Journal of *Marine Science and Engineering*

MDPI

Article

Feasibility, Advantages, and Limitations of Machine Learning for Identifying Spilled Oil in Offshore Conditions

Seong-Il Kang [1], Cheol Huh [1,*], Choong-Ki Kim [2], Meang-Ik Cho [3] and Hyuek-Jin Choi [4]

[1] Department of Convergence Study on the Ocean Science and Technology, Ocean Science and Technology School, Korea Maritime and Ocean University, 727 Taejong-ro, Youngdo-gu, Busan 49112, Republic of Korea; kis4824@g.kmou.ac.kr
[2] Division for Natural Environment, Korea Environment Institute, Sejong 30147, Republic of Korea; ckkim@kei.re.kr
[3] Offshore Industries R & BD Center, Korea Research Institute of Ships and Ocean Engineering, Geoje 53201, Republic of Korea; micho@kriso.re.kr
[4] Maritime Safety and Environment Research Center, Korea Research Institute of Ships and Ocean Engineering, Yuseong-daero 1312, Daejeon 34103, Republic of Korea; hjchoi@kriso.re.kr
* Correspondence: cheolhuh@kmou.ac.kr

Abstract: A rapid identification of oil would facilitate a prompt response and efficient removal in the event of an oil spill. Traditional chemical methods in oil fingerprinting have limitations in terms of both time and cost. This study considers machine learning models that can be applied immediately upon measurement of oil density and viscosity. The main objective was to compare models generated from various combinations of features and data. Under five different algorithms, the resulting models were evaluated in terms of their feasibility, advantages, and limitations (FAL). The extra tree (ET) and histogram-based gradient boosting (HGB) models, which incorporated physical features, their rates of change, and environmental features, were found to be the most accurate, achieving 88.55% and 88.41% accuracy, respectively. The accuracy of the models was further enhanced by adjusting the features. In particular, incorporating the rate of change in oil properties led to an enhancement in the accuracy of ET to 92.83%. However, further inclusion of secondary features led to a reduction in accuracy. The effect of input imprecision was analyzed. A 10% of inherent error reduced the accuracy of the HGB model to 60%. Comparing these FAL, machine learning can be a simple, rapid, and cost-effective auxiliary for forensic analysis in diverse spill environments.

Keywords: oil spills; oil spill classification; oil properties; machine learning; predictive modeling; feature selection; offshore spill detection

check for updates

Academic Editor: Merv Fingas

Received: 13 March 2025
Revised: 8 April 2025
Accepted: 14 April 2025
Published: 16 April 2025

Citation: Kang, S.-I.; Huh, C.; Kim, C.-K.; Cho, M.-I.; Choi, H.-J. Feasibility, Advantages, and Limitations of Machine Learning for Identifying Spilled Oil in Offshore Conditions. *J. Mar. Sci. Eng.* **2025**, *13*, 793. https://doi.org/10.3390/jmse13040793

Copyright: © 2025 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction and Literature Review

1.1. Introduction

Oil is an essential resource in modern industry and daily life. It plays a crucial role in transportation, industrial applications, power generation, and as a fundamental raw material [1,2]. Ships serve as the primary means of oil transportation due to their cost-effectiveness and capacity to carry large quantities over long distances [3]. Maritime transport handles more than 60% of the global oil exports and imports, making it a crucial component of the global oil trade [4]. However, despite its advantages, transporting oil by tankers presents notable challenges and risks [5]. Oil spills resulting from tanker collisions can have devastating environmental consequences, contaminating ecosystems and coastal

areas [6,7]. Both large-scale spills and smaller, localized spills pose serious environmental concerns [8].

The weathering process of spilled oil involves chemical and physical changes over time. This process primarily occurs through evaporation, dispersion, dissolution, emulsification, sedimentation, and biodegradation [9,10]. The rate and extent of weathering vary depending on factors, such as oil type, wave conditions, wind speed, and water temperature. As the oil weathers, it spreads and absorbs water, increasing its water content and altering its density and viscosity [11]. Over time, both density and viscosity increase, hindering response efforts [12–14]. Consequently, an immediate and effective response that considers the type of oil in question is essential for mitigating environmental damage following an oil spill.

Oil spill response strategies typically include removal, detection, response planning, and oil spill forensic analysis. Removal techniques involve physical methods such as booms, floating barriers designed to prevent the spread of oil, and skimmers, which mechanically recover oil from the water surface. Additionally, chemical dispersants can enhance the biodegradation process, while in situ burning allows for rapid oil removal by combustion.

Early and accurate detection is essential for prompt response efforts. Aerial photography and drone-based imaging technologies effectively manage and monitor extensive areas affected by oil spills. Furthermore, response planning involves predictive modeling to forecast the movement and dispersion patterns of spilled oil. Simulation and numerical models that predict oil spill trajectories significantly contribute to minimizing response costs and reducing environmental damage.

Oil spill environmental forensic analysis of is employed to determine the origin of the spill, identify the oil, and assess its environmental impact [15–17]. This process involves oil fingerprinting, which examines the unique chemical composition of the oil [18,19]. Gas chromatography (GC) has been widely used to analyze the chemical composition of spilled oil samples, enabling identification of the spill's source [20,21]. Additionally, various techniques have been developed to provide further insights into the chemical structures of separated compounds. Gas chromatography–mass spectrometry (GC-MS) further expands upon this capability by providing detailed molecular profiles, including the identification of n-alkanes, polycyclic aromatic hydrocarbons (PAHs), benzene derivatives, and biomarkers [15,22,23]. Despite their widespread use, both GC and GC-MS face limitations when analyzing highly weathered oil samples, as key chemical markers degrade, making detection difficult.

To address these analytical limitations, several supplementary methods have been developed. Flame ionization detection (FID), often coupled with GC, enhances detection sensitivity for petroleum hydrocarbons [24,25]. Infrared spectroscopy (IR) rapidly analyzes oil samples by identifying functional groups [26]. Fluorescence spectroscopy method have offered robust characterization capabilities but may also struggle with distinguishing chemically similar oils [27–29]. Additionally, stable isotope mass spectrometry has complemented GC-MS analyses effectively. Other methods, including nuclear magnetic resonance (NMR) spectroscopy, and laser- and ultraviolet-induced fluorescence spectroscopy, have enabled the precise classification of spilled oil [30,31].

However, despite technological advancements, oil fingerprinting in oil spill forensics still faces several challenges [18,23,32–34]. The chemical composition of oil samples can change rapidly, complicating source determination. Accurate and reliable oil analysis requires specific quality and quantity standards for applying certain technologies. Additionally, step-by-step fingerprint identification methods may require specialized expertise, posing further challenges in the analysis of weathered oil samples at sea. Therefore, simple,

rapid, and cheap alternatives are needed in oil spill environmental forensic and developing solutions to address these challenges is important.

This study aimed to investigate the feasibility, advantages, and limitations (FAL) of machine learning in the identification of spilled oil using physical properties. A comprehensive assessment of the advantages, disadvantages, and limitations of five models was conducted. The following inquiries were addressed to determine the FAL of ML models:

- Which model was the most effective?
- What was the model's accuracy for a specific oil type?
- What was the contribution of input parameters (features) to the model's accuracy?
- Can the performance of the model be improved by manipulating the feature?
- How did model performance change in response to uncertainty in input values?

1.2. Literature Review

Various sophisticated methods have been integrated with oil fingerprinting techniques to improve efficiency and enhance identification accuracy. Techniques such as Fourier transform infrared spectroscopy, partial least squares discrimination analysis (PLS-DA), and principal component analysis have been studied, enabling cost-effective and efficient analysis while minimizing sample usage [35–39]. Recent advancements in machine learning techniques have demonstrated promising results across various fields, including oil spill detection and classification [40–44]. These advancements offer the potential for rapid and more accurate identification of spilled oil, addressing some limitations of traditional analytical methods.

Machine learning has been widely applied to oil spill detection using various remote sensing technologies. Li et al. [45] combined histogram of oriented gradients (HOG) and support vector machine (SVM) classifiers in a thermal infrared imaging approach using drones to monitor spills at sea. The accuracy of the traditional SVM model was improved from 86% to 91%. Zhang et al. [46] combined hyperspectral imaging with convolutional neural networks (CNNs), achieving up to 100% detection accuracy in a real-time engine oil spill monitoring application.

Research identifying spilled oil using machine learning has been effectively demonstrated over the past five years. It has been effective in classifying various targets and has been used to handle complex relationships involving large datasets and numerous variables. It also offers classification possibilities to non-experts without relying on visual or computational expertise. Chen et al. [47] introduced a binary classification framework using machine learning to distinguish between weathered crude oil (WCO) and chemically dispersed oil (CDO) for the first time. The model showed accuracies ranging from 68% to 90% depending on the biomarker, demonstrating classification of WCO. Ekpe et al. [48] classified oil compounds in groundwater samples using gas chromatography–high-resolution mass spectrometry (GC-HRMS). The PLS-DA model achieved 90% accuracy but required training on compounds with great mobility. Xie et al. [49] demonstrated the superiority of a transformer-based model over CNN using fluorescence excitation–emission matrix (EEM) data to classify five types of oils. The model showed stable prediction performance with accuracies between 89% and 99% for each oil. However, the overall process from sample to result still required considerable time. Meanwhile, several studies have enabled the generation of analysis results from samples on-site using devices. Chung et al. [50] used classification models such as SVM, principal component analysis (PCA), and linear discriminant analysis (LDA). Oil types were identified as one of five categories using capillary flow velocity through microfluidic channel lengths and related parameters. The system demonstrated excellent performance, classifying crude oil with up to 90% accuracy in under one minute. However, the density and viscosity of spilled oil are among the most important properties and must account for weathering, which changes these properties over time.

Sonsnowski et al. [51] also used fluorescence spectroscopy combined with various machine learning models to classify spilled oils. These methods achieved impressive accuracy of approximately 86% across a wide range of target samples. Although the models were trained considering weathering of oil, limitations remained due to inconsistent effects or small sample sizes. Bills et al. and Loh et al. [52,53] presented portable devices for rapid oil type screening using EEM and laser-induced fluorescence (LIF), achieving 95% accuracy as an alternative to time-consuming oil analysis. Most previous studies have classified oils based on chemical analysis. Only a few studies used physical properties, and even then, they were not used as key features.

2. Methods

2.1. Dataset Preparation and Feature Selection

2.1.1. Model Target Selection

Preparing a dataset that reflects recent logistic data and accident history is essential [54]. South Korea heavily depends on imported oil, sourcing oil from various countries [55]. Moreover, it is estimated that approximately 300 oil spill incidents take place on an annual basis in the coastal waters of South Korea [56,57]. To incorporate relevant data, oil import data from the Korea National Oil Corporation's petroleum supply statistics dataset for three years (2018–2020) were utilized [58]. Most oils imported to South Korea have an American Petroleum Institute gravity (API gravity) between 21° and 40°. Nine oils were selected as model targets based on import quantities. These oil types account for approximately 65% of the total imported volume (in barrels). Other oil types were excluded as model targets due to their low import volumes or infrequent usage. Additionally, two oils with a history of spillage in Korean maritime areas were included [17]. One low-sulfur oil was also selected due to its increasing demand, driven by enhanced environmental regulations set by the International Maritime Organization (IMO) [59–61]. These twelve oils and their corresponding API gravity are presented in Table 1. The selection covered a broad range of API gravity values, ranging from 21.3 to 40.5, encompassing a diverse array of oils.

Table 1. The twelves oils selected for the model target and their API gravity.

Oil	API Gravity
Maya (MA)	21.3
Basrah Heavy (BH)	24.7
Arabian Medium (AM)	29.5
Iranian Heavy (IH)	30.0
Kuwait (KU)	30.6
Arabian Light (AL)	32.2
Hout (HO)	32.4
Basrah Light Mobil Oil Australia (BL)	34.2
West Texas Intermediate (WTI)	36.4
Arabian Extra Light Aramco (AELA)	37.0
Ultra-low Sulfur Diesel (ULSD)	38.0
Murban Shell Oil (MSO)	40.5

2.1.2. Oil Spill Simulation

Training an ML model requires a large amount of oil data [62]. This spilled oil data undergoes a weathering process, causing properties such as density and viscosity to change over time. However, obtaining these data through experiments consumes significant time and cost [11,63,64]. Therefore, the Web-based General National Oceanic and Atmospheric Administration Oil Modeling Environment (WebGNOME) was utilized as a simulation tool.

WebGNOME, developed by the National Oceanic and Atmospheric Administration (NOAA), has been widely used for oil spill simulation and response planning [65]. It simulates processes, such as oil dispersion, movement, and evaporation based on various parameters. Among its multiple modes, WebGNOME utilizes Automated Data Inquiry for Oil Spills 2 (ADIOS2) as its oil weathering model [66]. The model allows simulation of the oil weathering process according to environmental conditions and elapsed time [67]. The reliability has been demonstrated through various studies and real-world cases. Röhrs et al. (2018) examined the influence of horizontal oil spill movement using data obtained through ADIOS2 [64]. Zhong et al. (2022) reported that ADIOS2 was effective in predicting the weathering processes of various oils and is widely applied [68].

2.1.3. Data Collection and Preprocessing

The pre-training process was conducted in three steps: data collection from scenario-based simulation, data preprocessing, and feature selection. These pre-training process was shown as in Figure 1. Scenario parameters were specified as input variables for the simulation to generate the dataset. The input parameters included wind speed, wave height, water temperature, spilled oil type, and elapsed time since the spill. The elapsed time was applied at 1 h intervals from the moment of the spill up to 120 h, resulting in 121 time points. Water temperature was set at 6, 16, and 30 °C, wind speed at 0.88 and 8 m/s, and wave height at 0 and 0.2 m. For each oil type, 1452 data points were generated by considering all parameter combinations, yielding a total of 17,424 data points.

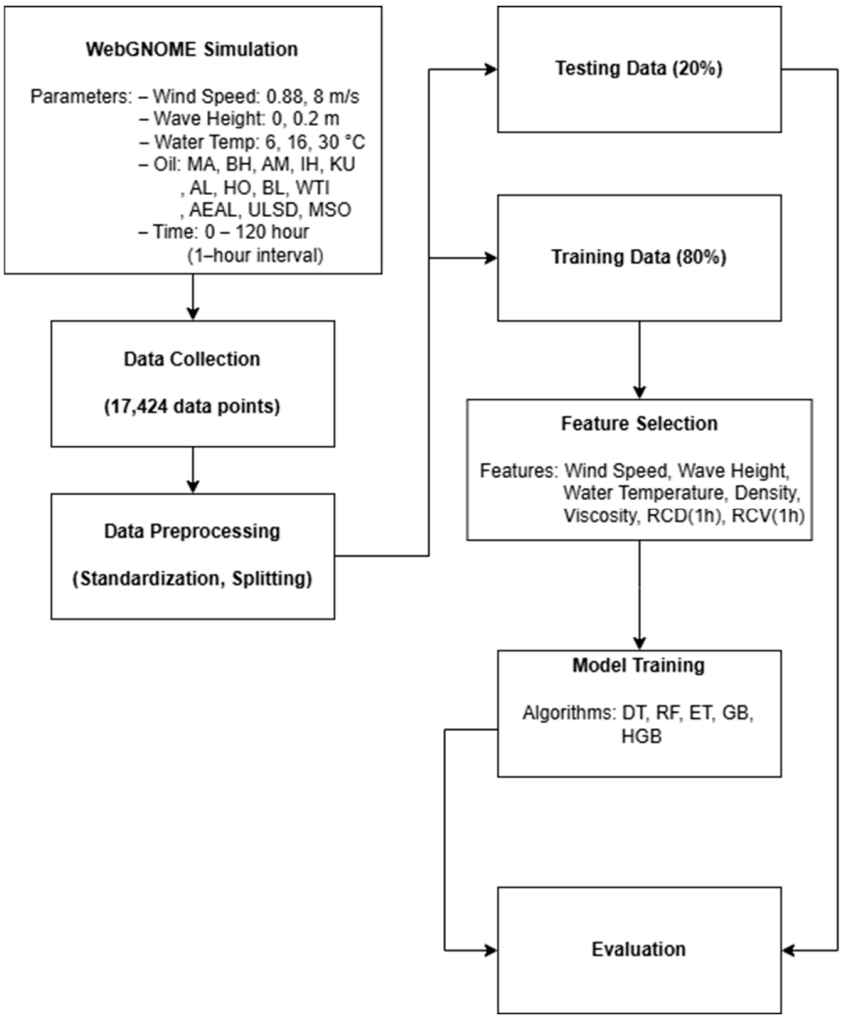


Figure 1. Flowchart of data collection, data processing, feature selection, model training, and evaluation.

The parameters used in the scenarios and their corresponding results were combined into a single dataset. This dataset included oil type, API gravity, wind speed, wave height, water temperature, evaporation, dispersion, sedimentation, water content, density, viscosity, and elapsed time. The data were split into training and test sets at an 8:2 ratio and standardized using StandardScaler from the Scikit-learn Python library. The version of Scikit-learn and Python was 1.0.2 and 3.9.13.

2.1.4. Feature Selection

Features were selected for model training. Features, which are independent variables in ML and data analysis, are essential for enabling models to learn patterns and make predictions [69]. The selection of features and data represents a critical step that significantly influences model performance and predictive accuracy [70].

The model was constructed to predict spilled oil using seven features: wind speed, wave height, water temperature, density, viscosity, the rate of change in density ($RCD_{\Delta t}$), and the rate of change in viscosity ($RCV_{\Delta t}$). API gravity, a property of spilled oil, was not considered in the model as it is measured under specific conditions. Water content, evaporation, floating, sedimentation, and dispersion were excluded from model training due to the difficulty of obtaining reliable measurements from field samples in spill environments [71]. $RCD_{\Delta t}$ and $RCV_{\Delta t}$ were derived features calculated using simulation data and equations. By incorporating both density and viscosity alongside their rates of change, the model aimed to leverage a broader range of features for enhanced predictive performance. These features were calculated using the following equations:

$$RCD_{\Delta t} = (\rho_t - \rho_{t-\Delta t})/\Delta t \quad (1)$$

$$RCV_{\Delta t} = (\mu_t - \mu_{t-\Delta t})/\Delta t \quad (2)$$

where ρ, μ, t, and Δt represent density, viscosity, elapsed time, and the difference in elapsed time for the rate of change, respectively. $RCD_{\Delta t}$ and $RCV_{\Delta t}$ varied depending on the difference in elapsed time as in Equations (1) and (2). The difference in elapsed time was set to 1, and the prediction results from models incorporating these features were analyzed in Section 3.1.

The difference in elapsed time could be set to values other than 1, suggesting the potential for enhancing predictive accuracy by incorporating additional features. This is examined in Section 3.3, where the impact of adding rate-of-change features with elapsed time differences of 2 and 3 on model performance is analyzed in detail. Additionally, the advantages and limitations of incorporating these features are evaluated.

This study primarily aimed to assess the FAL of the spilled oil prediction model. Feasibility was evaluated by adjusting features and modifying data to develop models applicable to various environments. The advantages and limitations of these models were analyzed to support feasibility by comparing their results. Section 3.1 presents this comparison using baseline models as a reference.

Sections 3.2 and 3.3 examined the FAL of models with additional or excluded features beyond the seven baseline features. Section 3.4 analyzed the FAL of models trained on limited data, evaluating their effectiveness in predicting oil within specific data ranges. Section 3.5 investigated the FAL of models by assessing accuracy variations caused by inherent measurement errors in the input data.

2.2. ML Algorithms

In this study, the FAL of the ML model was evaluated using five algorithms: decision tree (DT), random forest (RF), extra tree (ET), gradient boosting (GB), and histogram-based

gradient boosting (HGB). This section provides an overview of the principles underlying these algorithms and details their hyperparameters.

2.2.1. Decision Tree

DT is one of the most widely used methods for classification tasks [72]. It was developed to predict target variables based on rules and conditions among input variables. These rules and conditions are represented in a tree structure, enabling both classification and the determination of dependent variables [73]. This study utilized the classification and regression tree (CART) algorithm [74]. CART constructed the decision tree by recursively performing binary splits on the input space in a way that minimizes the impurity at each node. For a given subset, the algorithm split the data into left and right child nodes based on a selected feature and a corresponding threshold. The quality of each split was evaluated using the Gini index. The Gini index for a specific node t was calculated using the following equation:

$$\text{Gini}(t) = 1 - \sum_{k=1}^{k} p_k^2 \tag{3}$$

where p_k represents the proportion of samples in node t that belong to class k. A lower Gini value indicated a purer node, meaning the samples are more likely to belong to a single target class. CART recursively selected the split that minimizes this Gini index until a predefined stopping condition is met. These conditions may include factors such as the minimum number of samples in a node, the class distribution within the node, or the maximum depth of the tree. The DT algorithm facilitated the visual interpretation of data and results, providing insights into the decision-making process. Additionally, significant variables were identified through the constructed tree structure and basic functions, enhancing interpretability [75]. Consequently, the model's transparent structure promoted easy interpretation of the classification results [76].

2.2.2. Random Forest

RF is an ensemble technique composed of multiple decision trees (DTs) used for prediction tasks. RF efficiently handles large datasets and identified significant variables, even in high-dimensional data, enhancing both accuracy and interpretability [77]. Each tree was constructed using datasets derived from the training data through bootstrap sampling, where sampling with replacement is employed [73]. During the construction of each tree, node splitting was performed not across the entire feature set but within a randomly selected subset. The splitting criterion was typically based on impurity measures, such as the Gini index, with the feature that maximally reduced impurity being selected for the split. Predictions are determined through aggregation of individual tree outputs: via majority voting in classification tasks. For a given input (x), the prediction (y) is obtained by aggregating the outputs of all individual trees (B) as shown in Equation (4):

$$y = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{4}$$

2.2.3. Extra Tree

ET is an ensemble learning method that shares similarities with. Unlike RF, it differs notably in its sampling and splitting strategies. ET did not use duplicated training data. While RF determined optimal splits among randomly selected features to divide nodes, ET randomly selected both the number of samples and features for node division [53]. Among these, a split rule was selected to divide the node into two nodes. This splitting

was applied to each node until a leaf node was reached. A leaf node was defined as a terminal node with no further splits. Although single decision trees can pose challenges in prediction accuracy, the ensemble structure of multiple trees in ET mitigated the risk of overfitting. Furthermore, ET addressed issues related to permutation importance by preventing excessive reliance on specific features, enhancing the model's robustness and interpretability [78].

2.2.4. Gradient Boosting

GB is an algorithm that improves the learning capability of weak learners to minimize the model's loss [79]. Weak learners were trained sequentially, with subsequent learners assigned weights based on the correctness of previous predictions. This iterative process allowed trees to be constructed to compensate for prior errors, enhancing model performance while minimizing the loss function [80]. At iteration m, the model was updated by adding a newly trained weak learner $h(x)$, weighted by a learning rate γ, to the existing model, as shown in Equation (5):

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \tag{5}$$

where $h_m(x)$ is trained to approximate the gradient of the loss function with respect to the current model's predictions. The learning rate controls the contribution of each weak learner and can be used to regulate the complexity of the overall model.

2.2.5. Histogram-Based Gradient Boosting

HGB follows the same principle as GB by minimizing residuals from previous learners. Unlike the traditional approach that uses individual decision trees, HGB approximates the underlying data distribution using histograms. It utilized 256 bins for all features to obtain optimal splits at tree nodes [81]. This approach allows the model to train based on an approximation that closely reflects the original data distribution. Due to this property, HGB is well suited for datasets with outliers. In addition, HGB is capable of processing large datasets more efficiently, leading to faster training times [82].

2.3. Hyperparameter Tuning

Hyperparameters are variables that defined the model structure or controlled the learning process, set prior to training, and directly impacted model performance and complexity [83–85]. For tree-based algorithms, the number of trees (n_estimators) and maximum depth (max_depth) determined the branching structure, while the minimum number of samples required for a split (min_samples_split) and the minimum number of samples per leaf (min_samples_leaf) regulated tree complexity [86]. GB and HGB incorporated these parameters along with the learning rate (learning_rate) to prevent overfitting and enhance predictive accuracy [87]. In this study, hyperparameter tuning was conducted individually for all five classification algorithms. The tuning process was performed using the entire training dataset and a set of seven selected features, with classification accuracy used as the performance metric. As shown in Table 2, a range of values was tested for each hyperparameter, and the optimal values were selected based on their accuracy. The hyperparameter test ranges and tuning results obtained in this study are presented in Table 3. For all models except gradient boosting (GB), the default settings provided by the Scikit-learn package resulted in better performance than modified configurations. For GB, the optimal learning rate and number of trees were set to 0.2 and 400, respectively. The same set of hyperparameters was maintained throughout subsequent experiments, even when the feature set or data composition was modified. This approach

allowed the study to focus on analyzing the impact of feature configuration changes while eliminating performance variation caused by hyperparameter differences.

Table 2. Search range of hyperparameter for DT, RF, ET, GB, and HGB models.

Algorithm	N_Estimators	Max_Depth	Min_Samples_Split	Min_Samples_Leaf	Learning_Rate
DT	10–1000	None, 1–30	2–14	1–10	NaN
RF	1–500	None, 1–7	2–12	1–5	NaN
ET	1–1000	None, 1–28	2–14	1–5	NaN
GB	1–700	None, 1–20	2–12	1–5	0.01–0.3
HGB	1–400	None, 1–28	NaN	1–5	0.01–0.3

Table 3. Hyperparameter values for DT, RF, ET, GB, and HGB models.

Algorithm	N_Estimators	Max_Depth	Min_Samples_Split	Min_Samples_Leaf	Learning_Rate
DT	100	None	2	1	NaN
RF	100	None	2	1	NaN
ET	100	None	2	1	NaN
GB	400	None	2	1	0.2
HGB	100	None	2	1	0.1

2.4. Evaluation Measures

Evaluation metrics are essential components in assessing the performance of predictive models. Model evaluation was conducted using metrics derived from the confusion matrix, which categorized predicted values against actual values. The predicted classes obtained from the simulation-generated test data were compared with the actual values and recorded accordingly. Model performance was evaluated using a confusion matrix. As the confusion matrix in this study was multiclass, elements along the diagonal represented correctly classified instances, while off-diagonal elements indicated misclassifications [88,89]. The comparison between the predicted and actual values across the 12 classes was used to calculate the number of true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP). TP represented cases where the model correctly identifies positive instances, while TN indicated cases where the model correctly classifies negative instances. FP reflected cases where the model incorrectly classifies positive instances, and FN denoted cases where the model incorrectly classifies negative instances.

Model performance was assessed using accuracy, precision, recall, and the F1-score [90,91]. Accuracy indicates the proportion of total instances that are correctly classified (TP) by the model. Precision refers to the proportion of predicted positive cases that are truly positive. Sensitivity, also known as recall, represents the proportion of actual positive cases that are accurately identified. The F1-score, calculated as the harmonic mean of precision and sensitivity, provides a balanced measure that reflects both false positives and false negatives. The formulas used for evaluation are summarized in Table 4.

Table 4. Formulas of accuracy, precision, sensitivity, and F1-score for model evaluation.

Metric	Formula
Accuracy	$\frac{\text{Number of true positive cases}}{\text{Total number of predicted cases}}$
Precision	$\frac{TP}{TP+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
F1-score	$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision}+\text{recall}}$

3. Results

3.1. Prediction Performance and Contribution of Features

The performance metrics of the five algorithms discussed in Section 2.2 are presented in Table 5. Overall, the models exhibited strong predictive performance for spilled oil. The ET model achieved the highest accuracy (88.55%), making it the most accurate model. The HGB model performed similarly, with an accuracy of 88.41%. The RF and GB models followed, achieving accuracies of 86.4% and 85.45%, respectively. The DT model demonstrated the lowest accuracy among the five algorithms. For all models, precision, sensitivity, and F1-score closely aligned with their respective accuracy values.

Table 5. Performance metrics of DT, RF, ET, GB, and HGB models for oil spill classification.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
DT	83.41	83.60	83.44	83.50
RF	86.40	86.69	86.41	86.48
ET	88.55	88.70	88.56	88.59
GB	85.45	85.62	85.49	85.49
HGB	88.41	88.55	88.39	88.43

The permutation importance of the features used in each algorithm is illustrated in Figure 2. This method was employed to evaluate the relevance of individual features in the predictive model, representing the contribution of each feature to model performance [92,93]. Density and viscosity exhibited significant contributions, indicating their importance in model performance. Specifically, density showed the highest contribution in the HGB model (72%) and the lowest in the GB model (38%). Water temperature and wind speed played more influential roles in the ET model, contributing 39% and 34%, respectively, which corresponded to the model's highest accuracy among all algorithms. However, environmental features, such as wind speed, wave height, and temperature contributed less to the model's performance. Additionally, utilizing ML models without environmental features could serve as a simple alternative. Since environmental features did not significantly affect accuracy. Section 3.2 presented a comparison of prediction model results with and without environmental features. RCD_1 and RCV_1 were shown to be as important as density and viscosity, suggesting that the rate of change in properties over time could enhance model accuracy. The relationship between model performance and the number of features was analyzed to determine whether performance continued to improve as additional features were included. In Section 3.3, rates of change over 2 and 3 h periods (RCD_2, RCV_2, RCD_3, and RCV_3) were incorporated, and their impact on model accuracy was evaluated by analyzing permutation importance and the associated limitations.

The study examined the accuracy of the model for a particular oil type. Figure 3 presents the number of prediction failures for all models and oils. AM and HO oils had the highest average number of prediction failures, with 93 and 73 failures, respectively. MSO and AL followed, with 55 and 52 failures. In contrast, BH, MA, and ULSD exhibited significantly fewer prediction failures, with approximately nine, eight, and one, respectively. The number of prediction failures varied across algorithms. The ET model, which had the highest accuracy, demonstrated fewer prediction failures for oils other than AM and HO compared to other algorithms. The HGB model, the second-most accurate, showed fewer prediction failures for HO than the ET model, but a higher number for AL. In contrast, the GB and DT models exhibited an increase in prediction failures for oils other than HO and AM. Therefore, understanding the causes of these failures was essential, prompting a corresponding investigation.

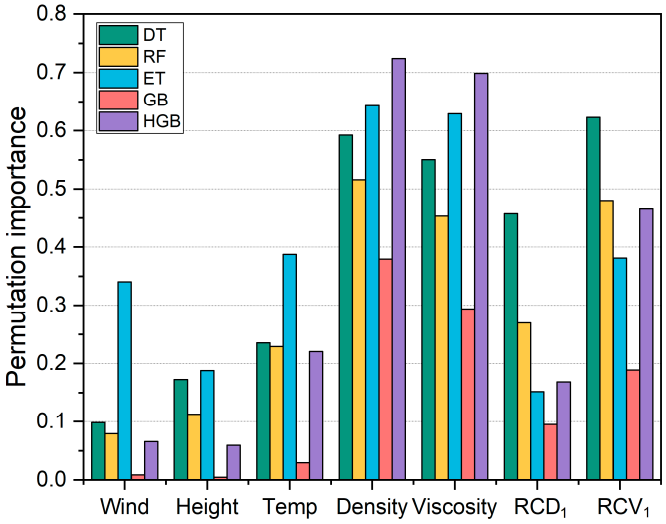


Figure 2. Permutation importance in DT, RF, ET, GB, and HGB models for oil spill prediction.

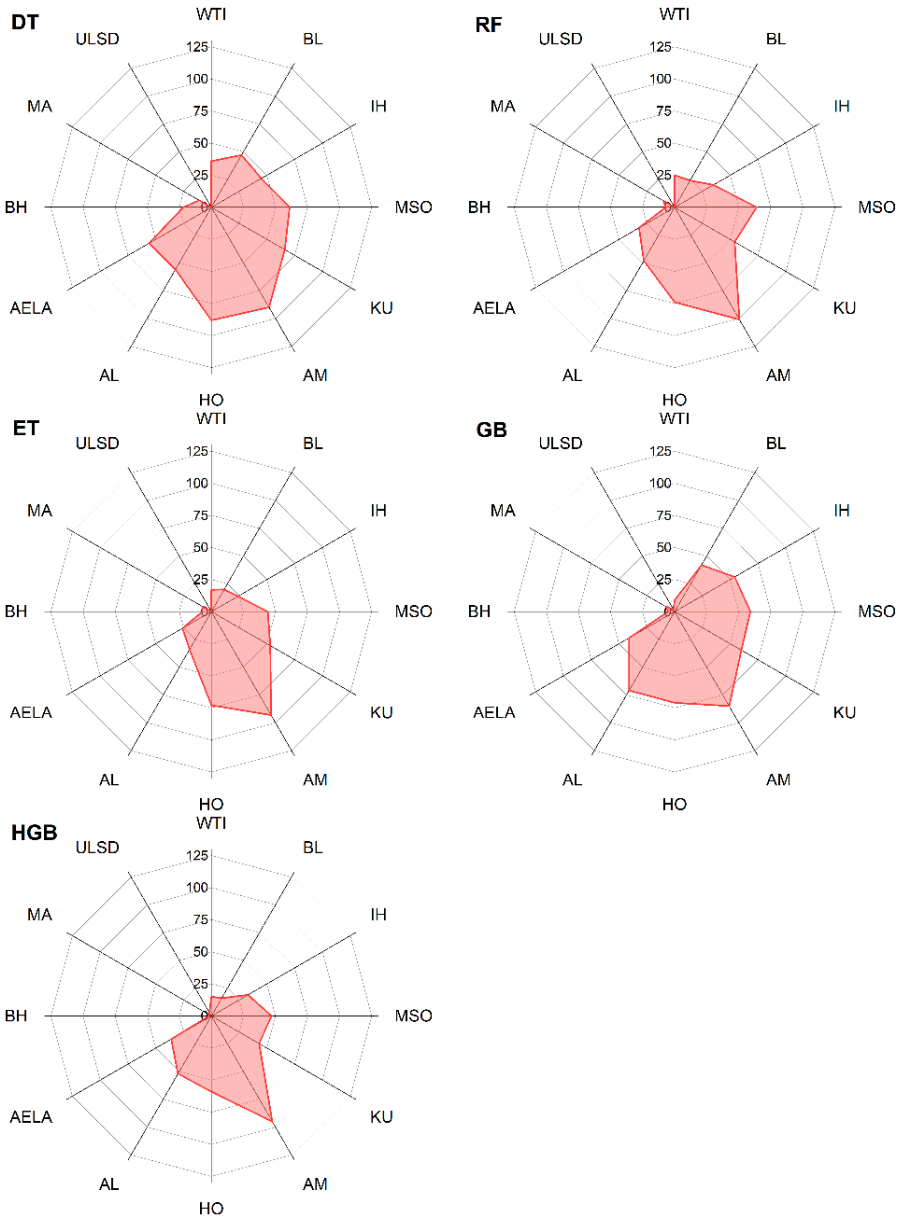


Figure 3. Number of prediction failures in DT, RF, ET, GB, and HGB models based on twelve oils.

Figure 4 presents the confusion matrices of the five algorithms used to evaluate the performance metrics in Table 5. The highest misclassification rates occurred in AM and HO, with frequent misclassifications between these two oils. In the ET model, AM was misclassified as HO in 53 instances, while HO was misclassified as AM in 50 instances. Similarly, in the HGB model, AM was misclassified as HO in 50 instances, and HO as AM in 43 instances. Additionally, major misclassifications were observed between KU and AM, BL and ALEA, and ALEA and MSO. This pattern indicates that misclassification errors were not evenly distributed across all oils but were concentrated among specific ones. Furthermore, the frequency of these misclassifications varied across algorithms. In contrast, misclassifications among other oils were minimal or nonexistent. The GB model exhibited this trend to a lesser extent than other models. However, as misclassifications decreased for certain oils, they increased for others, contributing to the GB model's lower accuracy compared to the ET and HGB models.

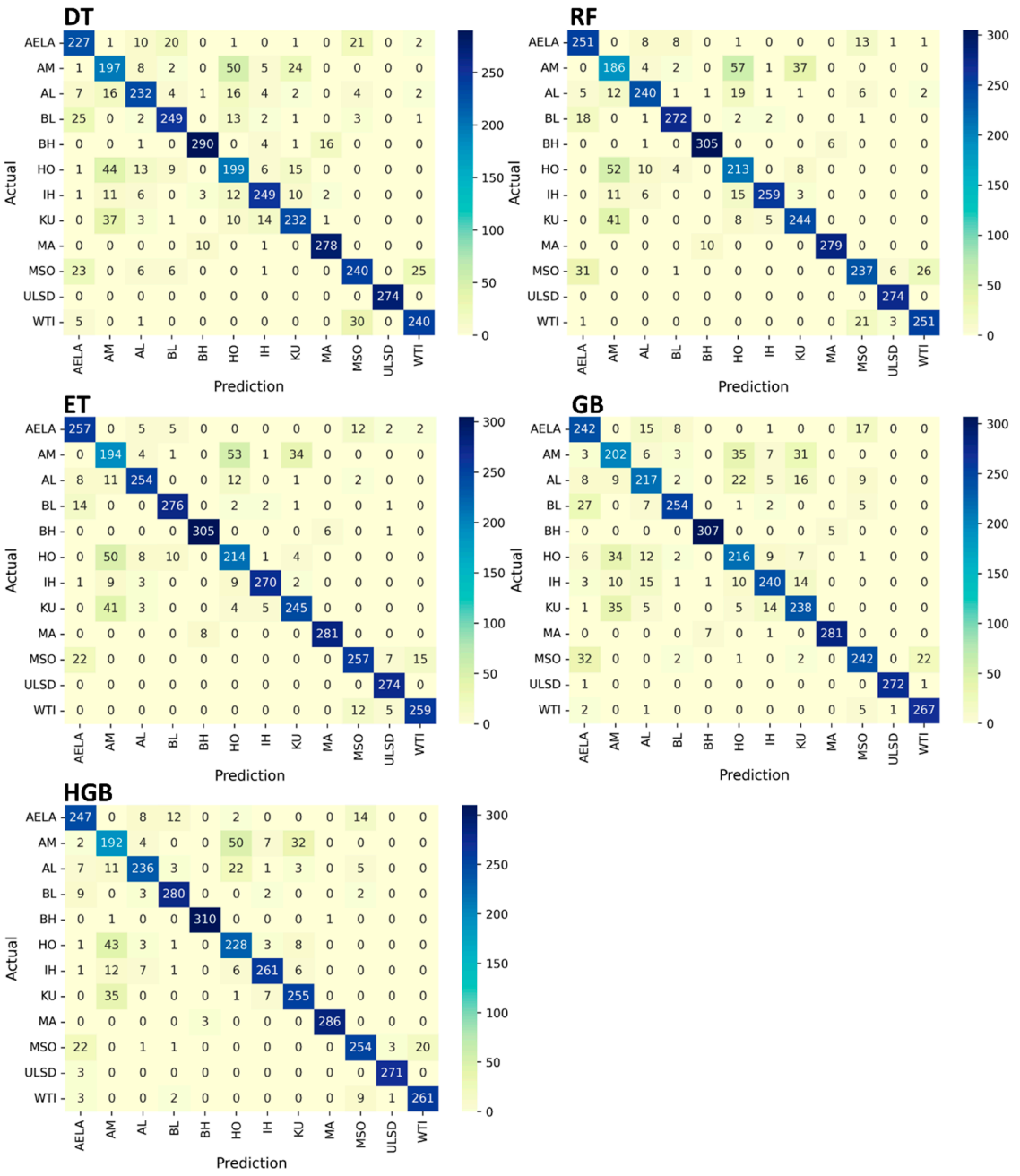


Figure 4. Confusion matrices for DT, RF, ET, GB, and HGB models in twelve-class oil type classification.

Although the ML model demonstrated high overall accuracy, variations in prediction accuracy across different oils led to misclassifications. These issues should be identified during model training, making it essential to determine where and why they occurred. Among the features used for training, viscosity and density exhibited high permutation importance. These variables significantly contributed to the model's performance, and their impact on misclassifications was analyzed.

Figure 5 presents the number of test data points and the number of prediction failures per oil within the density range of 800 kg/m^3 to 1050 kg/m^3. Similarly, Figure 6 illustrates the number of test data points and prediction failures per oil within the viscosity range of 0 cSt to 300,000 cSt. The distribution of density data, excluding ULSD, was most concentrated around 1020 kg/m^3. Based on the analysis of Figures 5 and 6, BL, MSO, and ALEA oils exhibited relatively fewer prediction failures in the ET model within the density range of approximately 1020 kg/m^3 and the viscosity range of 0–500 cSt. This contrasts sharply with the DT model, highlighting how differences in ML performance within these data ranges contribute to variations in model accuracy. Meanwhile, ULSD exhibited a distinct data distribution pattern at significantly low viscosity, as shown in Appendix A, Table A1. Its lower viscosity compared to other oils explained its lower prediction failure rate. Similarly, Figure 6 illustrates the higher prediction failure rates observed in AM, HO, MSO, and AL. The concentration of data in the low-viscosity range (0–1000 cSt) reduced accuracy, as the similarity among these data points made classification more challenging. A similar trend was observed in the density distributions of these oils. BH and MA exhibited either a wider distribution compared to other oils or a relatively limited distribution, similar to that of ULSD. These findings indicate that data distribution and the training range can significantly influence model performance. In Section 3.4, the model was trained using 10 oils, excluding ULSD and BH, to analyze the results of a model focusing on the remaining oils.

3.2. Performance Changes with Fewer Features

In Section 3.1, the permutation importance related to model prediction performance was analyzed. The results indicated that environmental features such as wind speed, wave height, and water temperature had considerably lower importance compared to density, viscosity, and their rate of change. Based on these findings, the impact of excluding low-importance features on the model was investigated. The performance of models using all features from Section 3.1 was compared with models that excluded environmental features.

The accuracy of the models without environmental features was compared, as shown in Figure 7, which was based on data from the confusion matrix in Appendix A, Figure A1. The DT, RF, and ET models demonstrated improved accuracy even when environmental features were excluded, with increases ranging from 2.1% to 2.8%. The ET model exhibited the best performance, achieving an accuracy of approximately 90.65%. In contrast, the GB and HGB models experienced reductions in accuracy of up to 3.3%. Overall, ensemble methods such as RF and ET demonstrated improved performance. The RF model showed the greatest improvement, achieving an accuracy of 89.21%, while the DT model recorded a higher accuracy of 85.94% compared to the GB model when environmental features were included. However, sensitivity to feature reduction led to decreased accuracy in the GB and HGB models. The GB model showed the lowest accuracy at 82.18%, and the HGB model dropped to 85.8%, which was lower than the accuracy of the DT model.

Figure 8 presents a comparison of the permutation importance of models trained with and without environmental features. The DT, RF, and ET models demonstrated improved accuracy despite the reduction in the number of features due to the removal of three environmental variables. In the ET model, the RCD_1 and RCV_1 features were assigned

lower importance, with RCV_1 being the least important in most models. As shown in Figure 7, models that exhibited performance improvements (DT, RF, and ET) showed an increase in the permutation importance of density and viscosity features, whereas models with decreased performance experienced either a reduction or minimal change in their importance.

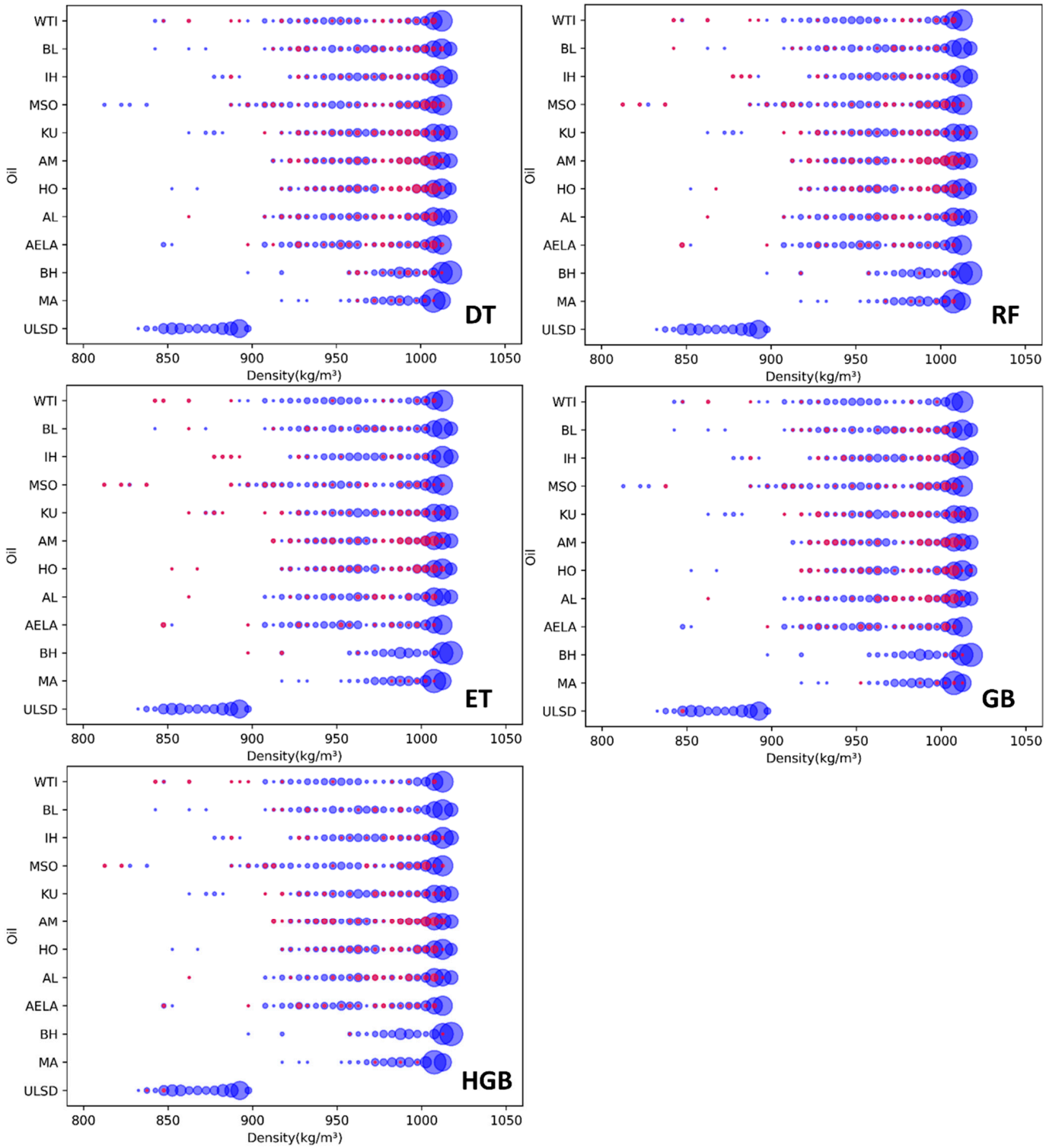


Figure 5. The distribution of test data and prediction failures in the density domain for DT, RF, ET, GB, and HGB models across twelve oils. (Blue: number of test data, red: number of prediction failures. The size of the circle is indicative of the number of the data.).

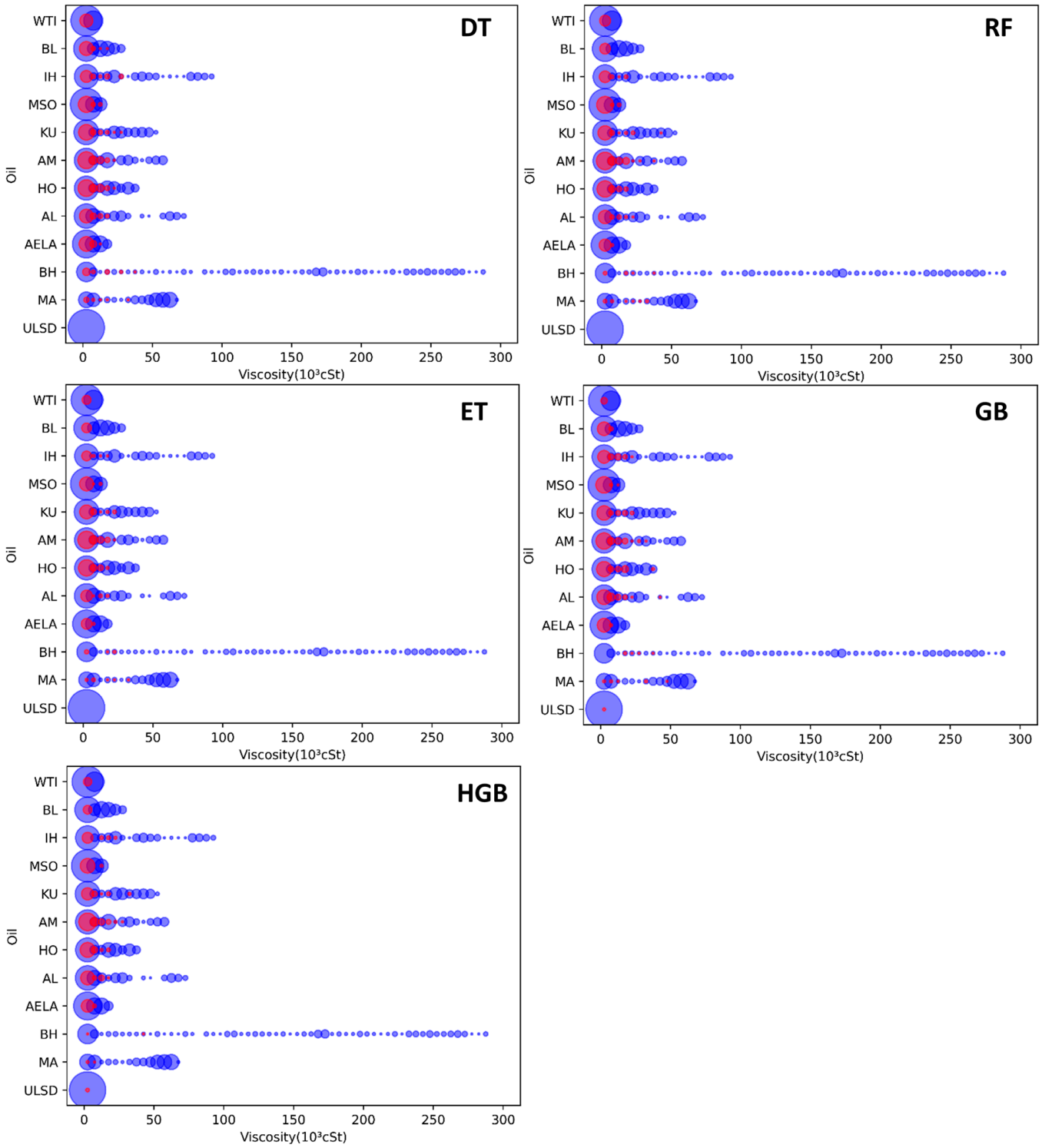


Figure 6. The distribution of test data and prediction failures in the viscosity domain for DT, RF, ET, GB, and HGB models across twelve oils. (Blue: number of test data, red: number of prediction failures. The size of the circle is indicative of the number of the data).

As a result, the ET and RF models were suitable choices with fewer features, as their accuracy was driven by the increased contribution of viscosity and density features. The HGB model, which had the second-highest accuracy after the ET model in Section 3.1, experienced a decline in permutation importance and accuracy due to the reduction in features, negatively impacting its performance.

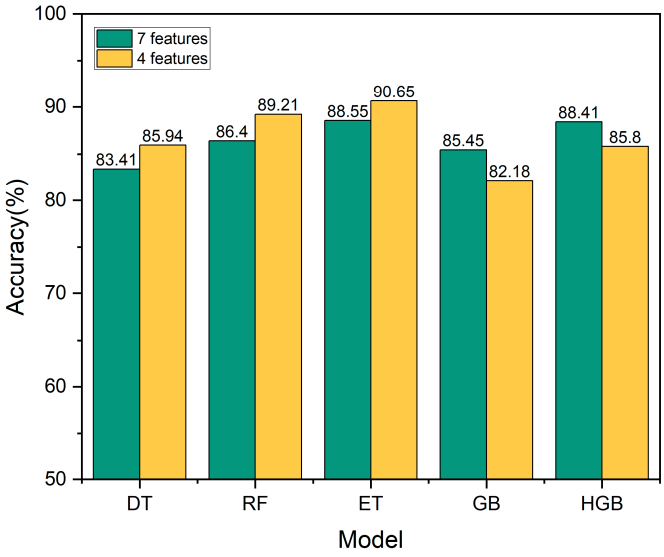


Figure 7. Accuracy comparison of DT, RF, ET, GB, and HGB models trained with and without environmental features.

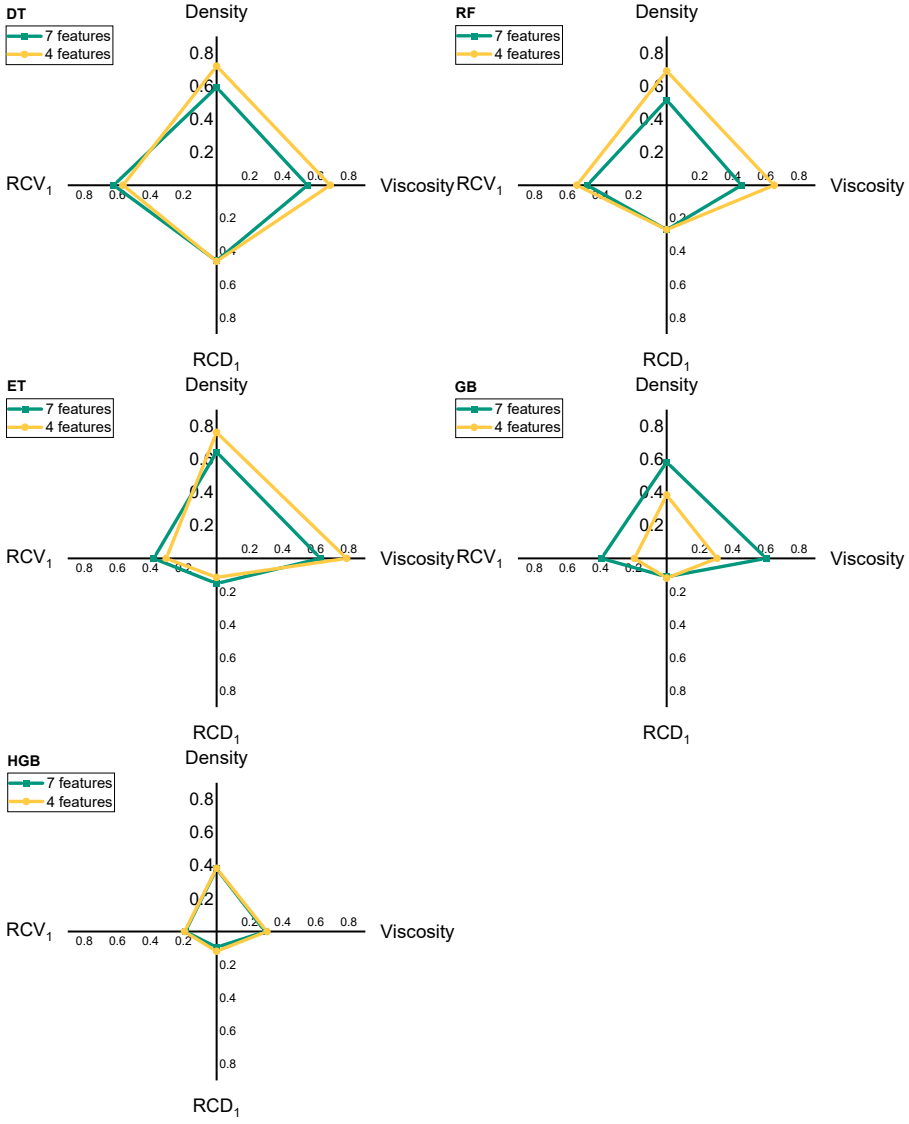


Figure 8. Permutation importance comparison of DT, RF, ET, GB, and HGB models trained with environmental features excluded.

3.3. Performance Changes with More Features

Based on Section 3.1, features related to the rate of change over multiple time intervals were investigated. The importance of RCD_1 was lower than that of density and viscosity but higher than that of other environmental features. Similarly, RCV_1 was found to be as important as viscosity. The study examined whether increasing the number of these secondary features proportionally improved performance. Models incorporating multiple $RCD_{\Delta t}$ and $RCV_{\Delta t}$ features, calculated using extended elapsed time differences (Δt = 2 and Δt = 3 h), were compared with the models explored in Section 3.1. This comparison aimed to assess how increasing these features affected model performance.

To facilitate comparison, three different cases were considered: (1) models trained with the features used in Section 3.1, (2) models incorporating the rate of change in density and viscosity over a 2 h interval (RC_2), and (3) models incorporating both RC_2 and the rate of change over a 3 h interval (RC_3).

Figure 9 presents the accuracy of the five algorithmic models for these three cases: (1) models with RC_1, (2) models with RC_2, and (3) models with both RC_2 and RC_3. The accuracy of the models in the second and third cases was derived from the confusion matrix in Appendix A, Figures A2 and A3. As more rate-of-change features were added, the accuracy of the GB and HGB models declined to 84.76% and 88.24%, respectively. In contrast, the DT model exhibited a gradual but steady increase in accuracy from 83.41% to 83.96%. The permutation importance of these models is shown in Figure 10. The addition of RC_2 and RC_3 resulted in a consistent decline in permutation importance across all features. In contrast to models based on other algorithms, the permutation importance in the GB and HGB models exhibited relatively minor changes, and their corresponding accuracies remained unaffected. In terms of permutation importance, the DT model maintained the relative importance of each feature as additional rate-of-change features were introduced. The increase in DT model accuracy suggests that these added features positively contributed to model performance. The RF and ET models demonstrated a significant improvement in accuracy from the first case, reaching 90.62% and 92.83%, respectively, but exhibited a slight decline in the third case.

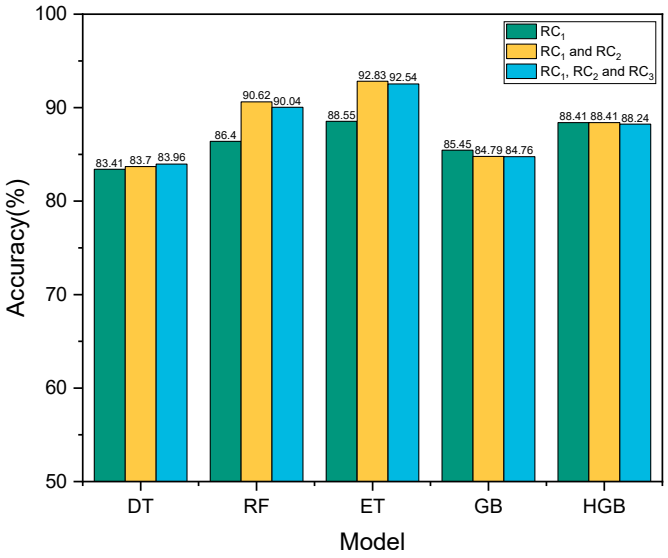


Figure 9. An accuracy comparison of the DT, RF, ET, GB, and HGB models with the addition of $RCD_{\Delta t}$ and $RCV_{\Delta t}$ across three cases (green: incorporating RC_1, yellow: incorporating RC_1 and RC_2, and blue: incorporating RC_1, RC_2, and RC_3).

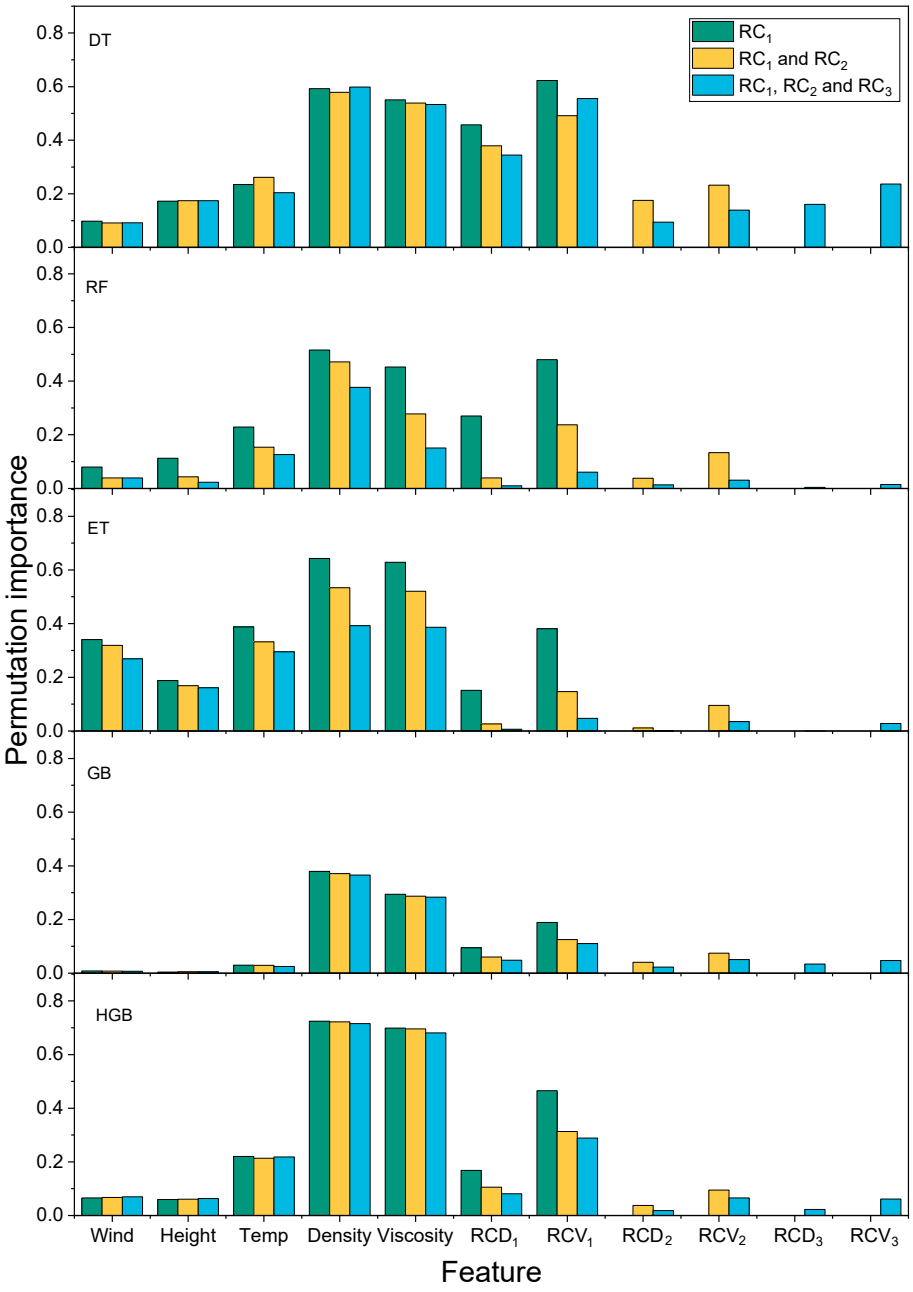


Figure 10. A permutation importance comparison of the DT, RF, ET, GB, and HGB models with the addition of $RCD_{\Delta t}$ and $RCV_{\Delta t}$ across three cases (green: incorporating RC_1, yellow: incorporating RC_1 and RC_2, and blue: incorporating RC_1, RC_2, and RC_3).

3.4. Effect of Widespread Data in Property Features

The results in Section 3.1 indicated relatively few failed predictions for BH, MA, and ULSD. This strong prediction performance could be attributed to the dataset distribution in the feature domain, as shown in Figures 5 and 6. In contrast, oils with a high number of prediction failures were concentrated in specific data regions. This section analyzes the results of models trained on a reduced data range, focusing on low-viscosity oils with limited training data. In modifying the data composition, ULSD and BH were excluded, thereby narrowing the training data range. However, MA was retained, as its exclusion did not alter the overall data range. The accuracy of the model trained on 10 oils was compared to the results in Section 3.1, with particular attention to accuracy changes in viscosity ranges where prediction failures were frequent. The feature set used remained consistent with Section 3.1.

The accuracy of the model trained without ULSD and BH was compared to the results in Section 3.1, as shown in Figure 11, which is based on the confusion matrix data in Appendix A, Figure A4. The model trained without ULSD and BH exhibited lower accuracy than the model trained with the full dataset. A reduction in accuracy was observed across models developed using all algorithms. The ET model experienced the smallest loss of accuracy, decreasing from 88.55% to 88.22%, while the DT model showed the most significant decline, from 83.41% to 81.61%. With the exception of these two algorithmic models, the remaining models demonstrated a reduction in accuracy of approximately 1% to 2%. However, this accuracy evaluation encompassed the entire viscosity range, meaning that excluding ULSD and BH—originally associated with high accuracy—may have resulted in a greater increase in prediction failures than the gain from increased true positives (TP) in the concentrated test data. Therefore, an analysis was conducted to assess whether the reduced training range affected the prediction performance of oils within the concentrated test data.

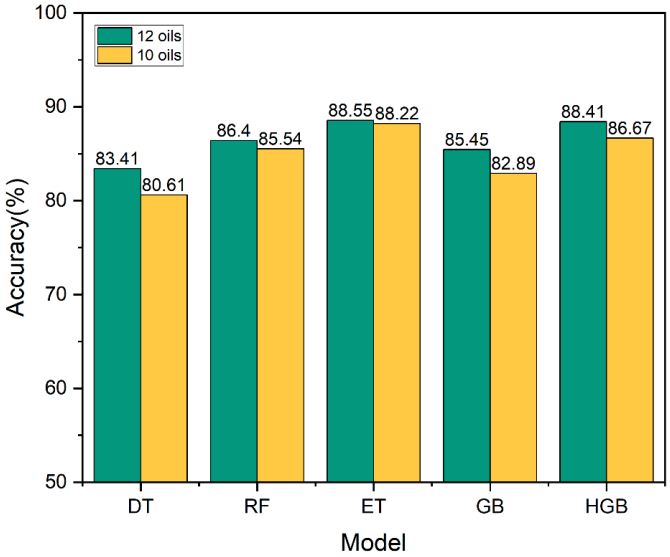


Figure 11. Accuracy comparison of DT, RF, ET, GB, and HGB models with and without ULSD and BH oils in training data.

Table 6 presents the prediction failure rates for test data within the 1000 cSt range and across the entire viscosity range, comparing models trained with and without ULSD and BH. For the ET model, the prediction failure rate for data below 1000 cSt significantly decreased from 24.97% to 21.08% after excluding ULSD and BH. Similarly, the RF model exhibited a reduction from 27.97% to 24.41%. The decrease was less pronounced for the HGB, GB, and DT models. These results indicate that although most models exhibited a reduction in prediction failures across the entire viscosity range, their improvements were particularly notable within the concentrated viscosity range.

Table 6. A comparison of prediction failure rates (%) within the 1000 cSt viscosity range and across the entire viscosity range for DT, RF, ET, GB, and HGB models with and without ULSD and BH oils in training data.

Model	DT	RF	ET	GB	HGB
12 oils	28.46	27.97	24.97	27.59	24.34
10 oils	28.17	24.41	21.08	25.16	21.40

3.5. Effect of Input Uncertainty

Oil weathering affects spilled oil differently over time, leading to significant changes in physical properties such as viscosity. During the sampling of spilled oil, the measurement process may be subject to inherent errors, thereby introducing uncertainties in the assessment of viscosity and other weathering-related properties. These uncertainties increase input variability in the model, directly impacting the accuracy of predictions and feasibility of ML models. This section examined how prediction accuracy was affected by uncertainty in viscosity data used as input. The models were trained using five algorithms and five features, excluding RCV_1 and RCD_1. Since RCV_1 and RCD_1 were derived from uncertain viscosity data, they were omitted from the analysis. The training data were not subject to uncertainty, whereas the test data contained uncertainties.

Figure 12 sequentially illustrates how the test data were utilized as inputs while incorporating uncertainty. The test data were duplicated 20 times. In each duplicated test dataset, the viscosity data were adjusted to reflect viscosity at a specific time before or after the original measurement. The viscosity uncertainty corresponding to the change at time t was applied at 1 h intervals, ranging from -10 to +10 h. Data earlier than 9 h or later than 111 h were removed, as they could not be adjusted using viscosity values from 10 h before or after. Following this, viscosity and other relevant data were grouped into a single dataset based on symmetric time shifts (-1 and +1 h, -2 and +2 h, …, -10 and +10 h). The data were then arranged according to normalized viscosity, with values ranging from -0.85 to 0.85 in increments of 0.05. The test data, categorized by normalized viscosity intervals, were used as input for the model to evaluate prediction performance. Normalized viscosity represents the ratio of the difference between the manipulated and unmanipulated viscosity to the unmanipulated viscosity, as shown in Equation (6).

$$\text{Normalized viscosity} = (\mu_t - \mu_{t+u})/\mu_t \tag{6}$$

where μ, t, and u represent viscosity, elapsed time, and uncertainty, respectively.

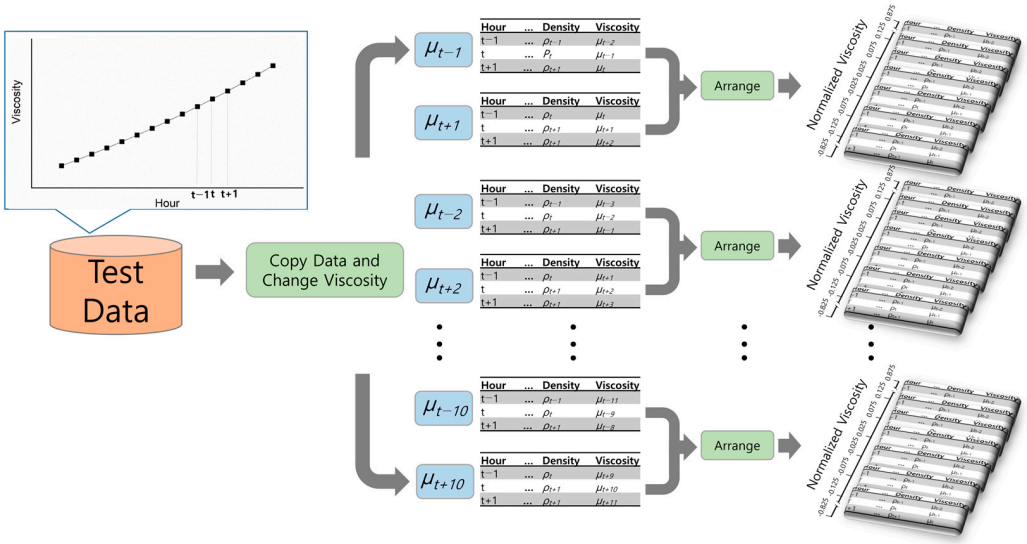


Figure 12. Process of preparing time-shifted data with viscosity uncertainty for input into predictive models.

The accuracy of five algorithmic models (DT, RF, ET, GB, and HGB) trained on unmanipulated data was 81.18%, 73.4%, 75.95%, 76.59%, and 79.66%, respectively. These accuracies were then compared to those obtained using inputs with manipulated data. The accuracy values were derived from the confusion matrix presented in Appendix A, Figure A5.

Figure 13 illustrates changes in accuracy across normalized viscosity intervals when incorporating viscosity uncertainty. In the figure, uncertainty represents the extent of the shift in time *t* that alters viscosity. As uncertainty increased, the range of normalized viscosity intervals expanded by 0.2. All algorithmic models exhibited a sharp decline in accuracy, dropping from 93% to 5% as normalized viscosity increased. Notably, even a 0.1 normalized viscosity resulted in a reduction in accuracy to approximately 60–75%. However, the HGB model maintained relatively high accuracy, ranging from approximately 60% to 85%, across all normalized viscosity intervals when viscosity uncertainty corresponded to a 1 h time difference.

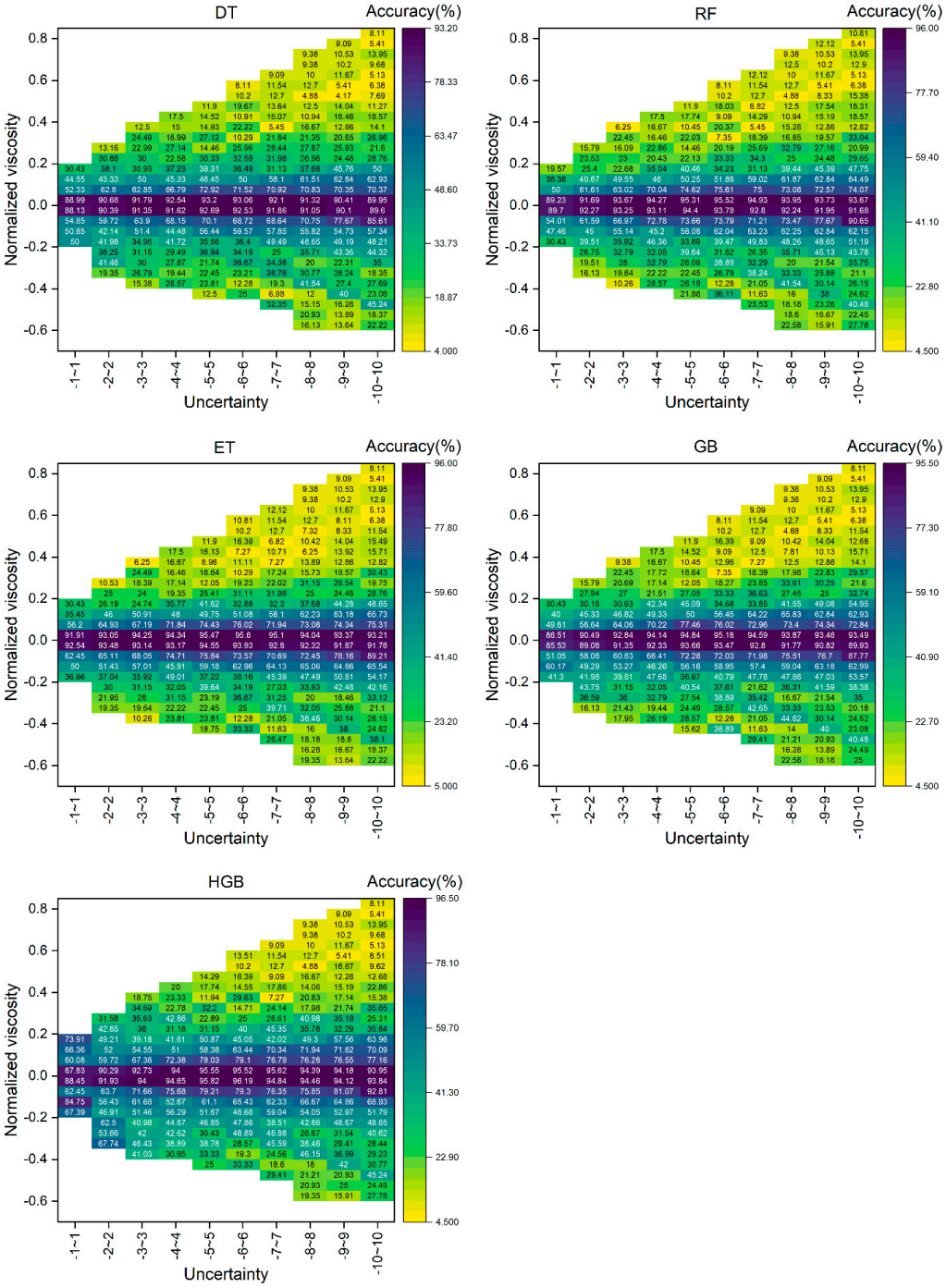


Figure 13. Accuracy changes in the DT, RF, ET, GB, and HGB models under increasing uncertainty levels and normalized viscosity intervals.

In the normalized viscosity intervals of −0.05 to 0 and 0 to 0.05, accuracy exhibited an increasing trend, reaching approximately 95% as uncertainty increased. However, beyond a 7 h uncertainty difference, accuracy began to decline. This decline was associated with the expansion of the normalized viscosity range as uncertainty increased, causing some previously misclassified data near 0 normalized viscosity to shift to farther intervals. Additionally, since mostly correctly predicted data remained within the initial interval, accuracy temporarily increased. In contrast, beyond the 7 h uncertainty, accuracy declined as correctly predicted data also shifted to farther intervals, resulting in an increase in prediction failures.

4. Discussion

This study compares the oil spill prediction performance of models that combine five different algorithms with varying configurations of input data and features. While all models generally demonstrated high classification accuracy across twelve oil types, substantial variation was observed depending on the specific oil and the selected feature set. The comparisons are based on models incorporating seven combinations of features, including environmental variables, physical properties of the oils, and the rate of change in those physical properties over one hour.

The ET and HGB models achieved the highest accuracy. Permutation importance analysis revealed that density and viscosity made the most significant contributions to model performance. Environmental features showed limited importance across all models. When these environmental features were excluded, the accuracy of the RF and ET models improved, suggesting that removing irrelevant data can enhance model performance in certain cases. GB and HGB models were negatively affected, and this decline was likely due to the impact of feature exclusion on its sequential learning process.

The integration of additional rate-of-change features over longer time intervals demonstrated that the ET model is a strong candidate for model selection. Additionally, an approach involving two sequential predictions—first using the rate-of-change data and then incorporating a longer time difference in the rate of change could be considered. However, more extensive incorporation led to a decrease in accuracy.

A substantial number of prediction failures were attributed to the model's inability to accurately classify AM and HO oils. These failures were particularly evident in cases where viscosity was below 1000 cSt. ULSD and BH oils exhibited high prediction accuracy, as their viscosity data remained below 25 cSt or covered a broader range than other oils. When ULSD and BH oils were excluded, the model trained on a narrower data range showed a decline in overall accuracy. However, it demonstrated the advantage of reducing prediction failures in the viscosity range below 1000 cSt. These findings underscore an alternative approach for selecting multiple models based on different training data ranges. In other words, an approach that categorizes oils by viscosity and develops specialized models for each range could further enhance predictive performance.

Errors inherent in the measurement process may introduce uncertainties in the data. These uncertainties had a direct impact on the accuracy of predictions and the feasibility of ML models. A 10% difference in viscosity input resulted in an accuracy decline of up to 60%. In the context of the models examined, the HGB model exhibited comparatively robust performance in the presence of uncertainty. Despite a decline in accuracy, the findings underscored the capacity to employ data with inherent errors by accounting for such variations.

Compared to previously analyzed studies, this research offers several notable advantages. The model developed by Chen et al. [47] achieved an accuracy ranging from 68% to 90%, whereas the present study demonstrated higher and more consistent performance.

Similarly, the accuracy obtained in this study exceeds the 90% reported by Chung et al. [50], who utilized SVM, PCA, and LDA. Moreover, this study accounts for property changes due to oil weathering and targets a broader range of oil types. On the other hand, Bills et al. and Loh et al. [52,53] presented portable devices based on SVM and PLS-DA that achieved 95% accuracy with rapid prediction times. In comparison, the present study still has room for improvement in terms of field applicability and real-time prediction capabilities.

5. Conclusions

This study evaluated the feasibility and accuracy of machine learning models for identifying spilled oil under offshore conditions. Weathered oil data were obtained hourly through simulations incorporating various parameters. The predictive performance of five models—decision tree (DT), random forest (RF), extra tree (ET), gradient boosting (GB), and histogram-based gradient boosting (HGB)—was assessed based on different feature and data compositions. Among these, the ET model was identified as the most promising candidate due to its high accuracy and robustness. In addition, comparative analyses of feature selection strategies contributed to further improvements in predictive accuracy. From a data perspective, the findings suggest the feasibility for adopting model groups specialized for specific viscosity ranges, particularly in cases where viscosity undergoes substantial changes due to weathering. One of the major strengths of this study lies in its ability to achieve high prediction accuracy within a short timeframe using only field-measurable features such as environmental variables, density, and viscosity. This makes the proposed approach a promising complementary tool to traditional oil fingerprinting methods in oil spill forensic analysis. The approach also demonstrates cost-efficiency by reducing reliance on licensed chemical fingerprinting databases and minimizing the need for additional sample collection for model retraining. This study also evaluated how uncertainty inherent in viscosity measurements impacts prediction performance, thereby highlighting the model's practical applicability when using field-acquired data.

Nevertheless, the study is limited by the number of oil types and the diversity of environmental scenarios considered. More extensive data are required to enhance model generalization, and field validation has not yet been conducted. In particular, water temperature, which has a significant impact on oil weathering, should be examined under more diverse conditions. Additionally, exploring alternative algorithms may further improve predictive performance. Taken together, future research should consider the integration of novel algorithms, along with refinements in data and feature selection strategies, to enhance model robustness and applicability.

Author Contributions: Conceptualization, C.H.; methodology, S.-I.K. and C.H.; validation, C.-K.K.; formal analysis, S.-I.K. and C.H.; investigation, S.-I.K. and C.H.; data curation, M.-I.C.; writing—original draft preparation, S.-I.K.; writing—review and editing, C.H.; visualization, S.-I.K.; supervision, C.H.; project administration, C.H.; funding acquisition, H.-J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the projects "Development of Bunkering technologies for Net-Zero Ship Fuel (RS-2025-02304029)" and "The Development and Demonstration of Low-Carbon Offshore Platform Repurposing Technologies (RS-2025-02314766)".

Data Availability Statement: Data and code for analysis are available on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

FAL	Feasibility, advantages, and limitations
ET	Extra tree
HGB	Histogram-based gradient boosting
GC	Gas chromatography
GC-MS	Gas chromatography–mass spectrometry
PAHs	Polycyclic aromatic hydrocarbons
FID	Flame ionization detection
IR	Infrared spectroscopy
NMR	Nuclear magnetic resonance
PLS-DA	Partial least squares discrimination analysis
HOG	Histogram of oriented gradients
SVM	Support vector machine
CNNs	Convolutional neural networks
WCO	Weathered crude oil
CDO	Chemically dispersed oil
GC-HRMS	Gas chromatography–high-resolution mass spectrometry
EEM	Fluorescence excitation–emission matrix
PCA	Principal component analysis
LDA	Linear discriminant analysis
LIF	Laser-induced fluorescence
API gravity	American Petroleum Institute gravity
IMO	International Maritime Organization
MA	Maya
BH	Basrah heavy
AM	Arabian medium
IH	Iranian heavy
KU	Kuwait
AL	Arabian light
HO	Hout
BL	Basrah Light Mobil Oil Australia
WTI	West Texas Intermediate
AELA	Arabian Extra Light Aramco
ULSD	Ultra-low Sulfur Diesel
MSO	Murban Shell Oil
ADIOS2	Automated Data Inquiry for Oil Spills 2
RCD	Rate of change in density
RCV	Rate of change in viscosity
CART	Classification and regression tree
DT	Decision tree
RF	Random forest
GB	Gradient boosting
TN	True negatives
FP	False positives
FN	False negatives
TP	True positives

Appendix A

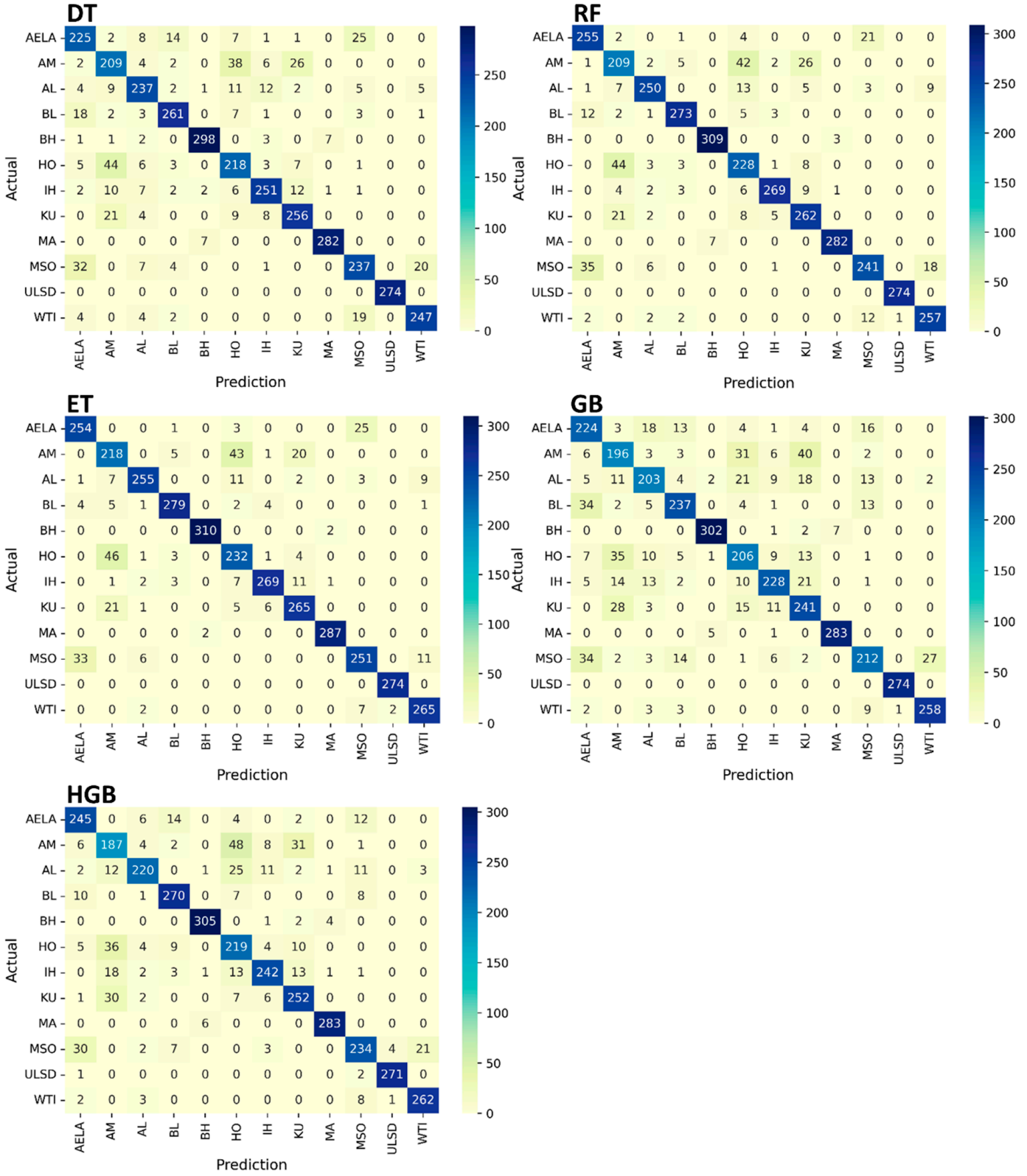


Figure A1. Confusion matrices of DT, RF, ET, GB, and HGB models trained with environmental features excluded.

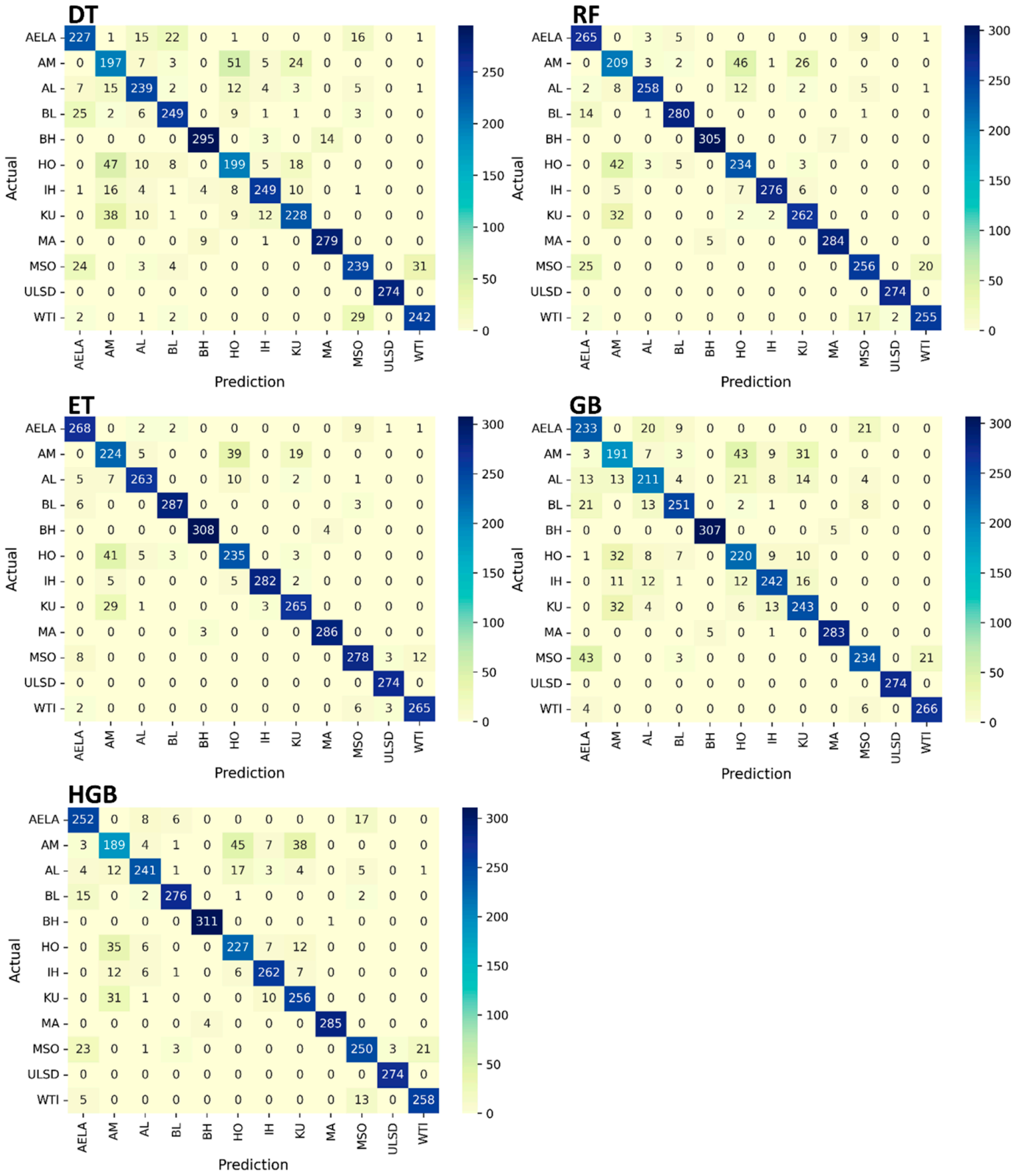


Figure A2. Confusion matrix of DT, RF, ET, GB, and HGB models incorporating RCD_2 and RCV_2.

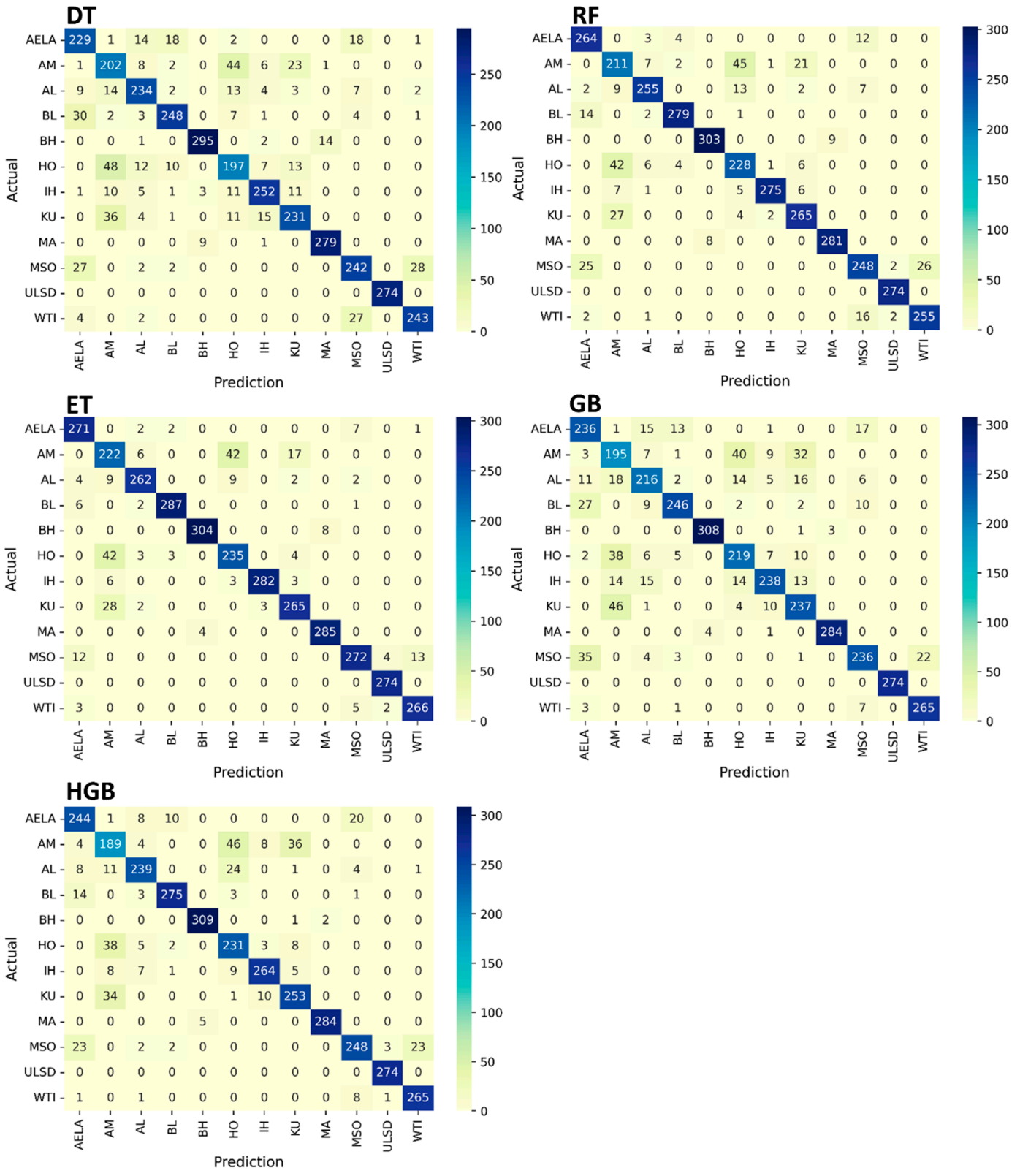


Figure A3. Confusion matrix of DT, RF, ET, GB, and HGB models incorporating RCD_2, RCV_2, RCD_3, and RCV_3.

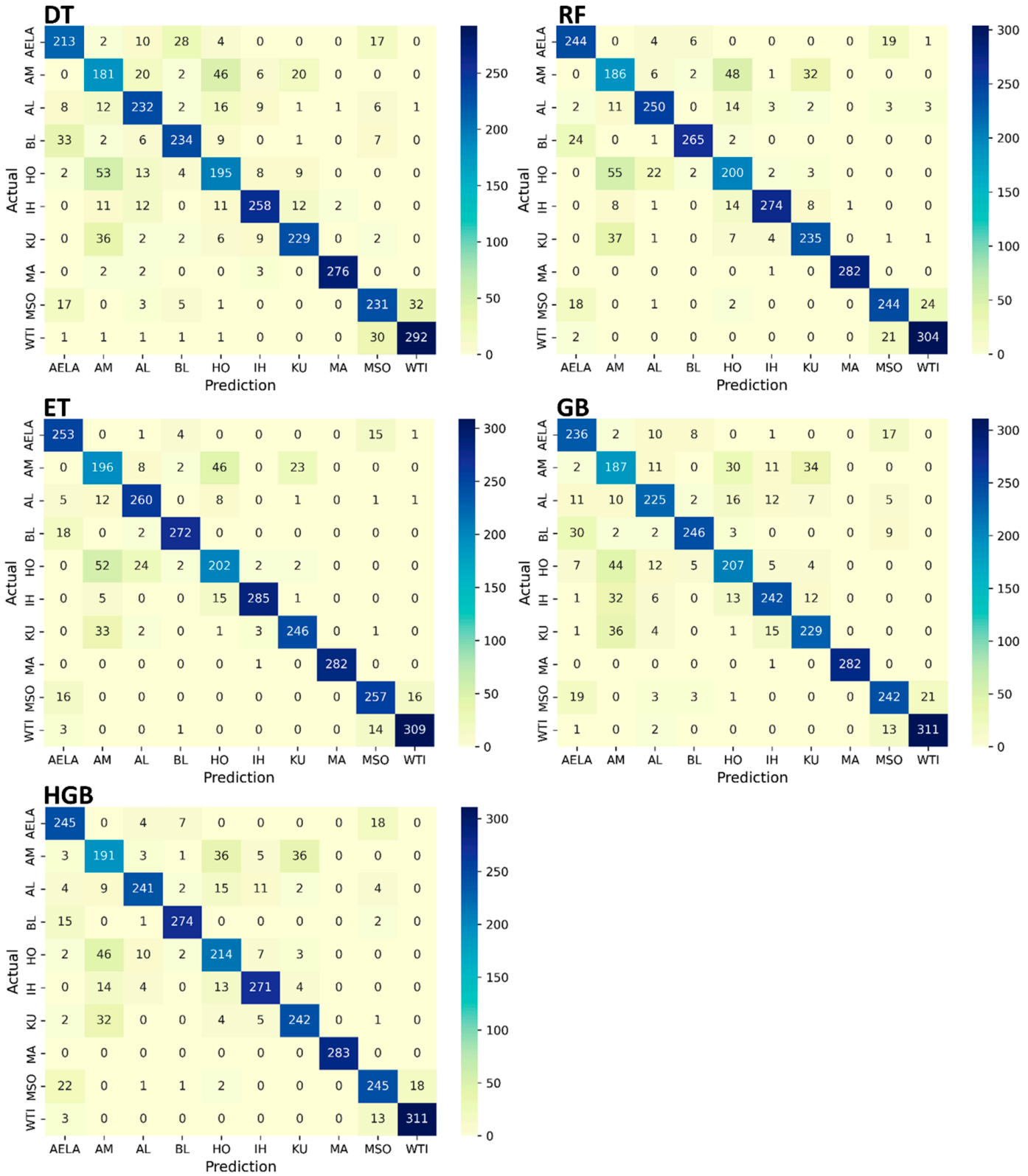


Figure A4. Confusion matrix of DT, RF, ET, GB, and HGB models with ULSD and BH oils excluded.

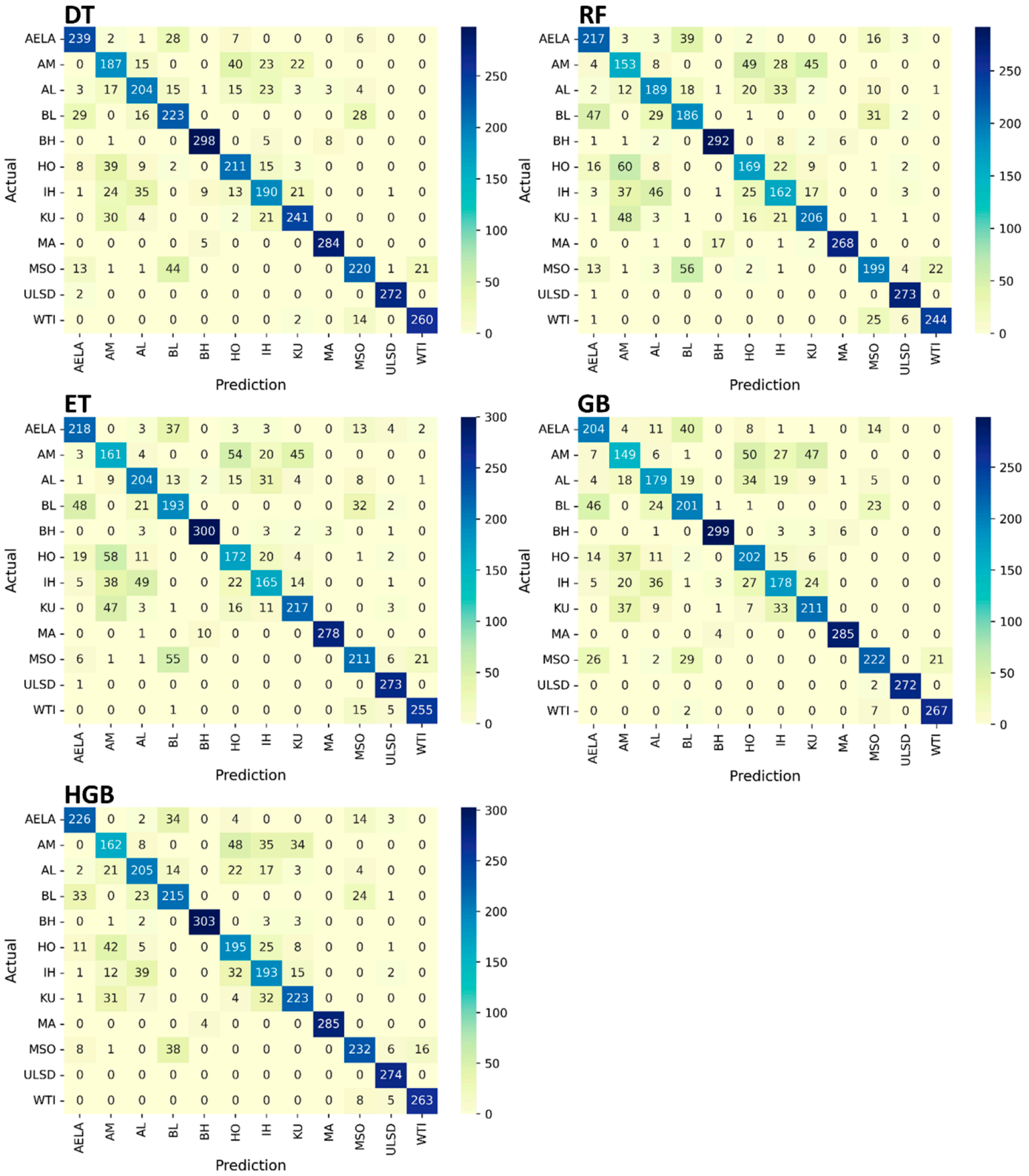


Figure A5. Confusion matrix of DT, RF, ET, GB, and HGB models with $RCD_{\Delta t}$ and $RCV_{\Delta t}$ excluded.

Table A1. Number of oil test data points for each oil type across viscosity bin centers (cSt).

Viscosity Bin Center (cSt)	AELA	AM	AL	BL	BH	HO	IH	KU	MA	MSO	ULSD	WTI
25	16	5	8	12	1	4	7	8	0	45	242	50
75	36	11	13	25	0	10	5	3	0	22	32	27
125	17	11	7	19	3	12	7	10	0	18	0	7
175	17	7	14	7	0	7	8	15	1	3	0	3
225	7	11	9	7	0	13	10	4	1	8	0	2
275	5	0	8	6	0	9	4	11	1	1	0	2
325	1	5	7	3	0	7	4	5	0	2	0	1
375	3	7	1	6	0	4	6	8	0	0	0	2
425	1	3	4	3	1	4	4	4	0	3	0	1
475	3	6	2	0	1	1	3	2	0	3	0	4
525	2	7	1	0	1	5	4	8	0	1	0	0
575	2	2	2	4	1	3	3	5	0	2	0	0
625	1	2	2	0	0	4	3	4	0	2	0	1
675	1	0	0	1	0	0	5	1	0	3	0	1
725	1	1	3	1	0	1	1	0	0	1	0	4
775	1	0	3	1	1	1	4	4	1	2	0	2
825	3	3	4	0	2	2	0	2	0	1	0	1
875	0	0	1	1	2	0	3	2	0	4	0	2
925	1	2	0	2	1	0	0	3	1	3	0	1
975	0	2	0	0	0	0	0	0	1	0	0	2

References

1. Zhao, P.; Zhao, Y.; Zou, C.; Gu, T. Study on Ultrasonic Extraction of Kerogen from Huadian Oil Shale by Solvents. *Oil Shale* **2013**, *30*, 491. [CrossRef]
2. Sulaiman, M.A.; Oni, A.O.; Fadare, D.A. Energy and Exergy Analysis of a Vegetable Oil Refinery. *Energy Power Eng.* **2012**, *04*, 358–364. [CrossRef]
3. Kowalski, R.; Baj, T.; Kowalska, G.; Pankiewicz, U. Estimation of Potential Availability of Essential Oil in Some Brands of Herbal Teas and Herbal Dietary Supplements. *PLoS ONE* **2015**, *10*, e0130714. [CrossRef]
4. Tian, H.; Liao, Z.Z. Experimental Study on the Effect of Ultrasonic Cavitation on Pyrolysis Characteristics of Oil Shale. *Appl. Mech. Mater.* **2013**, *295–298*, 3117–3123. [CrossRef]
5. Perhar, G.; Arhonditsis, G.B. Aquatic Ecosystem Dynamics Following Petroleum Hydrocarbon Perturbations: A Review of the Current State of Knowledge. *J. Great Lakes Res.* **2014**, *40*, 56–72. [CrossRef]
6. Jha, M.; Levy, J.; Gao, Y. Advances in Remote Sensing for Oil Spill Disaster Management: State-of-the-Art Sensors Technology for Oil Spill Surveillance. *Sensors* **2008**, *8*, 236–255. [CrossRef]
7. Li, P.; Cai, Q.; Lin, W.; Chen, B.; Zhang, B. Offshore Oil Spill Response Practices and Emerging Challenges. *Mar. Pollut. Bull.* **2016**, *110*, 6–27. [CrossRef] [PubMed]
8. O'Rourke, D.; Connolly, S. Just oil? The distribution of environmental and social impacts of oil production and consumption. *Annu. Rev. Environ. Resour.* **2003**, *28*, 587–617. [CrossRef]
9. Mishra, A.K.; Kumar, G.S. Weathering of Oil Spill: Modeling and Analysis. *Aquat. Procedia* **2015**, *4*, 435–442. [CrossRef]
10. Neff, J.M.; Ostazeski, S.; Gardiner, W.; Stejskal, I. Effects of Weathering on the Toxicity of Three Offshore Australian Crude Oils and a Diesel Fuel to Marine Animals. *Enviro. Toxic. Chem.* **2000**, *19*, 1809–1821. [CrossRef]
11. Freeman, D.H.; Niles, S.F.; Rodgers, R.P.; French-McCay, D.P.; Longnecker, K.; Reddy, C.M.; Ward, C.P. Hot and Cold: Photochemical Weathering Mediates Oil Properties and Fate Differently Depending on Seawater Temperature. *Environ. Sci. Technol.* **2023**, *57*, 11988–11998. [CrossRef] [PubMed]
12. Aeppli, C.; Mitchell, D.A.; Keyes, P.; Beirne, E.C.; McFarlin, K.M.; Roman-Hubers, A.T.; Rusyn, I.; Prince, R.C.; Zhao, L.; Parkerton, T.F.; et al. Oil Irradiation Experiments Document Changes in Oil Properties, Molecular Composition, and Dispersant Effectiveness Associated with Oil Photo-Oxidation. *Environ. Sci. Technol.* **2022**, *56*, 7789–7799. [CrossRef] [PubMed]
13. Ward, C.P.; Armstrong, C.J.; Conmy, R.N.; French-McCay, D.P.; Reddy, C.M. Photochemical Oxidation of Oil Reduced the Effectiveness of Aerial Dispersants Applied in Response to the Deepwater Horizon Spill. *Environ. Sci. Technol. Lett.* **2018**, *5*, 226–231. [CrossRef] [PubMed]

14. Oliveira, R.C.G.; Gonçalves, M.A.L. Emulsion Rheology—Theory vs. Field Observation. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 2–5 May 2005.
15. Boehm, P.D.; Douglas, G.S.; Burns, W.A.; Mankiewicz, P.J.; Page, D.S.; Bence, A.E. Application of Petroleum Hydrocarbon Chemical Fingerprinting and Allocation Techniques after the Exxon Valdez Oil Spill. *Mar. Pollut. Bull.* **1997**, *34*, 599–613. [CrossRef]
16. Mirnaghi, F.S.; Soucy, N.; Hollebone, B.P.; Brown, C.E. Rapid Fingerprinting of Spilled Petroleum Products Using Fluorescence Spectroscopy Coupled with Parallel Factor and Principal Component Analysis. *Chemosphere* **2018**, *208*, 185–195. [CrossRef]
17. Yim, U.H.; Kim, M.; Ha, S.Y.; Kim, S.; Shim, W.J. Oil Spill Environmental Forensics: The Hebei Spirit Oil Spill Case. *Environ. Sci. Technol.* **2012**, *46*, 6431–6437. [CrossRef]
18. Roman-Hubers, A.T.; McDonald, T.J.; Baker, E.S.; Chiu, W.A.; Rusyn, I. A Comparative Analysis of Analytical Techniques for Rapid Oil Spill Identification. *Enviro. Toxic. Chem.* **2021**, *40*, 1034–1049. [CrossRef]
19. Wang, Z.; Fingas, M. Developments in the Analysis of Petroleum Hydrocarbons in Oils, Petroleum Products and Oil-Spill-Related Environmental Samples by Gas Chromatography. *J. Chromatogr. A* **1997**, *774*, 51–78. [CrossRef]
20. Ehrhardt, M.; Blumer, M. The Source Identification of Marine Hydrocarbons by Gas Chromatography. *Environ. Pollut.* **1972**, *3*, 179–194. [CrossRef]
21. Wakeham, S.G.; Farrington, J.W.; Gagosian, R.B.; Lee, C.; DeBaar, H.; Nigrelli, G.E.; Tripp, B.W.; Smith, S.O.; Frew, N.M. Organic Matter Fluxes from Sediment Traps in the Equatorial Atlantic Ocean. *Nature* **1980**, *286*, 798–800. [CrossRef]
22. Gaines, R.B.; Frysinger, G.S.; Hendrick-Smith, M.S.; Stuart, J.D. Oil Spill Source Identification by Comprehensive Two-Dimensional Gas Chromatography. *Environ. Sci. Technol.* **1999**, *33*, 2106–2112. [CrossRef]
23. Nelson, R.K.; Gosselin, K.M.; Hollander, D.J.; Murawski, S.A.; Gracia, A.; Reddy, C.M.; Radović, J.R. Exploring the Complexity of Two Iconic Crude Oil Spills in the Gulf of Mexico (Ixtoc I and Deepwater Horizon) Using Comprehensive Two-Dimensional Gas Chromatography (GC × GC). *Energy Fuels* **2019**, *33*, 3925–3933. [CrossRef]
24. Sun, P.; Bao, M.; Li, G.; Wang, X.; Zhao, Y.; Zhou, Q.; Cao, L. Fingerprinting and Source Identification of an Oil Spill in China Bohai Sea by Gas Chromatography-Flame Ionization Detection and Gas Chromatography—Mass Spectrometry Coupled with Multi-Statistical Analyses. *J. Chromatogr. A* **2009**, *1216*, 830–836. [CrossRef] [PubMed]
25. Wang, S.; Guo, G.; Yan, Z.; Lu, G.; Wang, Q.; Li, F. The Development of a Method for the Qualitative and Quantitative Determination of Petroleum Hydrocarbon Components Using Thin-Layer Chromatography with Flame Ionization Detection. *J. Chromatogr. A* **2010**, *1217*, 368–374. [CrossRef] [PubMed]
26. Lovatti, B.P.O.; Silva, S.R.C.; Portela, N.D.A.; Sad, C.M.S.; Rainha, K.P.; Rocha, J.T.C.; Romão, W.; Castro, E.V.R.; Filgueiras, P.R. Identification of Petroleum Profiles by Infrared Spectroscopy and Chemometrics. *Fuel* **2019**, *254*, 115670. [CrossRef]
27. Yuan, L.; Meng, X.; Xin, K.; Ju, Y.; Zhang, Y.; Yin, C.; Hu, L. A Comparative Study on Classification of Edible Vegetable Oils by Infrared, near Infrared and Fluorescence Spectroscopy Combined with Chemometrics. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *288*, 122120. [CrossRef]
28. Frank, U. A Review of Fluorescence Spectroscopic Methods for Oil Spill Source Identification. *Toxicol. Environ. Chem. Rev.* **1978**, *2*, 163–185. [CrossRef]
29. Christensen, J.H.; Tomasi, G. Practical Aspects of Chemometrics for Oil Spill Fingerprinting. *J. Chromatogr. A* **2007**, *1169*, 1–22. [CrossRef]
30. Silva, S.L.; Silva, A.M.S.; Ribeiro, J.C.; Martins, F.G.; Da Silva, F.A.; Silva, C.M. Chromatographic and Spectroscopic Analysis of Heavy Crude Oil Mixtures with Emphasis in Nuclear Magnetic Resonance Spectroscopy: A Review. *Anal. Chim. Acta* **2011**, *707*, 18–37. [CrossRef]
31. Wang, C.; Li, W.; Luan, X.; Liu, Q.; Zhang, J.; Zheng, R. Species Identification and Concentration Quantification of Crude Oil Samples in Petroleum Exploration Using the Concentration-Synchronous-Matrix-Fluorescence Spectroscopy. *Talanta* **2010**, *81*, 684–691. [CrossRef]
32. Aeppli, C.; Nelson, R.K.; Radović, J.R.; Carmichael, C.A.; Valentine, D.L.; Reddy, C.M. Recalcitrance and Degradation of Petroleum Biomarkers upon Abiotic and Biotic Natural Weathering of Deepwater Horizon Oil. *Environ. Sci. Technol.* **2014**, *48*, 6726–6734. [CrossRef] [PubMed]
33. Wang, C.; Chen, B.; Zhang, B.; He, S.; Zhao, M. Fingerprint and Weathering Characteristics of Crude Oils after Dalian Oil Spill, China. *Mar. Pollut. Bull.* **2013**, *71*, 64–68. [CrossRef] [PubMed]
34. Han, Y.; Clement, T.P. Development of a Field Testing Protocol for Identifying Deepwater Horizon Oil Spill Residues Trapped near Gulf of Mexico Beaches. *PLoS ONE* **2018**, *13*, e0190508. [CrossRef]
35. Gaweł, B.; Eftekhardadkhah, M.; Øye, G. Elemental Composition and Fourier Transform Infrared Spectroscopy Analysis of Crude Oils and Their Fractions. *Energy Fuels* **2014**, *28*, 997–1003. [CrossRef]
36. Palacio Lozano, D.C.; Orrego-Ruiz, J.A.; Cabanzo Hernández, R.; Guerrero, J.E.; Mejía-Ospino, E. APPI(+)-FTICR Mass Spectrometry Coupled to Partial Least Squares with Genetic Algorithm Variable Selection for Prediction of API Gravity and CCR of Crude Oil and Vacuum Residues. *Fuel* **2017**, *193*, 39–449. [CrossRef]

37. Prata, P.S.; Alexandrino, G.L.; Mogollón, N.G.S.; Augusto, F. Discriminating Brazilian Crude Oils Using Comprehensive Two-Dimensional Gas Chromatography–Mass Spectrometry and Multiway Principal Component Analysis. *J. Chromatogr. A* **2016**, *1472*, 99–106. [CrossRef] [PubMed]
38. González, M.; Gorziza, R.P.; De Cássia Mariotti, K.; Pereira Limberger, R. Methodologies Applied to Fingerprint Analysis. *J. Forensic Sci.* **2020**, *65*, 1040–1048. [CrossRef]
39. Zhang, L.; Du, F.; Wang, Y.; Li, Y.; Yang, C.; Li, S.; Huang, X.; Wang, C. Oil Fingerprint Identification Technology Using a Simplified Set of Biomarkers Selected Based on Principal Component Difference. *Can. J. Chem. Eng.* **2022**, *100*, 23–34. [CrossRef]
40. Chen, S.; Du, X.; Zhao, W.; Guo, P.; Chen, H.; Jiang, Y.; Wu, H. Olive Oil Classification with Laser-Induced Fluorescence (LIF) Spectra Using 1-Dimensional Convolutional Neural Network and Dual Convolution Structure Model. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *279*, 121418. [CrossRef]
41. Chen, X.; Hu, Y.; Li, X.; Kong, D.; Guo, M. Fast Dentification of Overlapping Fluorescence Spectra of Oil Species Based on LDA and Two-Dimensional Convolutional Neural Network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *324*, 124979. [CrossRef]
42. Gjelsvik, E.L.; Fossen, M.; Tøndel, K. Current Overview and Way Forward for the Use of Machine Learning in the Field of Petroleum Gas Hydrates. *Fuel* **2023**, *334*, 126696. [CrossRef]
43. Li, Y.; Jia, Y.; Cai, X.; Xie, M.; Zhang, Z. Correction to: Oil Pollutant Identification Based on Excitation-emission Matrix of UV-induced Fluorescence and Deep Convolutional Neural Network. *Environ. Sci. Pollut. Res.* **2022**, *29*, 89806. [CrossRef] [PubMed]
44. Raljević, D.; Parlov Vuković, J.; Smrečki, V.; Marinić Pajc, L.; Novak, P.; Hrenar, T.; Jednačak, T.; Konjević, L.; Pinević, B.; Gašparac, T. Machine Learning Approach for Predicting Crude Oil Stability Based on NMR Spectroscopy. *Fuel* **2021**, *305*, 121561. [CrossRef]
45. Li, K.; Yu, H.; Xu, Y.; Luo, X. Detection of Marine Oil Spills Based on HOG Feature and SVM Classifier. *J. Sens.* **2022**, *2022*, 3296495. [CrossRef]
46. Zhang, S.; Yuan, Y.; Wang, Z.; Wei, S.; Zhang, X.; Zhang, T.; Song, X.; Zou, Y.; Wang, J.; Chen, F.; et al. A Novel Deep Learning Model for Spectral Analysis: Lightweight ResNet-CNN with Adaptive Feature Compression for Oil Spill Type Identification. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *329*, 125626. [CrossRef]
47. Chen, Y.; Chen, B.; Song, X.; Kang, Q.; Ye, X.; Zhang, B. A Data-Driven Binary-Classification Framework for Oil Fingerprinting Analysis. *Environ. Res.* **2021**, *201*, 111454. [CrossRef]
48. Ekpe, O.D.; Choo, G.; Kang, J.-K.; Yun, S.-T.; Oh, J.-E. Identification of Organic Chemical Indicators for Tracking Pollution Sources in Groundwater by Machine Learning from GC-HRMS-Based Suspect and Non-Target Screening Data. *Water Res.* **2024**, *252*, 121130. [CrossRef]
49. Xie, M.; Xie, L.; Li, Y.; Han, B. Oil Species Identification Based on Fluorescence Excitation-Emission Matrix and Transformer-Based Deep Learning. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *302*, 123059. [CrossRef]
50. Chung, S.; Loh, A.; Jennings, C.M.; Sosnowski, K.; Ha, S.Y.; Yim, U.H.; Yoon, J.-Y. Capillary Flow Velocity Profile Analysis on Paper-Based Microfluidic Chips for Screening Oil Types Using Machine Learning. *J. Hazard. Mater.* **2023**, *447*, 130806. [CrossRef]
51. Sosnowski, K.; Loh, A.; Zubler, A.V.; Shir, H.; Ha, S.Y.; Yim, U.H.; Yoon, J.-Y. Machine Learning Techniques for Chemical and Type Analysis of Ocean Oil Samples via Handheld Spectrophotometer Device. *Biosens. Bioelectron. X* **2022**, *10*, 100128. [CrossRef]
52. Bills, M.V.; Loh, A.; Sosnowski, K.; Nguyen, B.T.; Ha, S.Y.; Yim, U.H.; Yoon, J.-Y. Handheld UV Fluorescence Spectrophotometer Device for the Classification and Analysis of Petroleum Oil Samples. *Biosens. Bioelectron.* **2020**, *159*, 112193. [CrossRef]
53. Loh, A.; Ha, S.Y.; Kim, D.; Lee, J.; Baek, K.; Yim, U.H. Development of a Portable Oil Type Classifier Using Laser-Induced Fluorescence Spectrometer Coupled with Chemometrics. *J. Hazard. Mater.* **2021**, *416*, 125723. [CrossRef] [PubMed]
54. Wang, Y.; Liu, X.; Yu, X.; Zheng, X. Assessing Response Capabilities for Responding to Ship-Related Oil Spills in the Chinese Bohai Sea. *Int. J. Disaster Risk Reduct.* **2018**, *28*, 251–257. [CrossRef]
55. Kim, E.-S.; An, J.-G.; Kim, G.-B.; Shim, W.-J.; Joo, C.-K.; Kim, M.-K. Identification of Major Crude Oils Imported into Korea using Molecular and Stable Carbon Isotopic Compositions. *J. Korean Soc. Mar. Environ. Energy* **2012**, *15*, 247–256. [CrossRef]
56. Hong, S.; Yoon, S.J.; Kim, T.; Ryu, J.; Kang, S.-G.; Khim, J.S. Response to Oiled Wildlife in the Management and Evaluation of Marine Oil Spills in South Korea: A Review. *Reg. Stud. Mar. Sci.* **2020**, *40*, 101542. [CrossRef]
57. Lee, M.; Jung, J.-Y. Risk Assessment and National Measure Plan for Oil and HNS Spill Accidents near Korea. *Mar. Pollut. Bull.* **2013**, *73*, 339–344. [CrossRef]
58. South Korea Crude Oil Imports by Type. Available online: https://kosis.kr/statHtml/statHtml.do?orgId=318&tblId=TX_31801_A009&conn_path=I2 (accessed on 14 February 2024).
59. Choe, Y.; Kim, H.; Huh, C. Estimation of Chemical Dispersion Amount Considering the Dosage of Dispersant and Change of Oil Properties by Weathering. *J. Korean Soc. Mar. Environ. Energy* **2018**, *21*, 260–269. [CrossRef]
60. Hazrat, M.A.; Rasul, M.G.; Khan, M.M.K. Lubricity Improvement of the Ultra-Low Sulfur Diesel Fuel with the Biodiesel. *Energy Procedia* **2015**, *75*, 111–117. [CrossRef]

61. Stanislaus, A.; Marafi, A.; Rana, M.S. Recent Advances in the Science and Technology of Ultra Low Sulfur Diesel (ULSD) Production. *Catal. Today* **2010**, *153*, 1–68. [CrossRef]
62. Krestenitis, M.; Orfanidis, G.; Ioannidis, K.; Avgerinakis, K.; Vrochidis, S.; Kompatsiaris, I. Early Identification of Oil Spills in Satellite Images Using Deep CNNs. In *MultiMedia Modeling*; Kompatsiaris, I., Huet, B., Mezaris, V., Gurrin, C., Cheng, W.-H., Vrochidis, S., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; Volume 11295, pp. 424–435, ISBN 978-3-030-05709-1.
63. Hole, L.R.; Dagestad, K.-F.; Röhrs, J.; Wettre, C.; Kourafalou, V.H.; Androulidakis, Y.; Kang, H.; Le Hénaff, M.; Garcia-Pineda, O. The DeepWater Horizon Oil Slick: Simulations of River Front Effects and Oil Droplet Size Distribution. *J. Mar. Sci. Eng.* **2019**, *7*, 329. [CrossRef]
64. Röhrs, J.; Dagestad, K.-F.; Asbjørnsen, H.; Nordam, T.; Skancke, J.; Jones, C.E.; Brekke, C. The Effect of Vertical Mixing on the Horizontal Drift of Oil Spills. *Ocean. Sci.* **2018**, *14*, 1581–1601. [CrossRef]
65. Beegle-Krause, C.J. GNOME: NOAA's Next-Generation Spill Trajectory Model. In Proceedings of the Oceans '99. MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings (IEEE Cat. No.99CH37008), Seattle, WA, USA, 13–16 September 1999; Volume 3, pp. 1262–1266.
66. Lehr, W.; Jones, R.; Evans, M.; Simecek-Beatty, D.; Overstreet, R. Revisions of the ADIOS Oil Spill Model. *Environ. Model. Softw.* **2002**, *17*, 189–197. [CrossRef]
67. Keramea, P.; Spanoudaki, K.; Zodiatis, G.; Gikas, G.; Sylaios, G. Oil Spill Modeling: A Critical Review on Current Trends, Perspectives, and Challenges. *J. Mar. Sci. Eng.* **2021**, *9*, 181. [CrossRef]
68. Zhong, X.; Li, P.; Lin, X.; Zhao, Z.; He, Q.; Niu, H.; Yang, J. Diluted Bitumen: Physicochemical Properties, Weathering Processes, Emergency Response, and Recovery. *Front. Environ. Sci.* **2022**, *10*, 910365. [CrossRef]
69. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
70. Shabtai, A.; Moskovitch, R.; Elovici, Y.; Glezer, C. Detection of Malicious Code by Applying Machine Learning Classifiers on Static Features: A State-of-the-Art Survey. *Inf. Secur. Tech. Rep.* **2009**, *14*, 16–29. [CrossRef]
71. Zhang, K.; Sun, Y.; Cui, Z.; Yu, D.; Zheng, L.; Liu, P.; Lv, Z. Periodically Spilled-Oil Input as a Trigger to Stimulate the Development of Hydrocarbon-Degrading Consortia in a Beach Ecosystem. *Sci. Rep.* **2017**, *7*, 12446. [CrossRef]
72. Charbuty, B.; Abdulazeez, A. Classification Based on Decision Tree Algorithm for Machine Learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
73. Breiman, L. *Random Forests—Random Features*; Technical Report for University of California: Berkeley, CA, USA, 1999.
74. Loh, W. Classification and Regression Trees. *WIREs Data Min Knowl Discov.* **2011**, *1*, 14–23. [CrossRef]
75. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
76. Zhao, Y.; Zhang, Y. Comparison of Decision Tree Methods for Finding Active Objects. *Adv. Space Res.* **2008**, *41*, 1955–1959. [CrossRef]
77. Pal, M. Random Forest Classifier for Remote Sensing Classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
78. Alfian, G.; Syafrudin, M.; Fahrurrozi, I.; Fitriyani, N.L.; Atmaji, F.T.D.; Widodo, T.; Bahiyah, N.; Benes, F.; Rhee, J. Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. *Computers* **2022**, *11*, 136. [CrossRef]
79. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
80. Yi, C.D. Investment, Export, and Exchange Rate on Prediction of Employment with Decision Tree, Random Forest, and Gradient Boosting Machine Learning Models. *Korea Trade Rev.* **2021**, *46*, 281–299.
81. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Proces. Syst.* **2017**, 3149–3157.
82. Hossain, S.M.M.; Deb, K. Plant Leaf Disease Recognition Using Histogram Based Gradient Boosting Classifier. In *Intelligent Computing and Optimization*; Vasant, P., Zelinka, I., Weber, G.-W., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1324, pp. 530–545. ISBN 978-3-030-68153-1.
83. Hinz, T.; Navarro-Guerrero, N.; Magg, S.; Wermter, S. Speeding up the Hyperparameter Optimization of Deep Convolutional Neural Networks. *Int. J. Comp. Intel. Appl.* **2018**, *17*, 1850008. [CrossRef]
84. Qian, D. Analysis of the Hyperparameter Selection in Machine Learning. *Appl. Comput. Eng.* **2024**, *104*, 122–128. [CrossRef]
85. Van Rijn, J.N.; Hutter, F. Hyperparameter Importance Across Datasets. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19 July 2018; pp. 2367–2376.
86. Probst, P.; Wright, M.N.; Boulesteix, A. Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
87. Hoque, K.E.; Aljamaan, H. Impact of Hyperparameter Tuning on Machine Learning Models in Stock Price Forecasting. *IEEE Access* **2021**, *9*, 163815–163830. [CrossRef]
88. Louche, U.; Ralaivola, L. Unconfused Ultraconservative Multiclass Algorithms. *Mach. Learn.* **2015**, *99*, 327–351. [CrossRef]
89. Kurniasari, D.; Warsono, W.; Usman, M.; Lumbanraja, F.R.; Wamiliana, W. LSTM-CNN Hybrid Model Performance Improvement with BioWordVec for Biomedical Report Big Data Classification. *Sci. Technol. Indones.* **2024**, *9*, 273–283. [CrossRef]

90. Markoulidakis, I.; Rallis, I.; Georgoulas, I.; Kopsiaftis, G.; Doulamis, A.; Doulamis, N. A Machine Learning Based Classification Method for Customer Experience Survey Analysis. *Technologies* **2020**, *8*, 76. [CrossRef]
91. Ho, M.-C.; Shen, H.-A.; Chang, Y.-P.E.; Weng, J.-C. A CNN-Based Autoencoder and Machine Learning Model for Identifying Betel-Quid Chewers Using Functional MRI Features. *Brain Sci.* **2021**, *11*, 809. [CrossRef] [PubMed]
92. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef]
93. Wu, Q.; Nasoz, F.; Jung, J.; Bhattarai, B.; Han, M.V. Machine Learning Approaches for Fracture Risk Assessment: A Comparative Analysis of Genomic and Phenotypic Data in 5130 Older Men. *Calcif. Tissue Int.* **2020**, *107*, 353–361. [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.