

Article

Operational Availability Assessment of Tidal Stream Turbines Using Environmental Data and Fuzzy Logic Inference

Ali Fituri ^{1,2,*}, Abdelouahed Gherbi ¹  and Hmeda Musbah ³¹ Electrical Engineering, École de Technologie Supérieure ÉTS, Montréal, QC H3C 1K3, Canada; abdelouahed.gherbi@etsmtl.ca² Electrical Engineering, Higher Institute of Science and Technology, Gharian, Libya³ Libyan Authority for Scientific Research, Tripoli P.O. Box 80045, Libya; hmeda.musbah@dal.ca

* Correspondence: ali-fituri-ali.fituri.1@ens.etsmtl.ca

Abstract

In hybrid energy systems, maintaining an optimal scheduling strategy for real-time distribution systems, particularly in triple hybrid power generation units, remains a critical challenge. The lack of an efficient real-time observability platform for off-grid hybrid units directly impacts scheduling priorities. In this work, a novel operational condition monitor that has a data-driven predictive mechanism for determining the instant states of each tidal stream turbine is proposed. Environmental variables are first preprocessed using a multivariate fuzzy logic system to generate informative features, which in turn are used by a machine learning classifier to identify the turbine availability states. The classifier is evaluated using K-fold cross-validation and robustness under increasing environmental noise levels. The main contributions of this work are the reduction in uncertainty and the association with real-time operating conditions, which enable optimal scheduling decisions. The baseline XGBoost classifier achieved an F1-score that increased after adding fuzzy-derived features. Comparative evaluation under noise-free and increasing noise levels demonstrates that the proposed framework consistently outperformed the baseline model while maintaining robust classification performance.

Keywords: tidal stream turbines; operational condition assessment; environmental monitoring; machine learning; fuzzy feature engineering

1. Introduction

Data-driven condition monitoring has become an important research topic in complex engineering systems. Machine learning and soft computing techniques have been widely applied for classification and fault diagnosis tasks. However, many existing approaches struggle to capture multivariate patterns under non-stationary disturbances. To address this challenge, this study proposes a fuzzy feature engineering approach integrated with a machine learning classification framework for operational condition assessment. Unlike conventional fault detection methods, which mainly identify the occurrence of faults, the proposed classifier determines predefined operational condition states, enabling continuous evaluation of system behavior across varying operational conditions. The proposed approach is validated using a sensor dataset from tidal stream turbine systems that operate under varying environmental conditions [1,2].

Tidal stream turbines represent a promising renewable resource for off-grid and hybrid power systems due to the high predictability of tidal currents. However, the process



Academic Editor: Decheng Wan

Received: 4 June 2026

Revised: 23 July 2026

Accepted: 23 July 2026

Published: 10 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of harvesting tidal stream power is strongly influenced by severe weather conditions and transient phenomena, such as unsteady waves, which include flow fluctuations and degraded power quality in the generated tidal stream power [3]. The authors in [4] highlighted the influence of the variable environmental conditions on the performance of tidal turbines. Due to variable environmental conditions, maintaining reliable operation for tidal power extraction remains challenging. Falcão, A.F. [5] explains challenges in real-time operation and reliability under environmental variability. High maintenance and repair costs, coupled with performance uncertainty associated with hydrodynamic events, limit the penetration of tidal energy in the power sector by reducing its economic feasibility [4].

Fuzzy logic is implemented in this work to address the challenges of constantly changing environmental data impacting the performance of tidal turbines. By considering tidal speeds along with other relevant measurements, the decision-making process related to the unit commitment problem in a multi-source generation system can reliably determine which turbines are ready to operate, without requiring detailed models or specialized sensors. This real-time, data-driven approach not only improves turbine availability assessment but also enables advanced scheduling strategies and enhances energy management in hybrid or off-grid systems [6,7].

The main contributions of this paper are as follows:

- A multivariate classification framework is proposed for operational condition assessment under non-stationary disturbances.
- A data-driven threshold-based labeling approach is developed to generate binary operational condition labels from four environmental variables.
- The framework's performance is compared with and without fuzzy feature engineering.

Existing condition monitoring approaches rely mainly on turbine health aspects such as temperature and vibration for predicting availability. In contrast, the proposed framework utilizes ambient environmental variables such as tidal stream and wave characteristics to assess operational conditions and varying system behavior. Environmental variables are represented using fuzzy membership functions rather than crisp values, enabling the model to capture uncertainty and gradual transition between operational states.

The novelty of the proposed approach lies in the fact that it does not require detailed system models or hardware modifications while providing real-time information on the readiness of individual units to participate in the islanded multi-source power system. Furthermore, the method supports optimal power management in off-grid autonomous systems and contributes to improved operational efficiency in smart grids.

2. Literature Review

2.1. Environmental Effects on Tidal Turbine Operation

Many renewable energy systems, including tidal and wind turbines, continue to experience operational disruptions primarily driven by environmental forces [8,9]. Tidal turbines are highly susceptible to unsteady ocean waves and fluctuating flow conditions. When operating under irregular sea states, a tidal turbine's ability to sustain power generation while maintaining acceptable power quality becomes limited [8]. Consequently, accurate observation and characterization of ambient environmental conditions are essential for reliable turbine operation. In this context, the proposed fuzzy classifier is designed to determine the instantaneous operational availability status of the tidal turbine. By incorporating environmental descriptors, the classifier enables reliable discrimination between the environmentally induced performance degradation and genuine turbine faults.

2.2. Condition Monitoring and Fault Diagnosis in Turbine Systems

Over the past few years, interest in tidal turbines has grown steadily as a source of reliable renewable energy. However, these turbines operate in challenging marine environments, where unsteady waves, turbulence, and fluctuating flows can significantly affect performance. These environmental factors make it difficult to tell whether a turbine's unusual behavior is due to normal operating variations or an actual fault. To address this, researchers have explored a variety of approaches, from signal-based monitoring to data-driven classifiers, and more recently, hybrid methods that combine operational data with environmental information. In the following review, we focus on techniques that aim to assess the operational condition of tidal turbines while distinguishing environmentally induced variations from genuine faults [10,11]. One popular approach in turbine condition monitoring has been the use of convolutional neural networks (CNNs). These networks are particularly good at identifying patterns in large datasets, such as vibration signals or SCADA measurements, and can automatically extract important features that might indicate a fault. In the context of wind turbines, CNNs have been applied successfully to detect bearing faults, gearbox issues, and other common failure modes. While direct applications to tidal turbines are fewer, early studies suggest that CNNs could be adapted to analyze the complex hydrodynamic signals present in marine environments, offering a promising route for automated condition assessment [12,13]. Building on purely data-driven methods, researchers have also explored hybrid approaches that combine CNNs or other machine learning models with additional sources of information, such as environmental measurements or physics-based models. These hybrid systems are particularly appealing for tidal turbines, where environmental factors like wave height, current speed, and turbulence can mask the presence of a fault. By incorporating environmental data, hybrid models can better separate genuine mechanical or electrical issues from variations caused by changing sea conditions. This not only improves classification accuracy but also reduces false alarms, which are critical in costly offshore operations [14]. Despite the advances in CNNs and hybrid models, one challenge remains: most machine learning classifiers operate with hard decision boundaries, which may struggle with the inherent uncertainty and overlapping conditions in tidal turbine operation. Fuzzy logic offers a complementary solution here. By allowing partial membership in different operating states, fuzzy classifiers can express uncertainty naturally and combine multiple environmental and operational inputs in an interpretable way. This makes them particularly suitable for distinguishing between a turbine that is genuinely faulty and one that is simply experiencing environmental stress, while providing results that operators can understand and act upon [15].

Despite the progress made in turbine condition monitoring and fault classification, several limitations remain when these methods are applied to tidal energy systems. Most existing approaches rely heavily on data-driven classifiers trained under specific operating conditions, which reduces their robustness when environmental variability increases. In highly dynamic marine environments, changes in flow speed, turbulence, and wave activity can produce operational signatures that resemble fault behavior, leading to false alarms or misclassification. Moreover, many advanced machine learning models provide limited interpretability, which restricts their practical use in offshore decision-making. These gaps highlight the need for classification frameworks that can explicitly account for environmental uncertainty while maintaining transparent decision logic. Motivated by these challenges, this study investigates a fuzzy logic-based classifier with environmental data-level fusion to assess the operational availability of tidal turbines and to distinguish environmentally induced performance variations from genuine fault conditions.

2.3. Data-Driven Classification Approaches

The increasing availability of high-resolution sensor data and SCADA records has driven the widespread adoption of data-driven classification techniques for turbine condition monitoring. Early studies focused on conventional machine learning methods, such as artificial neural networks, support vector machines, and decision trees, which rely on manually engineered features extracted from vibration, current, or power signals. These approaches demonstrated promising results for fault detection in wind turbines, particularly for components such as gearboxes, bearings, and generators [16,17].

More recently, deep learning techniques have gained prominence due to their ability to learn discriminative features directly from raw or minimally processed data. Convolutional neural networks (CNNs), in particular, have been shown to outperform traditional classifiers by capturing complex spatial and temporal patterns in vibration signals and time–frequency representations. In wind energy applications, CNN-based models have been successfully applied to fault diagnosis under varying operational loads, motivating their extension to marine energy systems [18,19].

Despite these advances, several challenges limit the direct application of data-driven classifiers to tidal turbines. First, tidal turbine deployments typically generate smaller datasets and labeled fault data are scarce due to the high cost and limited accessibility of offshore installations. Second, most data-driven models implicitly assume that training and operating conditions are statistically similar. In tidal environments, however, changes in flow speed, turbulence intensity, and wave activity can significantly alter signal characteristics, leading to reduced classification robustness. As a result, classifiers trained without explicit environmental context may incorrectly interpret environmentally induced variations as fault conditions, increasing risky false alarms [8,9].

2.4. Hybrid and Environment-Aware Classification Methods

To address the limitations of purely data-driven approaches, researchers have increasingly explored hybrid classification frameworks that integrate machine learning techniques with environmental measurements or physics-based insights. These methods aim to improve robustness by incorporating contextual information such as flow velocity, turbulence intensity, wave parameters, or operational regimes into the classification process. By doing so, hybrid models can better distinguish between abnormal behavior caused by environmental forcing and genuine component degradation [15,16].

Several studies in wind turbine monitoring have demonstrated that incorporating environmental or operational variables into classification models improves fault detection accuracy and reduces misclassification under changing conditions. Approaches include multi-input classifiers that fuse SCADA data with vibration features, as well as hierarchical models that adapt decision thresholds based on operating regimes. Such strategies are particularly relevant to tidal turbines, where environmental variability is inherently stronger and less predictable [9]. Within this context, uncertainty-aware techniques such as fuzzy logic have attracted attention due to their ability to represent gradual transitions between operating states. Unlike hard classifiers with fixed decision boundaries, fuzzy systems allow partial membership across multiple conditions, which better reflects the continuous nature of environmental influence in marine systems. When combined with environmental data-level fusion, fuzzy classifiers provide a transparent and flexible framework for operational condition assessment, enabling more reliable discrimination between environmentally constrained operation and fault-driven behavior. These characteristics make hybrid fuzzy-based approaches well suited for tidal turbine monitoring under highly variable sea states [15].

2.5. Summary and Research Gap

The review of prior work highlights that while conventional and deep learning-based classifiers have demonstrated strong performance in wind turbine fault detection, their direct application to tidal turbines is limited by environmental variability, data scarcity, and the difficulty of distinguishing between operational disturbances and genuine faults. Hybrid approaches that integrate environmental or operational data improve robustness, yet many existing models still rely on rigid decision boundaries and offer limited interpretability. These observations underscore the need for a classification framework that can explicitly account for uncertain and continuously varying environmental conditions while providing transparent and actionable assessments of turbine operational status. Motivated by these gaps, the following sections present a fuzzy logic-based classification approach with environmental data-level fusion, designed to reliably monitor tidal turbines under highly variable sea states.

Recent studies have investigated reliability, availability, and condition monitoring of tidal turbines. Walker and Thies [4] reviewed the reliability challenges associated with the deployment of tidal turbines and highlighted the effects of harsh weather on system availability. Likewise, Grosvenor et al. [20] and Elasha et al. [21] examined condition monitoring approaches with more focus on turbine performance and fault diagnosis. While these studies provide valuable contributions to the field, they mainly rely on turbine operational data and focus on turbine health. Previous studies have also shown that environmental conditions, including tidal stream characteristics and site condition variability, can significantly influence turbine performance [22]. This highlights the need for operational assessment methods that utilize ambient environmental conditions to evaluate turbine availability, specifically when turbine data is unavailable.

- The similarities between wind turbines and tidal turbines

Both systems share similar operating principles, including the components of rotating machinery, fault detection methodologies, and pitch angle control strategies. As a result, condition monitoring and control techniques developed for wind turbines are frequently adapted and tested for tidal turbine applications (Table 1).

Table 1. Some methods in literature [6–8,14].

Data Source	Turbine Type	Best Precision	Approach	Methodology	Results	Research Gab
Scada data	Wind turbine (WT)	92.1%	CNN	using high-frequency SCADA data to classify multiple wind turbine faults	Well features extracted and accurate fault classification	Unknown faults study [6]
Scada data	Marine turbine	96.86%	SVM, DT and KNN	Comparing the three methods	Classification accuracy improved	Exploring non-stationary faults [14]
Experience-based data	Wind turbine (WT)	97.39%	CNN-LSTM	Multi-fault classification	Multiple fault detection with high accuracy	Multi-fault classification [7]
Experimental data	Wind turbine (WT)	99.5%	Variant—CNN	Multi-scale CNN MSFF and FOM	Accuracy improved with single and mixed loads	Adapting model to working condition [8]

3. Methodology

3.1. Data Description and Preprocessing

In this study, the datasets used were collected at the Eastern Passage, NS, Canada. The data collected consists of four different observation sets: tidal speeds, current directions, wave heights and wave directions. Tidal speed is the primary input variable, while current direction is considered a hydrodynamic factor. The wave-height and wave-direction observation sets are included as exogenous parameters in this work. These variables jointly capture both the intensity and directional characteristics of the marine environment and are therefore well suited for tidal condition classification. Input variables were synchronized in time, and missing values were removed to be in line with one-hour intervals prior to analysis. The target variable is a binary label representing two distinct tidal statuses. No temporal aggregation or filtering was applied, as the objective of this work is to evaluate classification robustness at the individual-observation level.

3.2. Fuzzy-Based Approach

Fuzzy logic is used in this study to handle the uncertainty and fluctuating environmental conditions that impact tidal Stream Energy systems. Unlike traditional methods that rely on fixed thresholds, fuzzy membership functions allow the main variables such as tidal speed, wave height, and flow/stream direction to alternate steadily between different operating states [23]. This approach matches well with ocean environments, where tidal and wave conditions usually evolve gradually rather than suddenly across all shifts [24]. Allowing measurements to partially belong to more than one category at the same time is a vital technique that enables the fuzzy framework to represent borderline and disturbed conditions more realistically and to reduce misclassification caused by instant disturbances or noisy measurements [23]. As a result, the proposed fuzzy feature engineering approach, integrated with machine learning classification, offers a physically meaningful representation and robust basis for assessing turbine availability and supporting decision-making strategies in tidal energy systems.

3.3. Noise Modeling and Motivation

In practical marine monitoring systems, sensor measurements are affected by environmental disturbances, calibration errors, and communication noise. To account for these effects, artificial Gaussian noise was added to the raw input features. For each feature x , zero-mean Gaussian noise was injected according to:

$$\tilde{x}_i = x_i + \epsilon_i, \epsilon_i \sim N(0, \sigma_x^2)$$

where the noise standard deviation is defined as: $\sigma_x = \alpha \cdot \text{std}(x)$ where α controls the noise intensity. Rather than being used as a data augmentation strategy, this noise injection serves a diagnostic purpose: it allows the evaluation of model stability under progressively degraded measurement conditions. By explicitly controlling the noise level, the sensitivity of different feature representations can be examined in a systematic manner.

3.4. Fuzzy Feature Representation

To model uncertainty more explicitly, a fuzzy logic-based feature transformation was introduced. Instead of relying solely on crisp numerical values, each continuous feature was mapped to three linguistic concepts: Low, Medium, and High. These concepts were represented using triangular membership functions.

For a given feature value x , the triangular membership function is defined as:

$$\mu(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

where a , b , and c denote the lower bound, center, and upper bound of the triangle, respectively.

For each feature, the parameters were selected as:

$$\text{Low : } (a = x_{\min}, b = x_{\min}, c = x_{\text{mean}})$$

$$\text{Medium : } (a = x_{\min}, b = x_{\text{mean}}, c = x_{\max})$$

$$\text{High : } (a = x_{\text{mean}}, b = x_{\max}, c = x_{\max})$$

As a result, each raw feature produces three fuzzy membership values:

$$f_i = [\mu_{\text{Low}}(x_i), \mu_{\text{Medium}}(x_i), \mu_{\text{High}}(x_i)]$$

This representation allows a single observation to partially belong to multiple categories at the same time, which is particularly important for borderline cases where hard thresholds may be misleading. The fuzzy features were concatenated with the original noisy features to form a combined input representation:

$$Z_i = [\tilde{X}_i, f_i]$$

Environmental data, including tidal speed, current direction, wave height, and wave direction, are first collected and may include noise to reflect real-world conditions. These inputs are transformed into fuzzy membership features representing Low, Medium, and High states, which are then combined with the original measurements. The resulting input vector is fed into the XG-Boost classifier to predict the turbine's operational state (Available or Not Available).

3.5. Class Imbalance Treatment

The dataset exhibits a moderate imbalance between the two tidal classes. To prevent the classifier from favoring the majority class, class weighting was incorporated directly into the learning algorithm. The imbalance ratio was computed as: $\omega = \frac{N_0}{N_1}$, where N_0 and N_1 denote the number of samples in classes 0 and 1, respectively. This weight was applied to scale the loss associated with the minority class, preserving the original data distribution while encouraging balanced decision boundaries.

3.6. Classification Model

Extreme Gradient Boosting (XGBoost) was selected as the classification model due to its strong performance on structured data and its ability to capture nonlinear interactions between features. The model minimizes a regularized logistic loss function of the form:

$$\gamma = \sum_i l(y_i, \hat{y}_i) + \sum_k \varphi(f_k)$$

where γ denotes the binary logistic loss and φ represents the regularization term controlling model complexity. Two feature configurations were evaluated:

A Baseline Model Using Noisy Raw Features Only

A fuzzy-enhanced model combining raw features with fuzzy membership features.

3.7. Cross-Validation

Model performance was evaluated using five-fold stratified cross-validation. This strategy ensures that class proportions remain consistent across training and testing splits, reducing variance in the estimated performance metrics. For each fold, the model was trained on four subsets and evaluated on the remaining subsets. Performance was quantified using macro-averaged precision, recall, and F1-score, defined as:

$$F1_{macro} = \frac{1}{C} \sum_{C=1}^C \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c}$$

where $C = 2$ represents the number of classes. Macro averaging was chosen to give equal importance to both classes, independently of their frequency in the dataset.

3.8. Error Analysis Using Confusion Matrices

In addition to scalar performance metrics, confusion matrices were generated for each cross-validation fold. For a binary classification problem, the confusion matrix is defined as:

$$CM = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$$

These matrices provide insight into the types of classification errors made by the model and allow for a more detailed interpretation of performance differences between the baseline and fuzzy-enhanced representations.

3.9. Robustness Analysis Under Increasing Noise

To further assess robustness, the evaluation was repeated across multiple noise levels $\alpha \in \{0, 0.1, 0.3, 0.5, 0.7\}$. For each level, the average macro F1 score was recorded:

$$F1(\alpha) = \frac{1}{K} \sum_{k=1}^K F1_k(\alpha), K = 5.$$

This analysis directly evaluates how classification performance degrades as uncertainty increases, demonstrating the added resilience of fuzzy-enhanced features for reliable turbine operational assessment under realistic marine conditions.

- Clarification of terminology and data leakage prevention:
 - Terminology of the proposed method. We agree with the reviewer that the proposed framework is more accurately described as a fuzzy feature engineering approach combined with supervised machine learning classification rather than a conventional fuzzy inference system. The objective of the fuzzy component in this work is to transform the original environmental variables into fuzzy feature representations using triangular membership functions. This transformation provides a more informative representation of uncertainty in the input data prior to the classification process. Unlike classical Mamdani or other fuzzy inference systems, the proposed method does not include components such as a fuzzy rule base, an inference engine, rule aggregation, or a defuzzification process, which are fundamental elements of a conventional fuzzy inference system [25,26]. Instead, the generated fuzzy membership values are directly used as additional input features for conventional machine learning classifiers, including XG-Boost, Random Forest, Support Vector Machine, Gradient Boosting, and Light-GBM. This design is consistent with fuzzy feature engineering approaches, where fuzzy logic is employed as a feature transformation technique rather than as a complete decision-making framework. Accordingly, the manuscript has been revised

throughout to consistently describe the proposed methodology as a fuzzy feature engineering approach integrated with machine learning classification, thereby eliminating any ambiguity regarding its methodological framework and aligning the terminology with established fuzzy systems literature [25,26].

- Prevention of data leakage during cross-validation. We agree that preventing data leakage is essential for obtaining unbiased and reproducible performance estimates. Since the triangular membership functions are defined using data-dependent statistics (minimum, mean, and maximum values), computing these parameters from the entire dataset prior to cross-validation would unintentionally introduce information from the test data into the feature construction process. As a result, this would lead to information leakage and potentially optimistic performance estimates [27,28]. To prevent this, the implementation follows a strict fold-aware feature engineering procedure. For each cross-validation fold, the fuzzy membership parameters are estimated exclusively from the training subset and are then applied, without modification, to generate the fuzzy feature representations for both the corresponding training and test subsets. Consequently, no information from the test fold contributes to estimating the membership functions or constructing the fuzzy features, ensuring complete separation between training and testing data throughout the evaluation process. This procedure is consistent with best practice for data-dependent preprocessing within cross-validation, whereby all preprocessing parameters are learned exclusively from the training data and subsequently applied to unseen test samples [27–30].

4. Results and Discussion

This section presents and discusses the experimental results obtained from the proposed tidal-condition classification framework. The analysis focuses on three main aspects: (i) the effect of fuzzy feature representation, (ii) the role of hyperparameter tuning, and (iii) model robustness under increasing noise measurement.

Figures 1–4 illustrate fuzzy membership functions constructed for tidal speed, wave height, mean current direction, and mean wave direction after noise injection. Each variable is smoothly mapped into Low, Medium, and High membership degrees using triangular functions. These visualizations confirm that the fuzzy representation preserves gradual transitions between physical regimes rather than imposing sharp decision boundaries. This behavior is consistent with the continuous nature of oceanographic processes, where changes in tidal conditions rarely occur abruptly.

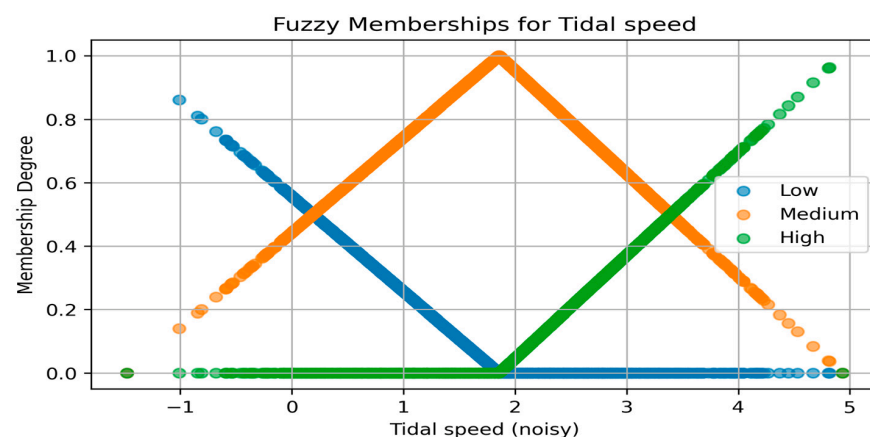


Figure 1. Fuzzy Memberships: Tidal speed.

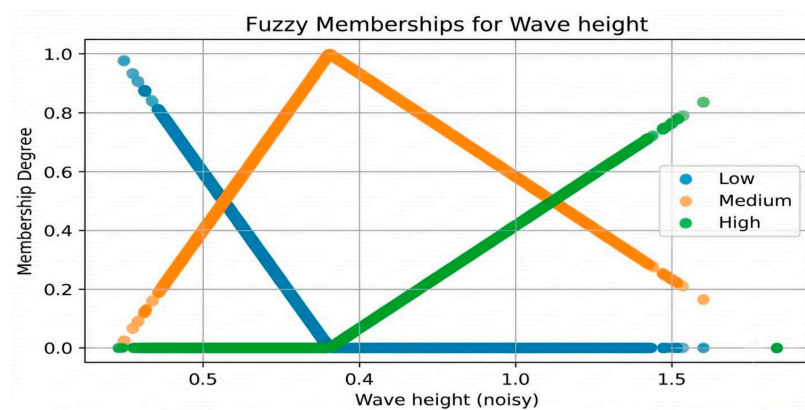


Figure 2. Fuzzy Memberships: Wave height.

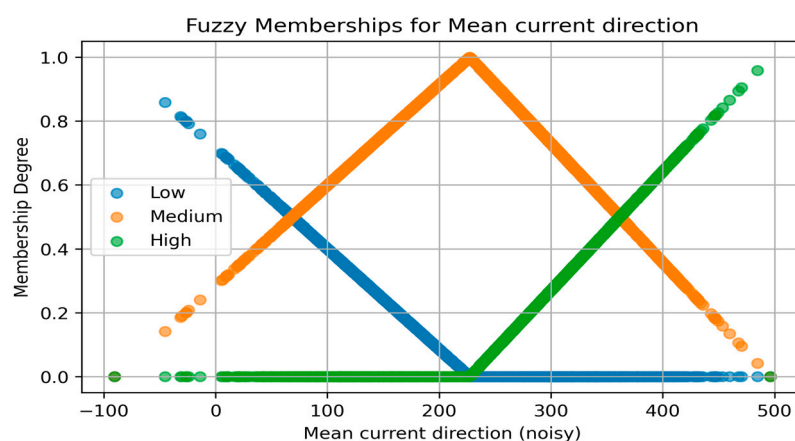


Figure 3. Fuzzy Memberships: Mean current direction.

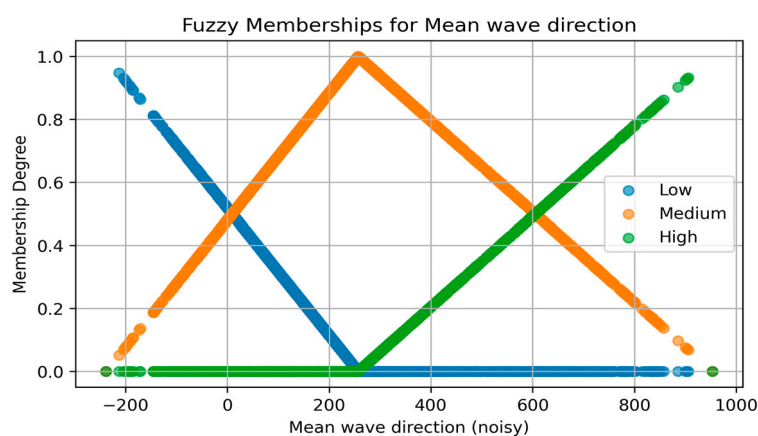


Figure 4. Fuzzy Memberships: Mean wave direction.

Importantly, individual observations can simultaneously belong to multiple fuzzy categories with different strengths, allowing ambiguous or borderline cases to be represented more faithfully than with crisp numerical values alone. Next are diverse scenarios for classifier performance evaluation:

4.1. Classification Scenarios

4.1.1. Raw Feature Baseline

In this classification scenario, model performance was evaluated using the macro F1 score as the primary summary metric, alongside precision and recall. Macro F1 was selected

due to its ability to equally weight all classes, providing a balanced assessment of performance under potential class imbalance. The baseline classifier trained on noisy raw features achieved stable performance across the five cross-validation folds (Figures 5–9 and Table 2). Macro F1 scores ranged between 0.8425 and 0.8721, with an average value of 0.8579. Precision and recall remained well balanced, indicating that the model did not systematically favor one class over the other. While this baseline performance is strong, noticeable variation across folds suggests sensitivity to data partitioning and noise, which is expected when relying solely on crisp sensor measurements.

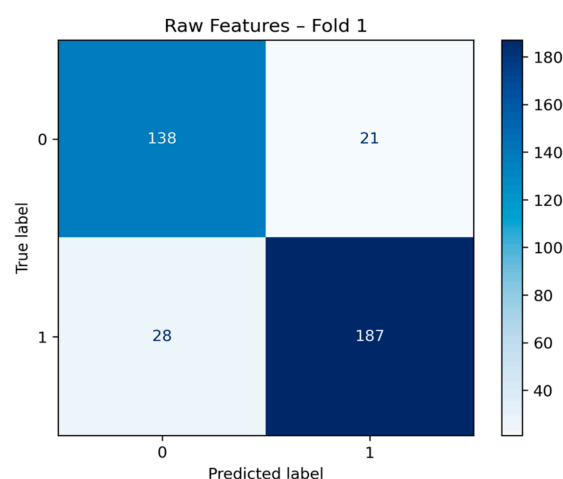


Figure 5. Cross-validation folds for baseline classifier with noisy raw features—Fold 1.

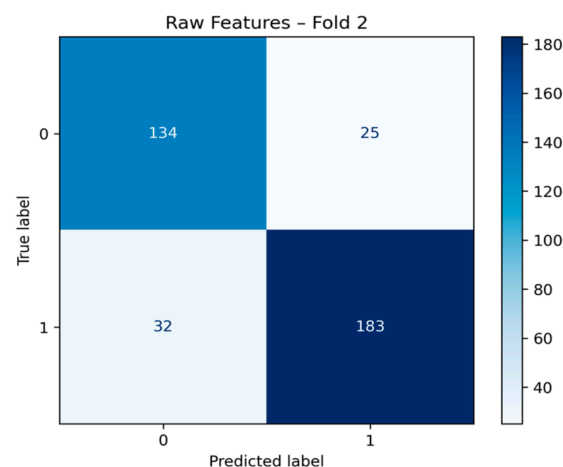


Figure 6. Cross-validation folds for baseline classifier with noisy raw features—Fold 2.

Table 2. The evaluation metrics for the five folds.

	Macro F1	Precision	Recall
Fold 1	0.8667	0.8652	0.8688
Fold 2	0.8449	0.8435	0.8470
Fold 3	0.8721	0.8706	0.8743
Fold 4	0.8634	0.8620	0.8653
Fold 5	0.8425	0.8460	0.8402

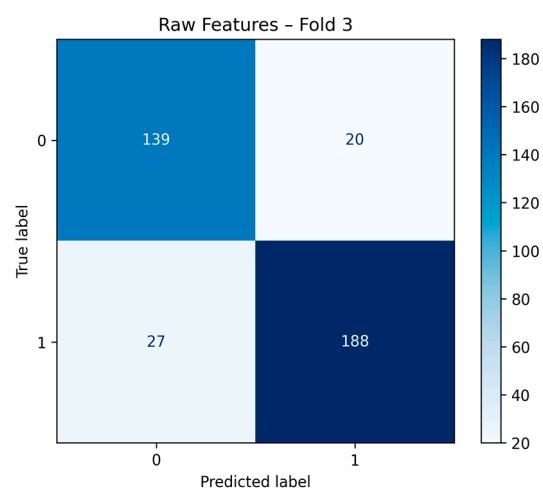


Figure 7. Cross-validation folds for baseline classifier with noisy raw features—Fold 3.

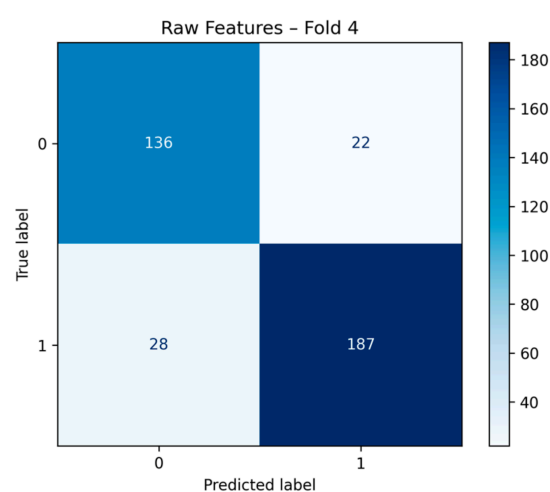


Figure 8. Cross-validation folds for baseline classifier with noisy raw features—Fold 4.

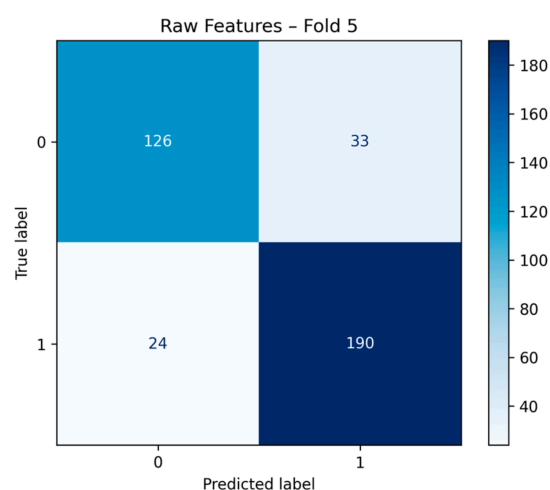


Figure 9. Cross-validation folds for baseline classifier with noisy raw features—Fold 5.

4.1.2. Effect of Adding Fuzzy Features

When fuzzy membership features were added as in Figures 10–14, the overall performance improved slightly, with an average macro F1 score of 0.8592. Although the absolute gain over the baseline is modest, two important qualitative improvements are observed. First, the fold-to-fold variability is reduced, indicating improved stability and

generalization. Second, recall values become more consistent across folds, suggesting that the fuzzy-enhanced model is better at handling borderline tidal conditions where class boundaries are less distinct.

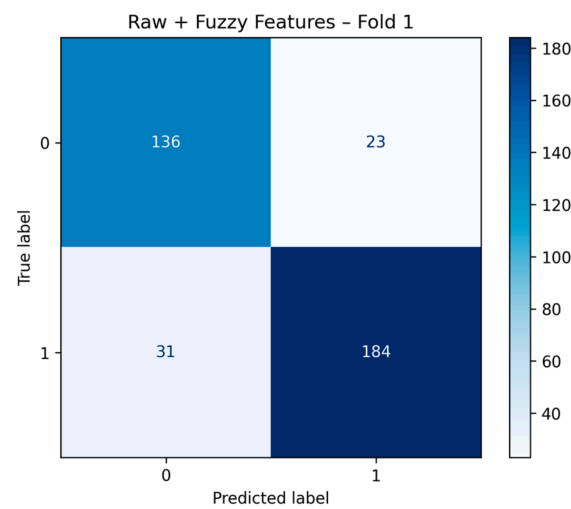


Figure 10. The performance after adding Fuzzy features—Fold 1.

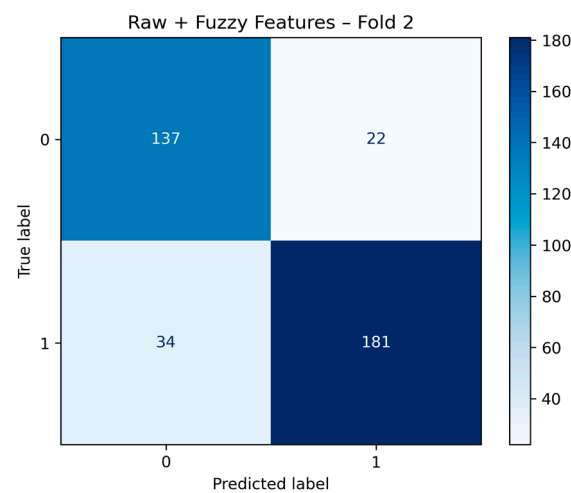


Figure 11. The performance after adding Fuzzy features—Fold 2.

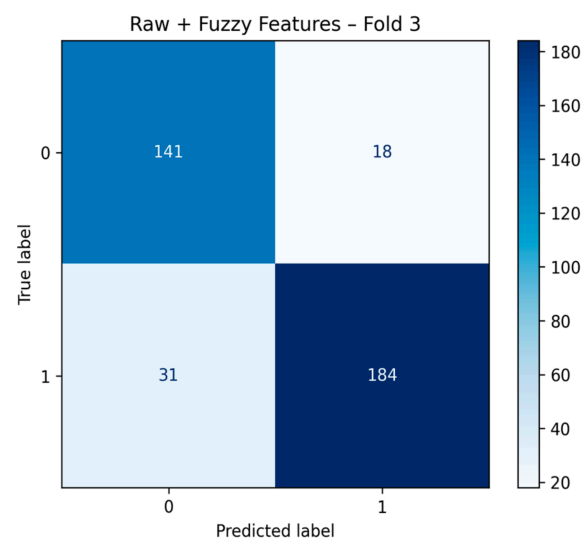


Figure 12. The performance after adding Fuzzy features—Fold 3.

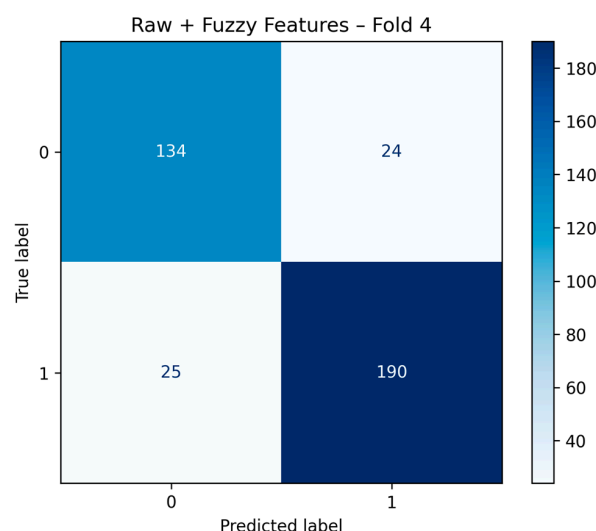


Figure 13. The performance after adding Fuzzy features—Fold 4.

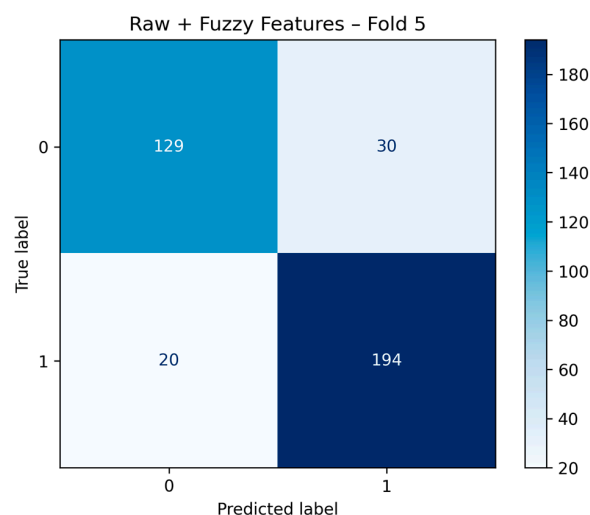


Figure 14. The performance after adding Fuzzy features—Fold 5.

The overall performance summary confirms that the fuzzy-enhanced model outperforms the raw-feature baseline across all reported metrics (macro F1, precision, and recall), although by a small margin. This indicates that fuzzy features contribute complementary information rather than replacing the predictive power of the original physical variables. Table 3 demonstrates the evaluation after adding fuzzy effect:

Table 3. The evaluation metrics for the five folds (2nd scenario).

	Macro F1	Precision	Recall
Fold 1	0.8532	0.8516	0.8556
Fold 2	0.8482	0.8464	0.8517
Fold 3	0.8672	0.8653	0.8713
Fold 4	0.8656	0.8653	0.8659
Fold 5	0.8618	0.8659	0.8589

Figure 15 shows the effect of progressively increasing Gaussian noise on classification performance. As expected, both models experience performance degradation as noise levels increase. However, the fuzzy-enhanced model consistently maintains a higher

macro F1 score across all noise levels. This behavior suggests that fuzzy membership encoding acts as a form of uncertainty-aware smoothing, reducing the sensitivity of the classifier to perturbations in individual measurements. The benefit becomes more visible at moderate noise levels, reinforcing the motivation for fuzzy modeling in real-world sensor-driven applications.

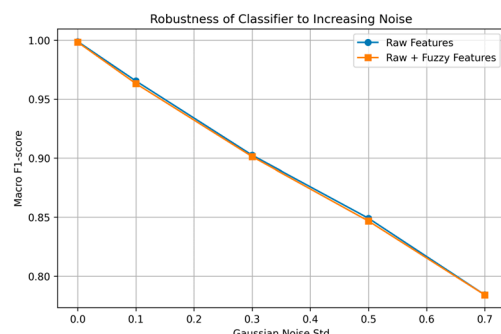


Figure 15. Robustness of classifier for increasing noise.

4.1.3. Tuned Raw Feature Model

After hyperparameter tuning, the raw feature model exhibits a clear decline in performance Figures 16–20 and Table 4, with the average macro F1 score dropping to 0.8331. This degradation indicates that tuning amplifies the model’s sensitivity to noise and data variability when only crisp features are used. In this case, optimization appears to push the model toward configurations that overfit noisy patterns, reducing its ability to generalize.

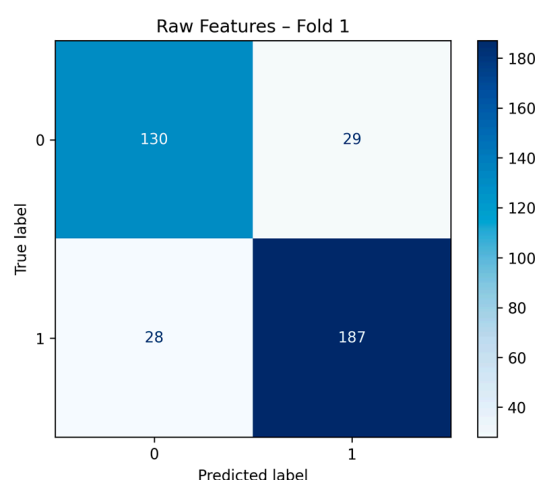


Figure 16. Fold 1.

Table 4. Cross-validation results for tuned raw feature model.

	Macro F1	Precision	Recall
Fold 1	0.8440	0.8443	0.8437
Fold 2	0.8327	0.8316	0.8386
Fold 3	0.8260	0.8246	0.8282
Fold 4	0.8415	0.8402	0.8433
Fold 5	0.8214	0.8223	0.8207

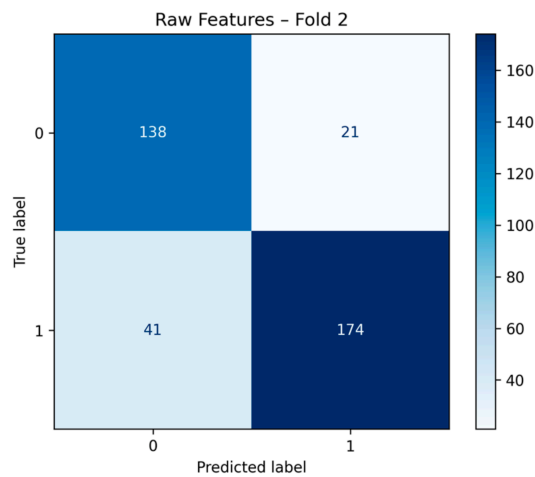


Figure 17. Fold 2.

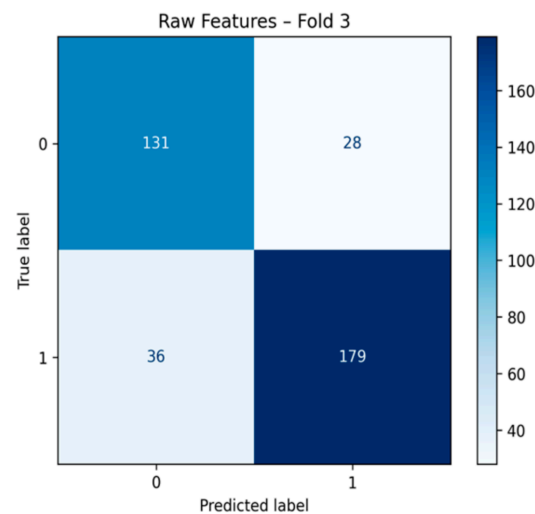


Figure 18. Fold 3.

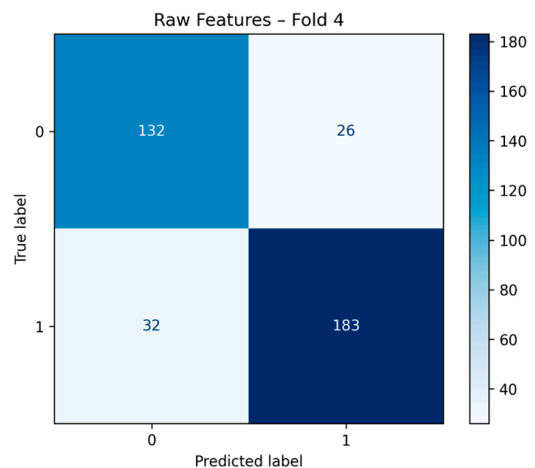


Figure 19. Fold 4.

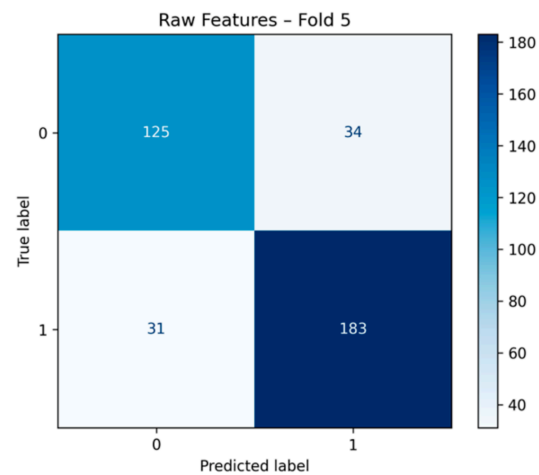


Figure 20. Fold 5.

4.1.4. Tuned Fuzzy Model

In contrast, the tuned fuzzy-enhanced model (Figures 21–25 and Table 5) achieves a substantially higher average macro F1 score of 0.8518. While this does not exceed the untuned fuzzy model's peak fold performance, it significantly outperforms the tuned raw model by nearly 2% of points. More importantly, the tuned fuzzy model maintains balanced precision and recall and shows less variation across folds. This indicates that fuzzy features provide a smoother and more structured feature space, making hyperparameter optimization more effective and less prone to instability.

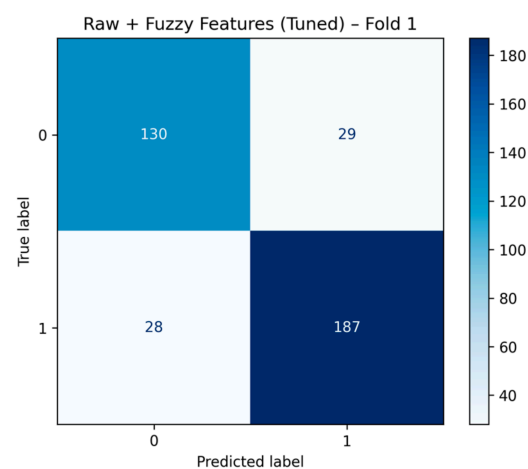


Figure 21. Fold 1.

Table 5. Results for tuned raw-fuzzy hybrid model.

	Macro F1	Precision	Recall
Fold 1	0.8440	0.8443	0.8437
Fold 2	0.8566	0.8548	0.8612
Fold 3	0.8584	0.8571	0.8602
Fold 4	0.8525	0.8511	0.8543
Fold 5	0.8476	0.8460	0.8506

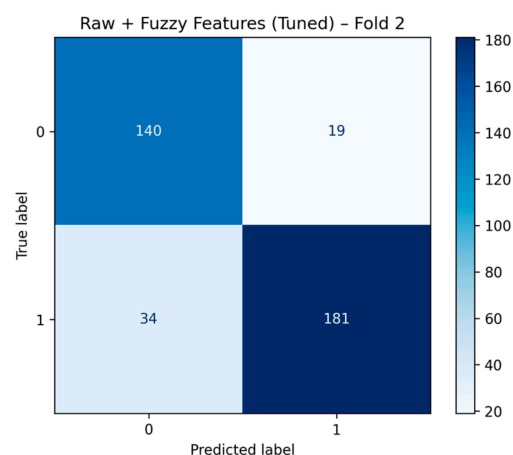


Figure 22. Fold 2.

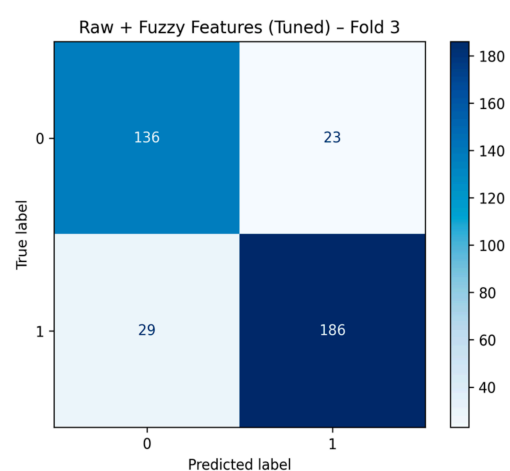


Figure 23. Fold 3.

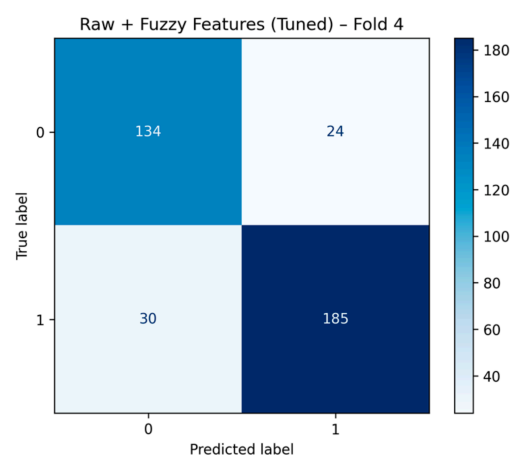


Figure 24. Fold 4.

Robustness to Increasing Noise after Tuning.

The robust analysis repeated after tuning, as shown in Figure 26, further highlights the advantage of fuzzy features. As noise increases, the tuned raw feature model degrades rapidly, whereas the tuned fuzzy-enhanced model retains higher performance across all noise levels. This confirms that the benefits of fuzzy representation persist even after optimization and are not limited to a specific parameter configuration. Figure 21 illustrates the overall Performance Summary.

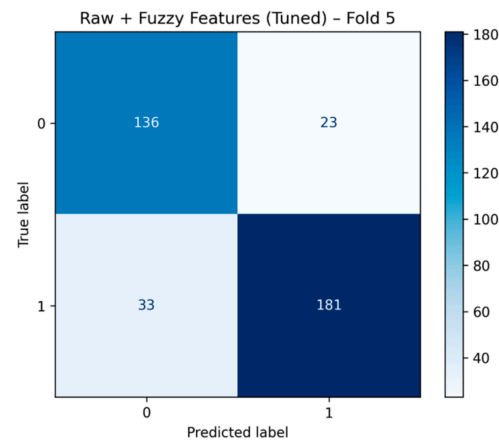


Figure 25. Fold 5.

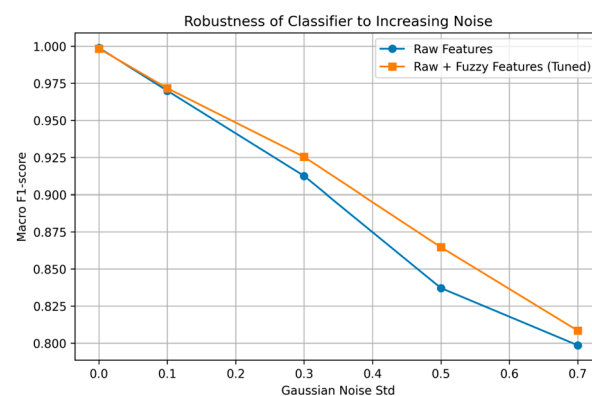


Figure 26. Robustness to increasing noise.

4.2. Overall Discussion and Model Ranking

Taken together, the results demonstrate that fuzzy feature augmentation does not primarily aim to maximize nominal accuracy, but rather to improve stability, robustness, and optimization reliability. Based on macro F1 score, consistency across folds, and robustness to noise, the models can be ranked as follows:

- Tuned Raw-Fuzzy Features. Best Overall Balance of Accuracy and Robustness
- Untuned Raw-Fuzzy Features. Strong and Stable Baseline
- Untuned Raw Features. Competitive but Noise-Sensitive

Tuned Raw Features. Lowest Performance and Least Robust

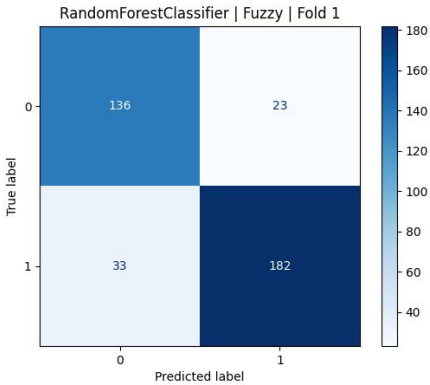
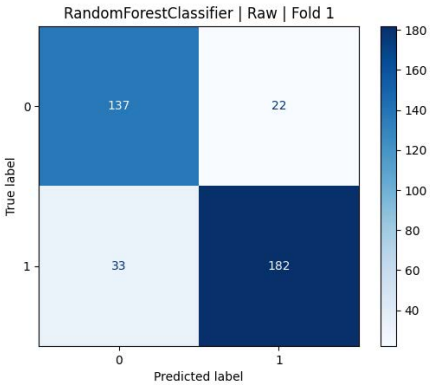
These findings highlight that fuzzy membership features are particularly valuable in noisy environments and play a crucial role in enabling effective hyperparameter tuning. For real-world tidal monitoring systems, where uncertainty is unavoidable, the fuzzy-enhanced approach offers a more reliable and interpretable solution than models relying solely on raw physical measurements.

To further evaluate the proposed fuzzy feature representation, additional experiments were conducted using multiple machine learning classifiers under identical experimental settings. Each classifier was trained and evaluated using both the original raw features and the proposed fuzzy-enhanced features through five-fold stratified cross-validation. The statistical significance of the observed differences was assessed using paired *t*-tests and Wilcoxon signed-rank tests based on the fold-wise macro F1 scores.

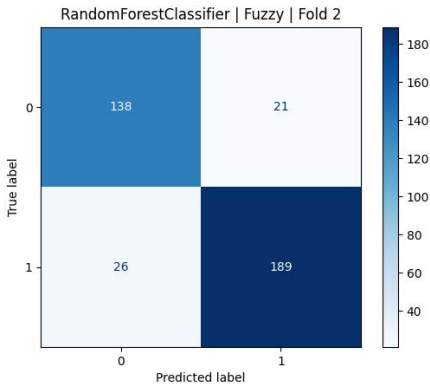
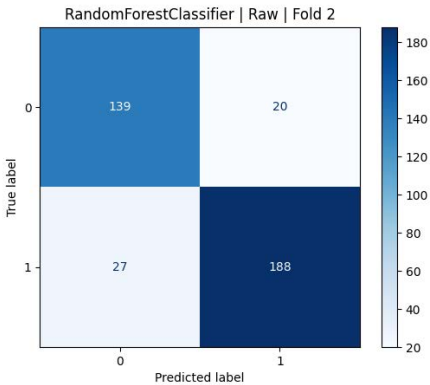
Confusion matrix and the results from the additional models are:

1. Random Forest Model [RF]:

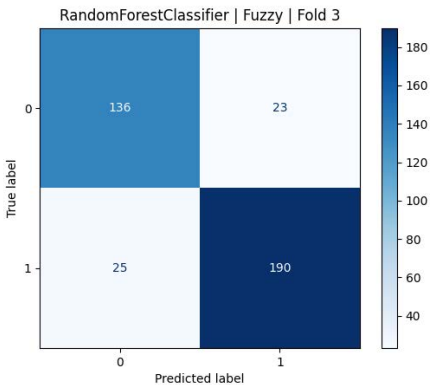
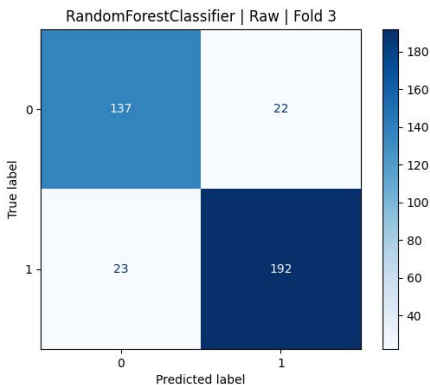
Fold 1



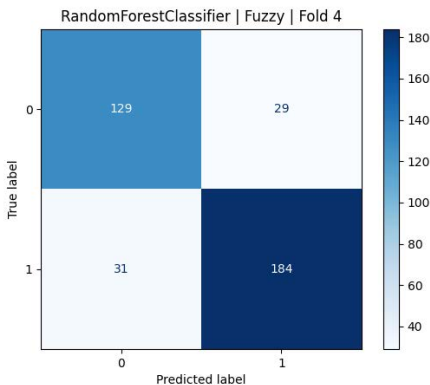
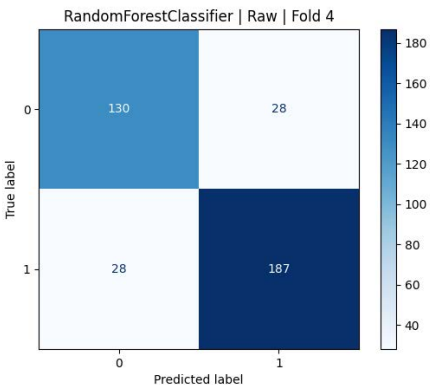
Fold 2



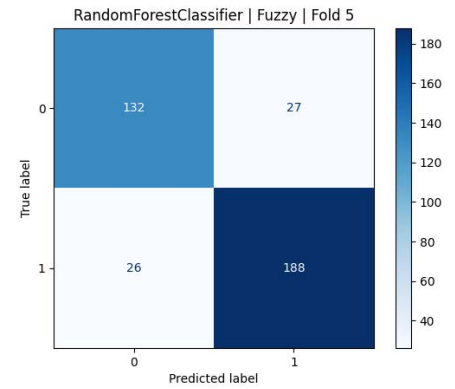
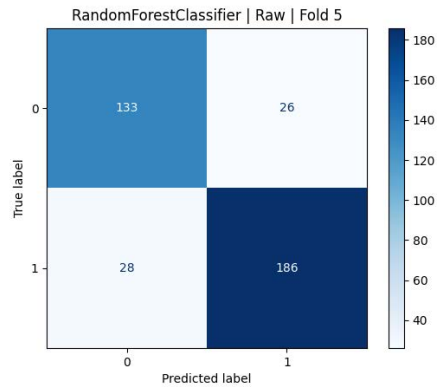
Fold 3



Fold 4

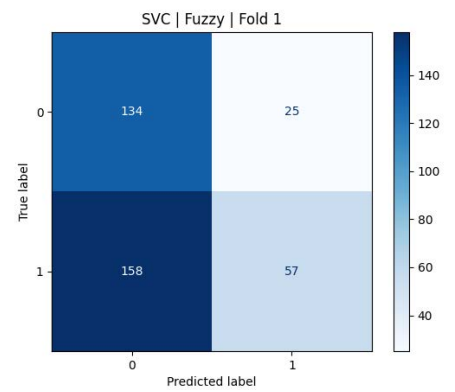
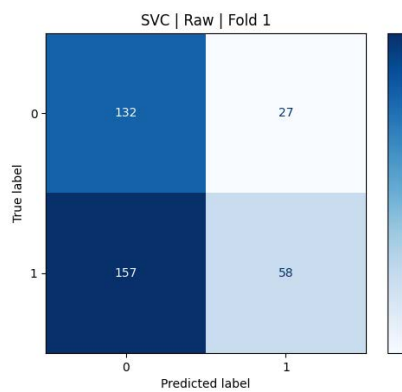


Fold 5

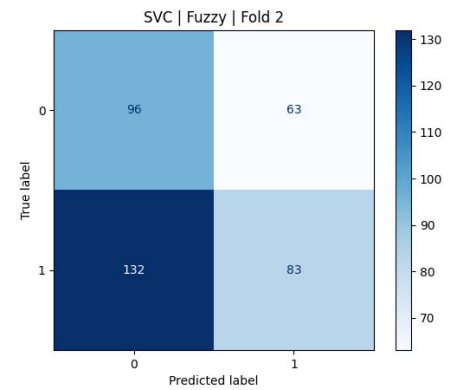
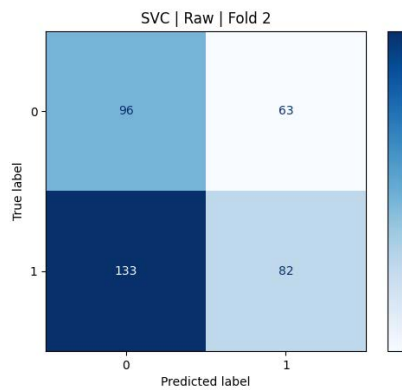


2. Support Vector Machine:

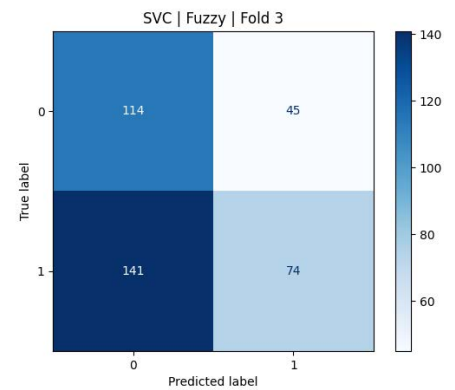
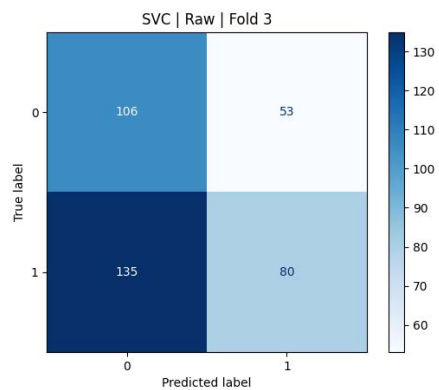
Fold 1



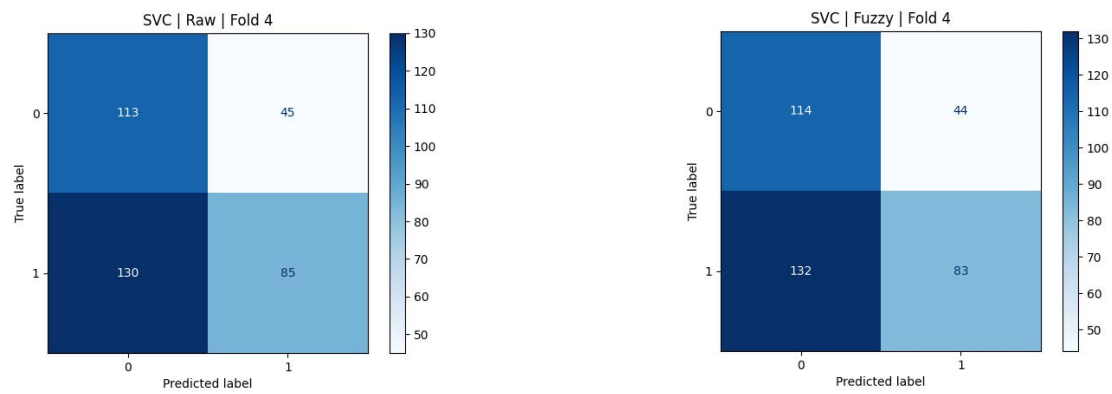
Fold 2



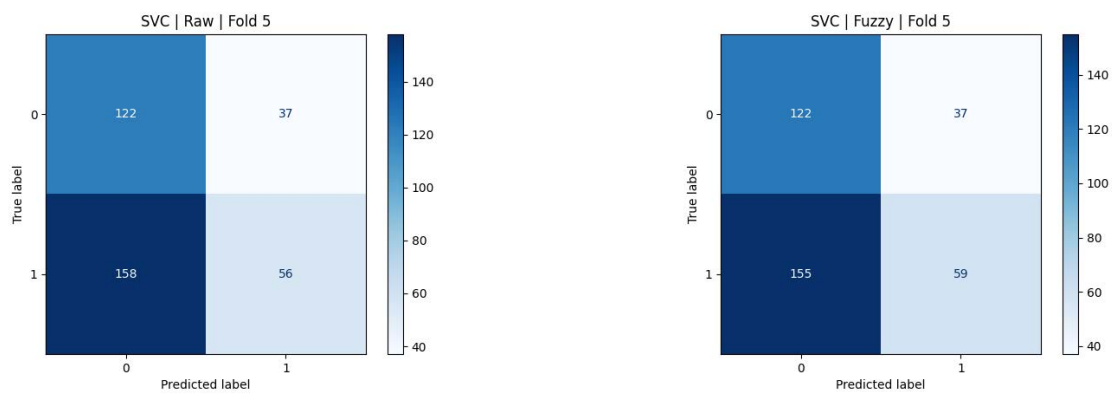
Fold 3



Fold 4

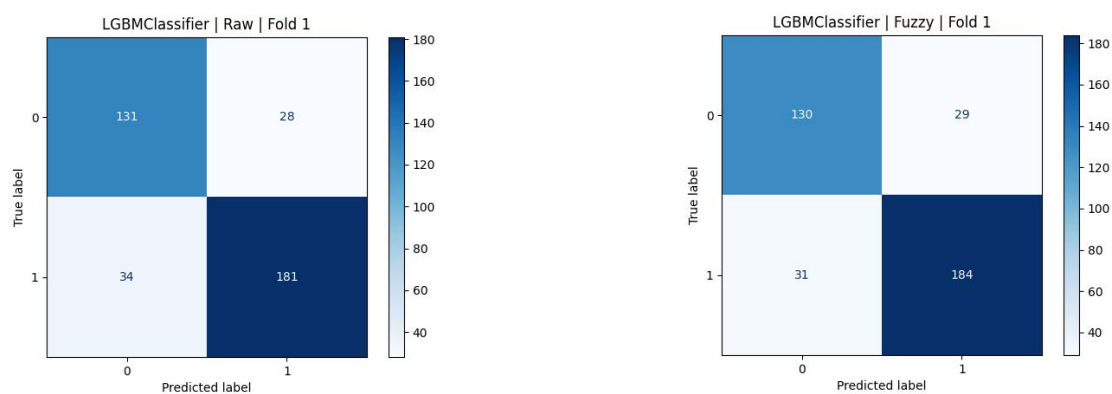


Fold 5

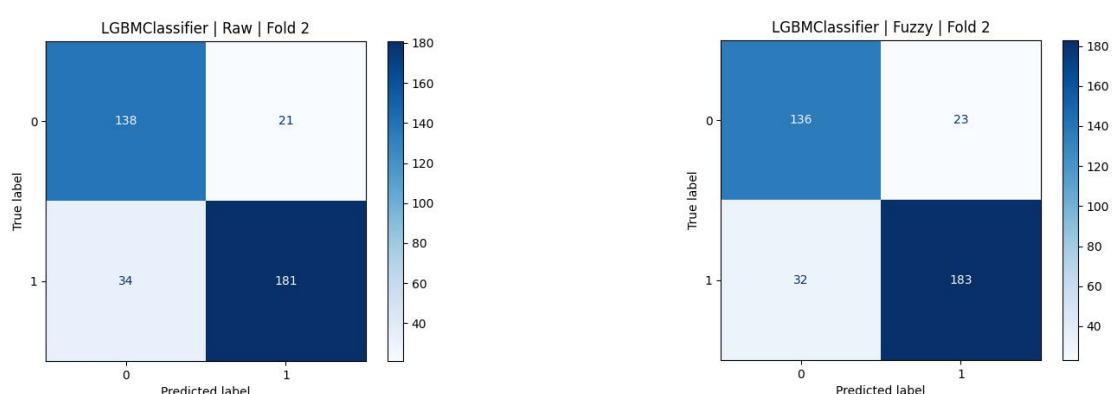


3. Light Gradient Boosting Machine [LGBM]

Fold 1



Fold 2



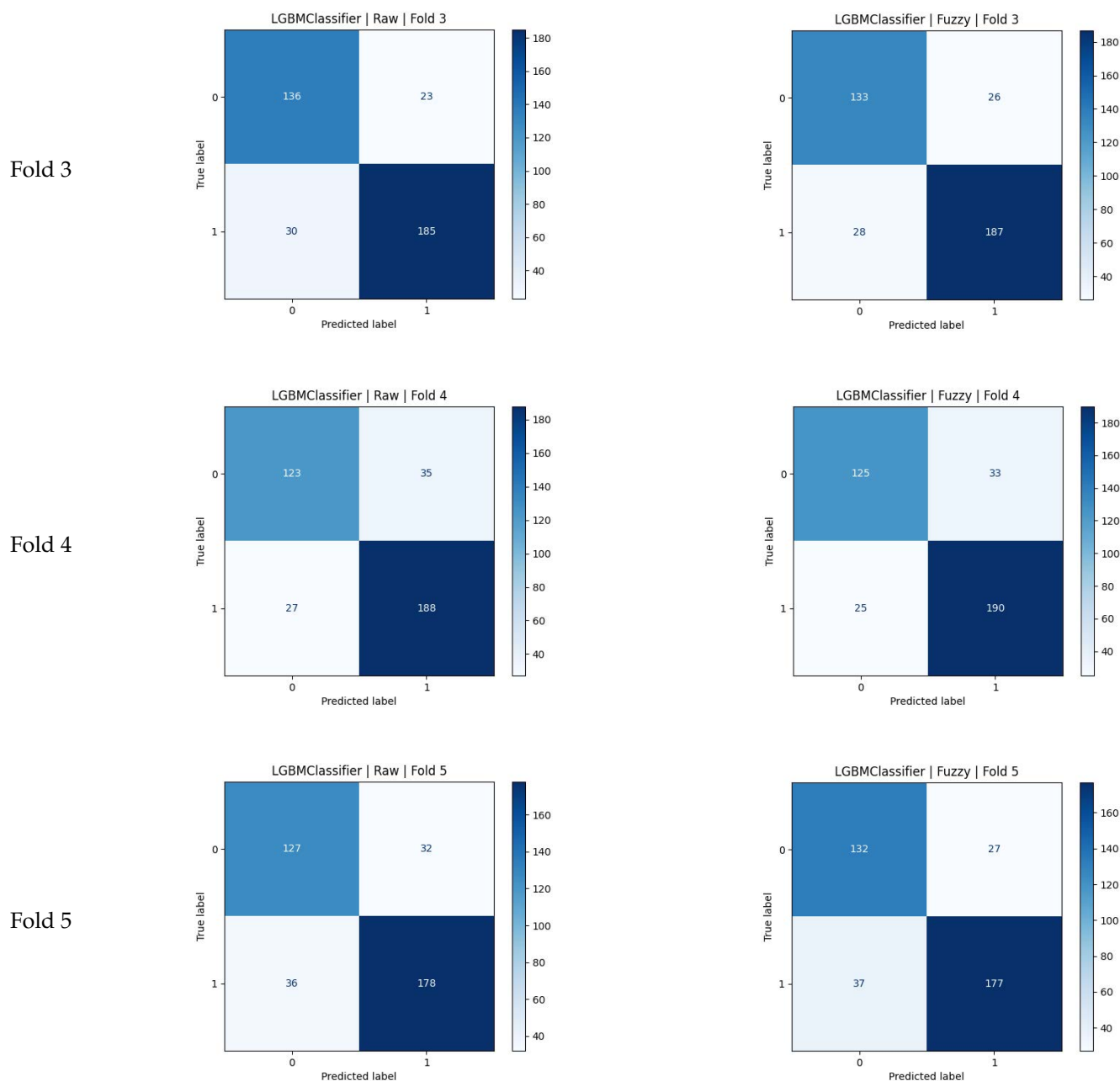


Table 6 illustrates F1 for the three machine learning classifiers:

- Statistical comparison of the four models:

	paired <i>t</i> -test	wilcoxon test
XG-Boost	<i>p</i> -value: 0.5694	<i>p</i> -value: 0.6250
RF	<i>p</i> -value: 0.1858	<i>p</i> -value: 0.1875
SVM	<i>p</i> -value: 0.3068	<i>p</i> -value: 0.4375
LGBM	<i>p</i> -value: 0.1832	<i>p</i> -value: 0.3125

Notice: No significant difference detected.

For the XGBoost classifier, the average macro F1 score increased slightly from 0.8579 to 0.8592. However, the paired *t*-test ($t = 0.6190$, $p = 0.5694$) and the Wilcoxon signed-rank test ($W = 5.0000$, $p = 0.6250$) indicated that the improvement was not statistically signifi-

cant. Similar results were obtained for the additional classifiers. Random Forest achieved macro F1 scores of 0.8597 and 0.8558, Support Vector Machine obtained 0.4893 and 0.4918, and LightGBM achieved 0.8362 and 0.8410 using the raw and fuzzy-enhanced features, respectively. For all three classifiers, both statistical tests produced p -values greater than 0.05, indicating that the observed performance differences were not statistically significant.

Table 6. The three machine learning classifiers.

	F1 Raw Mean	F1 Fuzzy Mean
RF	0.8597	0.8558
SVM	0.4893	0.4918
LGBM	0.8362	0.8410

Overall, the results show that the proposed fuzzy feature representation can be effectively integrated with different machine learning classifiers while maintaining comparable predictive performance. Although the improvements in macro F1 score are modest and not statistically significant, the fuzzy-enhanced models demonstrated greater robustness under increasing Gaussian noise. Therefore, the primary contribution of this work is not a statistically significant improvement in classification accuracy, but the development of a robust and interpretable environmental feature representation that improves model stability under uncertain and noisy operating conditions.

5. Conclusions

This paper has addressed the problem of classifying tidal turbine operational availability status as a function of varying ocean-state and underwater environmental conditions, rather than internal mechanical failures. A fuzzy-based classifier was employed in this study to evaluate the improvement in overall classification reliability compared to raw data-based models. Finally, the robustness of the classifier is further examined under increasing noise levels, and a comparative study with XG-Boost highlights the advantages of the fuzzy feature engineering approach integrated with machine learning classification over the raw-feature baseline. Overall, the findings show that the proposed classifier performs consistently well and remains robust in both normal and noise-injected conditions. Future work could focus on integrating the proposed classifier into hybrid autonomous and off-grid power systems to further explore its practical applicability. In addition, it would be valuable to compare the performance of the fuzzy-based classifier with other classification techniques to better understand its strengths and limitations.

Author Contributions: Conceptualization, A.F.; Methodology, A.F. and H.M.; Software, A.F. and H.M.; Validation, A.F. and A.G.; Formal analysis, A.F., A.G. and H.M.; Investigation, A.F.; Writing—original draft, A.F.; Writing—review & editing, A.F. and A.G.; Visualization, A.F. and H.M.; Supervision, A.G.; Project administration, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Calabrese, F.; Regattieri, A.; Bortolini, M.; Galizia, F.G. Data-driven fault detection and diagnosis: Challenges and opportunities in real-world scenarios. *Appl. Sci.* **2022**, *12*, 9212.
2. Qi, R.; Zhang, J.; Spencer, K. A review on data-driven condition monitoring of industrial equipment. *Algorithms* **2023**, *16*, 9.

3. Talebpour, N.; Ilbeigi, M. Modeling and Predicting Uncertainty in Tidal Turbine Power Output: A Data-Driven Time-Series Approach. *IEEE Access* **2025**, *13*, 82986–82994. [[CrossRef](#)]
4. Walker, S.; Thies, P.R. A review of component and system reliability in tidal turbine deployments. *Renew. Sustain. Energy Rev.* **2021**, *151*, 111495. [[CrossRef](#)]
5. Falcão, A.F. Wave energy utilization: A review of the technologies. *Renew. Sustain. Energy Rev.* **2010**, *14*, 899–918. [[CrossRef](#)]
6. Zhang, Z. Energy system optimization based on fuzzy decision support systems and unstructured data. *Energy Inform.* **2024**, *7*, 82. [[CrossRef](#)]
7. Flores, P.; Mendoza, E. A Fuzzy Logic Technique for the Environmental Impact Assessment of Marine Renewable Energy Power Plants. *Energies* **2025**, *18*, 272. [[CrossRef](#)]
8. Milne, I.A.; Day, A.H.; Sharma, R.N.; Flay, R.G.J. The characterisation of the hydrodynamic loads on tidal turbines due to turbulence. *Renew. Sustain. Energy Rev.* **2016**, *56*, 851–864. [[CrossRef](#)]
9. Pfaffel, S.; Faulstich, S.; Rohrig, K. Performance and reliability of wind turbines: A review. *Energies* **2017**, *10*, 1904. [[CrossRef](#)]
10. Sasinthiran, A.; Gnanasekaran, S.; Ragala, R. A review of artificial intelligence applications in wind turbine health monitoring. *Int. J. Sustain. Energy* **2024**, *43*, 2326296. [[CrossRef](#)]
11. Zhang, F.; Chen, M.; Zhu, Y.; Zhang, K.; Li, Q. A review of fault diagnosis, status prediction, and evaluation technology for wind turbines. *Energies* **2023**, *16*, 1125. [[CrossRef](#)]
12. Zhu, F.; Zhao, Q.; Yang, T.; Zhou, L.; Zeng, B. Condition monitoring of wind turbine based on deep learning networks and kernel principal component analysis. *Comput. Electr. Eng.* **2023**, *105*, 108538. [[CrossRef](#)]
13. Qi, L.; Zhang, Q.; Xie, Y.; Zhang, J.; Ke, J. Research on wind turbine fault detection based on CNN-LSTM. *Energies* **2024**, *17*, 4497. [[CrossRef](#)]
14. Jinjing, A.; Hu, X.; Gong, L.; Zheng, L.; Zou, Z. Fuzzy reliability evaluation and machine learning-based fault prediction of wind turbines. *J. Ind. Inf. Integr.* **2024**, *40*, 100606. [[CrossRef](#)]
15. Cross, P.; Ma, X. Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines. *Int. J. Autom. Comput.* **2015**, *12*, 25–34. [[CrossRef](#)]
16. Yin, S.; Li, X.; Gao, H.; Kaynak, O. Data-Based Techniques Focused on Modern Industry: An Overview. *IEEE Trans. Ind. Electron.* **2015**, *62*, 657–667. [[CrossRef](#)]
17. Hameed, Z.; Hong, Y.S.; Cho, Y.M.; Ahn, S.H.; Song, C.K. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renew. Sustain. Energy Rev.* **2009**, *13*, 1–39. [[CrossRef](#)]
18. Tavares, T.M.; Giesbrecht, M. Deep Learning-Based Fault Diagnosis in Wind Turbine Bearings and Gearboxes Using Vibration Signals: Survey, Challenges, and Recommendations. *IEEE Access* **2025**, *13*, 207013–207032. [[CrossRef](#)]
19. Jiang, L.; Yin, H.; Li, X.; Tang, S. Fault diagnosis of rotating machinery based on convolutional neural networks and time–frequency representations. *IEEE Sens. J.* **2019**, *19*, 5731–5740.
20. Grosvenor, R.I.; Prickett, P.W.; Frost, C.; Allmark, M. Performance and Condition Monitoring of Tidal Stream Turbines. In *Proceedings of the 2nd European Conference of the Prognostics and Health Management Society*; PHM Society: Rochester, NY, USA, 2014.
21. Elasha, F.; Mba, D.; Teixeira, J.A. Condition Monitoring Philosophy for Tidal Turbines. *Int. J. Perform. Eng.* **2014**, *10*, 521–534.
22. Frost, C.; Morris, C.E.; Mason-Jones, A.; O'Doherty, D.M.; O'Doherty, T. The effect of tidal flow directionality on tidal turbine performance characteristics. *Renew. Energy* **2015**, *78*, 609–620. [[CrossRef](#)]
23. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
24. Mason-Jones, A.; O'Doherty, D.M.; Morris, C.E.; O'Doherty, T. Influence of a velocity profile & support structure on tidal stream turbine performance. *Renew. Energy* **2013**, *52*, 23–30. [[CrossRef](#)]
25. Ross, T.J. *Fuzzy Logic with Engineering Applications*, 4th ed.; Wiley: Hoboken, NJ, USA, 2017.
26. Jang, J.-S.R.; Sun, C.-T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
27. Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data* **2012**, *6*, 15.
28. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
29. Shahzad, M.M.; Khan, M.H.F.; Lee, D. Analyzing the structural integrity of retrofitted jacket substructures for an offshore wind turbine under nonlinear dynamic loading. *Appl. Ocean Res.* **2025**, *165*, 104862. [[CrossRef](#)]
30. Shahzad, M.M.; Khan, M.H.F.; Lee, D. Parametric sensitivity analysis of soil properties for retrofitted offshore wind turbine jacket substructures under combined environmental loading. *Ocean Eng.* **2025**, *341*, 122690. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.