Explainable AI models for predicting drop coalescence in microfluidics device

Hu, Jinwei; Zhu, Kewei; Cheng, Sibo; Kovalchuk, Nina; Soulsby, Alfred; Simmons, Mark; Matar, Omar K.; Arcucci, Rossella

DOI:
10.1016/j.cej.2023.148465

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Hu, J, Zhu, K, Cheng, S, Kovalchuk, N, Soulsby, A, Simmons, M, Matar, OK & Arcucci, R 2024, 'Explainable AI models for predicting drop coalescence in microfluidics device', *Chemical Engineering Journal*, vol. 481, 148465. https://doi.org/10.1016/j.cej.2023.148465

Chemical Engineering Journal 481 (2024) 148465

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Explainable AI models for predicting drop coalescence in microfluidics device

Jinwei Hu [a,1], Kewei Zhu [b,1], Sibo Cheng [c,*], Nina M. Kovalchuk [d], Alfred Soulsby [d], Mark J.H. Simmons [d], Omar K. Matar [e], Rossella Arcucci [a]

[a] Department of Earth Science & Engineering, Imperial College London, UK
[b] Department of Computer Science, University of York, UK
[c] Data Science Institute, Department of Computing, Imperial College London, UK
[d] School of Chemical Engineering, University of Birmingham, UK
[e] Department of Chemical Engineering, Imperial College London, UK

ARTICLE INFO

Keywords:
Explainable AI
Drop coalescence
Machine learning
LIME
SHAP value

ABSTRACT

In the field of chemical engineering, understanding the dynamics and probability of drop coalescence is not just an academic pursuit, but a critical requirement for advancing process design by applying energy only where it is needed to build necessary interfacial structures, increasing efficiency towards Net Zero manufacture. This research applies machine learning predictive models to unravel the sophisticated relationships embedded in the experimental data on drop coalescence in a microfluidics device. Through the deployment of SHapley Additive exPlanations values, critical features relevant to coalescence processes are consistently identified. Comprehensive feature ablation tests further delineate the robustness and susceptibility of each model. Furthermore, the incorporation of Local Interpretable Model-agnostic Explanations for local interpretability offers an elucidative perspective, clarifying the intricate decision-making mechanisms inherent to each model's predictions. As a result, this research provides the relative importance of the features for the outcome of drop interactions. It also underscores the pivotal role of model interpretability in reinforcing confidence in machine learning predictions of complex physical phenomena that are central to chemical engineering applications.

1. Introduction

Microfluidic technologies have caused a significant paradigm-shift in the manipulation and analysis of fluids across a range of fields, including chemistry, biology, and material science [1]. One of the key operations in flow microfluidics is coalescence of dispersed phase droplets. This operation has been widely studied using both passive methods, where it is facilitated by the device geometry, and active methods, involving, for instance, the use of electric fields [2–10]. Microfluidics serve as a unique platform for studying coalescence under well-controlled conditions related to various industrial applications, such as emulsion formulation, which is crucial in sectors like food processing, cosmetics, pharmaceutics, transport, and separation processes [11–14]. While on industrial scale drop break up and coalescence are often studied under conditions of turbulent flow on large volumes of emulsion by tracking changes in drop size distribution and number of drops in the unit of volume over time [15], microfluidic approach enables monitoring the large amounts of individual pairs of drops with well-defined positions and velocities for each pair.

Coalescence is also broadly used in microfluidic synthesis and analysis [16–18]. Despite considerable research in this domain [19,20], reliable prediction of passive coalescence is still impeded [21,22] due to the large number of parameters involved, such as device geometry, chemical composition and viscosity of the continuous and dispersed phase, interfacial tension, temperature, droplet approach velocities and contact time, making it difficult to achieve desirable outcomes without external interventions [23]. For instance, higher temperatures can amplify the coalescence frequency, as detailed by Bera et al. [24]. Similarly, changes in fluid viscosity or flow rate can affect droplet sizes and coalescence occurrence, requiring real-time adjustments to maintain desired outcomes [25]. Therefore, the development of methods enabling prediction and control of drop coalescence in microfluidic devices is of great importance for multiple applications.

Traditionally, the analysis of factors influencing droplet coalescence relies on trial-and-error methods or analytical models [26,27]. Serving as the antithesis to conventional methodologies, Machine Learning

* Corresponding author.
E-mail address: sibo.cheng@imperial.ac.uk (S. Cheng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cej.2023.148465
Received 6 October 2023; Received in revised form 14 December 2023; Accepted 29 December 2023
Available online 2 January 2024

Main Notations

Notation	Description
L	Channel width at the entrance to chamber
H	Channel height
W	Chamber wall to wall width
Q	Flow rate
V	Average velocity in the channel
D	Equivalent drop diameters
S	Measured areas of the drops
$\mathbf{x}_{\frac{D}{W}1+\frac{D}{W}2}$	Size of doublet as related to the chamber size
$\mathbf{x}_{\lvert\frac{D}{W}1-\frac{D}{W}2\rvert}$	Disparity in the sizes of droplets
$\mathbf{x}_{dt}$	Time interval between the successive entrance of droplets into chamber
$\mathbf{x}_{\text{flow}}$	Total flow rate in the each of input into chamber
$\mathbf{x}_{\text{Heff}}$	The effective height of the channel
$\mathbf{x}_{\text{scaled}}$	Scaled feature results after normalizing
f	Function of original predictive model
g	Function of interpretable surrogate model
ϕ_0	Base value of all predictions
ϕ_i	SHapley Additive exPlanations (SHAP) value for the i-th feature
N	Total set of features
B	Subset of N that includes selected features
Z	The family of interpretable models
π_x	Proximity measure
ξ	Final explanation model in LIME
$\mathcal{L}$	Loss function in LIME
Ω	Measure of complexity of the explanation model

(ML) functions as a powerful tool for modeling intricate systems and proffering predictions for dimensional data [28–30]. Owing to their ability to assimilate information from data, machine learning algorithms excel in capturing sophisticated relationships that traditional methodologies are impotent to attain [31]. In the domain of chemical engineering, ML algorithms are being increasingly utilized. They provide robust support for addressing complex and critical problems, such as the capability to generate synthetic data to balance inherently imbalanced datasets which cannot be attained in physical experiments [32]. The gambit of applications spans not only quantum chemistry research and molecular reaction kinetics, but also extends to process optimization and control, which is imperative for enhancing efficiency and safety of chemical processes [33–35]. Our hypothesis here is that starting from initial conditions in a microfluidics device, ML algorithms can furnish reliable predictions of coalescence and insights into the conditions underlying this phenomenon; this can complement data and observations obtained from physical experiments. .

The application of ML algorithms to coalescence, however, brings forth challenges tied to model transparency. An example of these challenges is provided by the use of deep neural networks (often referred to as "black boxes" due to their opaque nature) which can yield remarkably accurate predictions yet but no clarity on their underlying decision-making mechanisms [36,37]. In critical domains like drop coalescence, where a comprehensive understanding of the dynamics is vital for both research and industrial applications, this opacity has significantly negative implications. The ensuing gap in transparency and interpretability can inhibit the wider adoption of these models, affecting the trust they garner among practitioners [38]. Based on this, the demand for more interpretable ML models, which do not compromise on predictive efficiency, is on the rise across engineering disciplines. In response to this overarching need, attention has been shifting towards the domain of Explainable Artificial Intelligence (XAI) [39–41]. These explainable models not only strengthen trust in the predictive outcomes of ML models but also assist researchers in discerning the pivotal factors influencing complex engineering phenomena.

In the realm of microfluidic applications, the relevance of XAI stands out sharply. Intricate behaviors are often observed in microfluidic experiments, such as the optimization of membraneless microfluidic fuel cells, the fusion dynamics of coalescing droplets, shear-induced phase transitions, and the nuanced mechanisms of droplet breakup at T-junctions [42–45]. Given that each of these applications involves a wide array of parameters and features, gaining a comprehensive understanding of key influencing factors and learning how to adeptly adjust operational parameters become increasingly indispensable. For example, while the study by Seemann et al. [46] provides a comprehensive understanding of how geometry and wetting properties influences droplet generation in microfluidic channels, the application of XAI could offer additional layers of insight. XAI could break down the complex computational models into interpretable components, allowing researchers to pinpoint how subtle changes in variables such as channel roughness or surface wettability interact to affect droplet formation rates. Through visualizations and interpretive algorithms, XAI can facilitate a more nuanced understanding of these relationships, thereby complementing traditional research methods. As research intensifies in microfluidic processes, the clarifications brought forth by explainable models are instrumental in guiding improved experimental designs, ensuring more dependable predictions leading to the development of more reliable microfluidic applications.

In this work, we employ a two-pronged strategy that integrates explainability into ML models to investigate the coalescence of aqueous droplets in oil within microfluidic devices. The first phase involves the design and construction of a suite of ML models, specifically Random Forest, XGBoost, and Multilayer Perceptrons (MLPs), each fine-tuned through hyperparameter optimization. These models are employed to predict the coalescence behavior of droplets under varying experimental conditions. The subsequent phase is dedicated to augmenting explainability by scrutinizing the model's predictive outcomes and associated features. The goal is to offer a clear understanding of which specific attributes or conditions, such as channel geometry or flow rates, most effectively contribute to the successful coalescence of aqueous drops in oil that can optimize resource allocation in subsequent experiments and minimize superfluous operations. This is achieved through the deployment of widely-used post-hoc explainability methodologies, such as SHAP and Local Interpretable Model-agnostic Explanations (LIME) [47–49]. Additionally, feature ablation testing is utilized to validate the influence and relevance of each feature on the droplet coalescence phenomena.

Our contribution lies in the novel integration of explainability within ML models specifically targeting the study of droplet coalescence in microfluidic systems. By employing XAI techniques, this integration enhances the trustworthiness and practical utility of our machine learning models [50]. Hence, our work does more than just provide accurate predictive models; it also offers actionable insights that are poised to catalyze advances in both academic research and industrial applications concerning droplet coalescence in microfluidic systems.

2. Experiments and dataset

In this section, we delve into the experimental procedures for droplet coalescence conducted within microfluidic devices and examine the distribution of the resulting dataset. Additionally, we provide a clear overview of the data pre-processing techniques employed.

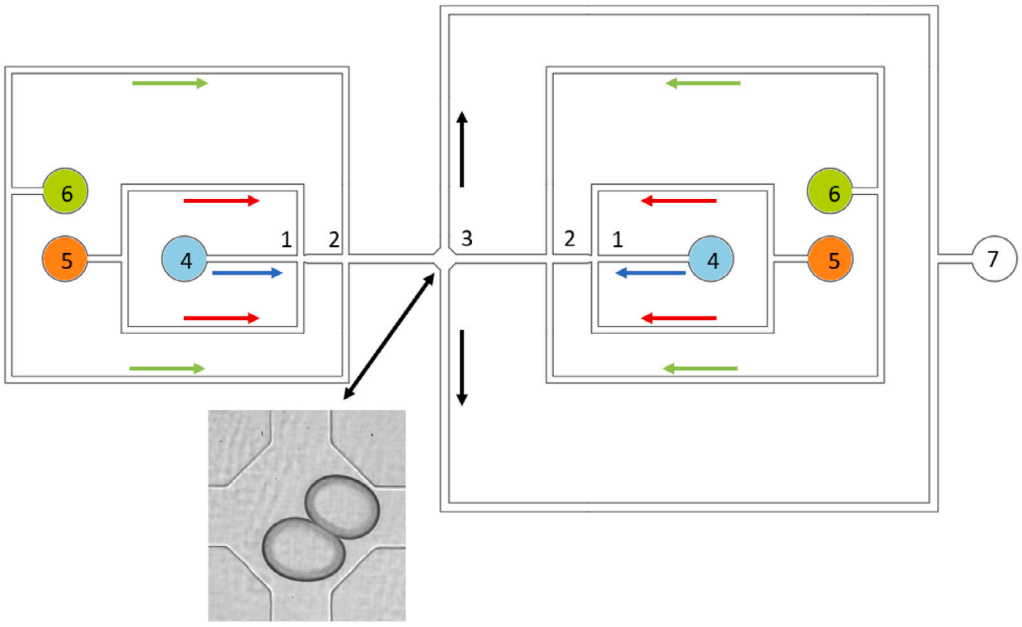


Fig. 1. Microfluidic device used in experimental studies: 1 — cross-junctions for drop formation, 2 — additional oil inputs through side channels, 3 — coalescence chamber, 4 - water inlets, 5 — main oil inlets, 6 — additional oil inlets, 7 — outlet.

2.1. Experimental procedure

Experimental studies of drop coalescence are carried out in microfluidic devices with rectangular channels made of PDMS using standard soft lithography [51]. The outline of the device is shown in Fig. 1. The device has two symmetrical cross-junctions where drops of dispersed phase, double distilled water from water still Aquatron A 4000 D, are formed by hydrodynamic flow-focusing [52] using a continuous phase of mineral oil (Sigma). After formation, the drops flow along the straight channels towards the coalescence chamber. There are side channels for an additional oil input enabling control of the distance between drops and the total flow rate at the chamber inlet. The drops meet within a coalescence chamber, which has a square shape with two symmetrical inlets and two outlets at the vertices of the square, as shown in the inset in Fig. 1.

The type of chamber described above has been shown to be beneficial for studying coalescence between pairs of drops in microfluidic device [8,29,32,53]. Three devices with some differences in sizes, mostly in channel height, are used. The first one has channel width at the entrance to chamber $L = 414 \pm 4$ μm, channel height $H = 226 \pm 13$ μm, and chamber wall-to-wall width $W = 1085 \pm 4$ μm. The second and third devices have $L = 428 \pm 23$ μm, $H = 160 \pm 5$ μm, $W = 1073 \pm 76$ μm, and $L = 394 \pm 1$ μm, $H = 117 \pm 3$ μm, $W = 985 \pm 2$ μm respectively. The presented sizes are the averages from several sets of experiments, each set comprising 100 measurements. The drop size is always larger than the channel height, therefore drops have a pancake shape enabling a sizeable contact area with the top/bottom wall. To minimize wetting issues on drop formation and coalescence, the devices are hydrophobised with Aquapel.

The liquids are supplied to 3 pairs of symmetrical inlets of the microfluidic device (4, 5 and 6 in Fig. 1) by syringe pumps Al-4000, World Precision Instruments. In this study, the flow rate of the dispersed phase is kept at 2 μL/min, that of the main flow of the continuous phase is kept at 10 μL/min, whereas the flow rate for additional oil input is varied between 1.5 and 8 μL/min. Considering that syringe pumps provide an oscillatory flow [54,55], the pumps equipped by two plastic syringes (5 mL, Fisher) are used for each pair of symmetrical inlets to synchronize the drop production. Nevertheless, flow oscillations as well as inevitable deviations in channel sizes and occasional drop coalescence in the channels result in a certain distribution of drop sizes advantageous for this study. In particular, the normalized drop diameters are used in the machine learning models. The deviations in the channel sizes inevitable in PDMS devices result also in a time difference between the drops arrival to the chamber. This time difference will be another feature of the models.

It is well known that most organic liquids, including mineral oil used in this study, cause PDMS swelling, resulting in the change of channel sizes. To minimize the effect of the swelling, the devices are primed with mineral oil for several days before their use. As there are still changes in the channel sizes due to oil exposure during the study, the wall-to-wall chamber width is measured for each series of experiments, and drop diameters from the series are normalized by this value. So the drop sizes used in the model are the sizes relative to the chamber size. Beside the width of the chamber, which is easily measurable, there are also changes in the chamber height that affect the drop contact area with the chamber. It is impossible to measure the height on the working device but as will be shown below this is an important feature of drop coalescence. Therefore, the effective height is introduced as one of the features. This is calculated as $H_{eff} = Q/(V \times L)$, where Q is the flow rate determined as sum of readings from the syringe pumps supplying the liquids to device and V is the average velocity in the channel. Considering that the drop length in the channel in this study is in the range of (1–1.5)L, we accept that the drop velocity in the channel near the entrance to the chamber is given by V to be a good approximation. The velocity is calculated as an average from at least 10 drops (5 from the each of input channels) with a standard deviation of no more than 10%.

Drop movement and coalescence are recorded using a high-speed video camera (Photron SA-5) connected to an inverted microscope (Nikon Eclipse Ti2-U) at 1000 frames per second and an exposure of 0.05 ms with a 1024 × 1024 pixel field of view. The lens (Nikon Plan Fluor 10×) provides a spatial resolution of 2 μm per pixel. The images are processed using ImageJ [56]. The equivalent drop diameters, D, are calculated from the measured areas of the drops in the field of observation, S. The drop velocities are calculated as the distance the leading edge of the drop moves inside the channel to the entrance to the chamber divided by the corresponding time. Considering the contraction/expansion structure of the flow field in the chamber [32], drops are first brought together, form a doublet, then rotate and can be detached from each other when the angle between the line connecting

Table 1
Distribution of instances of coalescence and non-coalescence among the training and testing datasets.

	Coalescence	Non-coalescence	Balance ratio (BR)	Total
Total dataset	782	719	1.09	1501
Training dataset	625	575	1.09	1200
Testing dataset	157	144	1.09	301

the drop centers and the axis of output channels becomes smaller than 45°. Drops can coalesce either in the compression or expansion stages. In this study, we do not distinguish between these modes of coalescence, but the vast majority of coalescence occurs in the expansion stage, often at the instant when drops are about to detach. This is in line with previous observations of drop coalescence in extensional flow [8,53].

2.2. Dataset

This study investigates an experimental tabular dataset comprising 1501 samples, inclusive of five features and one label, y:

$$\{x_{\frac{D}{W}1+\frac{D}{W}2}, x_{|\frac{D}{W}1-\frac{D}{W}2|}, x_{dt}, x_{\text{flow}}, x_{\text{Heff}}\}$$

The probability of coalescence depends on many factors, such as rheological properties of continuous and dispersed phase, their density, interfacial tension, shear stress, flow conditions etc [15]. Here, the conditions of laminar extensional flow with fixed densities, viscosities and interfacial tension are used. Effect of the last three parameters is a subject of an ongoing study, while the focus of the present study is the effect of stresses applied to drops due to position of drops encounter, flow intensity and drop confinement. These experimental factors are represented by the following features:

- $x_{\frac{D}{W}1+\frac{D}{W}2}$ refers to the sum of two droplet diameters (D1 and D2) normalized by the width of the coalescence chamber between two walls (W). This feature shows the size of doublet as related to the chamber size.
- $x_{|\frac{D}{W}1-\frac{D}{W}2|}$ is the absolute value of the difference between two normalized diameters. This is an indicator of the disparity in the sizes of droplets.
- x_{dt} represents a temporal element in the experiment, i.e. the time interval between the successive entrance of droplets into chamber.
- x_{flow} refers to the total flow rate in the each of input into chamber.
- x_{Heff} is the effective height of the channel explained in experimental section.

Furthermore, the label y is classified into two categories: "Coalescence" and "Non_coalescence". The experimental dataset in this study is a balanced dataset, comprising 782 instances of "Coalescence" and 719 instances of "Non-Coalescence". For robust evaluation of the ML models, we partition the original dataset into two subsets: a training dataset and a testing dataset, which are stratified by approximately 1.09 balance ratio (i.e., the ratio between "Coalescence" and "Non-Coalescence" samples in the dataset) and the details are shown in Table 1. This partitioning is performed using a shuffle strategy, meticulously maintaining the original label distribution ratio, thus ensuring the same stratification. Furthermore, it is unnecessary to create a separate validation dataset because k-fold cross-validation is utilized in the training process. This strategy can assess the generalization ability of predictive models and prevent over-fitting during training [57,58].

The distribution of all features for Coalescence and Non-coalescence is displayed in Fig. 2. Figs. 2(a)–2(e) illustrate the distribution of the five features within the dataset, specifically categorized under the labels "Coalescence" and "Non-Coalescence". These distributions are evaluated using the kernal density estimation method [59–61], with the solid lines in each plot representing the estimated distribution trends accordingly. Fig. 2(f) demonstrates the distribution ratio of instances for this binary classification task. It attests to a near-equal distribution of instances, with "Coalescence" constituting 52.1% and "Non-Coalescence" making up 47.9% of the data. This near-parity distribution proves that we have a well-balanced dataset, which ensures a fair representation of both classes, thereby eschewing biases and facilitating an objective evaluation of the proposed machine learning models.

2.3. Data pre-processing

The raw tabular data, pertaining to the coalescence phenomena of aqueous droplets in a mineral oil, need pre-processing to ensure its amenability for the ensuing analytical phase. A key pre-processing step involves the execution of min–max normalisation thereby rescaling data to a predefined range of $[0, 1]$.

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad x \in \{x_{\frac{D}{W}1+\frac{D}{W}2}, x_{|\frac{D}{W}1-\frac{D}{W}2|}, x_{dt}, x_{\text{flow}}, x_{\text{Heff}}\} \tag{1}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum value of the feature x, respectively, and x_{scaled} is the scaled results after normalization. This normalization technique functions not merely to mitigate potential discrepancies in the scale across different features, but rather it can preserve the relative relationships amongst individual sample points in the feature space [62].

3. Methodology

In this section, the construction and evaluation of machine learning models involved in the first phase, as well as methods falling under the domain of XAI, are elaborated in details. The contents encompass two kinds of tree-based models, Random Forest and XGBoost, a type of Deep Neural Networks (DNNs), Multilayer Perceptrons (MLPs), hyperparameter space search methods, and interpretability techniques such as SHAP and LIME.

3.1. Predictive models

3.1.1. Random forest and xgboost

Random Forest and XGBoost are both ensemble machine learning algorithms that utilize decision trees as their fundamental building blocks, albeit with different methods [63]. Random Forest creates an series of decision trees, each constructed independently through the utilization of bootstrapped samples from the dataset in a concurrent process [64]. Conversely, XGBoost constructs trees in a sequential manner, whereby each successive tree seeks to ameliorate the errors perpetrated by its predecessor [65]. Consequently, XGBoost frequently attains better performance; however, it is more computationally demanding and necessitates meticulous tuning of hyperparameters in comparison to random forest. In contrast, random forest is typically more flexible in training, exhibits robustness, and often delivers satisfactory performance with default configurations [66]. Moreover, due to these two tree-based methods always showing a powerful ability to process tabular data in small or medium-sized datasets, they are applied in this research with the expectation that they can demonstrate better performance than neural networks [67]. The intuitive representation of these two models is shown in Fig. 3.

3.1.2. Deep neural networks and multilayer perceptron

Deep Neural Networks (DNNs), inclusive of the widely utilized Multilayer Perceptron (MLPs), are marked by their layered structure and copious count of artificial neurons. The depth of these networks, coupled with the nonlinear activation functions employed within their hidden layers, equips them with the capability to discern and replicate highly complex patterns within data, thus demonstrating their proficiency in modeling nonlinear relationships [68]. However, their

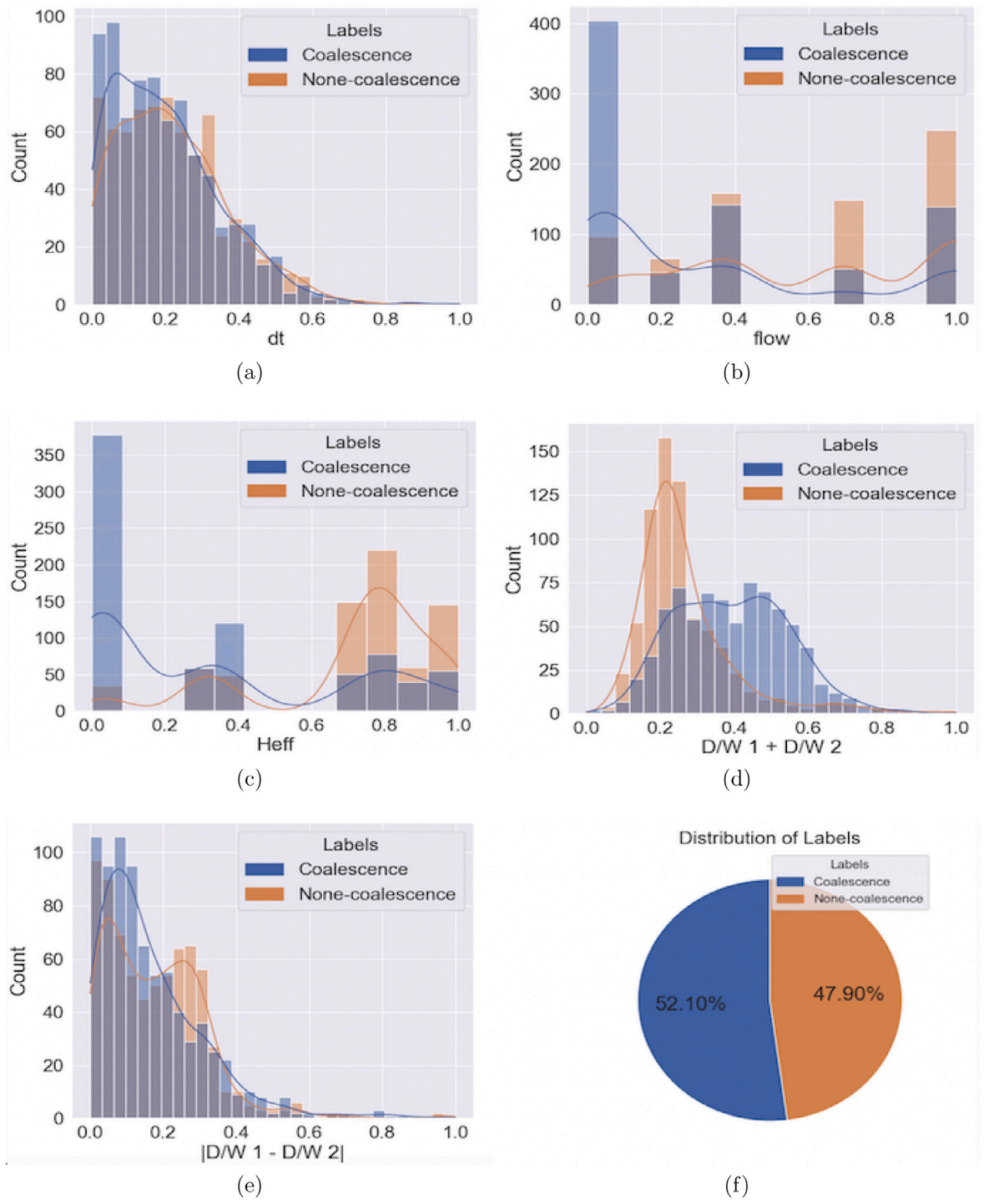


Fig. 2. Feature distribution comparison between Coalescence and Non-coalescence; see text for the definition of the features.

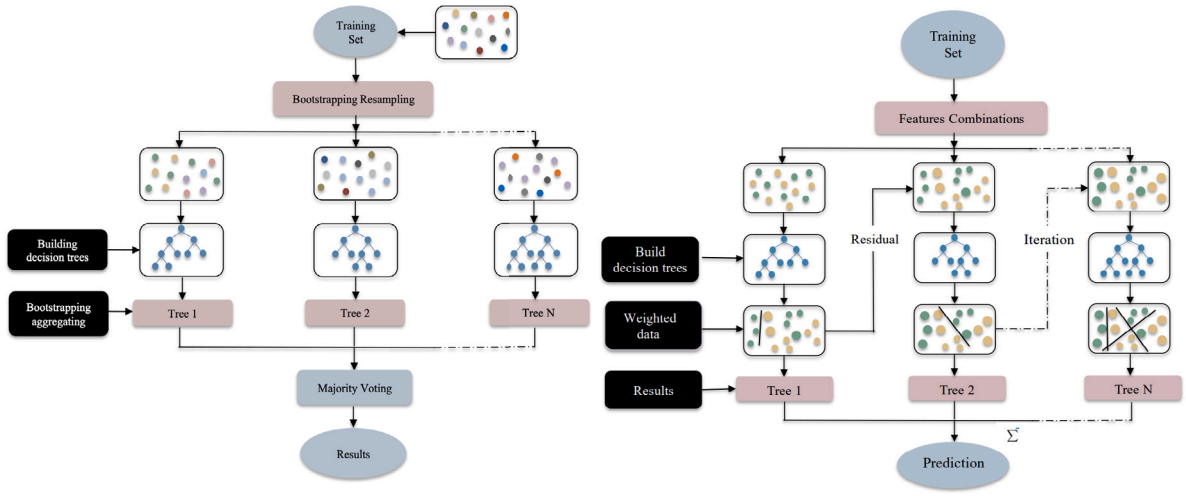
large parameter space renders them computationally demanding [69]. Nevertheless, their capacity to adeptly handle various types of data has fostered their wide application across diverse tasks, such as machine translation, sentiment analysis, image and speech recognition, anomaly detection, text classification, as well as various regression and classification problems [70,71].

3.2. Grid search method

Grid search is a technique for exploring a predefined set of hyperparameters to optimize a machine learning algorithm [72]. It can carefully traverse multiple combinations of hyperparameter configurations, performing cross-validation to determine the configuration that yields the best performance. In this study, the Grid Search method is employed to optimize both tree-based models and MLPs, utilizing 5-fold cross-validation for model training.

3.3. SHapley Additive exPlanations (SHAP)

SHAP values serve to interpret the influence of features on a specific prediction by computing the average marginal contribution of each feature across all conceivable permutations. This method is based on



(a) Random Forest (b) XGBoost

Fig. 3. Visualization of tree-based predictive models.

cooperative game theory and is utilized for interpreting the outcomes of machine learning models [73,74]. In the context of SHAP, features are treated as "players" that "contribute" to the prediction. The accumulated contribution of all features constitutes the ultimate prediction results of the model [75]. The outcome of the original predictive model, defined as a function of the input vector x, is given by

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \tag{2}$$

which can be a high-dimensional feature vector and $g(x')$ represents an interpretable surrogate model that approximates $f(x)$ but is expressed in terms of x', a simplified or lower-dimensional version of x. The transformation between x and x' is often achieved through feature selection or dimensionality-reduction techniques, making g more straightforward to interpret than f. The parameter ϕ_0 acts as a base value from which the contributions of individual features are added or subtracted, and ϕ_i is the SHAP value for the ith feature:

$$\phi_i = \sum_{B \subseteq N \setminus \{i\}} \frac{|B|!(|N| - |B| - 1)!}{|N|!} [f(B \cup \{i\}) - f(B)] \tag{3}$$

where ϕ_i is the Shapley value, representing the contribution of feature i to the prediction; N is the total set of features, which corresponds to all the elements in the feature vector x; B is a subset of N that includes selected features from the original feature vector x. The subset B does not contain the elements of feature represented by i; $f(B)$ is the predictive function with features in B. The term $\frac{|B|!(|N|-|B|-1)!}{|N|!}$ is the weighting factor, representing the number of permutations that include feature i. For any given sample, the feature possessing the larger absolute Shapley value wields a more substantial influence on the prediction result for that sample. The magnitude and sign of these Shapley values give insight into how significantly, and in what direction each feature influences a given prediction [76]. Specifically, a positive (negative) Shapley value suggests that the corresponding feature contributes positively (negatively) towards the model's predicted value [77,78].

In terms of global interpretability, SHAP provides an aggregated view across all samples, allowing researchers to discern the overall importance of each feature in the model. By analyzing the distribution of SHAP values for a particular feature, people can visualize not only the magnitude of its importance but also the direction of its effect on model predictions. Features with higher absolute SHAP values are typically more influential, and their consistent positive or negative values indicate a systematic increase or decrease in the model's prediction, respectively. Consequently, SHAP's global interpretability improves the identification of potential feature interactions and nonlinear dependencies [79]. The details of black-box model's and explainable model's interactions are shown in Fig. 4.

3.4. Local Interpretable Model-agnostic Explanations (LIME)

LIME is another explanatory method designed to clarify the predictions provided by any classifier or regressor in an understandable and faithful way [80]. It achieves this objective by approximating the model with a local surrogate model that is inherently interpretable, thus improving comprehension of the model's decisions in the vicinity of the instance under consideration [81]. It is worth noting that although both SHAP and LIME are designed to enhance the interpretability of machine learning models, they employ distinct approaches to achieve this objective. SHAP provides a global interpretability method with a theoretical foundation grounded in cooperative game theory. In contrast, LIME focuses on local interpretability by approximating the model's behavior near each individual prediction, typically by constructing a local linear surrogate model to explain each prediction. The mathematical formulation of LIME is provided by Eq. (4) [80]:

$$\xi(x) = \arg\min_{z \in Z} \mathcal{L}(f, z, \pi_x) + \Omega(z) \tag{4}$$

where Z generally refers to the family of models that are considered "interpretable" and can act as local surrogate models to approximate the behavior of the more complex model f. In the specific context of this study, which focuses on drop coalescence classification tasks, 'LimeTabularExplainer' is utilized. Consequently, Z is restricted to linear models. Specifically, each $z \in Z$ is a logistic regression model that is trained to provide a faithful local approximation of f's decision-making process in a localized region surrounding the data instance x; π_x is the proximity measure between the instance x and the data instances used to learn the explanation model. In this implementation, the measure metric is the Euclidean distance; $\xi(x)$ represents the explanation model for the instance x. Essentially, $\xi(x)$ is the optimized local linear surrogate model z that best approximates the original model f within a predefined local neighborhood around x, according to the minimization of the loss function $\mathcal{L}(f, z, \pi_x)$ and the complexity term $\Omega(z)$; $\mathcal{L}(f, z, \pi_x)$ is a measure of how unfaithfully z approximates f in the vicinity of instance x, defined by π_x; $\Omega(z)$ is a measure of complexity of the explanation model z, which aims to keep the explanation as simple as possible.

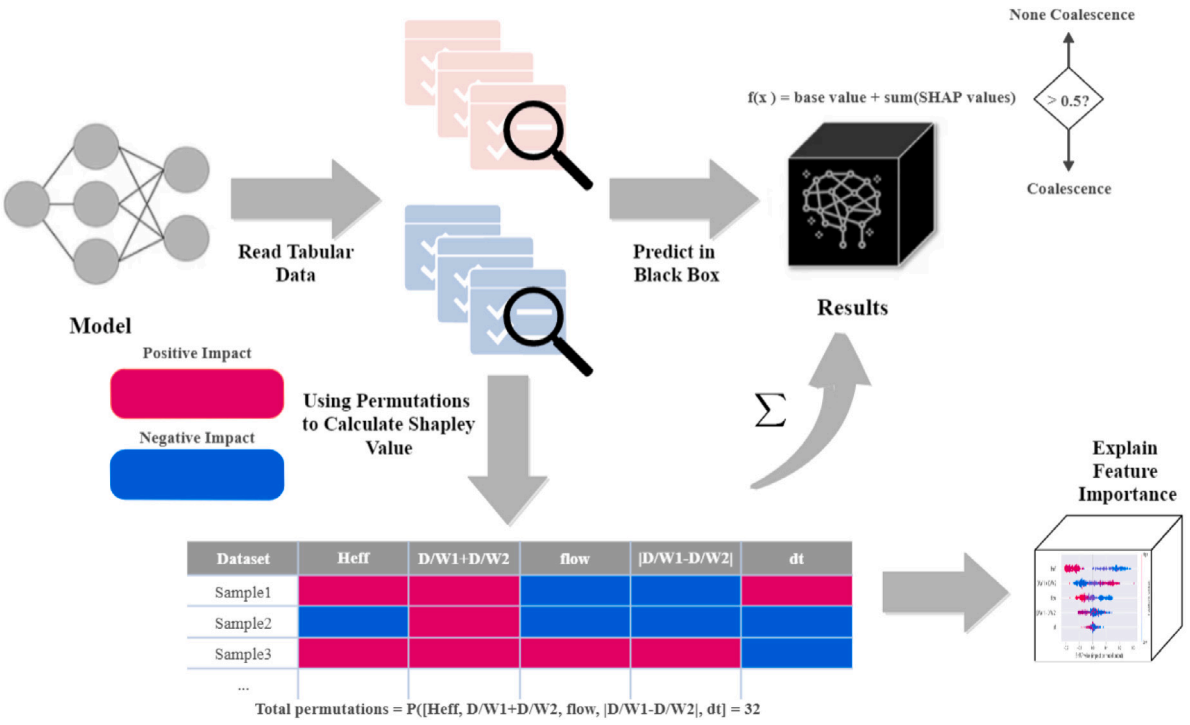


Fig. 4. Visual representation of the SHapley Additive exPlanations (SHAP).

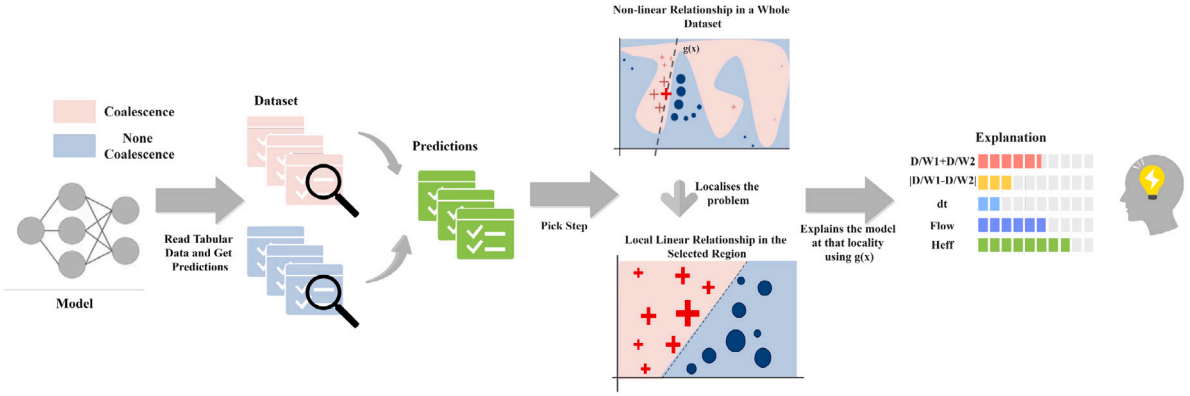


Fig. 5. Visual representation of ocal Interpretable Model-agnostic Explanations (LIME).

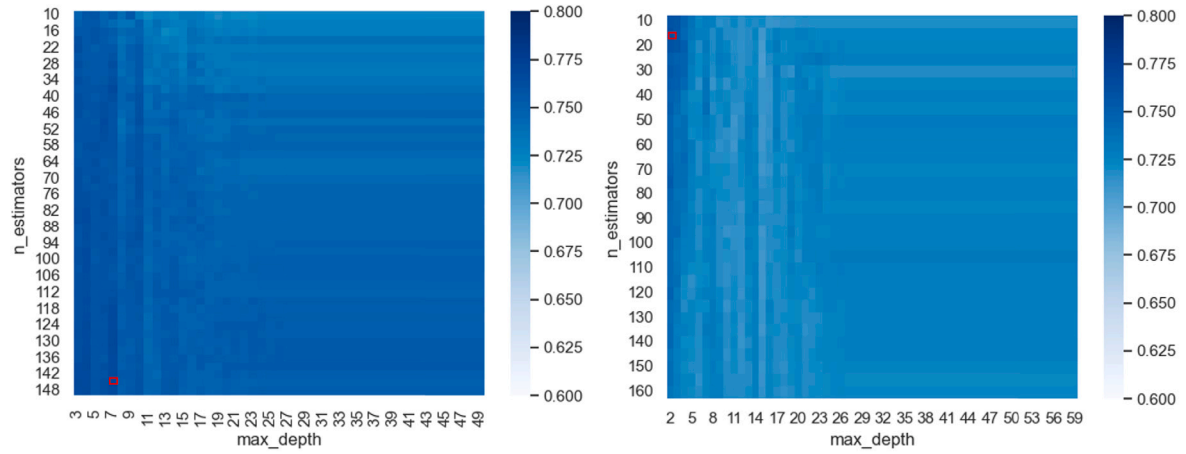
LIME learns the explanation model z by minimizing the loss function $\mathcal{L}(f, z, \pi_x)$ and the complexity measure $\Omega(z)$, effectively ensuring that z is locally faithful to f and is interpretable [80]. The Fig. 5 intuitively illustrates schematic representation of LIME's methodology, i.e. exploiting local linearity within a complex, globally nonlinear dataset. Despite the global distribution of data points depicts a nonlinear relationship, through zooming in a specific subset or a selected local region of the dataset, a simplified linear relationship can be observed. The transition from nonlinear complexity to linear simplicity underlines the versatility of LIME in deciphering the decision boundaries set by complex models. By focusing on the local view, rather than the global view, LIME can generate reasonable explanations that align with the simpler linear relationship, which is more comprehensible to humans and can keep local fidelity.

3.5. Performance metrics

The confusion matrix serves as a visualization tool for evaluating the performance of machine learning models which are mentioned in Section 3.1. A confusion matrix, in its simplest form, is a two-dimensional matrix that visualizes the performance of a supervised learning algorithm. For binary classification task, it has four entries:

- True Positives (TP): Occurrences in which both the actual outcome and the model's prediction are positive.
- True Negatives (TN): Occurrences in which both the actual outcome and the model's prediction are negative.
- False Positives (FP): Occurrences where the model erroneously classifies a negative instance as positive.
- False Negatives (FN): Occurrences where the model erroneously classifies a positive instance as negative.

These entries can be used to calculate various performance metrics [82]. The primary metric, accuracy, gauges the proportion of correct predictions. However, precision, recall, and the F1 score are also crucial for a more nuanced understanding of the model's performance. Precision focuses on the correctness of positive predictions, while recall assesses how well the model captures all actual positive cases. The F1 score harmonizes these two metrics, offering a balance between them. Four



(a) Random Forest (b) XGBoost

Fig. 6. Validation heatmaps for the tuning hyperparameters of tree-based predictive models. The selected hyperparameters are highlighted by the red frames.

performance metrics are defined as Eqs. (5)–(8):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

4. Results and analysis

In this section, we implement the three models mentioned in Section 3 and train them on the pre-processed dataset to calculate the accuracy scores for both validation and testing. Subsequently, we apply XAI techniques to conduct ablation tests aiming to comparatively analyze which features exert a substantial impact on the coalescence of aqueous droplets in oil.

4.1. Implementation details

In the modeling phase, the raw microfluidic droplet coalescence dataset is pre-processed following the pipelines delineated in Section 2.3. Due to this pre-processing, the dataset is divided into two balanced subsets with same balance ratio: a training set and a test set. Three distinct machine learning algorithms, Random Forest, XGBoost, and a Multi-Layer Perceptron (MLP) are then optimized through Grid Search method. This method is strategically employed to identify an optimal set of hyperparameters in predefined hyperparameter space, ensuring the most favorable performance on the extant dataset. Throughout the whole training process, a five-fold cross validation technique is consistently applied. Both the F1-Score and accuracy are adopted as the principal performance metrics. After a series of calculated iterations adjusting the hyperparameters, the optimal combination that yields the peak validation accuracy is selected for the final model's architecture.

The key hyperparameters of random forest and XGBoost are **n_estimators** and **max_depth**. In this study, we investigate the n_estimators parameter in the range of 10 to 151 and the max_depth parameter in the range of 3 to 50 for the random forest model. For the XGBoost model, due to its mechanism of calculating the next layer through weights, we set the range for n_estimators as 10 to 160 and for max_depth as 2 to 60. Finally, we set the optimal parameters for both the random forest and XGBoost models as [n_estimators = 145, max_depth = 7] and [n_estimators = 15, max_depth = 2], respectively. As illustrated in Fig. 6, the outcomes of hyperparameter tuning for both Random Forest and XGBoost algorithms are presented.

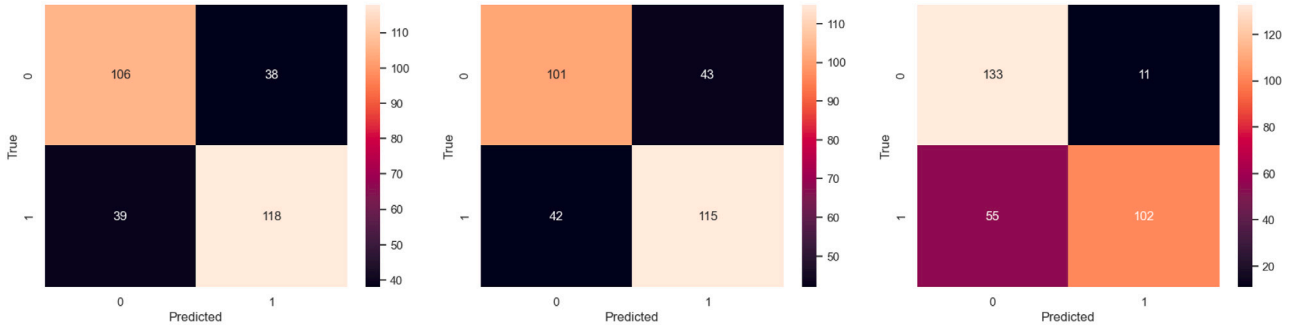
In the presented visualizations, the deep blue regions signify areas where the model achieves higher validation accuracy. Comparing the visualization heatmaps of the two models, it is evident that the Random Forest model exhibits a more extensive deep blue region. This prominence of deeper shades in the Random Forest heatmap underscores its superior adaptability on the droplet coalescence dataset in comparison to XGBoost. The broader coverage of this high-accuracy zone suggests that Random Forest might be inherently more suited for the intricacies and nuances of this particular dataset because its bagging strategy can result in the model being more resilient to over-fitting than using boosting strategy.

In our quest to optimize the MLPs, we recognize the necessity of fine-tuning an extensive array of hyperparameters. This requirement arises due to the inherent complexity of the MLPs model, as compared to more traditional machine learning counterparts. Consequently, we explore a diverse set of hyperparameters to achieve optimal performance in Table 2. The hyperparameters that are ultimately selected for the final implementation are denoted in bold within Table 2.

4.2. Predictive results

To assess the predictive performance of the considered machine learning models on coalescence events after training with 5-fold cross-validation, we present the predictive results and their confusion matrices on testing dataset. The predictive results for all three models are shown in Table 3. From the highlighted figures in Table 3, it is evident that the Multilayer Perceptron (MLP) consistently outperforms the tree-based models in metrics like precision, accuracy, and F1-score on the testing dataset. This superior performance underscores MLP's effectiveness for the droplet coalescence dataset. It indicates that, for this specific dataset, the MLP is more adept at generalizing its learnings from the training data to unseen samples. The intricacies of the data might be captured better by the MLP model structure than by the tree-based counterparts. Within the context of this study, Coalescence is designated as the positive class (1) whilst Non-Coalescence serves as the negative class (0). Accordingly, Fig. 7 presents the respective metrics from the confusion matrix, offering an intuitive insight into the predictive capabilities of the models.

Upon evaluating the classification between coalescence and non-coalescence events using three models, distinct performance metrics are observed. The Random Forest model displayed in Fig. 7(a) yields



(a) Random Forest (b) XGBoost (c) MLP

Fig. 7. Confusion metrics of three predictive models.

Table 2
Hyperparameters tuning for MLP.

Hyperparameter	Options
Activation functions	**relu**, tanh, sigmoid, linear
Optimizers	**adam**, sgd, rmsprop
Learning rates	**0.001**, 0.01, 0.1
L2 rates	0.0, 0.001, **0.01**, 0.1
Dropout rates	0.0, **0.1**, 0.2, 0.3
Epochs	10, 50, 100, 150, 200, 250, 300, 400, 450, 500, 800, 1000, 1500, 2000, **2500**, 3000

Table 3
Model performance metrics for optimal hyperparameters on testing dataset.

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Random Forest	75.64	**75.16**	75.40	74.42
XGBoost	72.78	73.25	73.02	71.76
MLP	**90.27**	64.97	**75.56**	**78.07**

a recall of 75% and a precision of 76%, reflecting its capability to identify coalescence events with a balanced accuracy, as further evidenced by its F1 score. The XGBoost model presented in Fig. 7(b) exhibits a recall of 73% with a precision of 73%, indicating a consistent balance in its predictions. In contrast, in Fig. 7(c), the MLP model has the highest precision among the three models, yet exhibits lowest recall, indicating a model that is highly accurate but not exhaustive in capturing all coalescence events. Elaborating further, the model's precision is calculated as '$\frac{TP}{TP+FP} = \frac{102}{102+11} = 90\%$'. This high precision suggests that the model is reliable in its positive classifications, ideal for applications where false positives are costly. On the other hand, its recall is $\frac{TP}{TP+FN} = \frac{102}{102+55} = 65\%$. This could be a concern in scenarios where missing a true event is risky. Additionally, the model's specificity, calculated as $\frac{TN}{TN+FP} = \frac{133}{133+11} = 92\%$, reveals that it is also proficient in correctly identifying non-coalescence events, making it versatile in its predictions.

4.3. Selection criteria for interpretability tools

In this project, it is important to carefully consider the function of explainable AI technologies, focusing on their suitability for different types of analytical tasks. To this end, the distinction between SHAP and LIME in terms of their applicability to global versus local importance analysis becomes particularly relevant.

Regarding the strengths of SHAP, it is grounded in cooperative game theory, which is adept at allocating 'credit' to features in a model's prediction. By enabling the aggregation of SHAP values, it provides a clear global view of feature impact across a dataset. Additionally, its consideration of feature interactions allows for a comprehensive understanding of how combined features affect model predictions. This approach aligns with the needs for consistent and transparent global interpretation, making SHAP an ideal tool for global feature importance analysis [83,84].

Conversely, LIME is tailored for local importance analysis. It explains individual model predictions by creating a local surrogate model around a specific instance. By using perturbations and weighting them based on proximity to the original data point, LIME excels in providing flexible and effective explanations in specific regions of the feature space. This makes it particularly suitable for analyzing why a model made a particular decision for a given instance, focusing on the rationale behind individual predictions rather than overall model behavior [85,86].

Therefore, based on the aforementioned analysis, we use SHAP for the analysis of Global Interpretability and favor LIME for Local Interpretability in our subsequent tasks. This strategic decision leverages the distinct advantages of each tool: SHAP's proficiency in aggregating feature contributions for an overarching dataset perspective and LIME's capability in creating detailed, localized models. This approach guarantees a thorough and equitable strategy for interpretability in our AI applications within chemical engineering, adeptly addressing both the overarching and detailed aspects of global and local analysis.

4.4. Global interpretability

The SHAP framework is initialized in our computational environment to facilitate the subsequent analysis. For each pre-trained model, a corresponding SHAP explainer is instantiated. The TreeExplainer is employed for tree-based models, while the KernelExplainer is used for the MLP model [87]. These explainers are adept at decomposing model predictions into individual feature contributions. By deploying the respective SHAP explainers tailored to each pre-trained predictive model, the SHAP values for our entire testing dataset are computed. This computational step involves iteratively evaluating the change in the model's output attributable to the inclusion of each feature, relative to a baseline output. The result of this calculation is a comprehensive array of SHAP values that provide a granular view of how each feature shifts the model's prediction away from a base value, thus displaying the individual feature's predictive influence. These values are not

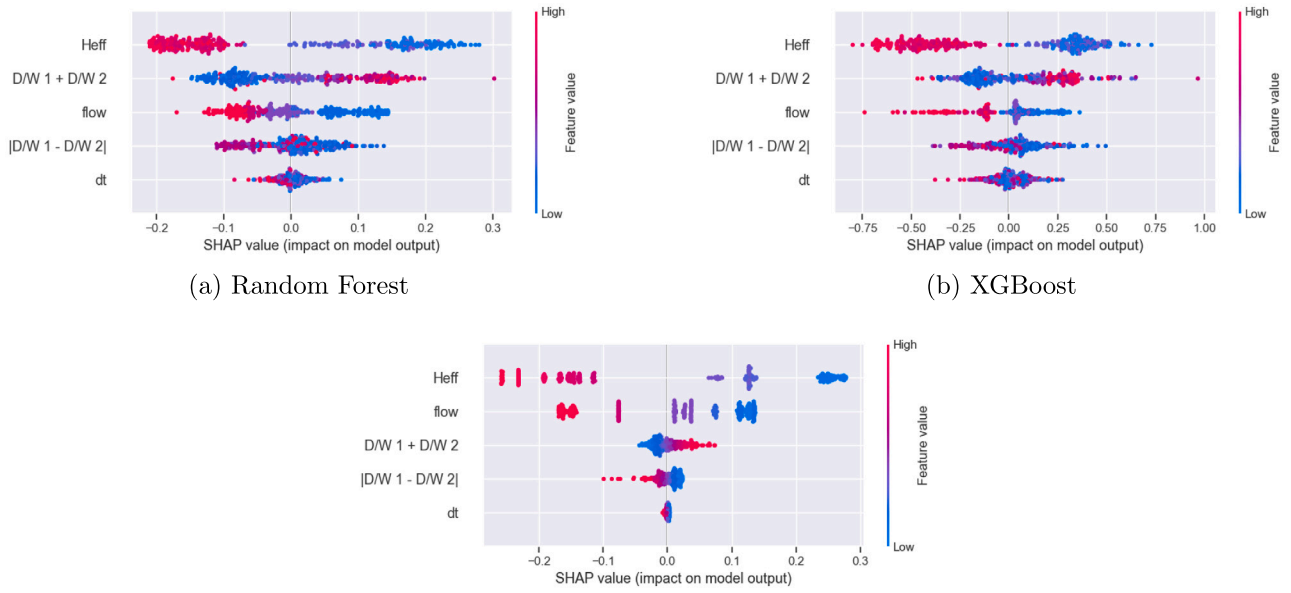


Fig. 8. SHAP plot for tree-based models.

only indicative of the feature's importance but also its directional impact, i.e., whether it increases or decreases the likelihood of droplet coalescence.

In assessing global feature importance using SHAP values, each feature's contribution to model predictions is quantified through the aggregation of every single SHAP values. By considering the absolute values of these SHAP values and calculating their mean across the dataset, one can robustly assess the overall impact of each feature, irrespective of whether it increases or decreases the model's output [88]. The features are then ranked based on these mean absolute SHAP values, with higher values indicating greater significance in affecting the model's predictions.

The SHAP summary plots in Fig. 8 provide global interpretability for both the tree-based models and the MLP, showing the influence of features on predictions. These images show the list of important features ranked from most significant to least significant (top to bottom). The abscissa, denoted as the X-axis, represents the impact on the label 1 (coalescence), where positive SHAP values connote a beneficial impact, whereas negative values signify a detrimental effect. The color bar indicates the quantity level of the original feature value, with red indicating high feature values, blue dots marking low feature values, and purple dots representing medium feature values. In these subfigures, each datum point within the plot corresponds to a specific sample from the test dataset, which is utilized for explanatory purposes.

In Figs. 8(a) and 8(b), the feature importance rank are analyzed as 'Heff', '$\frac{D}{W}1 + \frac{D}{W}2$', 'flow', '$|\frac{D}{W}1 - \frac{D}{W}2|$', and 'dt'. Furthermore, higher values of '$\frac{D}{W}1 + \frac{D}{W}2$' correspond to positive SHAP values for both models, indicating that as this ratio increases, the model's prediction is more likely to lean towards the positive class (Coalescence). Conversely, for the remaining features, lower values generally lead to positive SHAP values, suggesting an inverse relationship with positive predictions. A pellucid observation is the clustering of SHAP values for '$|\frac{D}{W}1 - \frac{D}{W}2|$' and 'dt' around the origin (zero value on the abscissa). This indicates that these features might have a neutral or varied impact on the predictions, reflecting their potential limited discriminative power in these models. The consistency in patterns across both the Random Forest and XGBoost models can be attributed to their shared tree-based structure. Both models employ a methodology of recursively splitting data based on feature thresholds, possibly leading to similar patterns in data. The ensemble nature of both models, which derive predictions from multiple decision trees, could also contribute to this similarity.

When compared with tree-based models, both the MLP and tree-based models unanimously identify 'Heff' as the most influential feature, indicating its universal significance across different modeling techniques. However, some notable discrepancies emerge in the ranking of other features and their impact partition pattern. In the MLP model, 'flow' is elevated to the second position, displacing '$\frac{D}{W}1 + \frac{D}{W}2$' which holds this rank in the tree-based models. This reshuffling highlights differing interpretations of feature importance between MLP and tree-based models. This variation in ranking could be due to the differing algorithmic mechanics: tree-based models use hierarchical partitioning of the feature space, while MLPs implement continuous, nonlinear transformations. As a result, features like 'flow' can gain prominence within the MLP's framework. Further distinguishing the MLP model's SHAP outcomes is its clear partition between features with positive and negative impacts. This is in contrast to tree-based models where such delineation is often less distinct. This sharp separation offers a more discerning view of feature contributions, which could be particularly useful when a detailed understanding of different feature impacts is required.

Moreover, there is a general agreement between the MLP and tree-based models on the directional influence of features on the output. Specifically, both types of machine learning models suggest that an increase in the value of '$\frac{D}{W}1 + \frac{D}{W}2$' is positively correlated with the outcome. On the other hand, for most other features, higher values are linked to a negative impact on the outcome. This consistency in feature influence across different modeling approaches strongly support the idea that these patterns are inherently stable and reliable in the data. Nonetheless, the nuanced disparities also demonstrate unique internal workings and sensitivities inherent to different model architectures. Simultaneously, MLP delineates the influence of SHAP values more distinctly compared to the other two models. This observation indicates that the MLPs model is more proficient in segregating and distinguishing the contributions of individual features.

4.5. Feature ablation testing

Feature ablation is used to determine the impact of specific features on ML model's performance. By systematically removing or altering

Table 4
Feature ablation testing (evaluated by accuracy).

Model	Baseline	w/o Heff	w/o $\frac{D}{W}1+\frac{D}{W}2$	w/o flow	w/o $\lvert\frac{D}{W}1-\frac{D}{W}2\rvert$	w/o dt
Random Forest	**74.42%**	72.09%	70.10%	74.09%	72.09%	70.10%
XGBoost	71.76%	66.11%	72.43%	72.09%	72.43%	**74.09%**
MLP	**78.07%**	69.77%	75.08%	76.08%	76.74%	75.75%

a feature and then evaluating the model's performance, we can gain insights into the true significance of that feature for the model's predictions. It essentially provides a way to validate feature importances derived from XAI techniques. The feature ablation testing results are shown in Table 4.

For the Random Forest model, the SHAP analysis placed 'Heff' as the primary influential feature. However, the results from the feature ablation suggest a modest performance drop when this feature is removed. This difference between the expected and observed impact of 'Heff' implies the inherent robustness of Random Forests. The ensemble nature of Random Forests, consisting of a multitude of decision trees, might allow it to adapt to the absence of even a critical feature. Each individual tree captures different facets of the dataset, and collectively, they may maintain performance levels even when a significant feature is missing. This resilience may cause a divergence between the perceived importance from techniques like SHAP and the empirical results from feature ablation. Nevertheless, it is notable that removing 'dt' leads to a more considerable performance decrease, from 74.42% to 70.10%. This discrepancy between the SHAP analysis and the feature ablation results suggests that 'dt' has a more intricate role than previously assumed.

In contrast, the XGBoost model exhibits a unique trend compared to the other models. While the removal of most features enhances its performance relative to the baseline, 'Heff' stands as a clear exception. The omission of 'Heff' incurs a significant drop in performance, highlighting its importance, which is consistent with the SHAP rankings. The improved performance upon the removal of '$\frac{D}{W}1+\frac{D}{W}2$' and 'flow', despite their indicated significance by SHAP, implies the possibility that these features, in the presence of other variables, introduce complexity, or ambiguity, which the XGBoost model finds hard to navigate. In simpler terms, the model might achieve clearer and more accurate decision boundaries when these features are absent. Even more intriguing is the noticeable enhancement in performance upon excluding 'dt'. This could suggest that 'dt', within the context of the XGBoost model, may be contributing a level of noise, or might be entangled with other features in a manner that hampers the model's predictive clarity. Moreover, according to Table 4 and Fig. 8(b), it is evident that XGBoost, although structurally similar to Random Forest as a tree-based model, reacts more sensitively to the omission of vital features. This distinction could be attributed to the boosting mechanism of XGBoost, which sequentially constructs trees to correct the errors of the preceding outputs. Consequently, each tree is more reliant on crucial features to correct the previous errors and enhance the model's predictive capacity.

Upon examining the MLP's results, one notes a distinct pattern that diverges from the tree-based models. First, the removal of 'Heff leads to a significant drop in accuracy, which stands in alignment with the SHAP results, denoting its importance. Nevertheless, when evaluating the performance implications of '$\frac{D}{W}1+\frac{D}{W}2$' and 'flow', it is observed that despite their altered rankings in SHAP importance, the exclusion of either feature from the MLP model causes merely a moderate impact on performance. This suggests that, within the MLP structure, these features may exhibit an interconnected influence, possibly sharing redundant information or compensating for one another when absent according to all three models' results. Lastly, the exclusion of 'dt' also presents an abnormal scenario. Although SHAP results suggest its relative insignificance, the model's accuracy actually declines when this feature is neglected. This counters the expectation based on the SHAP analysis, suggesting that 'dt' has a hidden or nonlinear contribution that is not fully captured by the importance ranking. This finding can also be proved by random forest's results and indicates the complex interplay and hidden dependencies that might exist between 'dt' and other features within the densely connected MLPs framework.

4.6. Local interpretability

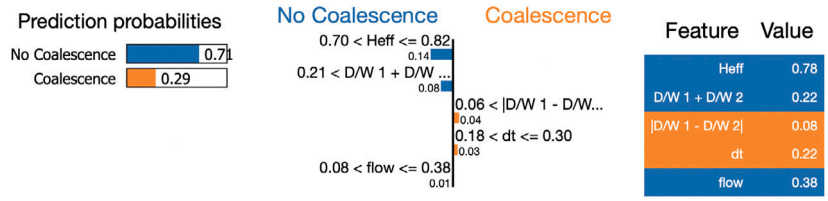
To gain a more nuanced understanding of our models' decision-making processes for specific instances, we employ LIME (Local Interpretable Model-agnostic Explanations), an approach which provides local model-agnostic explanations. LIME functions by generating a perturbed dataset around a chosen instance and learning a locally interpretable model from this new dataset. Specifically, LIME initiates its process by randomly sampling new data points in the vicinity of the instance under examination, introducing small variations to the original feature values. This sampling generates a localized region around the instance, capturing the behavior of the complex model in this constrained space.

Subsequently, LIME assigns weights to these newly generated samples based on their proximity to the original instance. Closer samples are given higher weights, indicating their greater relevance in approximating the local decision boundary. This weighting mechanism ensures that the explanations focus predominantly on the area immediately surrounding the instance of interest. The next critical step involves training a simple, interpretable model – typically, a linear model – on this weighted, perturbed dataset. The simplicity of this surrogate model is in stark contrast to the often opaque and complex nature of the original model (such as MLPs). The linear model attempts to mimic the behavior of the original model but only within this local, weighted context. It is crucial that the interpretability comes from the fact that this model is a local approximation and is not intended to capture the global dynamics of the original model [89].

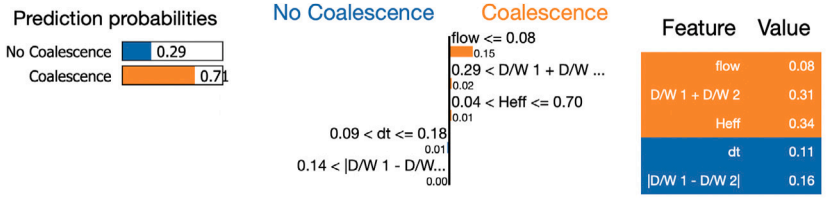
The rationale for using LIME to understand local interpretability is that while the machine learning models may exhibit complex, nonlinear behaviors globally in our drop coalescence dataset, they can exhibit approximately linear behaviors in the immediate vicinity of specific instances. This local linearity allows LIME to provide clear, interpretable insights into some specific predictions results.

Upon reviewing the LIME analyses for the two instances labeled 'Non-Coalescence' and 'Coalescence', several observations can be discussed in Fig. 9. Importantly, it is worth noting that the probabilities for 'Coalescence' (C) and 'Non-Coalescence' (NC) are generated using the LIME algorithm, rather than being directly an output from the original machine learning models, such as Random Forest, XGBoost, or MLP. For each specific instance under study, LIME creates a set of locally perturbed samples around the instance and obtains predictions from the original complex model for these samples. Subsequently, a weighted linear model is fitted to this localized set of samples. The weights are determined by how close each perturbed sample is to the original instance. This simpler, interpretable model is then used to estimate the probabilities for 'C' and 'NC', offering a localized explanation for the predictive decisions made by the original, more complex models.

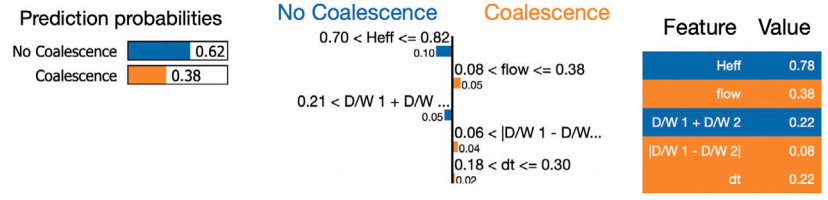
For the 'Non-Coalescence' instance, all three models – Random Forest, XGBoost, and MLP – produce predictions that align well with the true label, yielding probabilities of 0.71, 0.62, and 0.62, respectively. Although there are slight differences in how each model arrives at its prediction, there is a consistent emphasis on the features 'Heff', '$\frac{D}{W}1+\frac{D}{W}2$', and 'flow' across all models. In drawing parallels with previous SHAP analyses, it is noteworthy that the features 'Heff', '$\frac{D}{W}1+\frac{D}{W}2$', and 'flow' are similarly emphasized as significant contributors to prediction outcomes. This recurrent emphasis across different analytic methods substantiates the robustness of these features in the decision-making process. Moreover, the feature ablation tests conducted earlier reveal that when these features are individually removed from the model,



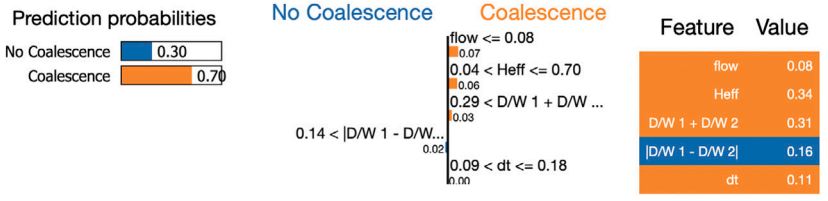
(a) Random Forest(NC)



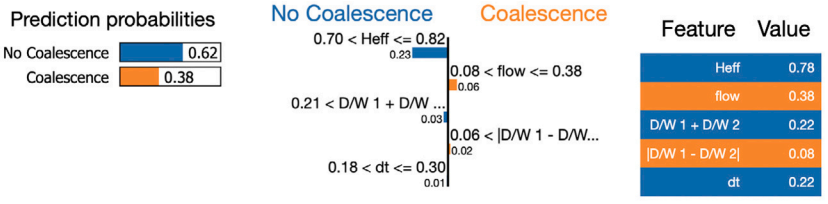
(b) Random Forest(C)



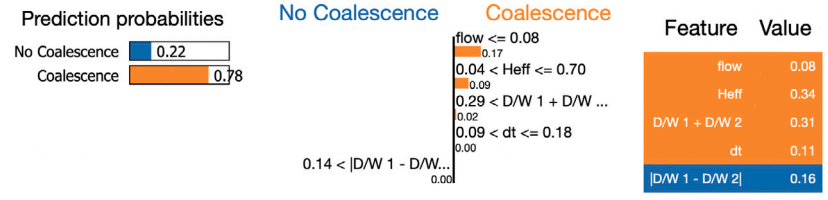
(c) XGBoost(NC)



(d) XGBoost(C)



(e) MLP(NC)

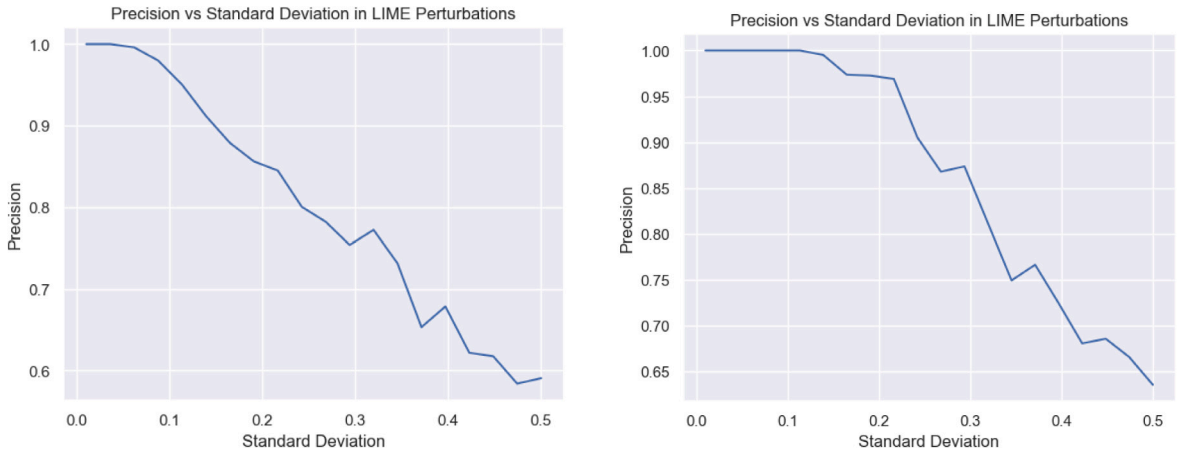


(f) MLP(C)

Fig. 9. LIME Plot for two instances. NC means the actual label for instance is Non-Coalescence, and C means the actual label for instance is Coalescence.

a marked degradation in prediction accuracy is observed. This aligns seamlessly with the feature importance indicated by both LIME and SHAP analyses.

The consistency in feature importance suggests a stable and strong relationship between these features and the predicted outcomes, and this matches the findings from the previous SHAP analysis. For the 'Coalescence' instance, the models again provide predictions that closely match the true label, with probabilities of 0.71, 0.70, and 0.78. Again, the features 'Heff', '$\frac{D}{W}1 + \frac{D}{W}2$', and 'flow' are highlighted as key determinants in the decision-making process across all models. Another



(a) XGBoost (b) MLP

Fig. 10. Local fidelity of LIME for ML models: Precision retention with low standard deviation perturbations.

point of note is the similarity in predicted probabilities across the three models. This consistent performance across different predictive model architectures indicates that, despite potential differences in how they process the dataset, their conclusions are consistent. it is worth recalling that the SHAP and feature ablation tests produced similar overarching themes, strengthening this claim. Such alignment across models suggests that the findings are not an outcome of a particular model but are likely reflective of the underlying data patterns.

4.6.1. Local fidelity and effectiveness

In evaluating the local effectiveness and fidelity of linear surrogate models derived from LIME, the provided Fig. 10 presents a clear depiction of their performance within the framework of LIME. When assessing the local fidelity, a point from the test set is randomly selected, around which, in line with LIME's methodology, five hundred perturbed points are generated to simulate LIME's computational approach. The surrogate model's weights and intercept, derived directly from the LIME explainer, are applied to calculate the outcomes for these perturbed instances. This allows for a direct comparison with the labels predicted by the original complex machine learning model, ensuring a rigorous evaluation of the surrogate model's interpretability. The Fig. 10 indicates that at lower ranges of standard deviation (0 to 0.1), there is a notable congruence between the precision of the surrogate linear model and the original machine learning model. Such high precision within this narrow perturbation scope reflects the linear surrogate model's capacity to accurately capture the original model's decision boundary, affirming the high local fidelity of LIME. This effective functioning of the linear model in close proximity to the selected instance underscores LIME's core principle: to approximate the complex model well within a local context.

Furthermore, Figs. 11(a) and 11(b) illustrate the differences between the probabilities predicted by the original models and those estimated by the LIME linear surrogate model. The distribution of residuals along the zero line (no difference) would indicate high accuracy of the surrogate model in approximating the original model's predictions. The narrow spread of residuals around the zero line, especially with a majority of data points clustering close to it, indicates that the LIME model's explanations are highly consistent with the original models' predictions.

The Figs. 11(c) and 11(d) complement this observation by quantifying the accuracy of the LIME model's classifications. An accuracy rate of 87% with XGBoost and 97% with MLP demonstrates that LIME not only approximates the overall probability distribution well but also maintains high accuracy in individual predictions, confirming its effectiveness for generating reliable local explanations in complex machine learning scenarios.

4.7. Observation and discussion

This section summarizes key insights from our XAI analysis on three machine learning models: Random Forest, XGBoost, and MLP, focusing on feature importance and model sensitivity to feature ablation.

4.7.1. SHAP feature rankings

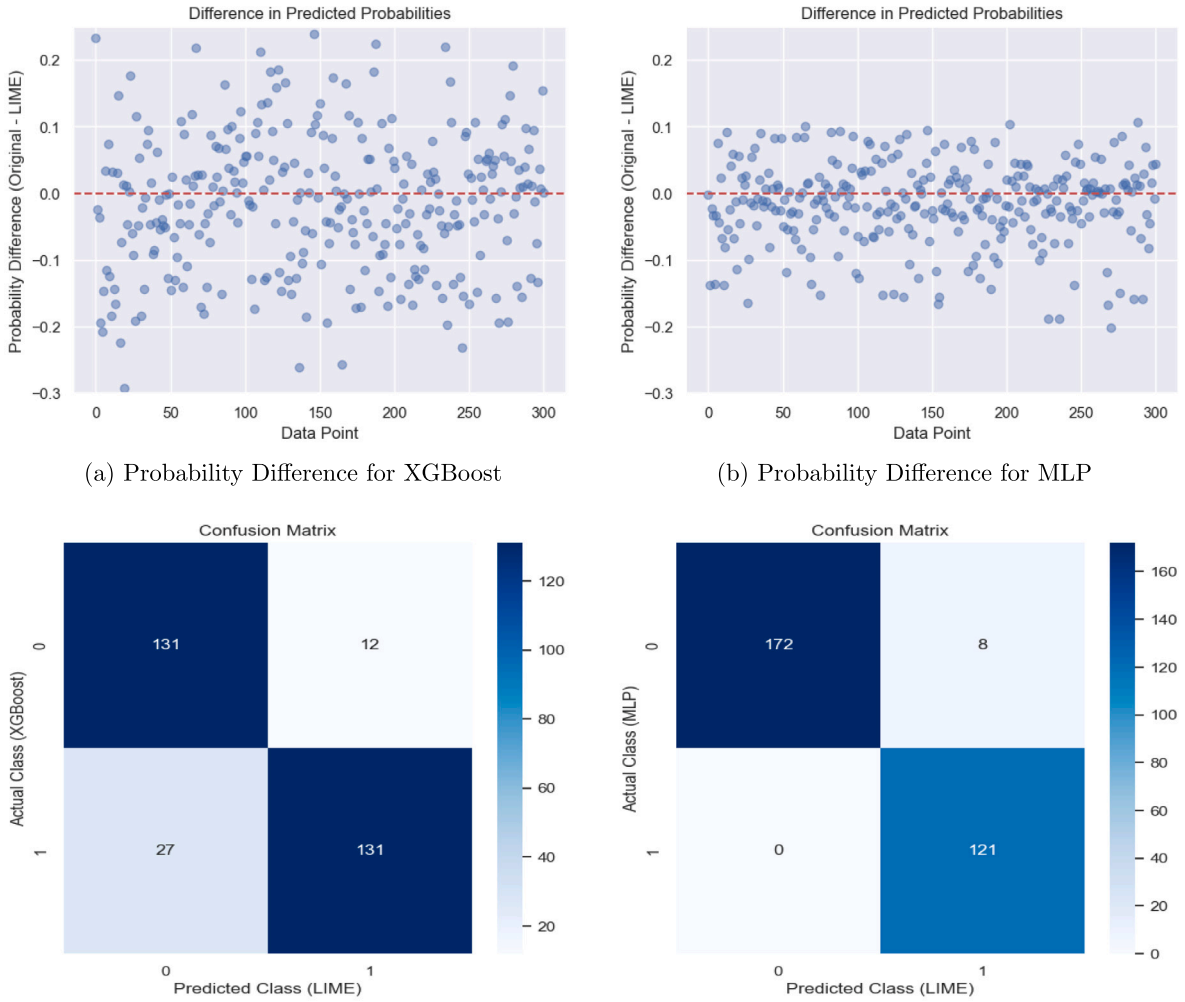
- **Random Forest**: Heff > $\frac{D}{W}1 + \frac{D}{W}2$ > flow > $|\frac{D}{W}1 - \frac{D}{W}2|$ > dt
- **XGBoost**: Heff > $\frac{D}{W}1 + \frac{D}{W}2$ > flow > $|\frac{D}{W}1 - \frac{D}{W}2|$ > dt
- **MLP**: Heff > flow > $\frac{D}{W}1 + \frac{D}{W}2$ > $|\frac{D}{W}1 - \frac{D}{W}2|$ > dt

4.7.2. General observations

- 'Heff' consistently appears as the most crucial feature across all models and interpretability methods, aligning with its top SHAP ranking in each model.
- '$\frac{D}{W}1 + \frac{D}{W}2$' and 'flow' are also frequently significant but their importance ranking varies between models, as indicated by SHAP.
- The importance of 'dt' is consistently lowest across all models in SHAP rankings, but its actual effect, particularly in the MLP model, suggests more complex relationships.
- Different models react differently to feature omission despite having similar SHAP rankings, highlighting their unique sensitivities and structural differences.
- There is a strong alignment between the features that are significant globally (via SHAP and feature ablation) and locally (via LIME), suggesting the robustness of these features.

4.7.3. Discussion

- **Feature Robustness**: 'Heff' consistently maintains its top SHAP ranking and shows significant impact in both local (LIME) and global (feature ablation) interpretability analyses, confirming its critical role. Similarly, the features '$\frac{D}{W}1 + \frac{D}{W}2$' and 'flow' are not only statistically significant but also practically significant, making substantial contributions in both local and global interpretability analyses.
- **Model Sensitivity**: Although Random Forest and XGBoost share similar SHAP rankings, they exhibit different resilience to feature omission, highlighting the nuanced differences between their tree-based architectures. Specifically, Random Forest's ensemble mechanism lends it robustness to the absence of critical features, while XGBoost's boosting technique makes it more sensitive to such omissions.



(a) Probability Difference for XGBoost

(b) Probability Difference for MLP

(c) Confusion Matrix for XGBoost and LIME

(d) Confusion Matrix for MLP and LIME

Fig. 11. Probability residuals and confusion matrix analysis for LIME and ML models.

- **Hidden Dependencies in MLP**: The lowest SHAP ranking of 'dt' in all models contrasts with its actual impact on model performance, hinting at complex, hidden dependencies. This discrepancy in MLP suggests that 'dt' may interact nonlinearly with other features, a behavior more easily captured by the densely connected architecture of MLPs. Therefore, 'dt"s role is more nuanced than linear methods like SHAP can reveal, reflecting the intricate feature interdependencies inherent to neural networks
- **Congruence Across Methods**: The congruence between LIME, SHAP, and feature ablation tests speaks to the reliability and validity of these critical features in the dataset. The consistency across different analytical methods and predictive models reaffirms that these features hold not just statistical but also practical significance in the phenomena being studied.
- **Local vs Global Interpretability**: The consistency between LIME and SHAP rankings for significant features suggests that these models, although complex, may behave linearly or at least predictably in the vicinity of specific instances. This adds a layer of trust to the predictive power and interpretability of these models.

Overall, the MLP model emerges as the most effective in predicting microfluidic droplet coalescence, outperforming its tree-based counterparts, Random Forest and XGBoost, in both accuracy and resilience to feature ablation. While the tree-based models share similar SHAP rankings, they display unique sensitivities to feature exclusion due to their internal architectures. MLP, however, stands out for its superior ability to capture complex feature interdependencies, as evidenced by its performance when the lowest-ranking feature 'dt' is removed.

5. Conclusions

In chemical engineering, understanding of the dynamics of drop coalescence in microfluidic devices is of considerable importance for understanding both the fundamentals and in optimization of process design. This study investigates the predictive capabilities of three prevalent machine learning models – Random Forest, XGBoost, and Multi-Layer Perceptron (MLP) – to elucidate the data patterns associated with 'Coalescence' and 'Non-Coalescence' events. Feature importance is first gauged using SHAP values. In this context, the effective height of the channel, the normalized sum of two droplet diameters relative to the size of coalescence chamber (both features reflect the doublet confinement), and the total flow rate into each channel consistently emerge as pivotal determinants across all models examined. These are important findings for engineers looking to optimize the design of processes involving droplet coalescence. Further scrutiny through feature ablation testing explores the sensitivity and robustness of each model when these key features are omitted. Complementing this, local interpretability through LIME not only corroborates the overarching importance of the identified features but also offers specific insights into each model's inferential logic, thereby reinforcing the global interpretability insights obtained from SHAP. Collectively, this multifaceted

approach integrates global and local interpretability with feature ablation, enhancing a profound understanding of the decision-making mechanics within the chosen models and amplified the trustworthiness of their predictions.

In looking ahead, the study identifies several avenues for future research. Future research can adapt this methodology to other fluid dynamics challenges in chemical engineering, such as emulsion stability and mass transfer. Additionally, exploring advanced neural networks may improve prediction accuracy. Importantly, tailor-made interpretability frameworks for chemical engineering could further enhance the practical utility of machine learning in the field, facilitating the development of increasingly transparent, accountable, and verifiable ML applications for complex engineering systems.

CRediT authorship contribution statement

Jinwei Hu: Data curation, Methodology, Software, Writing – original draft. **Kewei Zhu:** Methodology, Software, Writing – review & editing. **Sibo Cheng:** Conceptualization, Methodology, Software, Validation, Writing – original draft. **Nina M. Kovalchuk:** Conceptualization, Data curation, Formal analysis, Writing – review & editing. **Alfred Soulsby:** Data curation, Methodology, Writing – review & editing. **Mark J.H. Simmons:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Omar K. Matar:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Rossella Arcucci:** Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research is funded by the EP/T000414/1 PREdictive Modelling with QuantIfication of UncERtainty for MultiphasE Systems (PREMIERE), United Kingdom. This work is partially supported by the Leverhulme Centre for Wildfires, Environment and Society through the Leverhulme Trust, grant number RC-2018-023.

Data and code availability

The code of this study is available at https://github.com/DL-WG/Explainable-AI-for-Drop-Coalescence.

References

[1] E.A. Galan, H. Zhao, X. Wang, Q. Dai, W.T. Huck, S. Ma, Intelligent microfluidics: The convergence of machine learning and microfluidics in materials science and biomedicine, Matter 3 (6) (2020) 1893–1922.
[2] K. Muijlwijk, I. Colijn, H. Harsono, T. Krebs, C. Berton-Carabin, K. Schroën, Coalescence of protein-stabilised emulsions studied with microfluidics, Food Hydrocolloids 70 (2017) 96–104.
[3] B.M. Jose, T. Cubaud, Droplet arrangement and coalescence in diverging/converging microchannels, Microfluid. Nanofluidics 12 (2012) 687–696.
[4] N. Kovalchuk, J. Chowdhury, Z. Schofield, D. Vigolo, M. Simmons, Study of drop coalescence and mixing in microchannel using ghost particle velocimetry, Chem. Eng. Res. Des. 132 (2018) 881–889.
[5] M. Dudek, K. Muijlwijk, K. Schroën, G. Øye, The effect of dissolved gas on coalescence of oil drops studied with microfluidics, J. Colloid Interface Sci. 528 (2018) 166–173.
[6] M. Dudek, D. Fernandes, E.H. Herø, G. Øye, Microfluidic method for determining drop-drop coalescence and contact times in flow, Colloids Surf. A 586 (2020) 124265.
[7] N.M. Kovalchuk, M. Reichow, T. Frommweiler, D. Vigolo, M.J. Simmons, Mass transfer accompanying coalescence of surfactant-laden and surfactant-free drop in a microfluidic channel, Langmuir 35 (28) (2019) 9184–9193.
[8] H. Yi, T. Fu, C. Zhu, Y. Ma, Local deformation and coalescence between two equal-sized droplets in a cross-focused microchannel, Chem. Eng. J. 430 (2022) 133087.
[9] T. Leary, M. Yeganeh, C. Maldarelli, Microfluidic study of the electrocoalescence of aqueous droplets in crude oil, ACS Omega 5 (13) (2020) 7348–7360.
[10] I. Akartuna, D.M. Aubrecht, T.E. Kodger, D.A. Weitz, Chemically induced coalescence in droplet-based microfluidics, Lab Chip 15 (4) (2015) 1140–1144.
[11] E.D. Ngouémazong, S. Christiaens, A. Shpigelman, A. Van Loey, M. Hendrickx, The emulsifying and emulsion-stabilizing properties of pectin: A review, Compr. Rev. Food Sci. Food Saf. 14 (6) (2015) 705–718.
[12] C. Berton-Carabin, K. Schroën, Towards new food emulsions: Designing the interface and beyond, Curr. Opin. Food Sci. 27 (2019) 74–81.
[13] X. Sun, Y. Zhang, G. Chen, Z. Gai, Application of nanoparticles in enhanced oil recovery: a critical review of recent progress, Energies 10 (3) (2017) 345.
[14] O. Aarøen, E. Riccardi, M. Sletmoen, Exploring the effects of approach velocity on depletion force and coalescence in oil-in-water emulsions, RSC Adv. 11 (15) (2021) 8730–8740.
[15] G. Kelbaliyev, D.B. Tagiyev, S.R. Rasulov, Transport Phenomena in Dispersed Media, CRC Press, 2019.
[16] B. Zheng, R.F. Ismagilov, A microfluidic approach for screening submicroliter volumes against multiple reagents by using preformed arrays of nanoliter plugs in a three-phase liquid/liquid/gas flow, Angew. Chem., Int. Ed. Engl. 44 (17) (2005) 2520–2523.
[17] E. Um, D.-S. Lee, H.-B. Pyo, J.-K. Park, Continuous generation of hydrogel beads and encapsulation of biological materials using a microfluidic droplet-merging channel, Microfluid. Nanofluid. 5 (2008) 541–549.
[18] J.-T. Wang, J. Wang, J.-J. Han, Fabrication of advanced particles and particle-based materials assisted by droplet-based microfluidics, Small 7 (13) (2011) 1728–1754.
[19] T.M. Ho, A. Razzaghi, A. Ramachandran, K.S. Mikkonen, Emulsion characterization via microfluidic devices: A review on interfacial tension and stability to coalescence, Adv. Colloid Interface Sci. 299 (2022) 102541.
[20] N.M. Kovalchuk, M.J. Simmons, Review of the role of surfactant dynamics in drop microfluidics, Adv. Colloid Interface Sci. (2023) 102844.
[21] S. Feng, L. Yi, L. Zhao-Miao, C. Ren-Tuo, W. Gui-Ren, Advances in micro-droplets coalescence using microfluidics, Chin. J. Anal. Chem. 43 (12) (2015) 1942–1954.
[22] L. Mazutis, A.D. Griffiths, Selective droplet coalescence using microfluidic systems, Lab Chip 12 (10) (2012) 1800–1806.
[23] P. Ma, D. Liang, C. Zhu, T. Fu, Y. Ma, An effective method to facile coalescence of microdroplet in the symmetrical T-junction with expanded convergence, Chem. Eng. Sci. 213 (2020) 115389.
[24] B. Bera, R. Khazal, K. Schroën, Coalescence dynamics in oil-in-water emulsions at elevated temperatures, Sci. Rep. 11 (1) (2021) 10990.
[25] P. Dell'Aversana, J.R. Banavar, J. Koplik, Suppression of coalescence by shear and temperature gradients, Phys. Fluids 8 (1) (1996) 15–28.
[26] E. Chatzi, J.M. Lee, Analysis of interactions for liquid-liquid dispersions in agitated vessels, Ind. Eng. Chem. Res. 26 (11) (1987) 2263–2267.
[27] M.A. Hsia, L.L. Tavlarides, Simulation analysis of drop breakage, coalescence and micromixing in liquid-liquid stirred tanks, Chem. Eng. J. 26 (3) (1983) 189–199.
[28] Z. Ibrahim, P. Tulay, J. Abdullahi, Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa, Environ. Sci. Pollut. Res. 30 (2) (2023) 3621–3643.
[29] Y. Zhuang, S. Cheng, N. Kovalchuk, M. Simmons, O.K. Matar, Y.-K. Guo, R. Arcucci, Ensemble latent assimilation with deep learning surrogate model: application to drop interaction in a microfluidics device, Lab Chip 22 (17) (2022) 3187–3202.
[30] K. Nathanael, S. Cheng, N.M. Kovalchuk, R. Arcucci, M.J. Simmons, Optimization of microfluidic synthesis of silver nanoparticles: a generic approach using machine learning, Chem. Eng. Res. Des. 193 (2023) 65–74.
[31] R. Dong, H. Leng, J. Zhao, J. Song, S. Liang, A framework for four-dimensional variational data assimilation based on machine learning, Entropy 24 (2) (2022) 264.
[32] K. Zhu, S. Cheng, N. Kovalchuk, M. Simmons, Y.-K. Guo, O.K. Matar, R. Arcucci, Analyzing drop coalescence in microfluidic devices with a deep learning generative model, Phys. Chem. Chem. Phys. (2023).
[33] P.O. Dral, Quantum chemistry in the age of machine learning, J. Phys. Chem. Lett. 11 (6) (2020) 2336–2347.
[34] S. Stocker, G. Csányi, K. Reuter, J.T. Margraf, Machine learning in chemical reaction space, Nat. Commun. 11 (1) (2020) 5505.
[35] V. Venkatasubramanian, The promise of artificial intelligence in chemical engineering: Is it here, finally? AIChE J. 65 (2) (2019) 466–478.
[36] C.B. Azodi, J. Tang, S.-H. Shiu, Opening the black box: interpretable machine learning for geneticists, Trends Genet. 36 (6) (2020) 442–455.

[37] A.M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B.A. Becker, C. Mooney, Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review, Appl. Sci. 11 (11) (2021) 5088.

[38] E. Vorm, D.J. Combs, Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (ISTAM), Int. J. Hum.-Comput. Interact. 38 (18–20) (2022) 1828–1845.

[39] B. Esteki, M. Masoomi, M. Moosazadeh, C. Yoo, Data-driven prediction of Janus/Core–Shell morphology in polymer particles: A machine-learning approach, Langmuir 39 (14) (2023) 4943–4958.

[40] A. Sivaram, V. Venkatasubramanian, XAI-MEG: Combining symbolic AI and machine learning to generate first-principles models and causal explanations, AIChE J. 68 (6) (2022) e17687.

[41] Y. Yang, H. Gong, Q. Yang, Y. Deng, Q. He, S. Zhang, On the uncertainty analysis of the data-enabled physics-informed neural network for solving neutron diffusion eigenvalue problem, 2023, arXiv preprint arXiv:2303.08455.

[42] D.D. Nguyen, M. Tanveer, H.-N. Mai, T.Q.D. Pham, H. Khan, C.W. Park, G.M. Kim, Guiding the optimization of membraneless microfluidic fuel cells via explainable artificial intelligence: Comparative analyses of multiple machine learning models and investigation of key operating parameters, Fuel 349 (2023) 128742.

[43] C.N. Baroud, F. Gallaire, R. Dangla, Dynamics of microfluidic droplets, Lab Chip 10 (16) (2010) 2032–2045.

[44] Y. Wu, T. Fu, C. Zhu, X. Wang, Y. Ma, H.Z. Li, Shear-induced tail breakup of droplets (bubbles) flowing in a straight microfluidic channel, Chem. Eng. Sci. 135 (2015) 61–66.

[45] S. Afkhami, A. Leshansky, Y. Renardy, Numerical investigation of elongated drops in a microfluidic T-junction, Phys. Fluids 23 (2) (2011).

[46] R. Seemann, M. Brinkmann, T. Pfohl, S. Herminghaus, Droplet based microfluidics, Rep. Progr. Phys. 75 (1) (2011) 016601.

[47] S. Kwon, Y. Lee, Explainability-based mix-up approach for text data augmentation, ACM Trans. Knowl. Discov. Data 17 (1) (2023) 1–14.

[48] B.H. Misheva, J. Osterrieder, A. Hirsa, O. Kulkarni, S.F. Lin, Explainable AI in credit risk management, 2021, arXiv preprint arXiv:2103.00949.

[49] A. Salih, Z. Raisi-Estabragh, I.B. Galazzo, P. Radeva, S.E. Petersen, G. Menegaz, K. Lekadir, Commentary on explainable artificial intelligence methods: SHAP and LIME, 2023, arXiv preprint arXiv:2305.02012.

[50] C.I. Nwakanma, L.A.C. Ahakonye, J.N. Njoku, J.C. Odirichukwu, S.A. Okolie, C. Uzondu, C.C. Ndubuisi Nweke, D.-S. Kim, Explainable artificial intelligence (XAI) for intrusion detection and mitigation in intelligent connected vehicles: A review, Appl. Sci. 13 (3) (2023) 1252.

[51] P. Kim, K.W. Kwon, M.C. Park, S.H. Lee, S.M. Kim, K.Y. Suh, Soft lithography for microfluidics: a review, Biochip J. 2 (1) (2008) 1–11.

[52] N.M. Kovalchuk, E. Roumpea, E. Nowak, M. Chinaud, P. Angeli, M.J. Simmons, Effect of surfactant on emulsification in microchannels, Chem. Eng. Sci. 176 (2018) 139–152.

[53] H. Yi, C. Zhu, T. Fu, Y. Ma, Efficient coalescence of microdroplet in the cross-focused microchannel with symmetrical chamber, J. Taiwan Inst. Chem. Eng. 112 (2020) 52–59.

[54] P.M. Korczyk, O. Cybulski, S. Makulska, P. Garstecki, Effects of unsteadiness of the rates of flow on the dynamics of formation of droplets in microfluidic systems, Lab Chip 11 (1) (2011) 173–175.

[55] W. Zeng, I. Jacobi, D.J. Beck, S. Li, H.A. Stone, Characterization of syringe-pump-driven induced pressure fluctuations in elastic microchannels, Lab Chip 15 (4) (2015) 1110–1115.

[56] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nature Methods 9 (7) (2012) 671–675.

[57] M.N. Fienen, N.G. Plant, A cross-validation package driving netica with python, Environ. Model. Softw. 63 (2015) 14–23.

[58] D. Berrar, et al., Cross-validation, 2019.

[59] Q. Han, Z. Hao, T. Hu, F. Chu, Non-parametric models for joint probabilistic distributions of wind speed and direction data, Renew. Energy 126 (2018) 1032–1042.

[60] P. Carvalho, S. da Silva, E. Duarte, R. Brossier, G. Corso, J. de Araújo, Full waveform inversion based on the non-parametric estimate of the probability distribution of the residuals, Geophys. J. Int. 229 (1) (2022) 35–55.

[61] D. Jansson, O. Rosén, A. Medvedev, Non-parametric analysis of eye-tracking data by anomaly detection, in: 2013 European Control Conference, ECC, IEEE, 2013, pp. 632–637.

[62] D. Singh, B. Singh, Feature wise normalization: An effective way of normalizing data, Pattern Recognit. 122 (2022) 108307.

[63] R. Natras, B. Soja, M. Schmidt, Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting, Remote Sens. 14 (15) (2022) 3547.

[64] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, N. Villa-Vialaneix, Random forests for big data, Big Data Res. 9 (2017) 28–46.

[65] D. Elavarasan, D.R. Vincent, Reinforced XGBoost machine learning model for sustainable intelligent agrarian applications, J. Intell. Fuzzy Systems 39 (5) (2020) 7605–7620.

[66] M. Shahhosseini, R.A. Martinez-Feria, G. Hu, S.V. Archontoulis, Maize yield and nitrate loss prediction with machine learning algorithms, Environ. Res. Lett. 14 (12) (2019) 124026.

[67] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on tabular data? 2022, arXiv:2207.08815.

[68] P. Mamoshina, A. Vieira, E. Putin, A. Zhavoronkov, Applications of deep learning in biomedicine, Mol. Pharm. 13 (5) (2016) 1445–1454.

[69] Z. Wang, X. Liu, Y. Huang, P. Zhang, Y. Fu, A multivariate time series graph neural network for district heat load forecasting, Energy (2023) 127911.

[70] V. Sze, Y.-H. Chen, T.-J. Yang, J.S. Emer, Efficient processing of deep neural networks: A tutorial and survey, Proc. IEEE 105 (12) (2017) 2295–2329, http://dx.doi.org/10.1109/JPROC.2017.2761740.

[71] M. Xue, H. Wu, R. Li, DNN migration in IoTs: Emerging technologies, current challenges and open research directions, IEEE Consum. Electron. Mag. (2022).

[72] P. Liashchynskyi, P. Liashchynskyi, Grid search, random search, genetic algorithm: a big comparison for NAS, 2019, arXiv preprint arXiv:1912.06059.

[73] K. Lin, Y. Gao, Model interpretability of financial fraud detection by group SHAP, Expert Syst. Appl. 210 (2022) 118354.

[74] M.L. Martini, S.N. Neifert, E.K. Oermann, J.T. Gilligan, R.J. Rothrock, F.J. Yuk, J.S. Gal, D.A. Nistal, J.M. Caridi, Application of cooperative game theory principles to interpret machine learning models of nonhome discharge following spine surgery, Spine 46 (12) (2021) 803–812.

[75] M.V. García, J.L. Aznarte, Shapley additive explanations for NO2 forecasting, Ecol. Inform. 56 (2020) 101039.

[76] R. Stirnberg, J. Cermak, S. Kotthaus, M. Haeffelin, H. Andersen, J. Fuchs, M. Kim, J.-E. Petit, O. Favez, Meteorology-driven variability of air pollution (PM 1) revealed with explainable machine learning, Atmos. Chem. Phys. 21 (5) (2021) 3919–3948.

[77] J.-J. Liu, J.-C. Liu, Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization, Geofluids 2022 (2022) 1–15.

[78] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[79] X. Zhou, H. Wen, Z. Li, H. Zhang, W. Zhang, An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost, Geocarto Int. 37 (26) (2022) 13419–13450.

[80] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?" Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[81] J.A. Recio-García, B. Díaz-Agudo, V. Pino-Castilla, CBR-LIME: a case-based reasoning approach to provide specific local interpretable model-agnostic explanations, in: Case-Based Reasoning Research and Development: 28th International Conference, ICCBR 2020, Salamanca, Spain, June 8–12, 2020, Proceedings 28, Springer, 2020, pp. 179–194.

[82] R. Choudhary, H.K. Gianey, Comprehensive review on supervised machine learning algorithms, in: 2017 International Conference on Machine Learning and Data Science, MLDS, IEEE, 2017, pp. 37–43.

[83] E. Mosca, F. Szigeti, S. Tragianni, D. Gallagher, G. Groh, SHAP-based explanation methods: a review for NLP interpretability, in: Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 4593–4603.

[84] N. Pat, Y. Wang, A. Bartonicek, J. Candia, A. Stringaris, Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition, Cerebral Cortex 33 (6) (2023) 2682–2703.

[85] M. Robnik-Šikonja, M. Bohanec, Perturbation-based explanations of prediction models, in: Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent, Springer, 2018, pp. 159–175.

[86] J. Dieber, S. Kirrane, Why model why? Assessing the strengths and limitations of LIME, 2020, arXiv preprint arXiv:2012.00093.

[87] A. Vij, P. Nanjundan, Comparing strategies for post-hoc explanations in machine learning models, in: Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2021, Springer, 2022, pp. 585–592.

[88] I. Covert, S.M. Lundberg, S.-I. Lee, Understanding global feature contributions with additive importance measures, Adv. Neural Inf. Process. Syst. 33 (2020) 17212–17223.

[89] H. Prabhu, A. Sane, R. Dhadwal, N.R. Parlikkad, J.K. Valadi, Interpretation of drop size predictions from a random forest model using local interpretable model-agnostic explanations (LIME) in a rotating disc contactor, Ind. Eng. Chem. Res. (2023).