of the detection approach.

In addition to the two-stage approach, other studies have examined different machine learning architectures for the purpose of sensor fault detection. Deep learning methods have demonstrated significant potential; for instance, [18] utilized a Long Short-Term Memory autoencoder (LSTM-AE) for IoT sensor fault detection, employing multi-step-

ahead prediction. Furthermore, [19] attained an accuracy of 98.21% with a hybrid CNN-LSTM model, which integrates spatial feature extraction and temporal sequence modeling. Regarding time-series sensor data, [20] showed that Random Forest, alongside Explainable Artificial Intelligence, achieves an accuracy of 99.84% in the detection of anomalies within oil well sensors.

In aquatic settings, that mimick the conditions of Hydromast deployment, [21] developed ANNODE, an Artificial Neural Network-driven outlier detection technique designed for environmental sensors that monitor temperature and salinity; this method exhibited enhanced accuracy compared to conventional threshold-based strategies in challenging underwater environments. These investigations collectively validate that machine learning methodologies reliably attain accuracy levels surpassing 95% for sensor fault detection across a range of application areas.

3.2 Machine Learning Approaches for Sensor Problems

Machine learning provides two primary methods for detecting sensor failures. Supervised learning requires training data that includes known error labels encompassing both normal sensor behavior and various fault conditions. These approaches use techniques such as Support Vector Machines, Random Forests, and neural networks to differentiate between operational states. In industrial settings, these methods achieve over 95% accuracy when provided with sufficient fault examples [6].

Deep learning methodologies have demonstrated particularly strong results. By integrating Convolutional Neural Networks with Long Short-Term Memory architectures, these models can evaluate both instantaneous sensor readings and temporal trends. Recent research on environmental sensor systems achieved 99.93% accuracy using such methods on time-series data [7].

When fault examples are limited, unsupervised learning serves as a valuable alternative. Rather than learning from labeled faults, these approaches characterize normal operation and identify irregular patterns as potential anomalies. Techniques such as One-Class Support Vector Machines and Isolation Forest enable fault detection without requiring extensive fault databases.

3.3 Research Gap for Hydromast

While sensor fault detection has been extensively studied, bio-inspired mechanical devices like Hydromast present distinct challenges that existing research has not yet addressed.

The sensing technology differs fundamentally from conventional systems. Hydromast uses Hall effect sensors to measure mechanical deflection [15], whereas most water monitoring research focuses on acoustic or pressure-based sensors. Faults affecting these different sensor types likely produce distinct patterns in the measured data.

The failure mechanisms are also unique to Hydromast's design [8]. Elastic membrane fatigue and bio-fouling around the sensor structure are not typical failure modes encountered in industrial applications. These Hydromast-specific faults may generate characteristic signatures in Hall sensor data that differ from previously studied fault patterns.

The deployment environment introduces additional complexity [9]. Unlike conventional flow measurement devices such as ADVs and mechanical current meters, which are typically operated under supervised conditions for defined measurement campaigns, Hydromast devices operate unattended in natural waterways for extended periods, experiencing variable flow conditions, fluctuating water levels, and continuous biological and physical interference.

Finally, comprehensive fault databases for Hydromast do not yet exist, as the technology is relatively recent [15]. This limited labeled data availability suggests that methods requiring fewer training examples such as semi-supervised or unsupervised learning approaches may prove more practical for Hydromast fault detection.

This research fills these gaps by developing fault detection approaches designed specifically for Hydromast and its unique operational challenges.

4 Research Method

This chapter describes the methodology used to develop an automatic fault detection and classification system for the Hydromast device. The research follows a structured machine learning pipeline: data collection from field deployments, preprocessing of raw sensor measurements, engineering of statistical features, and training of a binary classification model for fault detection. Fault type classification is then performed using physics-based rules derived from Hall sensor signatures, while progressive membrane degradation is identified through statistical trend analysis using Spearman rank correlation. The system is validated across multiple deployment environments to assess generalization. Each stage is designed to produce a practical system capable of detecting faults automatically without manual inspection. Figure 3 provides an overview of the complete methodology.

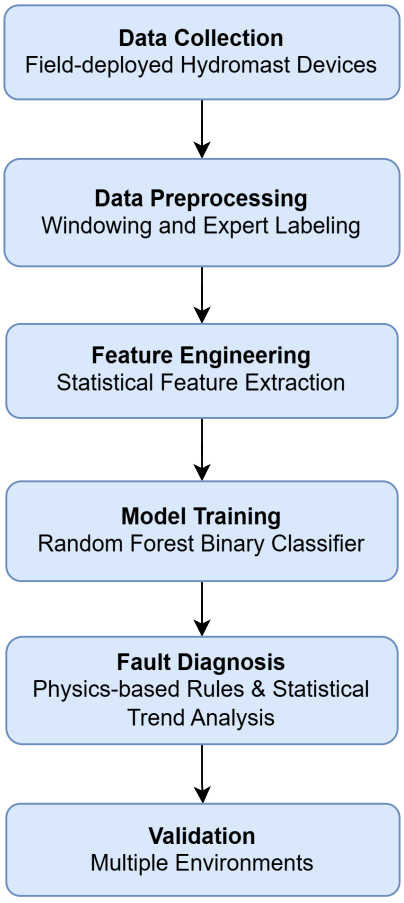


Figure 3. Overview of the research methodology for Hydromast fault detection

4.1 Data collection

The data used in this study was collected from Hydromast devices deployed in rivers at three locations in Estonia. Field deployments at Pohjaka provided training data with expert-identified fault progression, deployments at Särevere provided independent cross-site validation data, and deployments at Arbavere provided additional validation data including membrane fatigue. Additionally, a controlled laboratory experiment was conducted to test the pipeline's ability to distinguish between fault types.

4.1.1 Pohjaka deployment (primary training data)

The primary dataset was collected from a Hydromast device deployed at Pohjaka, Estonia, over a 50-day period from August 21 to October 9, 2025. The device was mounted in a river and recorded continuous measurements throughout the deployment. During this period, the device experienced a gradual fault progression that was identified through expert analysis of the velocity measurements. Domain experts analyzed the velocity time series and identified clear behavioral changes indicating device malfunction. These expert assessments provided the ground truth labels for training the fault detection model.

Each Hydromast device contains multiple sensors that work together to measure water flow characteristics, including a three-axis Hall effect sensor (TLV493D-A1B6) that measures changes in the magnetic field produced by a 5×5 mm neodymium magnet embedded in the silicone membrane [15]. As water flow deflects the mast, the Hall sensor registers this deflection across three axes: `hall_x` and `hall_y` capture the mast position corresponding to flow direction, while `hall_z` captures the mast tilt corresponding to flow velocity.

The Hall magnitude M is computed from the three-axis measurements as:

$$M = \sqrt{X^2 + Y^2 + Z^2} \tag{4.1}$$

where X, Y, and Z denote the magnetic field components measured along the respective axes. The magnitude M correlates linearly with the tilt angle of the mast.

The device also includes an accelerometer and magnetometer that together form an Inertial Measurement Unit (IMU), as well as a pressure sensor and a temperature sensor.

The device sampled data at 50 Hz, with raw measurements stored in CSV format. Devices equipped with the full IMU produce 14 data columns per sample (sensor identifier, timestamp in milliseconds, pressure, temperature, three Hall axes, three accelerometer axes, three magnetometer axes, and a checksum), while devices without the IMU produce 8 columns. For this study, only the Hall sensor data was used for fault detection, as it directly measures mast deflection and provides sufficient discriminative power for identifying faults.

4.1.2 Expert labeling of training data

Ground truth labels were established through domain expert analysis. Based on visual inspection of data trends and knowledge of environmental conditions, the 50-day deployment was divided into three distinct periods, as shown in Table 1.

Period	**Dates**	**Label**
Normal Operation	Aug 21 – Sept 18	0 (Normal)
Transition	Sept 19 – Sept 26	-1 (Excluded)
Confirmed Fault	Sept 27 – Oct 9	1 (Faulty)

Table 1. Expert-defined labeling periods for Pohjaka deployment data

The first period (August 21 – September 18) represents normal operation when the device was functioning correctly. The second period (19-26 September) is a transition phase in which the fault was developing gradually. This transition period was excluded from the training because it contains ambiguous data where the state of the device is unclear. Training on such ambiguous examples would confuse the model and reduce accuracy. The third period (September 27 – October 9) represents confirmed faulty operation. Expert analysis of the velocity data confirmed that the device behavior after September 26th was clearly abnormal, validating the classification of this period as faulty operation.

To validate these expert-assigned labels independently, an Isolation Forest algorithm [7] was applied to the entire dataset without any label information. This unsupervised anomaly detection method identifies data points that deviate from the statistical norm. The Isolation Forest achieved 74.5% agreement with the expert labels, confirming that the fault period is objectively detectable through statistical methods alone. The disagreement is attributed primarily to the transition period, where the fault was developing gradually and the boundary

between normal and faulty behavior is inherently ambiguous. This two-step validation approach strengthens confidence in the ground truth labels used for supervised model training.

4.1.3 Särevere deployment (cross-site validation data)

To test whether the pipeline generalises to different devices and locations, data from three Hydromast devices deployed at Särevere, Estonia was collected for validation. These devices were installed at different distances from the river bank: 3 meters (device N56), 6 meters (device N57), and 9 meters (device N58) from the left bank. Each device experienced varying levels of mechanical stress due to different flow conditions at their positions.

Expert assessment identified fault activity in devices N57 and N58, while device N56 remained largely clean. This Särevere data provides an independent test of whether the pipeline learned general fault patterns that transfer to new deployments, rather than memorizing location-specific artifacts from the Pohjaka training site.

4.1.4 Arbavere deployment (membrane fatigue validation data)

A third field deployment at Arbavere, Estonia provided additional validation data. Three Hydromast devices (N14, N20, N32) were deployed in the same river over a 56-day period from August 15 to October 9, 2025. Device N14 was positioned near the riverbank, while devices N20 and N32 were deployed further toward the centre of the river, where flow velocities are typically higher and the membrane experiences greater hydrodynamic stress.

This dataset was provided by the supervisor to evaluate whether the pipeline could detect and differentiate fault types under field conditions not seen during training. The co-located deployment is particularly valuable for validation, as devices sharing the same environment but exhibiting different fault patterns allow direct comparison between fault types within a single site and time period.

4.1.5 Laboratory validation setup

To evaluate the fault detection system under controlled conditions with known ground truth, laboratory experiments were conducted at the TalTech Centre for Biorobotics. Unlike field

deployments where fault timing is uncertain, the laboratory environment provides precise control over when each fault condition is introduced.

Equipment and Devices

Three Hydromast devices were used in the laboratory experiments:

- **LN58**: Healthy membrane in good condition
- **LN57**: Membrane with minor fatigue from previous deployment
- **LN56**: Membrane with significant fatigue from extended use

Using devices with different fatigue levels allowed evaluation of whether the model could detect membrane degradation independently of other fault types.

Fault Conditions

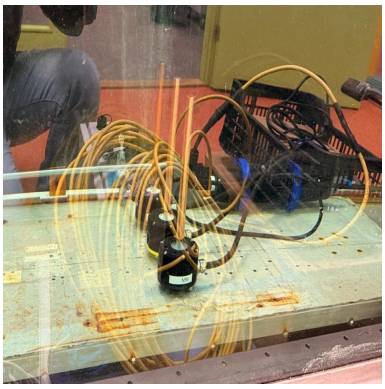
Four conditions were tested for each device:

1. **Clean**: Baseline condition with no faults introduced. The mast and membrane were inspected and cleaned before recording.
2. **Debris short**: Rubber strips attached horizontally across the mast, simulating debris such as leaves or plastic caught on the device. This configuration was expected to cause a sharp shift in Hall Z readings and increased signal variability.
3. **Debris long**: Rubber strips attached vertically along the mast length. This orientation was tested to determine whether debris position affects the fault signature.
4. **Bio-fouling**: A mixture of glue and sand applied to the mast surface, simulating the texture and weight of biological growth such as algae accumulation.

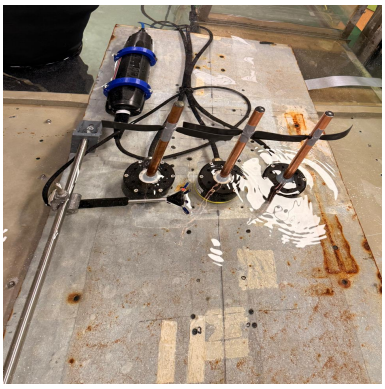
Figure 4 shows photographs of each fault condition as applied to the test devices.

Data Collection

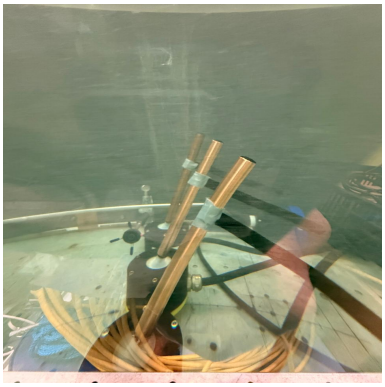
Data was recorded for each device-condition combination at 50 Hz, consistent with the field deployment configuration. Each fault condition was photographed to document the exact configuration. A Vectrino Profiler ADV (Nortek) was used as a reference device for flow velocity estimation during the laboratory experiments.



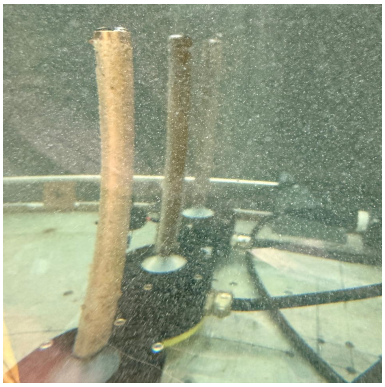
(a) Clean mast



(b) Debris short



(c) Debris long



(d) Bio-fouling

Figure 4. Laboratory fault conditions

4.2 Data preprocessing

Raw measurements require several preprocessing steps before they can be used for machine learning analysis.

4.2.1 Segmentation into Windows

The continuous 50 Hz data stream produces too many individual samples for direct machine learning analysis. The data was therefore segmented into non-overlapping 10-second windows, each containing 500 time series samples or data points. This windowing approach captures sufficient temporal context for calculating meaningful statistical features while keeping the dataset computationally manageable.

4.2.2 Handling Missing Data

Some windows contained missing measurements or sensor errors due to temporary communication issues or power fluctuations. Windows with more than 10% missing data were excluded from the analysis. For windows with minor gaps (fewer than 10% missing), the missing values were forward-filled using the last valid measurement, which is appropriate given the high sampling rate where consecutive measurements are nearly identical.

4.2.3 Excluding the Transition Period

As described in Section 4.1.2, the transition period (September 19–26) was excluded from model training, resulting in a clean binary classification task: normal operation versus confirmed fault.

4.2.4 Removing Temperature as a Feature

Initial experiments revealed that models trained with temperature features achieved artificially high accuracy because the normal period (summer) and the faulty period (early autumn) had systematically different water temperatures. The temperature was acting as a confounding variable, reflecting seasonal changes rather than device faults. Removing temperature from the feature set ensured that the model learned genuine fault-related patterns rather than seasonal artifacts.

4.3 Feature engineering

Machine learning models perform better with engineered features that summarise the statistical properties of each window rather than raw readings. Feature engineering transforms each 10-second window of 500 measurements into a fixed-length vector of descriptive statistics.

4.3.1 Statistical Features per Hall Sensor Axis

For each Hall sensor axis, 25 features were calculated from the 500 measurements within each window, organised into three categories:

Basic Statistical Features (14):

- Mean, standard deviation, minimum, maximum, range
- Median, 10th/25th/75th/90th percentiles, interquartile range (IQR)
- Skewness, kurtosis, root mean square (RMS)

Temporal Features (8):

- Delta mean, delta std, delta max (rate of change measures)
- Zero crossings (oscillation frequency)
- Trend slope (linear drift detection)
- First/second half means and their difference (within-window change)

Baseline Deviation Features (3):

- Deviation from baseline mean
- Z-score from baseline
- Percentage above baseline

These 25 features per axis capture signal strength, variability, temporal dynamics, and deviation from normal operation.

4.3.2 Derived Features: Hall Magnitude

In addition to features extracted from the individual axes, the Hall sensor magnitude was calculated following [15], as defined in Equation 4.1.

This derived quantity correlates directly with the mast tilt angle, and consequently, with flow velocity. The same 25 features (14 basic, 8 temporal, and 3 baseline features) were computed for the magnitude signal, thereby capturing overall flow behaviour independent of direction.

4.3.3 Total Feature Count

The model uses 100 features computed exclusively from Hall sensor data. Each source contributes 14 basic statistical features, 8 temporal features, and 3 baseline deviation features. The Hall sensors directly measure mast deflection caused by water flow, making

Table 2. Feature composition of the fault detection model

Source	**Features**
Hall X axis	25
Hall Y axis	25
Hall Z axis	25
Hall magnitude	25
Total	**100**

them the primary indicators of device health and fault conditions.

4.4 Model development

4.4.1 Algorithm Selection: Random Forest

Random Forest was selected as the classification algorithm for several reasons. First, it handles high-dimensional feature spaces well without requiring extensive feature selection. Second, it is robust to overfitting through its ensemble approach of training multiple decision trees on random subsets of both samples and features. Third, it provides built-in feature importance rankings that reveal which features contribute most to fault detection. Fourth, it requires minimal hyperparameter tuning compared to more complex methods such as deep neural networks, making it practical for deployment in environmental monitoring systems with limited computational resources. This choice aligns with findings from [7], who demonstrated Random Forest's effectiveness for predictive maintenance. To empirically confirm this selection, three machine learning algorithms Decision Tree, XGBoost and Random Forest, were trained and evaluated on the same feature set and held-out test set. The results are presented in Section 5.2.3.

4.4.2 Model Configuration

Initial experiments included features from both Hall sensors and the accelerometer (IMU). However, the accelerometer data introduced significant noise that degraded model performance. The accelerometer captures not only device movement caused by water flow, but also vibrations from the mounting structure, passing debris impacts, and environmental disturbances unrelated to fault conditions. This noise reduced the model's ability to

distinguish between normal operation and genuine faults.

After removing accelerometer features, the model was trained using 100 features extracted exclusively from Hall sensor data. This configuration achieved better accuracy and ensures compatibility across all Hydromast devices, regardless of whether they include the optional IMU sensors. The Hall sensors directly measure mast deflection caused by water flow, providing cleaner signals that reliably indicate device health and fault conditions.

4.4.3 Universal Model Training

To create a model capable of detecting faults in both field and laboratory environments, training data from both sources was combined. The Pohjaka field data provided examples of naturally occurring fault progression, while the laboratory data provided controlled examples of specific fault types with known ground truth.

Before any training, the lab data was split into 80% for training and 20% for testing using a stratified random split. This held-out 20% was locked away and never used during training, since each lab recording is an independent controlled experiment with no temporal dependency between samples.

Due to the smaller volume of lab data compared to field data, 50$\times$ oversampling was applied exclusively to the lab *training* portion. This resulted in lab data comprising approximately 12–15% of the combined training set, ensuring the model learned from both environments without being dominated by field data.

4.4.4 Training Process and Hyperparameters

The model was trained on the Pohjaka dataset after excluding the transition period. Since the data is a continuous time-series, a chronological split was used: the first 80% of days were used for training and the remaining 20% for testing. This approach avoids data leakage, as the model never sees future data during training. Table 3 lists the hyperparameter configuration

Table 3. Random Forest hyperparameter configuration

Parameter	**Value**	**Purpose**
Number of trees	200	Balance accuracy and computational cost
Maximum depth	25	Prevent overfitting
Min samples to split	5	Avoid branches based on noise
Min samples per leaf	2	Ensure statistical significance
Max features per split	$\sqrt{n}$	Promote tree diversity

These parameters balance model complexity with generalisation ability.

4.4.5 Fault Type Classification

Beyond binary fault detection (normal versus faulty), a physics-based classification approach was developed to differentiate between two faults categories: Debris/Bio-fouling and Membrane Fatigue. A physics-based rule engine was chosen over a second machine learning classifier because insufficient labeled examples of each fault type were available to train a supervised classifier; the training data contains only one fault progression at a single site. Additionally, a rule-based approach provides interpretable decisions grounded in the physical operating principles of the device, making the classification transparent and verifiable by domain experts.

Debris accumulation and biofouling are grouped into a single category because both represent external obstruction of the device, producing similar Hall sensor signatures (shifted mean position, altered vibration characteristics). In contrast, membrane fatigue is an internal degradation mechanism with fundamentally different data patterns. This classification follows the physical operating principles of the Hydromast device described by [15].

4.4.6 Physical Signatures of Fault Types

Each fault type produces a distinct pattern in the Hall sensor data due to different physical mechanisms:

- **Debris/Bio-fouling**: foreign material (leaves, branches) or biological growth (algae,

barnacles) attaches to the mast or membrane surface, physically displacing the mast from its resting position or restricting its movement. This produces a shifted `hall_z` mean (the mast is pushed to a new position), shifted `hall_x`/`hall_y` means (sideways displacement), and changed variability as the obstruction interacts with water flow or dampens mast vibration. The effect can be sudden (debris catching on the mast) or gradual (biological growth accumulating over days), but is characterised by abrupt changes in signal statistics relative to the established baseline.

- **Membrane fatigue**: the silicone membrane degrades over time, losing elasticity and altering the mast's restoring force. This produces a gradual, monotonic drift in Hall sensor readings (the mast no longer returns to the same resting position), detectable through consistent trend slopes and directional half-window changes across faulty windows. The effect is *permanent*: it cannot self-heal overnight. This physical basis is supported by Egerer et al. [15], who note that membrane properties directly affect the tilt calibration of the device, as the hand-made silicone membrane's elasticity governs the mast's restoring force and resting position.

4.4.7 Classification Indicators

Seven quantitative indicators were defined to distinguish between fault types:

1. **Hall Z mean shift** (z-score relative to baseline): a shift exceeding 3σ strongly indicates debris or bio-fouling physically displacing the mast.
2. **Baseline deviation magnitude**: quantifies how far current measurements deviate from the established normal baseline.
3. **Hall Z variability change**: increased or decreased variability (more than $2\times$ or less than $0.5\times$) suggests external obstruction altering mast vibration patterns.
4. **Cross-axis displacement**: lateral shifts in `hall_x` or `hall_y` indicate physical obstruction pushing the mast sideways.
5. **Progressive drift (trend slope)**: consistent non-zero trend slopes across faulty windows, with more than 70% of windows showing the same slope direction, indicate gradual membrane degradation rather than abrupt obstruction.
6. **Half-window directional change**: a consistent difference between the first and second half of measurement windows suggests monotonic within-window drift characteristic of fatigue.

7. **Magnitude deviation without Z shift**: a gradual change in Hall magnitude without a large abrupt Z shift distinguishes slow membrane degradation from sudden obstruction events.

Each indicator contributes a weighted score toward a fault type classification. A multi-day recovery detection rule was also implemented: if a device shows greater than 50% faulty windows on one day but fewer than 20% faulty windows the following day, membrane fatigue is ruled out (since permanent damage cannot self-heal), and the fault is reclassified as debris/Bio-fouling.

4.4.8 Baseline Drift Detection

The per-window fault classification described above detects acute signal anomalies but cannot identify slow progressive degradation spanning weeks or months. Membrane fatigue manifests as a gradual monotonic decline in the Hall Z baseline rather than sudden signal disruption. To detect this, a statistical trend analysis was applied to the sequence of daily Hall Z mean values across the deployment.

Spearman's rank correlation coefficient was selected as the trend detection method, following its established use in environmental monitoring time-series analysis [22]. Unlike Pearson correlation, Spearman's method is non-parametric and measures monotonic rather than linear relationships, making it robust to outlier days caused by temporary debris events. The Spearman rank correlation coefficient is defined as:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{4.2}$$

where n is the number of observation days, and $d_i = \text{rank}(x_i) - \text{rank}(y_i)$ is the difference between the rank of the day index and the rank of the corresponding daily mean Hall Z value. A coefficient of $r_s = -1$ indicates a perfectly monotonic decreasing trend, while $r_s = +1$ indicates a perfectly monotonic increasing trend.

The detection procedure operates as follows. First, daily mean Hall Z values are computed for days where fewer than 20% of windows were classified as faulty by the Random Forest model, ensuring that the trend is calculated from representative baseline readings rather

than days dominated by acute faults. The Spearman rank correlation coefficient ρ is then computed between the day index and the corresponding Z mean value. A statistically significant monotonic trend is identified when $|\rho| > 0.4$ and $p < 0.01$.

To confirm physical significance, the total Z shift between the first and last week of the deployment must exceed 3 mT. This threshold corresponds to approximately four standard deviations of the daily Z mean variability observed during confirmed normal operation (standard deviation ≈ 0.8 mT at Pohjaka and Särevere). This prevents flagging minor statistical trends that do not correspond to meaningful membrane degradation

Three additional checks refine the detection:

1. **Cross-sensor validation**: when multiple devices are deployed at the same site, the trend direction of each device is compared. If a majority of devices exhibit drift in the same direction with $|\rho| > 0.3$, the trend is attributed to environmental factors (such as seasonal changes, flooding or heavy rain) rather than individual device fatigue.
2. **Irreversibility check**: membrane fatigue is a permanent mechanical degradation that cannot self-heal. The algorithm identifies the last day on which the Z mean returned within two standard deviations of the initial baseline. Only days after this last recovery are flagged as drift days.
3. **Direction check**: if the overall trend is downward ($\rho < 0$), days where the Z mean exceeds the baseline are not flagged, as they contradict the expected degradation direction. The same logic applies in reverse for upward trends.

4.5 Validation

A four-stage validation strategy was designed to evaluate the fault detection system under progressively more challenging conditions.

4.5.1 Stage 1: Training Site Performance

The first stage used the held-out 20% test set from the Pohjaka deployment. This tests whether the model can correctly classify windows from the same device and location that it has not seen during training. High accuracy at this stage is necessary but not

sufficient, as the model might have learned site-specific patterns rather than generalisable fault signatures.

4.5.2 Stage 2: Laboratory Validation

The second stage used the controlled laboratory experiment described in Section 4.1.5 to evaluate fault detection and classification with known ground truth. Unlike field deployments where fault timing is uncertain, the laboratory environment provides precise control over when each fault condition is introduced, enabling direct assessment of detection accuracy for each fault type.

4.5.3 Stage 3: Cross-Site Validation (Särevere)

The third stage applied the pipeline to data from the three Särevere devices (N56, N57, N58). These devices operated in a different river, at different positions, and experienced different fault progression. This stage tests the critical question of whether the pipeline generalises beyond the training environment. The same Hall-only model trained on Pohjaka data was applied to all three Särevere devices.

4.5.4 Stage 4: Cross-Site Validation with Membrane Fatigue (Arbavere)

The fourth stage applied the pipeline to three co-located Hydromast devices (N14, N20, N32) at the Arbavere deployment site over 56 days. Unlike Särevere, this site includes devices with confirmed membrane fatigue, enabling validation of the Spearman-based drift detection stage (Section 4.4.8). Critically, the three devices at the same site exhibited different fault patterns, testing whether the pipeline can correctly differentiate between debris/bio-fouling and membrane fatigue under identical environmental conditions.

5 Results and Discussions

5.1 Overview

This chapter presents the results of the fault detection and classification pipeline following the four-stage validation strategy described in Section 4.5. The chapter is organised into three parts: model training and evaluation on the Pohjaka held-out test set (Section 5.2), laboratory validation under controlled conditions with known ground truth (Section 5.3), and field deployment results from three river sites (Section 5.4). Laboratory results are presented before field deployment results because the laboratory experiments provide verified fault conditions, establishing confidence in the pipeline before interpreting field data where fault labels were assigned by expert judgement and not directly verified.

5.2 Model Training and Evaluation

5.2.1 Training Setup

The model was trained using the methodology described in Chapter 4. The Random Forest classifier was evaluated on the held-out 20% test set from the Pohjaka deployment, with results shown in Table 4

5.2.2 Training Site Performance (Pohjaka)

On the held-out 20% test set from the Pohjaka deployment, the model achieved the results shown in Table 4.

The model achieved 100% recall, meaning it correctly identified every faulty window in the test set. The precision of 88.03% indicates that approximately 12% of windows flagged as faulty were actually normal, a conservative bias that is preferable in a monitoring context where missing a real fault is more costly than a false alarm. The ROC-AUC of 99.90% confirms strong separation between normal and faulty classes across all classification thresholds.

Table 4. Performance of the Random Forest model on the Pohjaka held-out test set

Metric	**Value (%)**
Accuracy	95.74
Precision	88.03
Recall	100.00
F1-score	93.63
ROC–AUC	99.90

5.2.3 Algorithm Comparison

To confirm that Random Forest was the right choice for this problem, two additional algorithms were evaluated on the same training and test data, Decision Tree and XGBoost. All three algorithms were trained on the 100 Hall sensor features and tested on the same held-out test set from the Pohjaka deployment. Table 5 summarises the performance across five metrics.

Table 5. Algorithm comparison on the Pohjaka held-out test set.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**ROC-AUC (%)**
Decision Tree	93.15	82.09	99.96	90.15	96.92
XGBoost	93.74	83.36	99.98	90.92	99.84
Random Forest	**95.74**	**88.03**	**100.00**	**93.63**	**99.90**

All three algorithms achieve very high recall, which shows that the fault patterns in the Hall sensor data are clearly learnable. The key differentiator is precision, where Random Forest achieves 88.03%, compared to 83.36% for XGBoost and 82.09% for Decision Tree. Random Forest is also the only algorithm that produces zero false negatives, meaning every faulty window in the test set was correctly identified, while the Decision Tree missed 8 faulty windows and XGBoost missed 4.

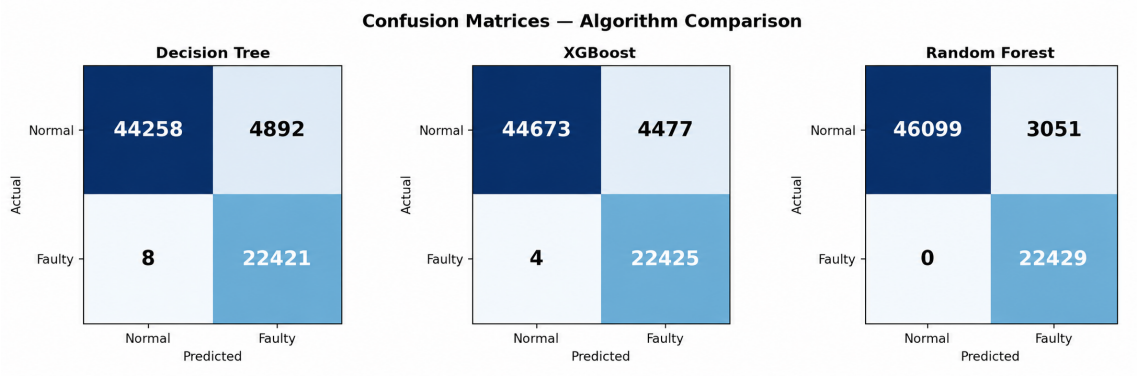


Figure 5. Confusion matrices for the three algorithms. Random Forest is the only algorithm with zero false negatives.

5.3 Laboratory Results

The laboratory experiments were conducted to validate the pipeline under controlled conditions where the fault type is known beforehand. Since the fault conditions were physically introduced by the experts, these results serve as a ground truth reference for evaluating the pipeline's detection and classification accuracy.

Three devices with different membrane fatigue levels (LN58: healthy membrane, LN57: less worn, LN56: more worn) were tested under four conditions: clean, debris (short), debris (long), and bio-fouling.

To understand the effect of membrane fatigue on sensor readings, the clean recordings from all three devices were compared using the healthy device LN58 as the baseline reference. When LN57 (less worn) was compared against the LN58 baseline, the pipeline produced 0% faulty windows; its readings were still within normal range. However, when LN56 (more worn) was compared against the same LN58 baseline, the pipeline flagged 100% of windows as faulty, with a Z mean shift of 9.36 mT. This demonstrates that significant membrane fatigue permanently alters the Hall sensor output to a degree that is indistinguishable from an external fault when compared to a healthy baseline. The device's worn membrane has shifted the resting position of the mast, changing the magnetic field readings even without any external obstruction. Table 6 summarises the comparison.

Table 6. Clean condition comparison using healthy device LN58 as baseline reference.

Device	Membrane Condition	Faulty (%)	Classification
LN58	Healthy	0	Normal
LN57	Less worn	0	Normal
LN56	More worn	100	Debris/Bio-fouling

For the standard laboratory evaluation, baselines were computed from each device's own clean recordings. Under this configuration, all three devices produced 0% faulty windows in the clean condition, confirming no false positives. For the three fault conditions (debris short, debris long, and bio-fouling), the model detected 95.6–100% faulty windows across all device-condition combinations (Figure 6). All fault conditions were correctly classified as Debris/Bio-fouling, as each condition involved external obstruction of the mast.

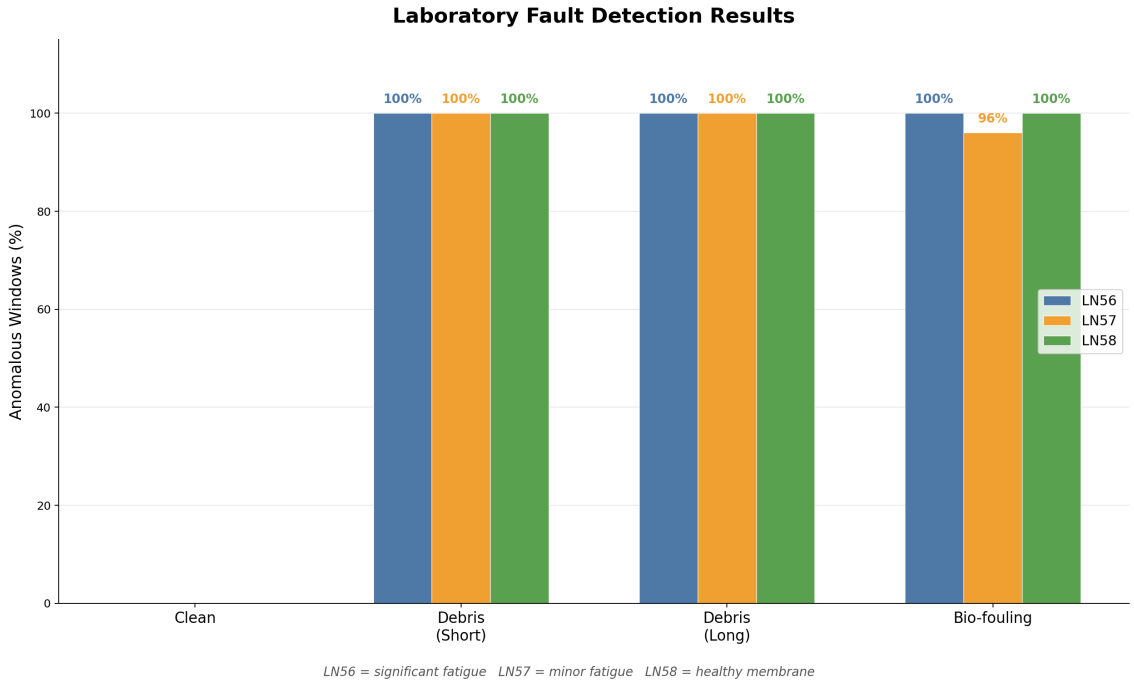


Figure 6. Laboratory validation results showing fault detection rates for each condition and device. Clean conditions produce 0% false positives, while all fault conditions are detected at 95–100%.

It is important to note that membrane fatigue could not be directly evaluated in the laboratory setting. Membrane fatigue is a progressive degradation that develops over weeks or months of continuous deployment in flowing water. A short-duration laboratory experiment cannot replicate this gradual process, which is why the Spearman-based drift

detection stage (Section 4.4.8) is validated exclusively through long-term field data at Arbavere (Section 5.4.3). However, the LN56 baseline comparison above shows that the cumulative effect of fatigue is measurable through the Hall sensor, supporting the approach of tracking progressive drift in field data.

5.4 Field Deployment Results

The pipeline was applied to data from three field sites: Pohjaka (the training site), Särevere, and Arbavere. It is important to highlight that the fault labels for the field data were assigned based on expert analysis of the sensor signals and deployment logs, rather than through direct physical inspection of the devices at the time of the fault. The devices were only examined after retrieval at the end of the deployment period, which confirmed their general condition but not the exact timing of the fault onset. The laboratory results presented in Section 5.3 serve as a verified reference point; since the pipeline accurately detects and classifies faults under known conditions, the classifications made in the field can be interpreted with greater confidence.

5.4.1 Pohjaka Results

When the trained model was applied to the full 50-day Pohjaka deployment, the diagnosis pipeline produced results consistent with the known fault timeline. Figure 7 shows the daily fault percentage over the deployment period.

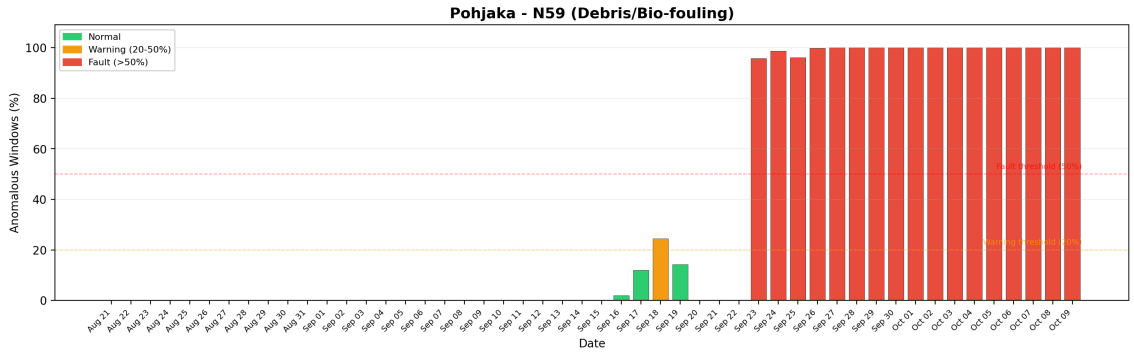


Figure 7. Pohjaka N59 fault diagnosis over 50 days. Days with 0% faulty windows show no bar. Green bars indicate low-level fault activity (below 20%), and red bars indicate confirmed fault days (above 50%). The orange dashed line marks the 20% warning threshold and the red dashed line marks the 50% fault threshold.

The first 26 days (August 21 – September 15) showed 0% faulty windows, correctly

identifying normal operation. Warning-level activity appeared on September 17–19 (1.1%, 15.4% and 2.4% faulty windows respectively), corresponding to the early stages of fault development. A brief drop on September 20–22 was followed by a sharp transition to confirmed fault status on September 23 (95.7% faulty), after which fault levels remained above 90% for the remainder of the deployment through October 9.

The fault classification system identified all faulty days as being caused by Debris or Bio-fouling. An expert analysis of the velocity time series confirmed this classification; toward the end of the deployment, the velocity signal changed sharply, indicating that an object had become lodged on the mast. The persistent nature of the fault, with no recovery after September 23, suggests that the debris remained attached for the remainder of the deployment. This finding demonstrates that the system accurately identifies both the presence and type of fault.

5.4.2 Särevere Results

The pipeline was applied to data from three Hydromast devices (N56, N57, N58) deployed at Särevere, a different river site from the training location. This evaluation tests whether the pipeline generalises to new devices and environments.

Device N57 exhibited the strongest fault activity, with 96% faulty windows on September 19, followed by elevated levels on September 18 and 20 (Figure 8). Device N58 showed moderate fault activity between September 15–17 (Figure 9), peaking at 54% on September 17. Device N56 remained largely clean throughout the deployment, with only a brief warning on September 11 (25% faulty windows).

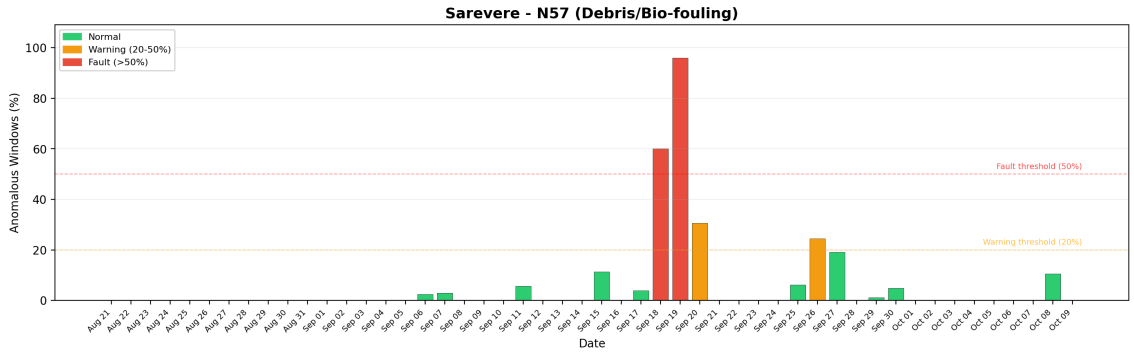


Figure 8. Särevere N57 diagnosis timeline. Peak fault activity of 96% on September 19, classified as debris/bio-fouling.

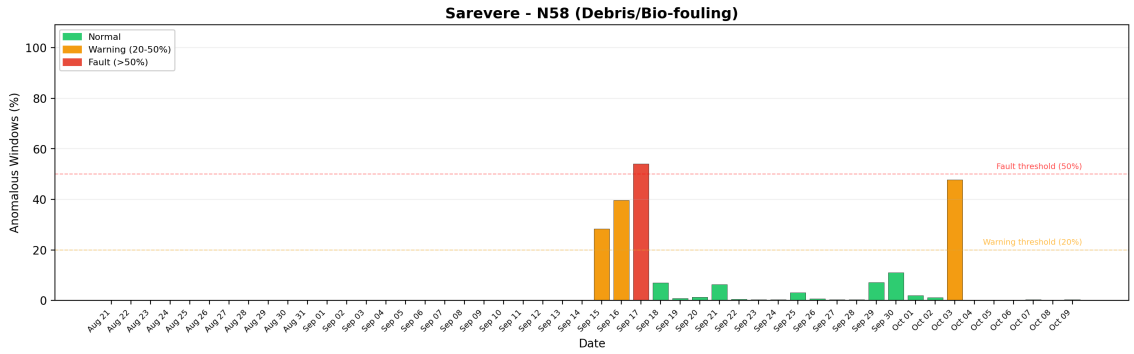


Figure 9. Särevere N58 diagnosis timeline. Fault episode on September 17 at 54%, with warning-level activity on surrounding days.

All detected faults were classified as Debris/Bio-fouling, consistent with the intermittent pattern of fault activity followed by recovery. No membrane fatigue was detected at this site. Notably, all three devices exhibited a gradual downward trend in Hall Z values over the deployment period. However, the cross-sensor validation identified that all three devices drifted in the same direction (Spearman ρ values of -0.88, -0.92, and -0.81 respectively), correctly attributing this shared trend to environmental factors rather than individual device degradation.

5.4.3 Arbavere Results

Data from three co-located devices at Arbavere (N14, N20, N32) provided the most informative validation of the fault classification pipeline, as the three devices exhibited distinctly different fault patterns despite operating in the same water body during the same 56-day period (August 15 – October 9, 2025).

Device N32: Membrane Fatigue

Device N32 was identified as exhibiting membrane fatigue. The Spearman rank correlation analysis detected a significant downward trend in daily Z means ($\rho = -0.72$, $p = 5.19 \times 10^{-8}$), with the baseline shifting from 21.2 mT in the first week to 14.7 mT in the final week, a total decline of 6.5 mT. The null measurement for this device recorded a Hall Z value of 38.3 mT at the zero (undeployed) position.

The irreversibility check identified September 12 as the last day the Z mean returned within two standard deviations of the initial baseline ($Z = 20.23$ mT), establishing a fatigue onset date of September 13. After this date, the Z baseline remained permanently below the

initial level. In total, 20 drift days were flagged between September 17 and October 9.

Several days also exhibited acute bio-fouling events (e.g., August 26 with 100% faulty windows, September 18 with 98.2%), demonstrating that a device can experience both bio-fouling episodes and underlying membrane fatigue simultaneously. The pipeline correctly identified the primary fault type as Membrane Fatigue despite the presence of intermittent bio-fouling events.

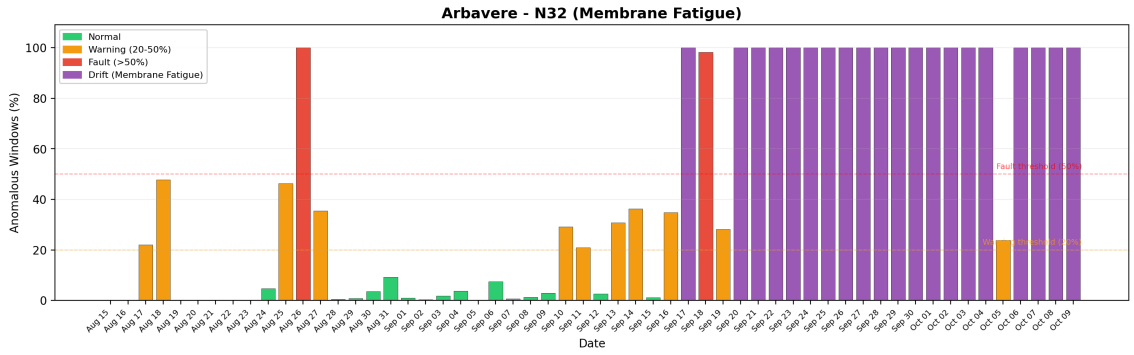


Figure 10. Arbavere N32 diagnosis timeline. The progressive downward drift in Hall Z values ($r_s = -0.72$) indicates membrane fatigue with onset on September 13.

Device N14: Debris/Bio-fouling

In contrast to N32, device N14 exhibited severe debris/bio-fouling from the second day of deployment. Across the 56-day period, 73.3% of all windows were classified as faulty. The device showed recurring cycles of high fault activity (often 100% faulty) followed by brief periods of partial recovery, characteristic of debris accumulating and being partially cleared by water flow or Experts. N14 was deployed near the river bank, where debris accumulation is expected to be more frequent.

Critically, the Spearman trend analysis did not detect membrane fatigue in N14. Although the Z values fluctuated widely (ranging from 8.98 to 26.93 mT), the pattern was intermittent rather than monotonic. The Z mean repeatedly returned toward the baseline before dropping again, ruling out irreversible degradation. The null measurement for N14 recorded $Z = 36.9$ mT.

Device N20: Membrane Fatigue

Device N20 exhibited a pattern similar to N32, with a significant downward trend detected by the Spearman analysis ($\rho = -0.755$, $p = 7.84 \times 10^{-9}$). The Z baseline declined from 16.7 mT to 11.0 mT over the deployment period. The irreversibility check established

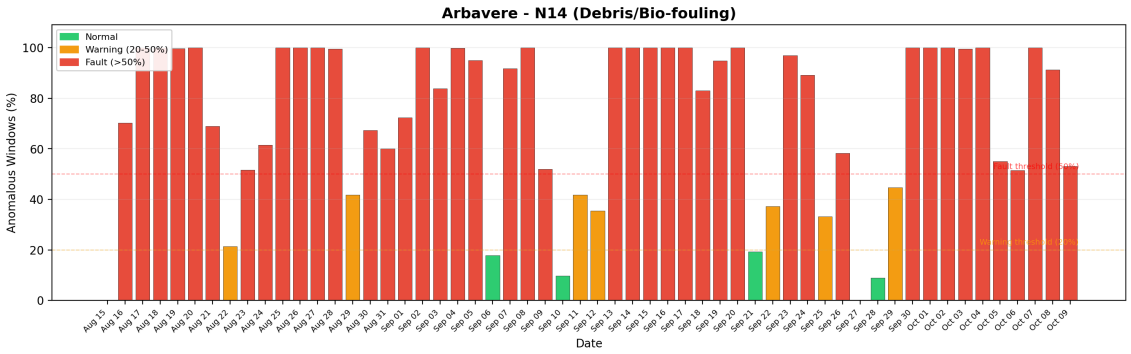


Figure 11. Arbavere N14 diagnosis timeline. Heavy intermittent debris/bio-fouling with 73.3% of days classified as faulty. Recovery days rule out membrane fatigue.

a fatigue onset date of September 18, after which 14 drift days were flagged. The null measurement recorded $Z = 34.3$ mT. N20 was deployed toward the centre of the river, in a similar position to N32.

Cross-Device Comparison

The Arbavere results provide strong evidence for the pipeline's ability to differentiate fault types. Three devices at the same site, during the same period, yielded three distinct diagnoses that align with the physical characteristics of each fault type:

- N32 and N20: progressive, irreversible Z decline consistent with membrane fatigue
- N14: intermittent, reversible Z fluctuations consistent with debris/bio-fouling

The cross-sensor validation confirmed that this issue was not due to environmental drift, as N14 did not show the same consistent downward trend as N32 and N20. This finding provides strong evidence that the classification pipeline is capable of identifying genuine device-level faults rather than merely reflecting shared environmental effects.

6 Conclusion and Outlook

In this work, data obtained from Hydromast water monitoring devices was collected, preprocessed, and analysed to develop an automatic fault detection and classification system. The data under investigation was collected from field deployments at three river discharge monitoring sites in Estonia (Pohjaka, Särevere, and Arbavere) as well as controlled laboratory experiments conducted at the TalTech Centre for Biorobotics. The raw Hall sensor measurements were segmented into 10-second windows, from which 100 statistical features were extracted across three axes and their magnitude. A Random Forest classifier was trained on these features for binary fault detection, followed by physics-based rules for fault type classification and Spearman rank correlation analysis for progressive membrane fatigue detection. The system was validated across four progressively challenging evaluation stages. This study aimed to investigate two research questions focusing on the identification and differentiation of faults in Hydromast devices. The following are the outcomes of this study in answering the anticipated research questions.

RQ1: Which statistical features can be used to identify the faults in recorded measurements?

This work utilises 100 statistical features computed exclusively from the Hall effect sensor data, organised into three categories: basic statistics (mean, standard deviation, range, percentiles, skewness, kurtosis, RMS), temporal features (trend slope, zero crossings, half-window differences, delta measures), and baseline deviation indicators (z-score, deviation from baseline, percentage above baseline). These features were computed for each of the three Hall sensor axes and the derived magnitude signal, resulting in 25 features per source. The Random Forest model trained on these features achieved 95.74% accuracy, 100% recall, and 99.90% ROC-AUC on the held-out test set from the Pohjaka deployment. Laboratory validation confirmed 0% false positives under clean conditions and 95.6–100% detection rates across all fault conditions. The model generalised successfully to unseen devices and sites at Särevere and Arbavere without retraining, demonstrating that the selected features capture general fault signatures rather than site-specific artifacts.

RQ2: How do the statistical features distinguish between membrane fatigue and debris or bio-fouling?

The pipeline differentiates fault types through two complementary approaches. Debris and bio-fouling produce abrupt shifts in Hall sensor readings across multiple axes, which are detected by physics-based classification rules using indicators such as Z mean shift, cross-axis displacement, and variability change. Membrane fatigue, in contrast, produces a gradual monotonic drift in the Hall Z baseline over weeks, which is detected through Spearman rank correlation applied to daily Z means. The Arbavere deployment provided the strongest validation of this differentiation: three devices operating in the same water body during the same period exhibited different fault patterns, devices N32 and N20 showed progressive irreversible Z decline ($\rho = -0.72$ and $\rho = -0.755$), classified as membrane fatigue, while device N14 showed intermittent reversible fluctuations, classified as debris/bio-fouling. The cross-sensor validation correctly distinguished these device-level faults from shared environmental drift, as observed at Särevere where all three devices drifted in the same direction due to seasonal factors. Additionally, laboratory comparison demonstrated that significant membrane fatigue permanently alters the Hall sensor output, with device N56 (most worn) producing 100% faulty readings when measured against a healthy device baseline, confirming that the cumulative effect of fatigue is measurable through the Hall sensor.

In summary, the combination of statistical features with a Random Forest classifier provides reliable fault detection, while the physics-based rules and Spearman trend analysis enable differentiation between fault types based on their distinct signal characteristics. The three-stage pipeline approach proved effective because each stage addresses a different aspect of fault behaviour: the machine learning model detects statistical anomalies, the rule engine interprets their physical cause, and the trend analysis captures slow degradation that is not visible within individual measurement windows

6.1 Limitations

The main limitation of this work is that field fault labels were assigned through expert analysis of sensor data and deployment logs, and were not independently verified. The laboratory experiments partially address this limitation by providing verified ground

truth for debris and bio-fouling conditions, but long-term membrane fatigue could not be replicated in the laboratory setting due to its progressive nature. Although cross-site validation at Särevere and Arbavere demonstrated generalisability, additional training data from more sites and fault types would strengthen the model further.

6.2 Future Outlook

Expanding the training dataset to include more deployment sites and fault types would further improve the generalisability of the model. A practical next step would be to implement the fault detection pipeline on an embedded system within the Hydromast device itself. Deploying the pipeline on the device's microcontroller or on edge computing devices interrogating the Hydromasts would enable real-time fault detection without requiring data transfer to an external server for post-processing. This would allow the device to automatically flag faults or adjust its operation during ongoing deployments, reducing the need for manual data inspection and enabling faster response to device issues in remote monitoring locations. All algorithm comparisons in this work used tree-based methods. Future work could also investigate alternative approaches, such as neural network-based methods, to determine whether they provide additional advantages for this type of data.

References

[1] Lukas Gudmundsson et al. „Globally observed trends in mean and extreme river flow attributed to climate change“. In: *Science* 371.6534 (2021), pp. 1159–1162.

[2] PJ Depetris. *The importance of monitoring river water discharge. Frontiers in Water, 3, 745912*. 2021.

[3] Asko Ristolainen et al. „Hydromast: A bioinspired flow sensor with accelerometers“. In: *Conference on Biomimetic and Biohybrid Systems*. Springer. 2016, pp. 510–517.

[4] Sheryl Coombs and John C Montgomery. „The enigmatic lateral line system“. In: *Comparative hearing: Fish and amphibians*. Springer, 1999, pp. 319–362.

[5] Laurent Delauney, Chantal Compere, and Michel Lehaitre. „Biofouling protection for marine environmental sensors“. In: *Ocean Science* 6.2 (2010), pp. 503–511.

[6] Ardalan F Khalil and Sarkawt Rostam. „Machine learning-based predictive maintenance for fault detection in rotating machinery: A case study“. In: *Engineering, Technology & Applied Science Research* 14.2 (2024), pp. 13181–13189.

[7] Saiprasad Potharaju et al. „A two-step machine learning approach for predictive maintenance and anomaly detection in environmental sensor systems“. In: *MethodsX* 14 (2025), p. 103181.

[8] Bauyrzhan Zhakanov et al. „Real-time estimation of river discharge with entropy based model“. In: *Available at SSRN 5737525* (2025).

[9] L. Piho et al. „Monitoring of coastal waves and velocities with a robotic platform“. In: *OCEANS 2024-Halifax*. 2024.

[10] Margit Egerer et al. „Exploring Hydrodynamic Patterns Using the Hydromast: Varna Port Case Study“. In: *OCEANS 2025 Brest*. 2025, pp. 1–7. DOI: 10.1109/OCEANS58557.2025.11104387.

[11] G. Voulgaris and J. H. Trowbridge. „Evaluation of the Acoustic Doppler Velocimeter (ADV) for Turbulence Measurements*“. In: *Journal of Atmospheric and Oceanic Technology* 15 (1 Feb. 1998), pp. 272–289. ISSN: 0739-0572. DOI: 10.1175/1520-0426(1998)015<0272:EOTADV>2.0.CO;2.

[12] Carlos M. García et al. „Turbulence Measurements with Acoustic Doppler Velocimeters“. In: *Journal of Hydraulic Engineering* 131 (12 Dec. 2005), pp. 1062–1073. ISSN: 0733-9429. DOI: 10.1061/(ASCE)0733-9429(2005)131:12(1062).

[13] A. Tazioli. „Experimental methods for river discharge measurements: comparison among tracers and current meter“. In: *Hydrological Sciences Journal* 56 (7 Oct. 2011), pp. 1314–1324. ISSN: 0262-6667. DOI: 10.1080/02626667.2011.607822.

[14] Adrián Delgado, Ciprian Briciu-Burghina, and Fiona Regan. „Antifouling Strategies for Sensors Used in Water Monitoring: Review and Future Perspectives“. In: *Sensors* 21 (2 Jan. 2021), p. 389. ISSN: 1424-8220. DOI: 10.3390/s21020389.

[15] M. Egerer et al. „Hall effect sensor-based low-cost flow monitoring device: Design and validation“. In: *IEEE Sensors Journal* (2024).

[16] A. Ristolainen et al. „Hydromorphological classification using synchronous pressure and inertial sensing“. In: *IEEE Transactions on Geoscience and Remote Sensing* 56.6 (2018), pp. 3222–3232.

[17] Xiangmin Li et al. „Internet of Things in Marine Environment Monitoring: A Review“. In: *Sensors* 19.7 (2019).

[18] Md. Nazmul Hasan, Sana Ullah Jan, and Insoo Koo. „Sensor Fault Detection and Classification Using Multi-Step-Ahead Prediction with a Long Short-Term Memory (LSTM) Autoencoder“. In: *Applied Sciences* 14 (17 Sept. 2024), p. 7717. ISSN: 2076-3417. DOI: 10.3390/app14177717.

[19] Adisu Mulu Seba, Ketema Adere Gemeda, and Perumalla Janaki Ramulu. „Prediction and classification of IoT sensor faults using hybrid deep learning model“. In: *Discover Applied Sciences* 6 (1 Jan. 2024), p. 9. ISSN: 3004-9261. DOI: 10.1007/s42452-024-05633-7.

[20] „Anomaly Detection Using Explainable Random Forest for the Prediction of Undesirable Events in Oil Wells“. In: *Applied Computational Intelligence and Soft Computing* 2022 (Aug. 2022), pp. 1–14. ISSN: 1687-9732. DOI: 10.1155/2022/1558381.

[21] Gonçalo Jesus, António Casimiro, and Anabela Oliveira. „Using Machine Learning for Dependable Outlier Detection in Environmental Monitoring Systems“. In: *ACM Transactions on Cyber-Physical Systems* 5 (3 July 2021), pp. 1–30. ISSN: 2378-962X. DOI: 10.1145/3445812.

[22] Thomas D. Gautheir. „Detecting Trends Using Spearman's Rank Correlation Coefficient“. In: *Environmental Forensics* 2.4 (2001), pp. 359–362. DOI: 10.1080/713848278. URL: https://www.tandfonline.com/doi/abs/10.1080/713848278.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Syed Muhammad Hamza Bacha

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "Hydromast Fault Detection", supervised by Asko Ristolainen and Laura Piho
    1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
    1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright
2. I am aware that the author also retains the rights specified in clause 1 of the nonexclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

11.05.2026

[1]The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive licence shall not be valid for the period.