Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project
Project Plan

Explainable AI: understand the black box of predictive models with chemical engineering applications

by

Jinwei Hu

Email: jinwei.hu22@imperial.ac.uk
GitHub username: acse-jh4322
Repository: https://github.com/ese-msc-2022/irp-jh4322

Supervisors:

Dr. Sibo Cheng
Dr. Rossella Arcucci

June 2023

Abstract

This plan presents a preliminary study on microfluidic coalescence of aqueous drops on oil, focusing on data-driven predictive modeling. Initial processing of raw tabular data employs Min-Max Normalisation, ensuring scale uniformity. Three machine learning models, Random Forest, XGBoost, and Deep Neural Network, are developed and optimized via Grid Search. Utilizing LIME and SHAP, an in-depth analysis elucidates key features influencing predictions. Subsequently, Supervised Contrastive Learning (SCL) is expected to be applied for discerning analysis between tabular and video data, fostering the development of an optimized model. Flow rate emerges as the pivotal feature, with D/W 1 and dt, ms also holding significance. This study contributes to the understanding of microfluidic coalescence mechanisms and augments predictive modeling capabilities in chemical engineering.

1 Introduction

1.1 Literature Review

Microfluidic technologies have engendered a significant paradigm shift in the manipulation and analysis of diminutive volumes of fluids. Such a shift has had a profound impact across many fields, including chemistry, biology, and materials science [9]. Nested within this wide-ranging scope, one pivotal process that stands out is the coalescence of aqueous droplets in an oil. This intricate process, which involves the application of an electric field to prompt the merging of droplets, has been studied in depth by researchers like Leary et al. [15]. Demonstrating its versatility, coalescence is utilised in a plethora of applications, from petroleum refining where it aids in separating water from crude oil to droplet-based microfluidics, in which it facilitates the combination of water droplets in oil, thus breaking emulsions [15, 8, 14]. Beyond these applications, the coalescence processes also hold a cardinal significance in the formulation of emulsions, playing a vital role in industries such as food processing, oil and gas, and separation processes [4]. Pivoting towards the field of digital microfluidics, the precise manipulation of discrete droplets through electrocoalescence serves to propel its advancement further. In turn, this enables the development of intricate lab-on-a-chip technologies, thereby contributing to both the simplification and enhancement of laboratory procedures [36].

The capacity to accurately control droplet coalescence is not merely a convenience but a critical necessity, particularly in the fabrication of complex materials such as Janus particles or core-shell particles [28]. These materials find applications in multiple areas like targeted drug delivery and catalysis [10, 16]. Furthermore, it is worth mentioning that the dynamics of coalescence can change under various experimental conditions and it is hard to keep stable. For instance, elevated temperatures can cause an increased frequency of coalescence, as highlighted in the insightful work of Bera et al [4]. Consequently, through the research and studies performed in this field, it becomes clear that the understanding and control of droplet coalescence in microfluidics persist as an area of significant interest. This is primarily due to its wide-ranging applications that span across multiple industries, attesting to its universal appeal and relevance [9, 4].

Historically, the scrutiny of parameters influencing drop coalescence was largely reliant on trial-and-error methodologies or analytical models [31, 23]. However, the convoluted interplay between factors such as interfacial tension, droplet size, viscosity, and flow rates renders the prediction and understanding of drop coalescence dynamics highly non-linear and difficult[1]. Serving as the antithesis to conventional methodologies, Machine Learning (ML) has burgeoned into a formidable instrument for modelling intricate systems and proffering predictions for dimensional data [13]. Owing to their ability to assimilate information from data, machine learning

algorithms excel in capturing sophisticated relationships that traditional methodologies are impotent to attain [5]. Furthermore, these algorithms are being increasingly enlisted in the domain of chemical engineering, providing robust support for solving some critical problems. The ambit 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 in chemical processes [6, 29, 32]. Consequently, there is cogent reason to postulate that, by employing the initial conditions of droplet microfluidic coalescence as input, machine learning algorithms possess the capability to furnish more precise predictions and search a more conducive set of experimental parameters, thereby contributing to the enhancement of experimental accuracy.

1.2 Problem Discription

The advent of ML has brought a plethora of benefits to the area of chemical engineering. However, a prevailing concern surrounding the utilization of ML models, especially the more complex ones like deep neural networks, is the inherent opacity in their internal workings, often referred to as ”black boxes” [3]. While these models are remarkably adept at making predictions, they are notably reticent in providing insights into the underlying processes or the reasoning behind a specific prediction [2]. This lack of transparency and interpretability can act as a significant barrier in gaining the trust and adoption of such models among researchers and the general citizens [33]. Such apprehensions underscore the need for developing more interpretable ML models that do not sacrifice the predictive power but enhance the understanding and confidence in their applications in chemical engineering.

1.3 Objectives

To address this problem, the focus has been shifting towards the realm of Explainable Artificial Intelligence (XAI). This study employs a two-pronged strategy, incorporating explainability within ML models for droplet microfluidic coalescence. The initial phase entails the construction of an array of ML predictive models through the optimisation of disparate parameters, with a view to ascertaining whether droplets coalesce under varying experimental settings. Subsequently, the second phase underscores explainability by meticulously scrutinising model outcomes and parameters to furnish researchers with a lucid comprehension of the attributes or parameters that are more conducive to droplet coalescence, thereby augmenting the resource utilisation in experiments and mitigating redundant operations. Through marrying explainability with ML models, this project aspires not only to realise precise predictions but also to offer reasonable insights to researchers and practitioners. These insights harbour the potential to not only deepen our understanding of latent phenomena within microfluidic systems but also to foster innovation. Moreover, through the integration of XAI, the credibility and trustworthiness of ML models are bolstered [22]. This enhancement, in turn, renders them indispensable assets, catalysing the advancement of research and applications in the domain of droplet microfluidic coalescence.

2 Progress to Date

This section succinctly delineates the work that has been finished in the first phase. The prerequisite knowledge pertaining to the respective method is explained in the Appendices. A.

2.1 Tabular Data preprocessing

In the ongoing Microfluidic Coalescence project, we have meticulously processed the raw tabular data pertaining to the coalescence of aqueous droplets in oil. Flow rates have been computed and min-max normalisation has been applied to rescale the data to a range $[0, 1]$. This normalization serves to mitigate scale discrepancies among different features, whilst judiciously preserving the relative relationships amongst sample points [27]. The rules followed by equation. 1:

$$x_{\text{scaled}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{1}$$

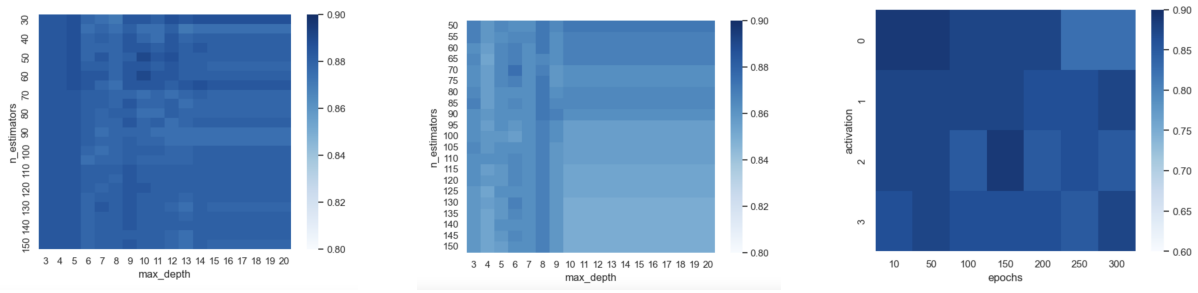
where x is the feature value of each sample in the dataset, x_{max} and x_{min} are the maximum and minimum value of the feature x.

2.2 Model Selection and Training

Upon the completion of data preprocessing, three machine learning models are constructed, namely Random Forest, XGBoost, and a Deep Neural Network. The dataset is partitioned into training, testing, and validation subsets by employing a random seed to ensure unbiased randomisation. During the training phase, we have opted to employ the Grid Search methodology due to its strong capability in exhaustively searching through a specified subset of the hyperparameter space. This method has proven efficacious in identifying the most favourable configuration for model parameters. Subsequent to the training process, all three models, now calibrated with optimal parameters, exhibit an accuracy exceeding 80% on the test dataset. Notably, the Deep Neural Network stands preeminent among the models with an accuracy of 86.14%. The performance matrices and validation heatmaps for the tuning hyperparameters are shown in the Table. 1 and Figure. 1.

	Accuracy(%)	**Precision(%)**	**Recall(%)**	**F1-Score(%)**
Random Forest	81.19	81	85	83
XGBoost	85.34	85	87	86
Neural Network	**86.14**	**86**	**89**	**87**

Table 1: Performance matrices of three ML models



Random Forest XGBoost Neural Networks

Figure 1: Validation heatmaps for the tuning hyperparameters of the three predictive models.

2.3 XAI Analysis of Predictive Results

To elucidate the underpinnings of the predictions rendered by the models, we have undertaken an incisive analysis employing LIME (Locally Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) values. Our analysis reveals that the flow rate commands the most significant impact on predictions, with D/W 1 and dt, ms holding relative importance. Furthermore, LIME proves instrumental in shedding light on the model's reasoning on an individual sample basis, it can explain why a sample can be predicted as coalescence in a clear manner. Comparablely, SHAP affords a comprehensive perspective of the influence exerted by different attributes. In essence, this rigorous analytical approach has furnished invaluable insights into the internal machinations of the machine learning models, thereby bolstering the confidence of the scholarly community in the integrity and credibility of the models' inferences.The details of analysis is shown in Figure. 2, Figure. 3 and Appendice. C.

Random Forest XGBoost Neural Networks

Figure 2: SHAP analysis of diverse ML models

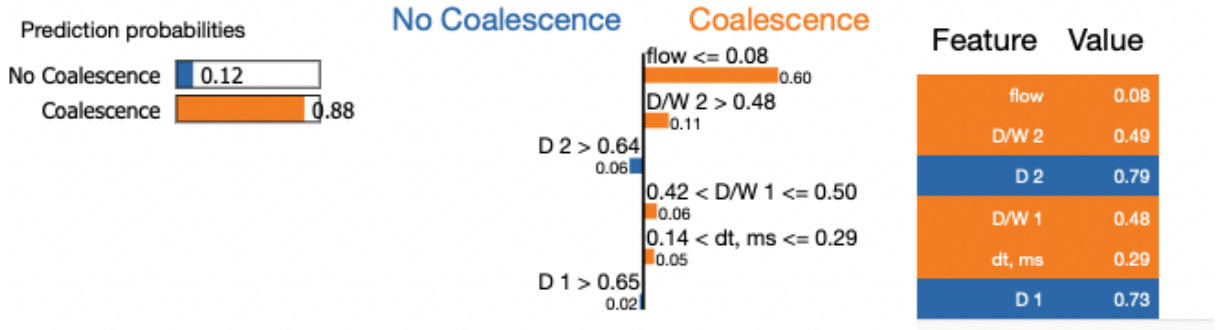


Figure 3: LIME analysis of neural network for a coalescence sample

3 Future Plan

Figure. 4 delineates the timelines and progression of this Microfluidic Coalescence Analysis project. In the initial phase, a foundational groundwork is meticulously established, employing SHAP (SHapley Additive exPlanations) and LIME (Locally Interpretable Model-agnostic Explanations) to rigorously analyse the multifarious models' outcomes pertaining to microfluidic coalescence.

Subsequently, the ensuing phase is characterised by a concerted focus on the adroit utilisation of Supervised Contrastive Learning (SCL). This advanced technique is employed to assiduously compare and contrast the characteristics of tabular and video data, thereby unearthing nuanced similarities and disparities. Through this extending analysis, an optimised model, exhibiting heightened precision and efficacy, is expected to be formulated.

Supervised Contrastive Learning techniques are currently being learned and deployed (June 17 - August 1) for the concurrent processing of video and tabular data. The development

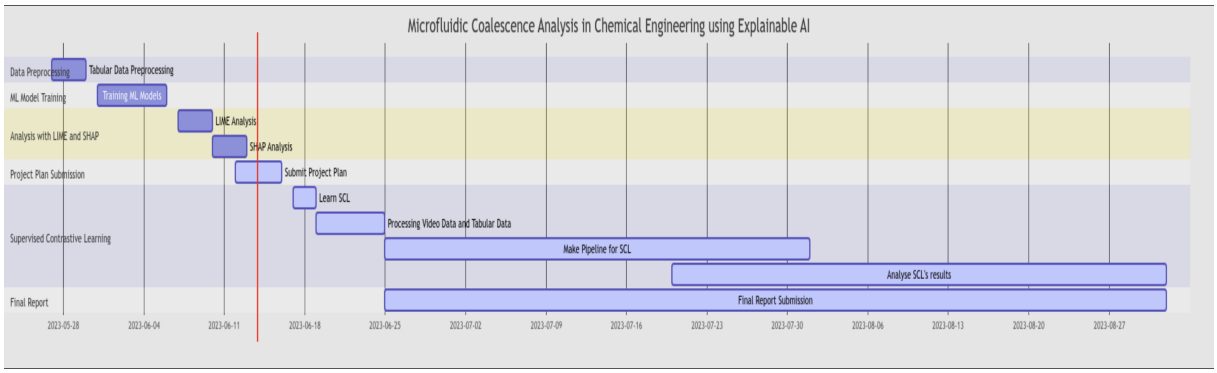


Figure 4: Gantt Chart illustrating the timeline and progress of the Microfluidic Coalescence Analysis project in Chemical Engineering using Explainable AI

of an pipeline tailored for SCL is ongoing, in conjunction with a systematic analysis of the results obtained through SCL (July 20 - September 1). The final report, encapsulating the comprehensive analysis and derived conclusions, is slated for submission on September 1, 2023.

References

[1] Ilke Akartuna, Donald M Aubrecht, Thomas E Kodger, and David A Weitz. Chemically induced coalescence in droplet-based microfluidics. *Lab on a Chip*, 15(4):1140–1144, 2015.

[2] Anna Markella Antoniadi, Yuhan Du, Yasmine Guendouz, Lan Wei, Claudia Mazo, Brett A Becker, and Catherine Mooney. Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review. *Applied Sciences*, 11(11):5088, 2021.

[3] Christina B Azodi, Jiliang Tang, and Shin-Han Shiu. Opening the black box: interpretable machine learning for geneticists. *Trends in genetics*, 36(6):442–455, 2020.

[4] Bijoy Bera, Rama Khazal, and Karin Schroën. Coalescence dynamics in oil-in-water emulsions at elevated temperatures. *Scientific reports*, 11(1):10990, 2021.

[5] Renze Dong, Hongze Leng, Juan Zhao, Junqiang Song, and Shutian Liang. A framework for four-dimensional variational data assimilation based on machine learning. *Entropy*, 24(2):264, 2022.

[6] Pavlo O Dral. Quantum chemistry in the age of machine learning. *The journal of physical chemistry letters*, 11(6):2336–2347, 2020.

[7] Dhivya Elavarasan and Durai Raj Vincent. Reinforced xgboost machine learning model for sustainable intelligent agrarian applications. *Journal of Intelligent & Fuzzy Systems*, 39(5):7605–7620, 2020.

[8] John S Eow and Mojtaba Ghadiri. Electrostatic enhancement of coalescence of water droplets in oil: a review of the technology. *Chemical Engineering Journal*, 85(2-3):357–368, 2002.

[9] Edgar A Galan, Haoran Zhao, Xukang Wang, Qionghai Dai, Wilhelm TS Huck, and Shaohua Ma. Intelligent microfluidics: The convergence of machine learning and microfluidics in materials science and biomedicine. *Matter*, 3(6):1893–1922, 2020.

[10] Unnati Garg, Swati Chauhan, Upendra Nagaich, and Neha Jain. Current advances in chitosan nanoparticles based drug delivery and targeting. *Advanced pharmaceutical bulletin*, 9(2):195, 2019.

[11] Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix. Random forests for big data. *Big Data Research*, 9:28–46, 2017.

[12] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on tabular data?, 2022.

[13] Zurki Ibrahim, Pinar Tulay, and Jazuli Abdullahi. Multi-region machine learning-based novel ensemble approaches for predicting covid-19 pandemic in africa. *Environmental Science and Pollution Research*, 30(2):3621–3643, 2023.

[14] Prashant S Kulkarni, Shagufta U Patel, and George G Chase. Layered hydrophilic/hydrophobic fiber media for water-in-oil coalescence. *Separation and purification technology*, 85:157–164, 2012.

[15] Thomas Leary, Mohsen Yeganeh, and Charles Maldarelli. Microfluidic study of the electrocoalescence of aqueous droplets in crude oil. *ACS omega*, 5(13):7348–7360, 2020.

[16] Tao Li, Junjun Wang, Fenglong Wang, Lishu Zhang, Yanyan Jiang, Hamidreza Arandiyan, and Hui Li. The effect of surface wettability and coalescence dynamics in catalytic performance and catalyst preparation: a review. *ChemCatChem*, 11(6):1576–1586, 2019.

[17] Petro Liashchynskyi and Pavlo Liashchynskyi. Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*, 2019.

[18] Kang Lin and Yuzhuo Gao. Model interpretability of financial fraud detection by group shap. *Expert Systems with Applications*, 210:118354, 2022.

[19] Polina Mamoshina, Armando Vieira, Evgeny Putin, and Alex Zhavoronkov. Applications of deep learning in biomedicine. *Molecular pharmaceutics*, 13(5):1445–1454, 2016.

[20] Michael L Martini, Sean N Neifert, Eric K Oermann, Jeffrey T Gilligan, Robert J Rothrock, Frank J Yuk, Jonathan S Gal, Dominic A Nistal, and John M Caridi. Application of cooperative game theory principles to interpret machine learning models of nonhome discharge following spine surgery. *Spine*, 46(12):803–812, 2021.

[21] Randa Natras, Benedikt Soja, and Michael Schmidt. Ensemble machine learning of random forest, adaboost and xgboost for vertical total electron content forecasting. *Remote Sensing*, 14(15):3547, 2022.

[22] Cosmas Ifeanyi Nwakanma, Love Allen Chijioke Ahakonye, Judith Nkechinyere Njoku, Jacinta Chioma Odirichukwu, Stanley Adiele Okolie, Chinebuli Uzondu, Christiana Chidimma Ndubuisi Nweke, and Dong-Seong Kim. Explainable artificial intelligence (xai) for intrusion detection and mitigation in intelligent connected vehicles: A review. *Applied Sciences*, 13(3):1252, 2023.

[23] Sadegh Poozesh and Ecevit Bilgili. Scale-up of pharmaceutical spray drying using scale-up rules: A review. *International Journal of Pharmaceutics*, 562:271–292, 2019.

[24] Juan A Recio-García, Belén Díaz-Agudo, and Victor 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*, pages 179–194. Springer, 2020.

[25] Marco Tulio Ribeiro, Sameer Singh, and Carlos 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*, pages 1135–1144, 2016.

[26] Mohsen Shahhosseini, Rafael A Martinez-Feria, Guiping Hu, and Sotirios V Archontoulis. Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*, 14(12):124026, 2019.

[27] Dalwinder Singh and Birmohan Singh. Feature wise normalization: An effective way of normalizing data. *Pattern Recognition*, 122:108307, 2022.

[28] Fotis Spyropoulos, David M Lloyd, Robin D Hancocks, and Aleksandra K Pawlik. Advances in membrane emulsification. part a: recent developments in processing aspects and microstructural design approaches. *Journal of the Science of Food and Agriculture*, 94(4):613–627, 2014.

[29] Sina Stocker, Gábor Csányi, Karsten Reuter, and Johannes T Margraf. Machine learning in chemical reaction space. *Nature communications*, 11(1):5505, 2020.

[30] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S. Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

[31] Priyanka Talukdar and Arindam Dey. Hydraulic failures of earthen dams and embankments. *Innovative Infrastructure Solutions*, 4:1–20, 2019.

[32] Venkat Venkatasubramanian. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2):466–478, 2019.

[33] ES Vorm and David JY Combs. Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (istam). *International Journal of Human–Computer Interaction*, 38(18-20):1828–1845, 2022.

[34] Zhijin Wang, Xiufeng Liu, Yaohui Huang, Peisong Zhang, and Yonggang Fu. A multivariate time series graph neural network for district heat load forecasting. *Energy*, page 127911, 2023.

[35] Min Xue, Huaming Wu, and Ruidong Li. Dnn migration in iots: Emerging technologies, current challenges and open research directions. *IEEE Consumer Electronics Magazine*, 2022.

[36] Michele Zagnoni, Charles N. Baroud, and Jonathan M. Cooper. Electrically initiated upstream coalescence cascade of droplets in a microfluidic flow. *Phys. Rev. E*, 80:046303, Oct 2009.

Appendices

A Preliminaries

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 specifically encompass two kinds of tree-based models, deep neural network, hyper-parameter space search methods, along with LIME and SHAP explanations.

A.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 [21]. Random Forest creates an series of decision trees, each constructed independently through the utilization of bootstrapped samples from the dataset in a concurrent process [11]. Conversely, XGBoost constructs trees in a sequential manner, whereby each successive tree seeks to ameliorate the errors perpetrated by its predecessor [7]. 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 [26]. Moreover, due to these two tree-based methods always show 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 [12].

A.2 Deep Neural Networks

Deep Neural Networks (DNNs) are composed of numerous layers of artificial neurons. The depth of these networks enables them to discern highly intricate patterns and relationships within data [19]. DNNs are distinguished by their adeptness at modeling non-linear relationships, which can be attributed to the activation function, and their large parameter space, rendering them computationally demanding [34]. DNNs are widely used in natural language processing such as machine translation and sentiment analysis, image and speech recognition, anomaly detection, and predictive analysis based on various kinds of data [30, 35].

A.3 Grid Search Method

The grid search is used to probe a manually specified subset of the hyperparameter space of the learning algorithm [17]. It can carefully traverse multiple combinations of hyperparameter configurations, performing cross-validation to determine the configuration that yields the best performance.

A.4 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 combinations. This method is based on cooperative game theory and is utilized for interpreting the output of machine learning model [18, 20].

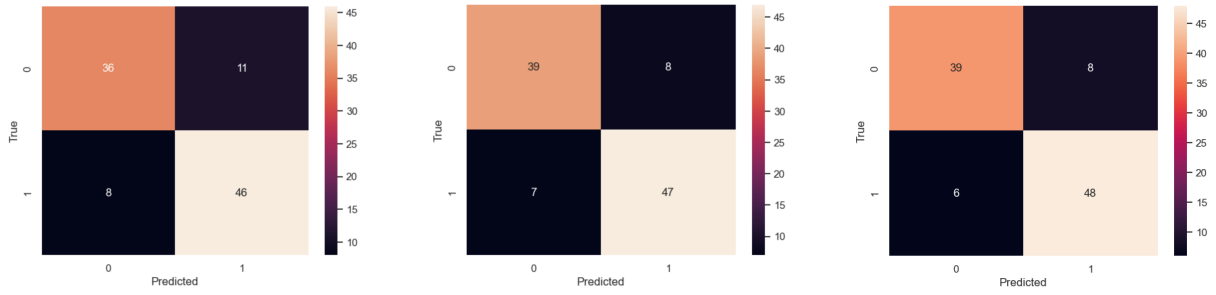
A.5 Local Interpretable Model-agnostic Explanations(LIME)

LIME is an explanatory method designed to clarify the predictions provided by any classifier or regressor in an understandable and faithful way [25]. 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 [24].

B Performance Metrics of ML models

B.1 Confusion Metrics

This section show the confusion metrics of three ML models.



Random Forest XGBoost Neural Networks

Figure 5: Confusion metrics of diverse ML models

C Analysis among three models

C.1 Overall Observation

1. Flow appears to be the most significant attribute in all three models.

2. The colors occupying 50% of the rectangles indicate that all features contribute equally to both classes, whether it is coalescence (label=1) or not (label=0).

3. SHAP rankings for the Random Forest (RF), XGBoost, and MLP models are as follows:

- $RF : Flow > D/W1 > D/W2 > D2 > dt, ms > D1$
- $XGBoost : Flow > dt, ms > D/W2 > D/W1 > D1 > D2$
- $MLP : Flow > dt, ms > D/W2 > D/W1 > D2 > D1$

4. The images of XGBoost and RF show more similarity to each other compared to MLP, indicating some consistency between the two tree-based models. However, MLP demonstrates better discriminatory ability overall. Interestingly, the feature ranking of XGBoost is similar to that of MLP, suggesting some agreement between the two models in identifying important features.

5. According to the final overall trend, higher values of flow, dt, ms, and D 2 have a negative impact on the classification of instances as coalescence. Conversely, higher values of D/W 1, D/W 2, and D 1 have a positive impact on the classification of instances as coalescence.

6. In terms of LIME's results, the image can show a good understanding that why the model make the final decision. For the example instance, Flow and D/W2 provide a strong support in comparison to other four attributes.

C.2 Reasons

1. Feature Importance: Across all models, "Flow" is identified as the most critical feature. This suggests that in predicting whether coalescence occurs or not, "Flow" bears the most substantial influence. However, the importance ranking of the remaining features differs across models, a variation that could be attributed to distinct learning mechanisms inherent in each model.

2. Contribution of Features: All features appear to contribute equally to predicting both classes (coalescence = 1 or non-coalescence = 0), indicated by the colors occupying 50% of the rectangles. This observation might suggest that within this specific dataset, no single feature completely dictates the prediction outcome; instead, a combination and interaction of all features should be considered.

3. Distribution of SHAP values: MLP delineates the influence of SHAP values more distinctly compared to the other two models. This observation might indicate that the MLP model is more proficient in segregating and distinguishing the contributions of individual features.

4. Comparison between Models: The results from RF and XGBoost exhibit more similarity to each other compared to MLP. This similarity can be attributed to RF and XGBoost both being tree-based models, while MLP is a neural network-based model that may employ different learning and prediction strategies. Additionally, MLP achieves the highest test accuracy of 86%, indicating its superior performance in accurately predicting the outcome.

5. LIME Results: LIME is a tool employed for explaining individual prediction results. For a given instance, "Flow" and "D/W2" provide stronger support for the final decision compared to other features. This further corroborates the critical role of "Flow" and "D/W2" in this context.