SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V12I12P103

Volume 12 Issue 12, 28-40, December 2025
© 2025 Seventh Sense Research Group®

Original Article

Performance and Generalization Analysis of Machine Learning Models for Potential Fishing Zone Classification

Nik Nur Shaadah Nik Dzulkefli[1], Norsuzila Ya'acob[2*], Mohd Azri Abdul Aziz[2], Azita Laily Yusof[2], Syila Izawana Ismail[1]

[1]*Faculty of Electrical Engineering, Universiti Teknologi MARA Cawangan Terengganu, Terengganu, Malaysia.*
[2]*Faculty of Electrical Engineering, Universiti Teknologi MARA, Selangor, Malaysia.*

[*]*Corresponding Author : norsuzila@uitm.edu.my*

Received: 03 October 2025 Revised: 05 November 2025 Accepted: 04 December 2025 Published: 27 December 2025

Abstract *- Potential Fishing Zones (PFZ) classification using satellite and environmental data remains challenging due to the complexity of marine environments. Accurate PFZ classification is needed for ensuring sustainability, reducing search time, and optimizing fishing resources. This study investigates the effectiveness of Machine Learning (ML) models in predicting PFZ using a dataset of satellite-derived oceanographic and climate features. An initial set of seven features was refined to six key features using Random Forest (RF) and Recursive Feature Elimination (RFE) to enhance model performance by focusing on the most important features. Five classification models, including Random Forest (RF), K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM), were tested using accuracy, precision, recall, and F1-score across the training, validation, and testing phases. RF showed strong generalization by achieving a perfect F1-score of 100% in training and 93% in testing. The key contribution of this study is showing that RF can outperform more complex models by achieving a better balance between interpretability when combined with feature selection. These findings provide practical implications for fisheries management by offering RF-based frameworks as reliable decision-support tools and guiding future PFZ research toward integrating feature selection to improve robustness.*

Keywords *- Feature Selection, Machine Learning (ML), Potential Fishing Zone (PFZ), Random Forest (RF), Remote Sensing.*

1. Introduction

Sustainable fisheries management depends on the ability to identify Potential Fishing Zones (PFZ) accurately, which areas are most likely to have productive and efficient fishing [1]. However, traditional fishing methods often depend on historical experience, manual observations, or localized ecological indicators, leading to inconsistent and uncertain PFZ identification [2]. The combination of satellite remote sensing and data-driven modeling in the last few years has enabled a more systematic understanding of the oceanographic and atmospheric factors influencing fish distribution. Features such as Sea Surface Temperature (SST), Nighttime Sea Surface Temperature (NSST), Chlorophyll-a (Chl-a), normalized Fluorescence Line Height (nFLH), diffuse attenuation coefficient (KD490), wind speed, and rainfall have been used to support more precise and sustainable fishing management [3]. Despite these advancements, several problems remain unresolved. Existing PFZ prediction studies have often focused on a single Machine Learning (ML) algorithm. They focused on prediction accuracy without properly testing the model's robustness, generalization ability, or feature significance. Furthermore, feature selection methods are often overlooked, leading to the inclusion of redundant variables that can reduce interpretability and increase computational cost. Another limitation in previous studies was limited feature combinations [1, 4, 5]. This study addresses this gap by using an updated multi-year (2021-2023) dataset that combines both oceanographic and atmospheric features to show more variation in PFZ classifications. These improvements provide a stronger base for evaluating model performance and ensure it can be used for real-world fisheries management. To address these gaps, this study performs a comprehensive evaluation and comparison of five ML architectures of Random Forest (RF), K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM) for PFZ classification using multi-source satellite-derived environmental features. The feature selection techniques based on RF and RFE were applied to refine input features and enhance predictive efficiency. Model performance is assessed across the training, validation, and testing phases using accuracy, precision, recall, and F1-score, with a focus on generalization and overfitting control. The novelty of this study lies in combining multi-model benchmarking with feature selection to establish a performance-generalization analysis framework for PFZ classification. Unlike previous studies, such as Tan and Mustapha [6], who applied RF for PFZ mapping, or Zhang et al. [7], who used ensemble learning for tuna fishing ground prediction, this work evaluates model robustness across different ML architectures and clearly measures generalization performance. This study introduces a more comprehensive framework for PFZ prediction that reflects marine environmental variability by integrating a multi-source environmental dataset with optimized feature selection. The results highlight the effectiveness of integrating feature selection with ensemble and distance-based methods for developing interpretable and robust PFZ frameworks. The remainder of this paper is organized as follows: Section 2 presents the literature review; Section 3 describes the methodology; Section 4 discusses the results and discussions, and Section 5 concludes the study and outlines future research directions.

2. Literature Review

Machine learning methods have been increasingly applied to predict potential fishing zones, offering significant improvements in prediction accuracy compared to conventional statistical or empirical-based methods [8]. These data-driven frameworks provide dynamic analysis of complex oceanographic and atmospheric features, allowing more adaptive and sustainable fisheries management [9]. The growing availability of satellite-based environmental data, such as SST, NSST, Chl-a, nFLH, and KD490, along with environmental data on wind speed and rainfall, has significantly improved PFZ modeling capabilities, leading to more spatially and temporally informed decision-making processes. This study evaluates five machine learning methods for potential fishing zone prediction, focusing on their classification accuracy, generalization, and suitability.

Among various ML methods, RF has been extensively used for modeling species distribution and PFZ prediction due to its ensemble-based structure and robustness, and can handle noisy, non-linear data [10]. RF effectively captures high-dimensional relationships and variable interactions, making it a good choice for fisheries and environmental prediction tasks. Its versatility extends beyond fisheries, including predicting fishing effort [11], tracking crop yields under drought conditions [12], identifying mangrove vegetation [3], and modeling ecological distributions using climate variables [14]. RF has consistently shown its ability in predictive matters. For example, Zhang et al. [7] integrated RF and XGBoost for yellowfin tuna (Thunnus Albacres) fishing ground prediction, achieving higher spatial accuracy than traditional regression-based methods. Similarly, Tan and Mustapha [6] applied RF for Rastrelliger Kanagurta PFZ identification along Peninsular Malaysia with 81% classification accuracy. These findings confirm the adaptability and interpretability of RF in ocean-based prediction frameworks.

In addition to RF, several other ML frameworks have been applied for PFZ prediction. LightGBM, based on the Gradient Boosting Decision Tree (GBDT) algorithm, has gained attention for its computational efficiency and scalability on large datasets [15, 16]. Its architecture allows it to capture temporal and non-linear patterns in environmental data more effectively [17]. Similarly, KNN has been used in PFZ prediction tasks, especially for analyzing large datasets [18]. KNN is valued for its simplicity and interpretability, classifying new samples based on their similarity to stored observations [19]. Another widely adopted algorithm is XGBoost, which applies an iterative boosting strategy to reduce errors and improve model robustness. XGBoost demonstrated strong accuracy in fisheries and environmental studies by combining multiple weak learners into a strong predictive framework [11]. However, most of the existing PFZ studies depend on limited combinations of datasets such as SST and Chl-a without integrating other oceanographic features like NSST, nFLH, and KD490, and environmental data of wind speed and rainfall. As a result, previous models may overlook multi-source feature interactions that better reflect the complex dynamics of marine ecosystems. In contrast, this study utilizes an updated multi-year dataset (2021-2023) combining seven satellite-derived features of SST, NSST, Chl-a, nFLH, and KD490 and environmental features of wind speed and rainfall, which enabled a more comprehensive representation of oceanic conditions affecting fish location.

While these studies highlight the potential of ML approaches in PFZ prediction, most prior works are limited to individual models or specific case studies. They often prioritize accuracy metrics without systematically evaluating robustness, generalization, or overfitting risk, which are critical for operational deployment. Moreover, feature selection methods based on RFE and RF remain underexplored in PFZ modeling despite their potential to enhance model interpretability and computational efficiency. To address these gaps, this study conducts a comprehensive benchmarking of five ML models (RF, KNN, XGBoost, LightGBM, and LSTM) and integrates feature selection using RF and RFE, and highlights generalization analysis to provide more reliable PFZ prediction frameworks. This combination of feature selection and expanded environmental datasets distinguishes the present study from earlier PFZ modeling by providing both analytical depth and improved generalization performance.

3. Methodology

This study evaluates and compares the effectiveness and generalization capabilities of five machine learning model architectures. They are Random Forest (RF), K-Nearest

Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM) for predicting potential fishing zones.

The models used essential environmental features obtained from remote sensing and environmental data to improve sustainable fisheries management. The methodology of this study referred to a structured approach, which is illustrated in Figure 1. The methodology includes the data collection, preprocessing the data, implementation of a classifier model/testing, model training, and evaluation analysis to determine the most effective architecture of machine learning for potential fishing zone classification.

3.1. Data Collection

Oceanographic environmental data were collected from 2021 to 2023 using satellite remote sensing products from NASA's (National Aeronautics and Space Administration) Ocean Color website https://oceancolor.gsfc.nasa.gov/. The seven features overall of the dataset were used in this study, representing various input features in predicting potential fishing zones, together with being categorized into a multi-class classification into three output levels of high, low, and medium classification. More than two class labels are included in the classification task for multi-class classification [20].

The dataset includes features related to remote sensing and environmental data, which are Sea Surface Temperature (SST), Nighttime SST (NSST), chlorophyll-a (Chl-a), normalized Fluorescence Line Height (nFLH), diffuse attenuation coefficient (KD490), wind speed, and rainfall. The target variable represents a potential fishing zone that is predicted based on these selected features. The study examines the Terengganu Sea in the South China Sea, situated between latitudes 5.844°N and 4.295°N and longitudes 104.594°E and 102.614°E. Table 1 provides a detailed overview of the 7 features used in this study.

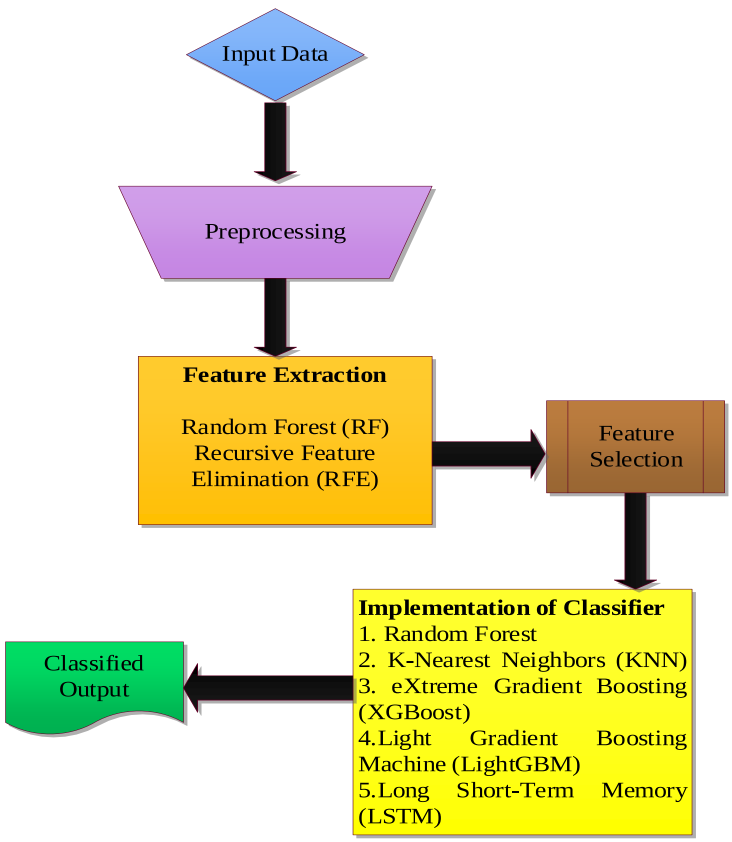

Fig. 1 Workflow of implementation of machine learning models

In this study, Random Forest (RF) and Recursive Feature Elimination (RFE) select six most significant features out of seven features overall. They are chlorophyll-a (Chl-a), Sea Surface Temperature (SST), Nighttime SST (NSST), diffuse attenuation coefficient (KD490), normalized Fluorescence Line Height (nFLH), and rainfall. By using feature selection methods of Random Forest (RF) and Recursive Feature Elimination (RFE), the model focuses only on the most significant features, which simplifies the input features. The reduction of features enhanced training efficiency, thus improving generalization by removing unwanted noise and redundancy from the dataset. The selected features kept the most important predictive signals, which are needed for a strong classification of the potential fishing zone model.

Table 1. Input features to the classifiers

Feature Number	Feature Description
Feature 1	Sea Surface Temperature (SST)
Feature 2	Nighttime Sea Surface (NSST)
Feature 3	Chlorophyll-a (Chl-a)
Feature 4	Normalized Fluorescence Line Height (nFLH)
Feature 5	Diffuse Attenuation Coefficient (KD490)
Feature 6	Wind Speed
Feature 7	Rainfall

3.2. Data Pre-Processing

A systematic data pre-processing workflow was carried out to ensure data quality and enhance its suitability for modeling. Initially, outliers were identified and handled using the Interquartile Range (IQR) method to reduce their impact on model performance. IQR was used to detect outliers in the data using the Winsorizing technique to handle them effectively [21]. Then, the duplicate data were identified and removed to eliminate redundancy and ensure data uniqueness. Furthermore, class imbalance was handled through the application of the Adaptive Synthetic Sampling (ADASYN) algorithm, where this method synthetically generated a minority class to introduce a more balanced class distribution across the dataset [22]. ADASYN was chosen over other imbalance-handling methods, such as random oversampling and the Synthetic Minority Over-sampling Technique (SMOTE), because it focuses on generating synthetic samples in regions that are harder to classify. This makes it suitable for PFZ data where minority patterns are often complex, thus improving the model's ability to generalize without bias towards majority classes.

3.3. Implementation of Classifier Model/Tested

The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections. Selected features were used to predict the potential fishing zone. The dataset was split into training (70%), validation (15%), and testing (15%) sets. The dataset was going into stratified sampling to ensure that the proportion of each class remains the same in each split as the original data to avoid imbalance problems. The numerical values of the dataset were standardized to a common scale using StandardScaler before modelling so that the learning process would be more stable and accurate. The RF model was trained with 200 trees, a maximum depth of 60, and other tuned hyperparameters. The XGBoost trained the data after the data had been scaled first to help the model learn patterns more effectively without being biased by different scales of the input features. The multi-class log loss (mlogloss) is used as its evaluation metric during the training session to measure how well the predicted possibilities align with the actual class labels in a multi-class setting. The model was evaluated on the validation set data after training to check performance and ensure it generalizes well. Finally, the separate test set was used to test the model to simulate the real-world unseen data to evaluate the model's final accuracy and robustness. The accuracy scores, classification reports, and confusion matrices were assessed to evaluate model performance.

A KNN classifier also implemented scikit-learn's KNeighborsClassifier to predict the model performance based on six features. This model also used StandardScaler to standardize the features. The KNN classifier (k=5) was trained based on a scaled training set, and hence it was evaluated on the validation and test set after training. Same as XGBoost, the model performance was assessed using accuracy scores, classification reports, and confusion matrices. K-fold cross-validation (5-fold) as a model validation technique is used in this model to ensure that the model performs consistently and robustly.

An LSTM model was developed using the TensorFlow machine learning framework and the Keras deep learning API to classify potential fishing zones based on six selected features. It used stratified sampling to maintain balanced class sizes and was followed by a process of feature standardization using StandardScaler to improve model convergence. The data was then reshaped into a 3D structure so that it could be processed as a sequence by LSTM. The class labels were converted into one-hot vectors to match the output format for multiclass classification. The architecture of the model consists of an LSTM layer with 64 units (tanh activation), followed by dropout (rate 0.3), a Dense layer (32 Neurons, ReLU), another Dropout (rate 0.2), and a softmax output layer with three classes. It was then compiled by Adam, a categorical cross-entropy loss. It was trained for 50 epochs (batch size 32). This model's performance was also evaluated by accuracy scores, classification reports, and confusion matrices.

A LightGBM classifier was used LightGBM to classify potential fishing zones based on six selected features. It used stratified sampling to maintain balanced class sizes and was followed by a process of feature standardization using StandardScaler. A LightGBM classifier with a multiclass objective and 300 estimators was trained on the scaled training set and validated using multi-class log loss (mlogloss) with early stopping after 20 rounds. The model performance was also assessed using accuracy scores, classification reports, and confusion matrices.

3.4. Model Testing and Validation

The model validation tests give a comprehensive assessment of the classification performance of the model and serve as a basis for analyzing the Random Forest (RF) model's effectiveness in experiments. The confusion matrix metrics, such as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), are important to analyse the key performance metrics of accuracy, precision, recall, and F1-score. It is a very important tool to evaluate the performance of a classifier [23]. By analysing these metrics, they offer a clear view of each model's ability to correctly differentiate between production classes, thus highlighting the strengths and weaknesses of the models applied.

The accuracy of a classifier is the proportion of correct predictions for both true positives and true negatives over the total number of predictions, as shown in Equation (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

Precision indicates the proportion of correctly predicted positive samples among all samples predicted as positive. It tells the reliability of the classifier, which can identify true positives, minimizing false alarms. The equation of precision is shown in Equation (2).

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (2)$$

Recall measures the model's ability to correctly detect actual positive instances. It represents the proportion of true positive samples that are successfully identified out of all actual positive samples, as shown in Equation (3).

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3)$$

The F1-score becomes a comprehensive metric that gives balances between precision and recall. It provides a single measure of the model's effectiveness in capturing both correct positive predictions and minimizing the missed detections. The equation of the F1-score is shown in Equation (4).

$$F1 - score = \frac{2TP}{2TP+FP+FN} \times 100 \quad (4)$$

These evaluations effectively evaluate and assess the performance of all classifiers to predict the potential fishing zone. It also ensures the model learns more efficiently, avoids overfitting, and achieves better accuracy and generalization through Random Forest (RF) and Recursive Feature Elimination (RFE) feature selection methods by choosing only the most important features. The results indicate that these methods improve the model efficiency by simplifying the feature set while maintaining the important predictive information. Thus, it leads to improved classification accuracy and stronger generalization performance.

4. Results and Discussions

This section presents the results and discussion on the comparative evaluation of five machine learning models for predicting the potential fishing zone. The analysis is organized into five main aspects: classification accuracy and generalization, the balance between precision and recall, learning stability and convergence behavior, the impact of feature selection using Random Forest (RF) and Recursive Feature Elimination (RFE) methods, and model robustness related to overfitting. Each subsection examines performance patterns during training, validation, and testing phases. It investigates the strengths and potential limitations of each model based on the dataset characteristics and their predictive reliability.

4.1. Classification Accuracy and Generalization

Table 2 shows the accuracy results for all tested models for both training and testing datasets. The classification accuracy examines models' predictive accuracy and their ability to maintain performance across different data splits, thus reflecting their generalization strength from training to unseen data. These values result from three evaluation phases: training, validation, and testing the data.

The RF achieved the most consistent performance with 100% accuracy in training and 93% accuracy for both validation and testing, as depicted in Figure 2. This minimal drop shows a strong generalization and efficient learning without overfitting, showing that RF successfully captured the relationships among all features. In contrast, KNN and XGBoost showed good and balanced performance with training accuracies of 91% and 89% respectively. Both only drops in validation and testing around 86%. This suggests that both models also generalized effectively. LightGBM achieved moderate performance with training accuracy at 83% and maintained about 81-82% in validation and testing. It reflected that the model had steady learning, but possibly underfitting.

The LSTM model recorded the lowest accuracy among all since it was maintaining scores around 77% for all three phases. While the model performance is stable, it can be said that it may not capture the data patterns as efficiently as the other models. This limitation is due to the non-sequential nature of the PFZ dataset. LSTM architecture could not handle the PFZ dataset, which led to weaker performance compared to ensemble methods such as RF. Nonetheless, including LSTM served as a useful benchmark to show that simpler ensemble methods can generalize better than deep sequential models when applied to environmental datasets.

From a practical perspective, these findings show the importance of choosing a model architecture that suits well with the characteristics of the dataset to achieve reliable generalization. In this study, RF, a model that is able to handle feature interactions among the data, performed strongly and showed superior performance through all phases, thus making it a robust choice for potential fishing zone prediction.

Meanwhile, other simple models like KNN and XGBoost maintained a good balance between accuracy and generalization. It shows that they can be effective alternatives for efficient solutions. These findings show that the interpretability and stability of ensemble-based models are advantageous in fisheries and environmental data analysis, where transparency is critical for decision support. The findings of this study aligned with previous work that reported the effectiveness of ML methods for PFZ prediction modeling. The study in [6] of applied RF to identify PFZ for Rastrelliger Kanagurta with an accuracy of 81.1%. While the study in [7] also used an ensemble learning model based on RF, Support Vector Machine (SVM), KNN, XGBoost, and Gaussian Process (GP) to forecast potential fishing grounds for albacore tuna (Thunnus alalunga) in the South Pacific Ocean with an accuracy of 86.92%. The study, which used the Empirical Cumulative Distribution Function (ECDF) algorithm to identify PFZ in the Bali Strait, only achieved 69.33% [1].

In comparison, the present study's RF model achieved a 93% F1-score on testing data, outperforming the mentioned studies by 6-24 percentage points. This improvement is mainly attributed to the integration of an expanded multi-source dataset (2021-2023), which combines oceanographic and atmospheric features, the application of RF-based feature selection and Recursive Feature Elimination (RFE), which reduced redundant inputs and improved model interpretability, and the implementation of a systematic benchmarking framework across multiple ML models emphasizing both performance and generalization. Unlike earlier studies that focused solely on accuracy, this study highlights that achieving a balance between accuracy, robustness, and interpretability is crucial for the deployment of PFZ prediction models in real-world fisheries management.

Table 2. Validation metrics results of accuracy for tested classifiers

Model	Accuracy (%)		
	Training	Validation	Testing
RF	100	93	93
KNN	91	87	87
XGBoost	89	86	86
LightGBM	83	81	82
LSTM	77	76	76

4.2. Precision Versus Recall

Table 3 shows the results of precision and recall for all tested models for both training and testing datasets. The results highlight important differences for each model's prediction behavior and generalization capabilities. The RF model had the highest performance with perfect precision and recall of 100% during training and 93% during testing, as depicted in Figure 3. However, the perfect training scores suggest it might be a bit overfitting. The KNN model showed good and steady performance. This model scored 91% for both precision and recall during training. A small drop happened on testing, scoring 87% for both precision and recall. This means that it can generalize well to new data. Moderate results were produced by LightGBM with training precision and recall scores 84% and 83% respectively. This model scored 82% on the testing set for both precision and recall. The results showed the model to be stable, but not as strong as RF and KNN. XGBoost is another model that performed well with a precision and recall of 89% for training and 86% for the testing set. These results indicate that it learns well and generalizes constantly to new data. The LSTM model only scores 78% precision and 77% recall during training, and decreases a bit to about 76% - 78% during testing. This shows that this model can still make a good prediction on unseen data without major drops in performance.

This trend indicates that some models, such as LSTM and KNN, maintained high recall but slightly lower precision, thus making them more likely to classify uncertain areas as potential fishing zones. At the same time, they might also suggest areas that are not good for fishing. On the other hand, RF achieved very high precision and recall, making it very accurate, but it might be slightly overfitting on the training data. XGBoost and LightGBM gave more balanced performance, showing stable precision and recall in both training and testing. These models can make more reliable predictions and function better in real fishing conditions.

These results focused on the importance of looking at precision and recall together, not separately. In potential fishing zone prediction, it is very important to avoid missing real productive fishing zones (low recall) and avoid wrongly predicting poor zones as good (low precision) because both faults can waste time, resources, and reduce fishing success. A model that balances for both metrics is crucial to make sure that the model is reliable for fishers. In this study, the RF model has strong and stable performance across all phases, which shows that it is suitable for real-world deployment. Overall, the findings highlight that successful classification models should not focus only on good accuracy but also maintain a good balance between correctly identifying true positive cases and minimizing false alarms.

Table 3. Validation metrics results of precision and recall for tested classifiers

Model	Precision (%)		Recall (%)	
	Training	Testing	Training	Testing
RF	100	93	100	93
KNN	91	87	91	87
LightGBM	84	82	83	82
XGBoost	89	86	89	86
LSTM	78	78	77	76

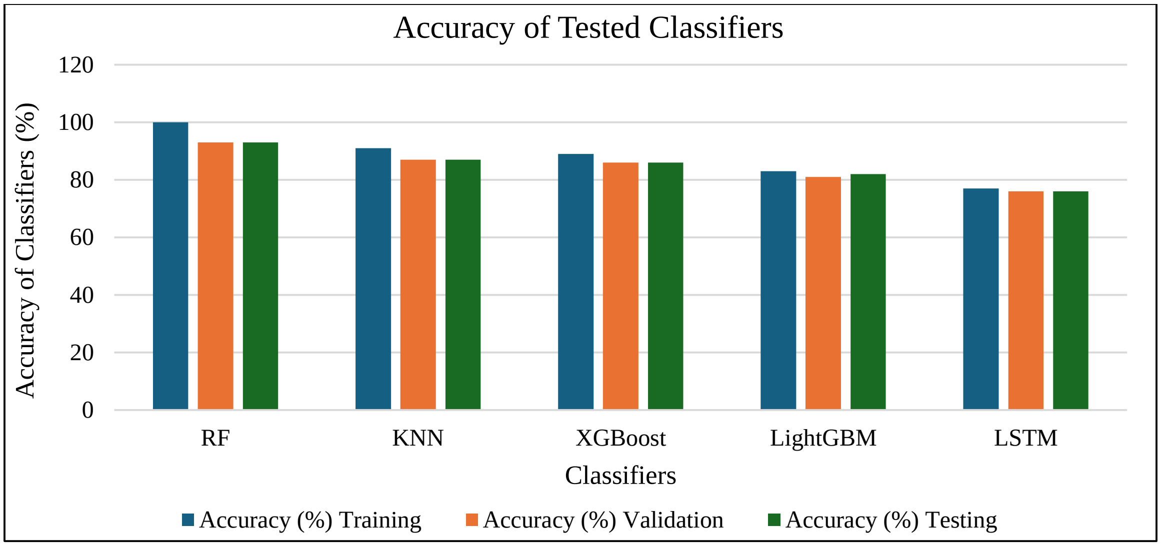

Fig. 2 Accuracy of the tested classifiers

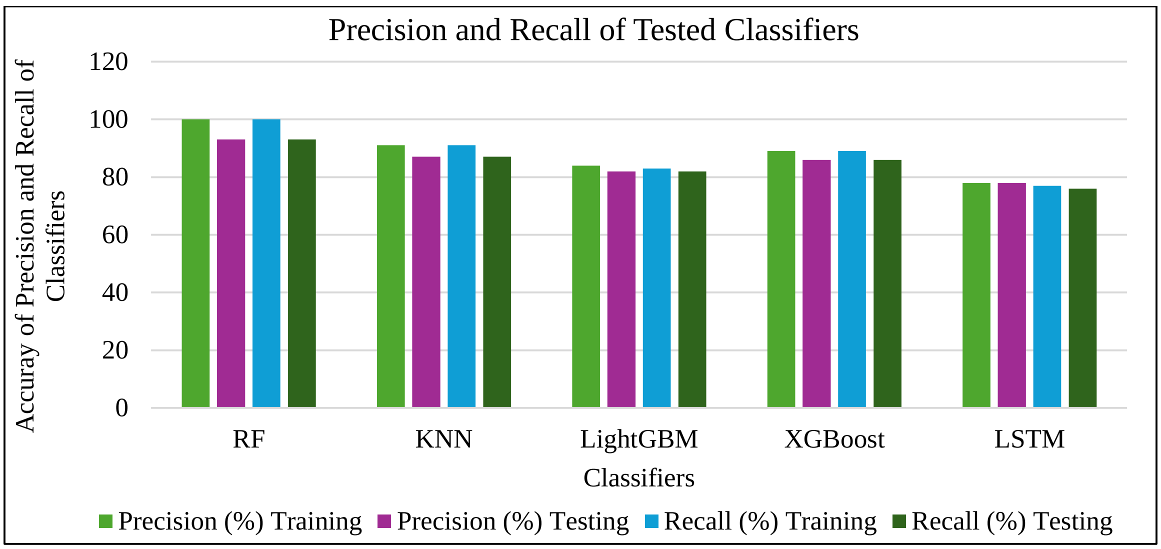

Fig. 3 Accuracy of precision and recall of the tested classifier

4.3. Learning Stability and Convergence

Table 4 and Figure 4 show the results of the F1-score for training, validation, and testing events. According to Table 4, RF achieved the highest training F1-score at 100% indicating that the model scores perfectly on the training data. However, the validation and testing F1-scores dropped to 93%. Although the model had a slight drop, it still shows a strong generalization to new data, which suggests that the model is not overfitting badly. This shows the model is robust and reliable in practice. KNN and XGBoost both showed slightly similar results when they achieved about 89% - 91% on training and then slightly decreased to 86% on validation and 85% - 86% on testing. These models learned well and maintained stable performance across all metrices. LightGBM scored a training F1-score of 83% but decreased to 81% in both validation and testing phases. This model shows moderate stability but less robust convergence than RF and KNN. In contrast, LSTM recorded the lowest F1-scores, at 75%, in all three phases of training, validation, and testing. The consistent phases indicated that LSTM did not overfit, but the model struggled to capture the key patterns in the dataset. Thus, its learning capabilities and convergence are weaker, making LSTM less suitable for the prediction task compared to other models. The small F1-score gaps for KNN and LightGBM around 2 – 5% indicated that these models produced good generalization, and the models can handle new data without significant performance loss. On the other hand, RF's perfect training performance, combined with a moderate drop, causes slight overfitting, but it still maintains high accuracy. This means RF can generalize well and make accurate predictions in practical scenarios. Therefore, even though it overfitted a bit, it still has strong accuracy and stability, making it a good choice for deployment in real-world applications in predicting potential fishing zones.

Table 4. Validation metrics results of F1-score for tested classifiers

Model	F1 – Score (%)		
	Training	Validation	Testing
RF	100	93	93
KNN	91	86	86
LightGBM	83	81	81
XGBoost	89	86	85
LSTM	75	75	75

From a practical aspect, these results highlight the importance of choosing models that demonstrate stable convergence across all evaluation phases, especially in potential fishing zone prediction, where data can be complex. RF and KNN showed minimal performance drop between training and testing, which showed that they not only learn patterns effectively but also maintain reliable accuracy when applied to new data. In contrast, the lower and less consistent performance of LSTM sets the limitations of using deep sequential models for this type of environmental prediction when the data does not fully support their complexity. Models like RF and KNN, which offer strong generalization and stable behavior, are suitable for potential fishing zone prediction where consistent, repeatable, and interpretable outputs are critical in guiding fishing activities.

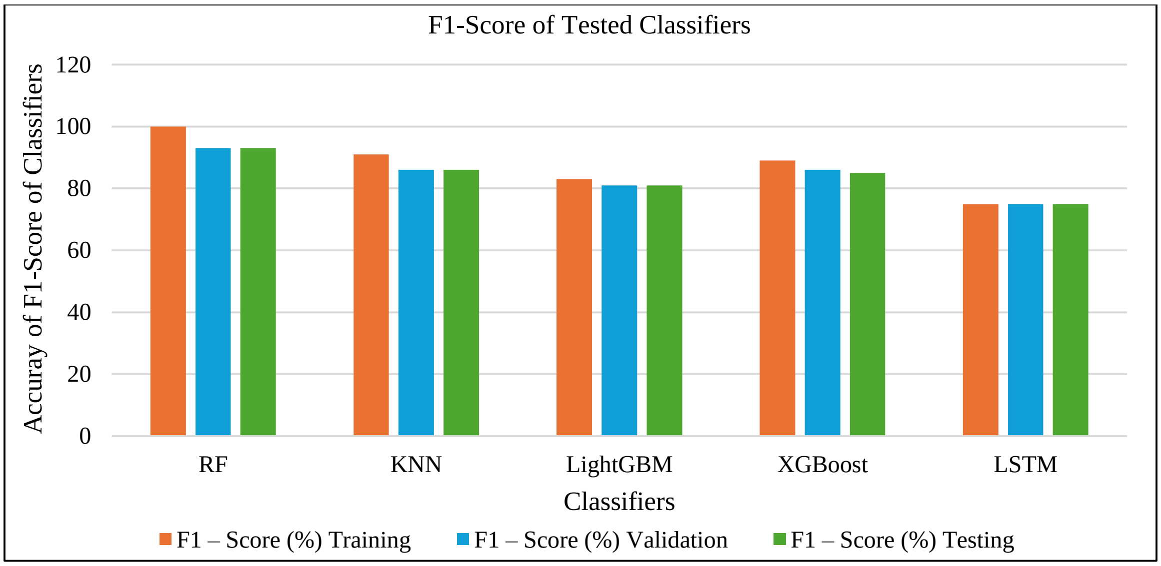

Fig. 4 Accuracy of the F1-score of tested classifiers

4.4. Effect of Feature Selection - Random Forest (RF) and Recursive Feature Elimination (RFE)

This study used a total of seven environmental features derived from satellite remote sensing and climate data, including Sea Surface Temperature (SST), Nighttime SST (NSST), chlorophyll-a (Chl-a), normalized Fluorescence Line Height (nFLH), diffuse attenuation coefficient (KD490), wind speed, and rainfall. Feature selection methods based on Random Forest (RF) and Recursive Feature Elimination (RFE) were applied to identify the most influential features to enhance the accuracy of the predictive model and reduce the complexity of potential fishing zone classification frameworks. These methods were selected due to their strong compatibility with non-linear and high-dimensional datasets, which is very crucial in complex environmental systems [24].

RF has many strengths, such as it works well even if the data is scaled or transformed, can handle extra or irrelevant features, and is good at finding complex patterns between features [25]. Moreover, RF also improves on regular decision trees by reducing the risk of overfitting and shows which features are most important to make predictions [25]. While RFE is a feature elimination method, it has become popular because it can find the best set of features based on model performance accuracy of classification prediction [26, 27].

This feature selection process established a refined set of important features that gave the essential predictive signal to predict potential fishing zones. After feature selection, the evaluation confirmed that model performance remained strong for RF and KNN with high F1-scores in both training and testing phases of 100% and 93% for RF, 91% and 86% for KNN, respectively, as shown in Figure 5. Although the total number of features was reduced, it does not affect the model accuracy, but it still enhances its generalization, thus supporting the applicability of these models for real-world predictions of potential fishing zones.

The integration of RF and RFE improved the performance compared to previous PFZ studies that did not apply feature selection because redundant or weakly correlated environmental features can obscure meaningful relationships in the data. By systematically removing less informative inputs, the model avoided noise accumulation, reduced computational overhead, and enhanced interpretability. In earlier PFZ research, where all variables were treated equally, redundant information often led to model overfitting and decreased generalization. In contrast, the present study demonstrates that applying RF-based feature ranking with RFE not only streamlined the model but also strengthened its predictive reliability across unseen datasets.

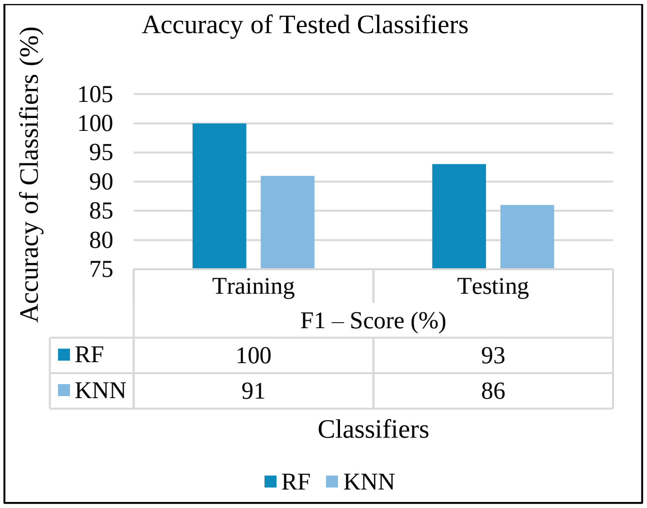

Fig. 5 Accuracy of tested classifiers for RF and KNN

4.5. Overfitting and Model Robustness

The differences in model performance during training, validation, and testing phases provide an important perception towards overfitting tendencies and model robustness. Table 5 presents the gap between training and testing F1-scores to evaluate these characteristics. RF achieved the highest training score (100%) with only 7% drop in testing, which shows strong learning ability and minimal overfitting risk. Its testing F1-score of 93% confirms that RF remains the most accurate and effective model overall. KNN and XGBoost showed smaller gaps of 5% and 4% respectively, with lower testing scores, suggesting that KNN and XGBoost are more stable but less powerful than RF. LightGBM showed a small gap of 2% but lower performance led to weaker convergence. LSTM showed perfect consistency with 0% gap, but a very low overall performance of 75% makes the model lack predictive strength. Overall, the gap analysis shows that a small drop from training to testing is acceptable if the model maintains high testing accuracy, as for RF. A small gap does not mean the model is weak, but it reflects the balance between learning from training data and generalizing well [28]. In contrast, models with minimal gaps but low predictive power, like LSTM, are less useful for real-world PFZ classification. Accuracy on real-world data matters the most in PFZ classification prediction, and RF clearly provides the most reliable and effective results among all. The robustness observed in this study, compared with earlier PFZ modeling efforts, can be attributed to the combination of an expanded multi-source dataset and the application of a feature selection method that minimized noise and redundancy. Previous PFZ studies, which depend on single-model frameworks or smaller feature sets, often demonstrated higher sensitivity to data variation and inconsistent validation accuracy across regions and time periods. In contrast, the present multi-model benchmarking approach and cross-phase evaluation ensured stable performance and controlled overfitting, with RF and KNN remaining consistent across validation folds. These results confirm that integrating environmental features with systematic feature selection enhances model robustness and generalization.

Table 5. Overfitting gap for tested classifiers

Model	F1 – Score (%)		Overfitting Gap (Train – Test) (%)
	Training	Testing	
RF	100	93	7
KNN	91	86	5
LightGBM	83	81	2
XGBoost	89	85	4
LSTM	75	75	0

4.6. K-Fold Cross Validation

The graph in Figure 6 illustrates the 5-fold cross-validation performance, showing the mean and standard deviation for each performance metric. Cross-validation is applied to validate the models' performance and provide a more reliable generalization ability as it reduces the risk of bias from a single train-test split and ensures that the results remain consistent across different subsets of data. Among all models, RF achieved the highest and most consistent performance with accuracy = 0.9290 ± 0.0015, precision = 0.9309 ± 0.0016, recall = 0.9288 ± 0.0015, and F1-score = 0.9282 ± 0.0016 as depicted in Table 6. This consistency shows that RF not only delivers strong results but also maintains its stability across folds. KNN was the second-best model with accuracy = 0.8793 ± 0.0020, precision = 0.8910 ± 0.0023, close to 0.88-0.89, showing only a minor drop, as it might be an alternative to RF. LightGBM and XGBoost produced mid-range results with accuracy 0.8165 ± 0.0041 and 0.7946 ± 0.0028, respectively, and F1-scores of 0.8102 ± 0.0044 and 0.7871 ± 0.0029, which showed less competitive performance. On the other hand, LSTM got the lowest scores through all metrics, with accuracy at 0.7526 ± 0.0049 and F1-score at 0.7440 ± 0.0060, which the model struggles to match the other models. The cross-validation results confirm that RF is the most accurate and stable model among those tested. Its small standard deviation across folds confirms low variance and strong generalization capability. KNN showed similar reliability with only minor drops across folds, while LightGBM and XGBoost showed moderate generalization ability. LTSM's weaker and inconsistent performance suggests that deep sequential models are not well-suited for static environmental data without temporal dependencies. Compared to earlier PFZ modeling works that validate performance on a single test split, this study's 5-fold cross-validation framework provides stronger evidence of generalization and robustness. The RF model maintained an average F1-score of 0.93 ± 0.0016 across all folds, outperforming earlier RF-based PFZ studies such as Tan and Mustapha [6] – 81% and Zhang et al. [7] – 86.9%. The improved performance can be credited to the combination of multi-source environmental features, optimized feature selection using RF and RFE, and a systematic benchmarking framework. These components minimized model variance, enhanced predictive stability, and ensured that the proposed framework remains reliable under varying environmental and data conditions. Figure 7 (a)-(d) presents the trend of model performance across all five folds for accuracy, precision, recall, and F1-score. From the results, RF clearly outperforms other models in every metric, which remains steady at around 0.93 across all folds, indicating its strong consistency and robustness. KNN is the second-best model, scoring at 0.88-0.89 with only small variations across folds. LightGBM and XGBoost performed at a moderate level with scores between 0.79 and 0.82. The performance is acceptable but less consistent and weaker compared to the top two models. Finally, LSTM records the lowest performance in all metrics with scores at 0.74-0.75, confirming its weaker generalization capability. Overall, the fold-wise trends validate the generalization consistency of RF and KNN, showing that these models are reliable to reproduce high performance across various data.

Table 6. 5-fold cross-validation performance (Mean ± Std)

Model	Accuracy	Precision	Recall	F1-Score
RF	0.9290 ± 0.0015	0.9309 ± 0.0016	0.9288 ± 0.0015	0.9282 ± 0.0016
KNN	0.8793 ± 0.0020	0.8910 ± 0.0023	0.8787 ± 0.0020	0.8751 ± 0.0021
LightGBM	0.8165 ± 0.0041	0.8238 ± 0.0033	0.8161 ± 0.0041	0.8102 ± 0.0044
XGBoost	0.7946 ± 0.0028	0.8019 ± 0.0027	0.7941 ± 0.0028	0.7871 ± 0.0029
LSTM	0.7526 ± 0.0049	0.7529 ± 0.0033	0.7522 ± 0.0050	0.7440 ± 0.0060

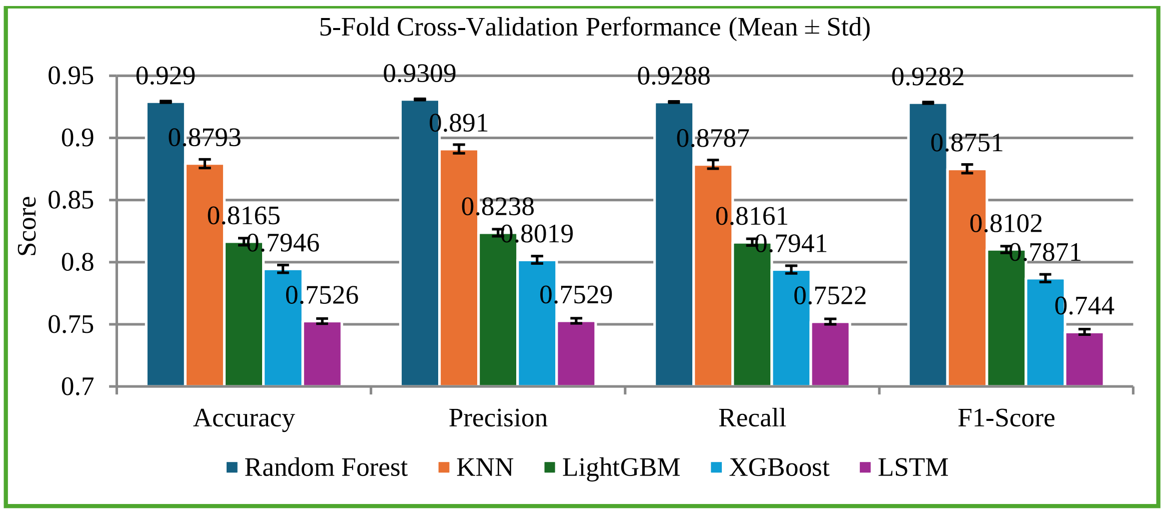

Fig. 6 5-fold cross-validation performance (Mean ± Std)

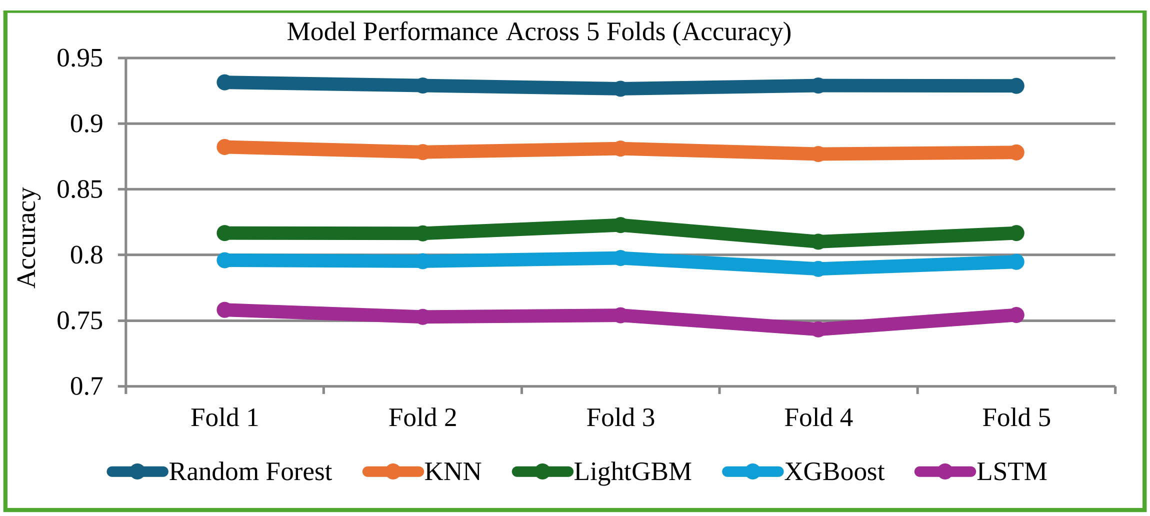

Fig. 7(a) Model performance across 5-folds cross-validation

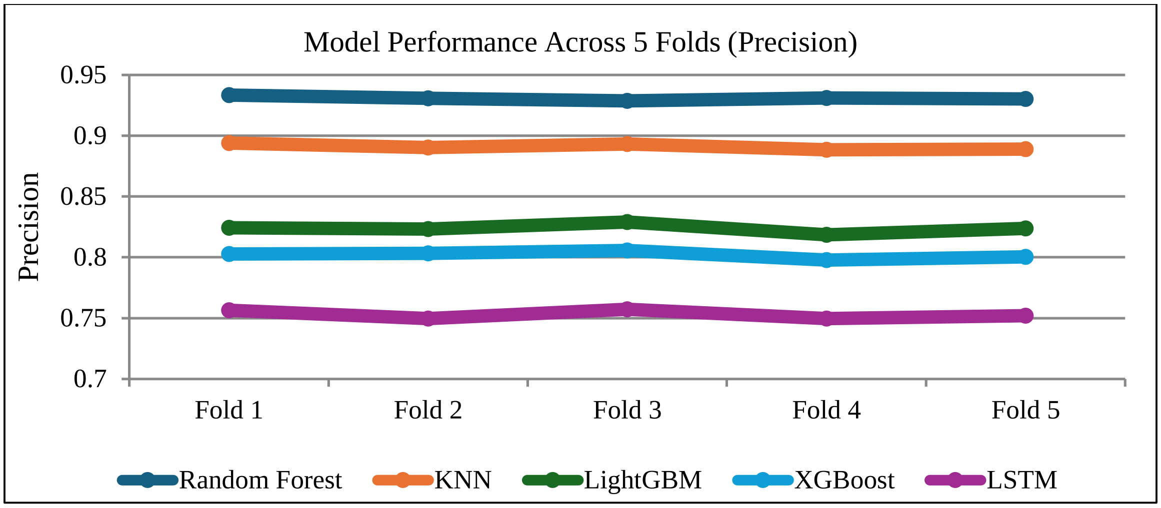

Fig. 7(b) Model performance across 5-folds cross-validation

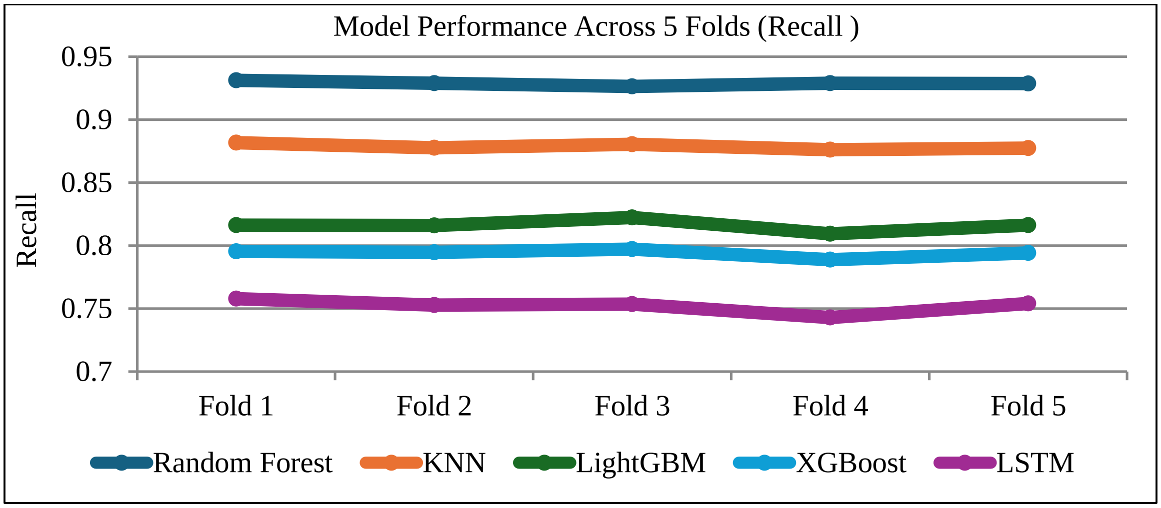

Fig. 7(c) Model performance across 5-Folds cross-validation

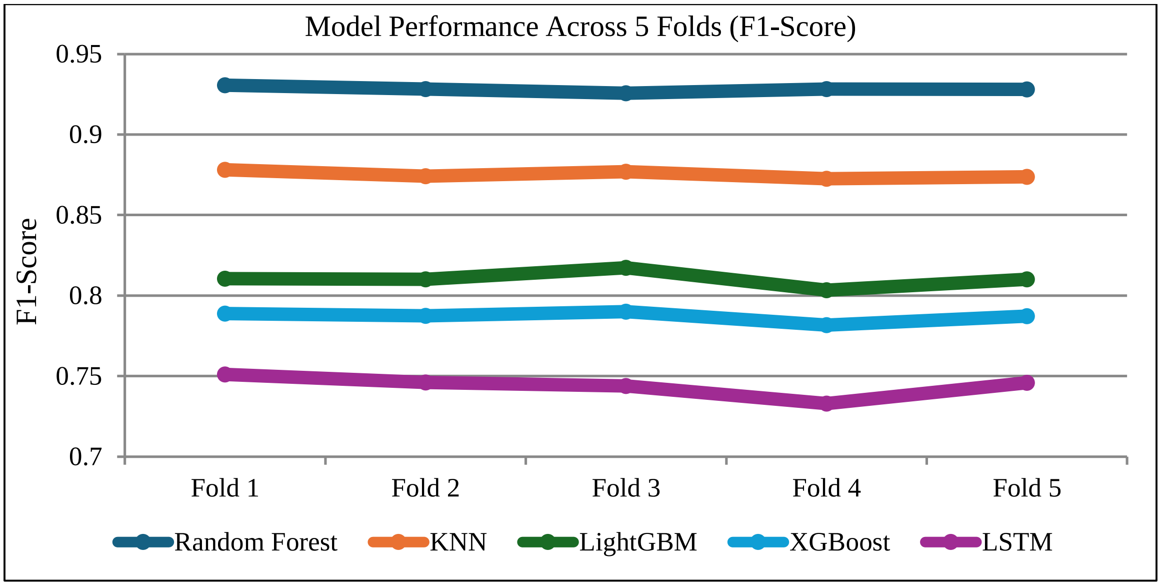

Fig. 7(d) Model performance across 5-Folds cross-validation

5. Conclusion

The Random Forest (RF) and K-nearest Neighbors (KNN) models consistently demonstrated excellent predictive accuracy, generalization, and robustness in PFZ classification. Their stable F1-scores across all phases indicate that both models effectively captured nonlinear relationships between environmental features while minimizing overfitting. In comparison, LightGBM and XGBoost achieved moderate but reliable results, whereas LSTM showed limited effectiveness due to the non-sequential nature of the PFZ dataset. The implementation of the RF and RFE feature selection method further enhanced computational efficiency and interpretability by removing redundant features without reducing accuracy. This study uses a new dataset that differs from previous studies by combining oceanographic and environmental features, offering a more comprehensive PFZ modeling. These findings also highlight that simpler ensemble architectures, when properly aligned with data characteristics, can outperform deeper models in environmental prediction tasks. Compared with earlier PFZ modeling studies, the present framework achieved higher accuracy and stronger generalization by using a multi-source dataset and an optimized feature selection strategy. This established a more adaptive and reliable ML framework for fisheries applications. Future work should focus on incorporating near-real-time satellite observations, spatiotemporal deep learning models, and hybrid ensemble techniques to forecast PFZ and support sustainable fisheries management.

Funding Statement

This research article was financially supported by Universiti Teknologi MARA and the Institute of Postgraduate Studies UiTM.

References

[1] Dinarika Jatisworo et al., "Bali Strait's Potential Fishing Zone of Sardinella Lemuru," *Indonesian Journal of Geography*, vol. 54, no. 2, pp. 254-262, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sukardi et al., "Prediction Potential Fishing Zone of the Yellowfin Tuna (Thunnus Albacares) in the Southern Flores Sea using Satellite Remote Sensing Data," *Egyptian Journal of Aquatic Biology and Fisheries*, vol. 28, no. 4, pp. 627-645, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Robinson Mugo, and Sei-Ichi Saitoh, "Ensemble Modelling of Skipjack Tuna (Katsuwonus Pelamis) Habitats in the Western North Pacific using Satellite Remotely Sensed Data; A Comparative Analysis using Machine-Learning Models," *Remote Sensing*, vol. 12, no. 16, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yeny Nadira Kamaruzzaman, Muzzneena Ahmad Mustapha, and Mazlan Abd Ghaffar, "Determination of Fishing Grounds Distribution of the Indian Mackerel in Malaysia's Exclusive Economic Zone Off South China Sea using Boosted Regression Trees Model," *International Journal of Marine Sciences*, vol. 37, no. 1, pp. 147-161, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] M. Sivasankari, R. Anandan, and Fekadu Ashine Chamato, "HE-DFNETS: A Novel Hybrid Deep Learning Architecture for the Prediction of Potential Fishing Zone Areas in Indian Ocean using Remote Sensing Images," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ming Kun Tan, and Muzzneena Ahmad Mustapha, "Application of the Random Forest Algorithm for Mapping Potential Fishing Zones of Rastrelliger Kanagurta of the East Coast of Peninsular Malaysia," *Regional Studies Marine Science*, vol. 60, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jie Zhang et al., "Forecasting Albacore (Thunnus Alalunga) Fishing Grounds in the South Pacific based on Machine Learning Algorithms and Ensemble Learning Model," *Applied Sciences*, vol. 13, no. 9, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Shota Kunimatsu et al., "Predicting Unseen Chub Mackerel Densities through Spatiotemporal Machine Learning: Indications of Potential Hyperdepletion in Catch-Per-Unit-Effort Due to Fishing Ground Contraction," *Ecological Informatics*, vol. 85, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Vishal Singh Rawat, Gubash Azhikodan, and Katsuhide Yokoyama, "Prediction of Fish (Coilia Nasus) Catch using Spatiotemporal Environmental Variables and Random Forest Model in a Highly Turbid Macrotidal Estuary," *Ecological Informatics*, vol. 86, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Andrés Flores et al., "Applying Machine Learning to Predict Reproductive Condition in Fish," *Ecological Informatics*, vol. 80, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] J. Samarão et al., "Improving Machine Learning Predictions to Estimate Fishing Effort using Vessel's Tracking Data," *Ecological Informatics*, vol. 85, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Guangpo Geng et al., "Random Forest Model that Incorporates Solar-Induced Chlorophyll Fluorescence Data Can Accurately Track Crop Yield Variations under Drought Conditions," *Ecological Informatics*, vol. 85, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Judith Aviña-Hernández et al., "Predictive Performance of Random Forest on the Identification of Mangrove Species in Arid Environments," *Ecological Informatics*, vol. 75, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Brice B. Hanberry, "Practical Guide for Retaining Correlated Climate Variables and Unthinned Samples in Species Distribution Modeling, using Random Forests," *Ecological Informatics*, vol. 79, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hyunsoo Kim et al., "Interpretable General Thermal Comfort Model based on Physiological Data from Wearable Bio Sensors: Light Gradient Boosting Machine (LightGBM) and SHapley Additive exPlanations (SHAP)," *Building Environment*, vol. 266, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Chenxi Ni et al., "Light Gradient Boosting Machine (LightGBM) for Forecasting Data and Assisting in the Defrosting Strategy Design of Refrigerators. Experimental Study and Mathematical Modeling," *International Journal of Refrigeration*, vol. 160, pp. 182-196, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yu Fu et al., "Prediction and Analysis of Sea Surface Temperature based on LSTM-Transformer Model," *Regional Studies Marine Science*, vol. 78, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Xue Ji, Bisheng Yang, and Qiuhua Tang, "Seabed Sediment Classification using Multibeam Backscatter Data based on the Selecting Optimal Random Forest Model," *Applied Acoustics*, vol. 167, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Nurnadiah Zamri et al., "River Quality Classification using Different Distances in k-Nearest Neighbors Algorithm," *Procedia Computer Science*, vol. 204, pp. 180-186, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Seyed Hossein Seyed Ebrahimi, "A Hybrid Principal Label Space Transformation-Based Binary Relevance Support Vector Machine and Q-Learning Algorithm for Multi-label Classification," *Arab Journal Science Engineering*, vol. 50, no. 2, pp. 851-875, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ch. Sanjeev Kumar Dash et al., "An Outliers Detection and Elimination Framework in Classification Task of Data Mining," *Decision Analytics Journal*, vol. 6, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Sheikh Muhammad Saqib et al., "Enhancing Electricity Theft Detection with ADASYN-Enhanced Machine Learning Models," *Electrical Engineering*, vol. 107, no. 8, pp. 10525-10542, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Amalesh Gope et al., "Multi-Class Identification of Tonal Contrasts in Chokri using Supervised Machine Learning Algorithms," *Humanity Social Science Communications*, vol. 11, no. 1, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Huanjing Wang et al., "Feature Selection Strategies: A Comparative Analysis of SHAP-Value and Importance-Based Methods," *Journal Big Data*, vol. 11, no. 1, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Reza Iranzad, and Xiao Liu, "A Review of Random Forest-Based Feature Selection Methods for Data Science Education and Applications," *International Journal Data Science Analytics*, vol. 20, no. 2, pp. 197-211, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ya-Han Hu et al., "A Novel Miss Forest-Based Missing Values Imputation Approach with Recursive Feature Elimination in Medical Applications," *BMC Medical Research Methodology*, vol. 24, no. 1, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Chang Wang et al., "Multi-Feature Fusion RFE Random Forest for Schizophrenia Classification and Treatment Response Prediction," *Science Reports*, vol. 15, no. 1, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Mengqing Huang, Hongchuan Yu, and Jianjun Zhang, "A Practical Generalization Metric for Deep Networks Benchmarking," *Scientific Reports*, vol. 15, no. 1, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]