भारतीय प्रौद्योगिकी संस्थान जम्मू
INDIAN INSTITUTE OF TECHNOLOGY JAMMU

Indian Institute of Technology Jammu

MINI PROJECT

ME-2-20(PO)

ML FOR BUBBLE GROWTH

Supervisor

Prof. Pothukuchi Harish

Manas Gupta(2022UEE0137)
Borisagar Meet(2022UCS0087)
Muthres Gurjar(2022UCS0097)
Vibhav Bhagat(2022UEE0155)

Introduction :

The study of bubble growth dynamics in flowing liquid systems plays a pivotal role in various industrial and scientific applications, ranging from chemical engineering processes to medical diagnostics. Understanding the behavior of bubbles in flowing liquids is crucial for optimizing processes such as fermentation, wastewater treatment, and multiphase flow in pipelines. In this mini-project, we explore the application of machine learning (ML) techniques to model and predict the dynamics of bubble growth in flowing liquids. By harnessing the power of ML algorithms, we aim to develop predictive models that can enhance our understanding of bubble dynamics and contribute to the advancement of multiphase flow research.

Objective :

The primary objective of this mini-project is to collect experimental data and develop a machine learning (ML) model that incorporates physical features such as pressure, mass flux, heat flux, subcooling, and other relevant parameters. By harnessing a diverse set of physical features, we aim to construct a comprehensive and robust ML model capable of accurately predicting bubble growth dynamics in flowing liquids.

Data Collection Approach :

Our data collection process involved two distinct approaches, each designed to ensure the acquisition of a comprehensive dataset suitable for training our machine learning (ML) model on bubble growth dynamics in flowing liquids.

Literature-Informed Data Extraction:

Initially, we conducted an extensive review of existing literature to identify relevant studies on bubble growth dynamics in flowing liquids. We scrutinized research papers[4], journals, and academic publications[5] to gather insights into experimental setups, methodologies, and results.In our first approach, we attempted to extract data directly from graphical representations presented in the literature. While this method provided a starting point, we encountered limitations due to the sparse nature of data points and insufficient coverage of the desired parameters. The extracted data points were not deemed sufficient to train a robust ML model for predicting bubble growth dynamics.

Relation-Based Data Generation:

Recognizing the limitations of the first approach, we adopted a more innovative strategy for data collection. Leveraging the information gleaned from the literature regarding the relationships and correlations between various physical parameters and bubble growth dynamics, we devised a systematic approach to generate synthetic data points.In this second approach, we meticulously analyzed the mathematical relations and empirical models proposed in the literature for describing bubble growth phenomena. By extrapolating these relations and considering the experimental conditions provided in the respective papers, we synthesized additional data points to augment our dataset.To ensure the integrity and diversity of the generated data, we employed multiple formulae and equations derived from theoretical frameworks and empirical observations. This iterative process enabled us to capture a wide range of scenarios and conditions while mitigating the risk of overfitting in our ML model.

By integrating these two approaches, we aimed to assemble a comprehensive dataset encompassing a diverse array of physical features and bubble growth dynamics. This hybrid data collection strategy not only enriched our dataset but also facilitated the development of a robust ML model capable of accurately predicting bubble behavior in flowing liquids.

Formule Used for Relation-Based Data Generation :

$$R_b = Ja\sqrt{\frac{12}{\pi}}$$

Relation 1 [1]

$$\frac{D_b}{D_{bm}} = 1 - 2^K \left| \frac{1}{2} - \left(\frac{t}{t_b}\right)^N \right|^K$$

Relation 2[2]

$$a(t) = \frac{2}{3}\frac{B^2}{A}\left[(t^+ + 1)^{\frac{3}{2}} - (t^+)^{\frac{3}{2}} - 1\right]$$

Relation 3[3]

Dataset :

	pressure(bar)	heat flux(kW/m2)	mass fluxkg/(m2·s)	sub cooling	channel dia(mm)	d/dMax	t/tMax
0	1.11	173.000	495.0	6.5	13.33	0.154977	0.004673
1	1.11	173.000	495.0	6.5	13.33	0.215351	0.009346
2	1.11	173.000	495.0	6.5	13.33	0.260160	0.014019
3	1.11	173.000	495.0	6.5	13.33	0.296914	0.018692
4	1.11	173.000	495.0	6.5	13.33	0.328518	0.023364
...	...	...	...	...	...	...	...
16609	1.00	0.045	30.0	24.0	3.75	0.916238	0.824840
16610	1.00	0.045	30.0	24.0	3.75	0.939426	0.880710
16611	1.00	0.045	30.0	24.0	3.75	0.953692	0.913930
16612	1.00	0.045	30.0	24.0	3.75	0.992806	0.960740
16613	1.00	0.045	30.0	24.0	3.75	0.545288	1.000000

Fig. 4 dataset (link: Bubble Growth)

Model Selection and Training Approach:

Exploration of Regression Models:

We commenced our model selection process by exploring a range of regression algorithms, including linear regression, polynomial regression, decision trees, random forest, and gradient boosting.
Each model was trained and evaluated using the Literature-Informed Data obtained from our initial data extraction approach.
While these models demonstrated varying degrees of predictive performance, none were able to sufficiently capture the complexity of bubble growth dynamics in flowing liquids.

Selection of XGBoost:

After thorough experimentation, we identified XGBoost (Extreme Gradient Boosting) as the most promising candidate due to its superior performance in terms of predictive accuracy and robustness.
XGBoost, a scalable and efficient implementation of gradient boosting, excelled in handling complex relationships and non-linearities present in our dataset.
Its ability to handle missing data, feature importance analysis, and parallel computation further solidified its suitability for our task.

Hybrid Training Approach:

To further enhance the performance of our XGBoost model, we adopted a hybrid training approach.
Initially, we trained the XGBoost model solely on the Literature-Informed Data, leveraging the insights obtained from our initial literature review.
Subsequently, we integrated the Relation-Based Data generated through mathematical relations from the literature into our dataset.
By combining both types of data, we aimed to enrich the model's understanding of bubble growth dynamics and improve its generalization capabilities.

Why XGBoost was Chosen:

Performance: XGBoost consistently outperformed other regression models in terms of predictive accuracy and generalization on unseen data.

Robustness: XGBoost demonstrated robustness against overfitting and noise in the dataset, thanks to its regularization techniques.

Feature Importance: XGBoost provided valuable insights into feature importance, allowing us to identify the most influential factors contributing to bubble growth dynamics.
Scalability: XGBoost's scalability and efficiency made it well-suited for handling large datasets and complex models.

Challenges with Other Models:

Linear Models: Linear regression and polynomial regression struggled to capture the non-linear relationships inherent in our dataset, resulting in suboptimal performance.

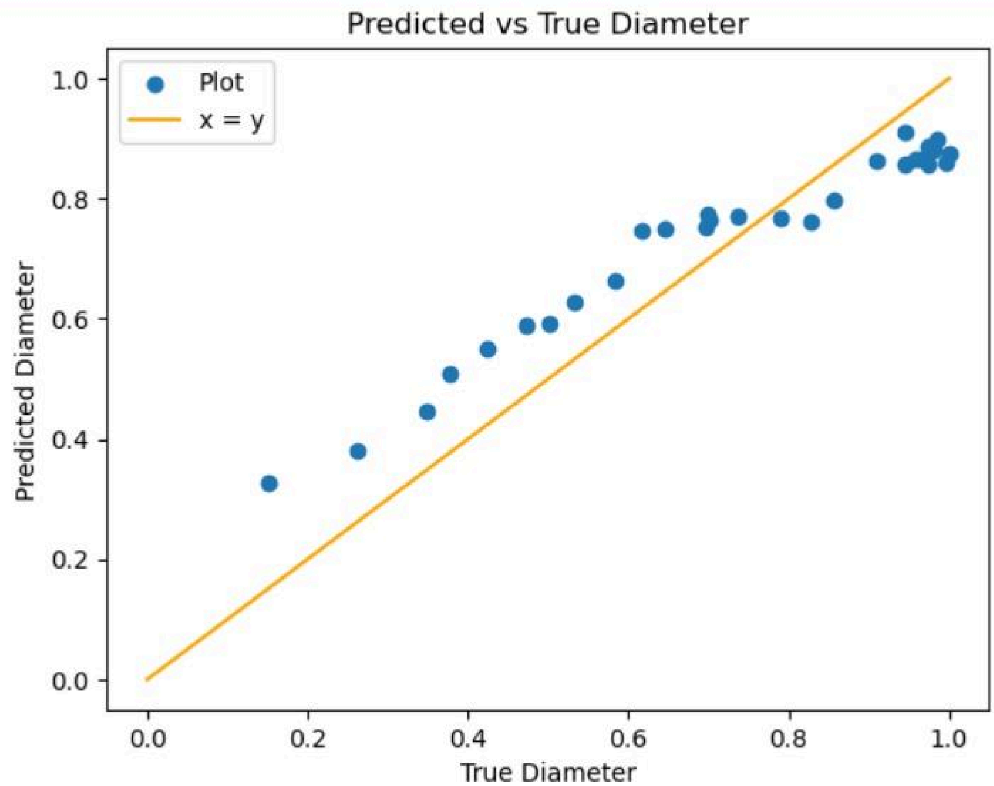
Decision Trees: While decision trees and random forest showed promise initially, they exhibited limitations in handling high-dimensional data and suffered from overfitting.

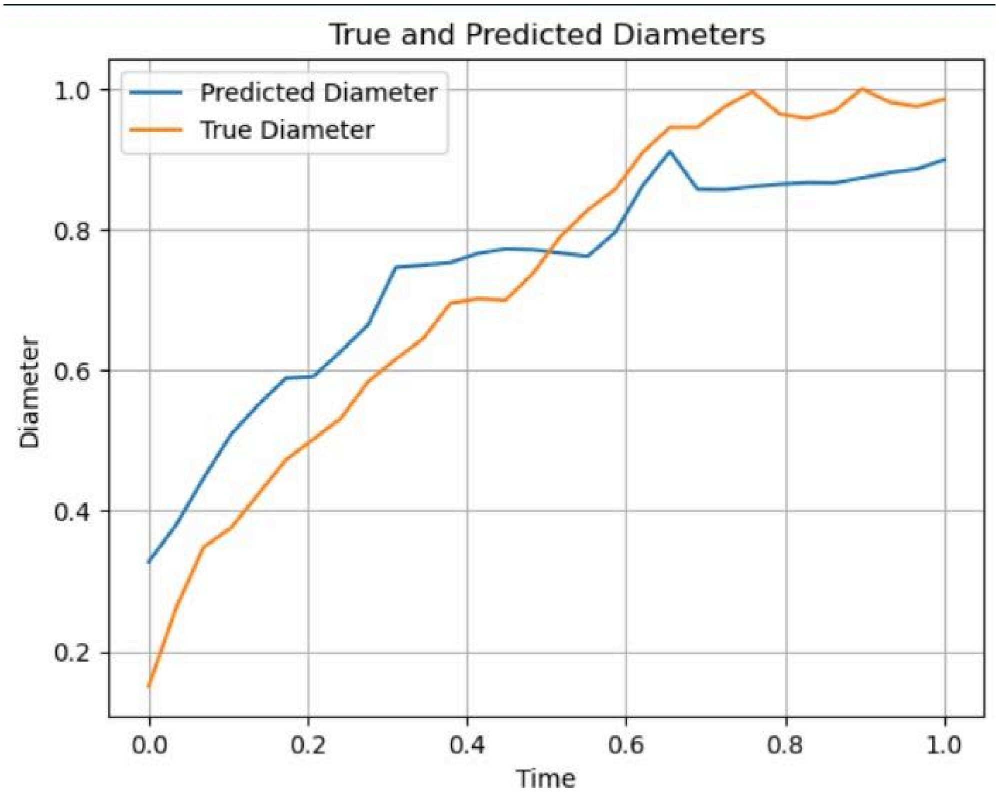
Gradient Boosting: While gradient boosting algorithms performed better than individual decision trees, they were comparatively slower and less efficient than XGBoost.

Model Evaluation:

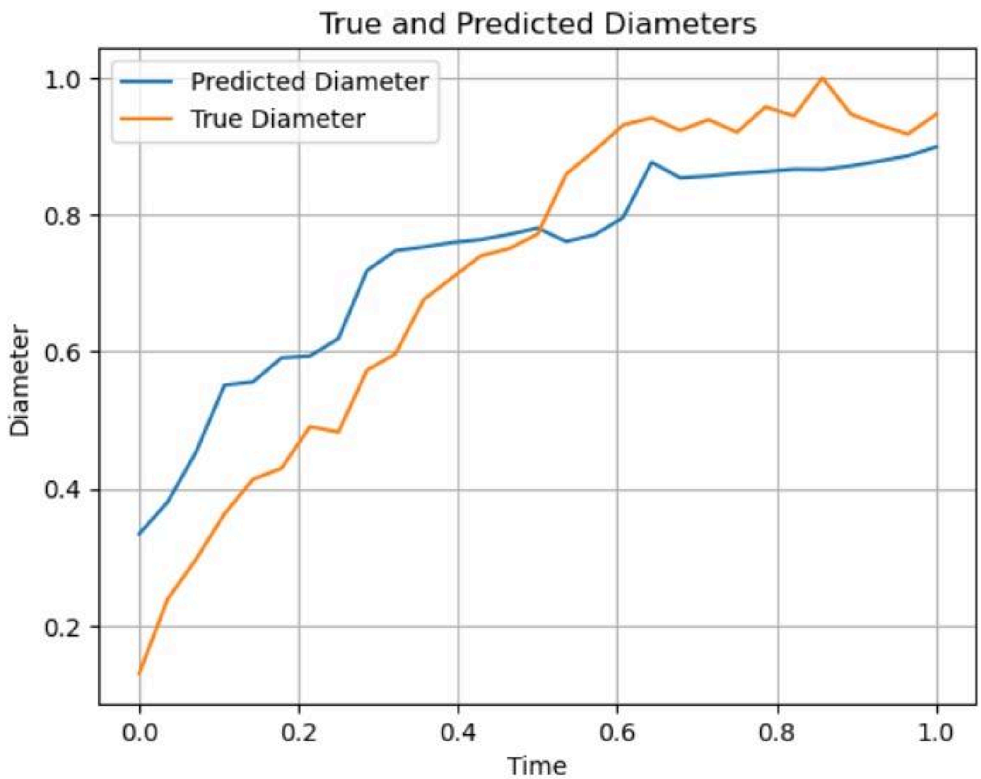
1) Trained on Literature-Informed Data

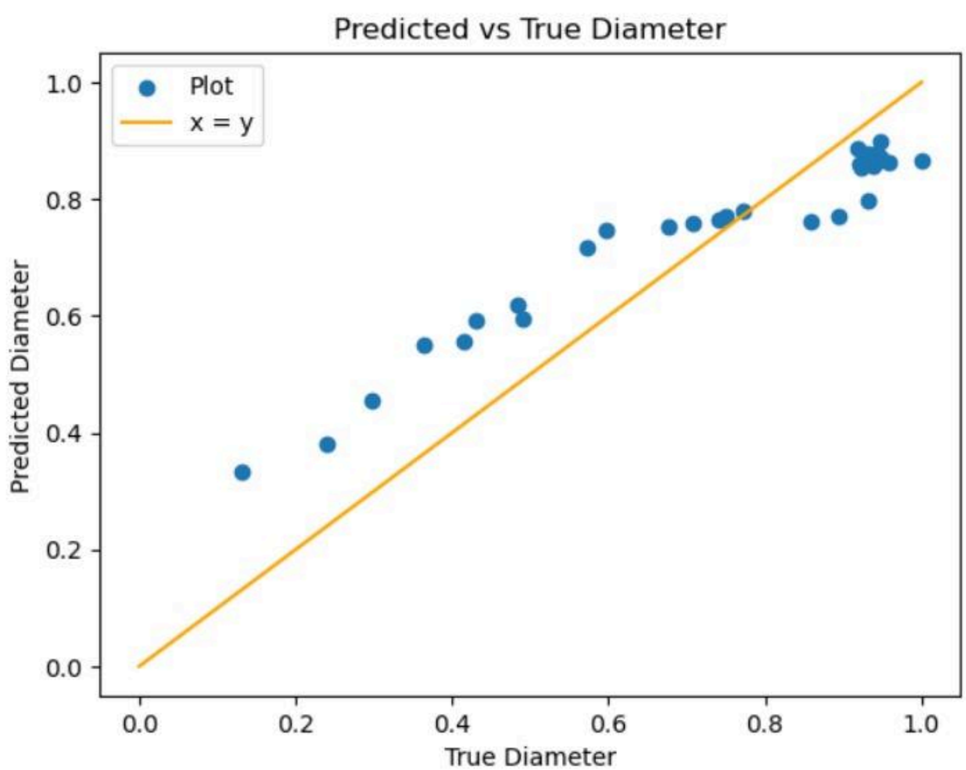
Condition 1: pressure - 3 bar , heat flux - 165 kW/m^2 , mass flux - 400 kg/m^2s , subcooling - 10K , channel diameter - 3.63 mm


Predicted vs True Diameter
Plot
x = y
Predicted Diameter
True Diameter
0.0
0.2
0.4
0.6
0.8
1.0

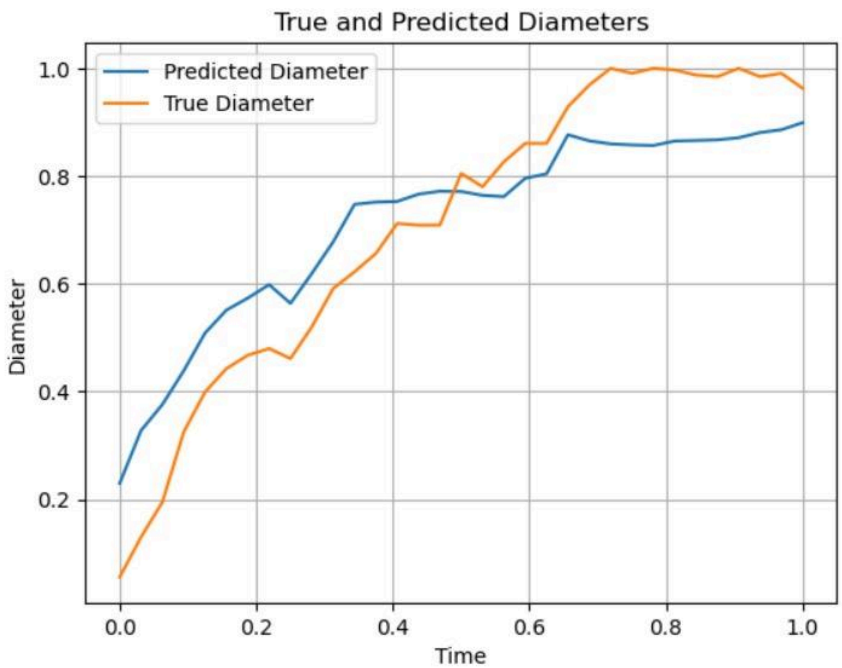

True and Predicted Diameters
Predicted Diameter
True Diameter
Diameter
Time
0.0
0.2
0.4
0.6
0.8
1.0

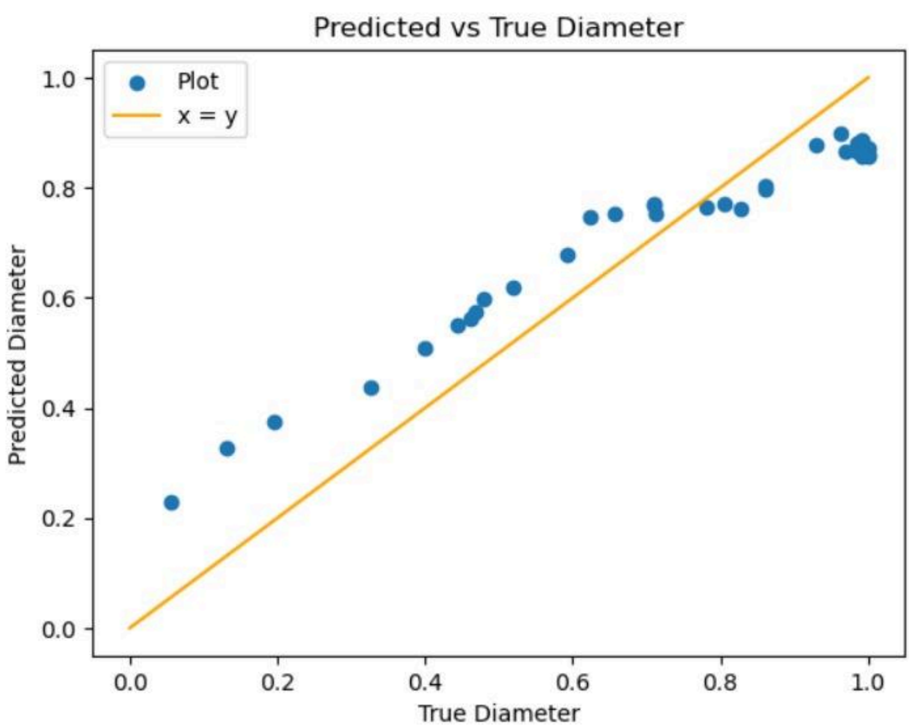
Condition 2 : pressure - 5 bar , heat flux - 165 kW/m^2 , mass flux - 400 kg/m^2s , subcooling - 10K , channel diameter - 3.63 mm





Condition 3 : pressure - 7 bar , heat flux - 165 kW/m^2 , mass flux - 400 kg/m^2s , subcooling - 10K , channel diameter - 3.63 mm





Evaluation matrix for the above three conditions :

- ```
  [13] evaluation(testingy1, yp1)                                      Python

  ... the mean absolute error is  0.0921967766290567
      the median absolute error is  0.09341824583876651
      the mean poisson deviance is  0.017078411852341022
      the r2_score is   84.17
      the explained_variance is  0.8426810626642378

  [14] evaluation(testingy2, yp2)                                      Python

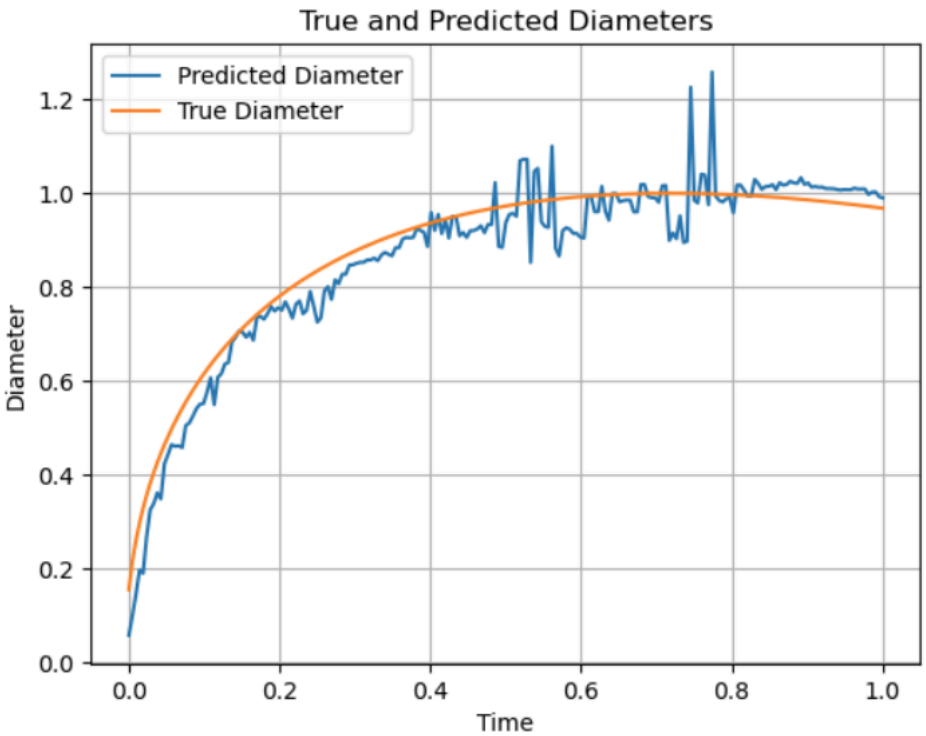
  ... the mean absolute error is  0.0985424080518318
      the median absolute error is  0.09457955436540533
      the mean poisson deviance is  0.024132951039125123
      the r2_score is   80.76
      the explained_variance is  0.8134929854530497

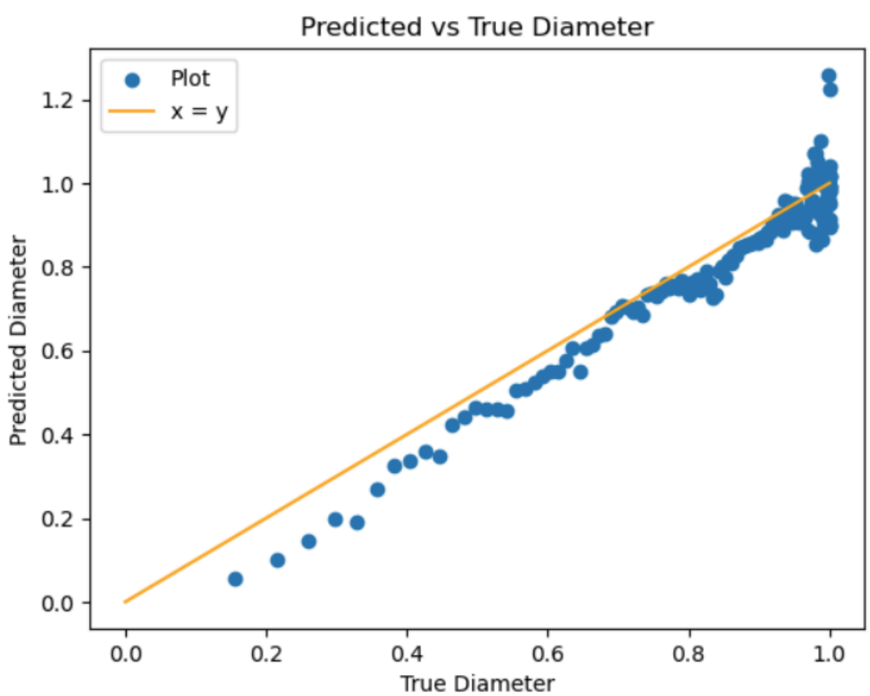
  [15] evaluation(testingy3, yp3)                                      Python

  ... the mean absolute error is  0.1017634808199585
      the median absolute error is  0.10467985069852292
      the mean poisson deviance is  0.026222351027837914
      the r2_score is   84.39
      the explained_variance is  0.8444240018504391
  ```

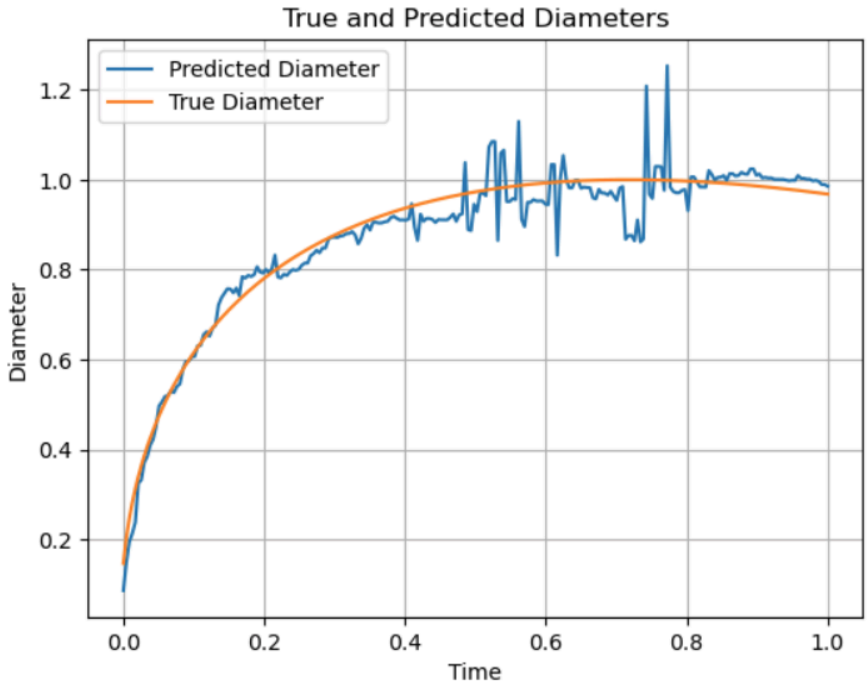
2) Trained on Hybrid data

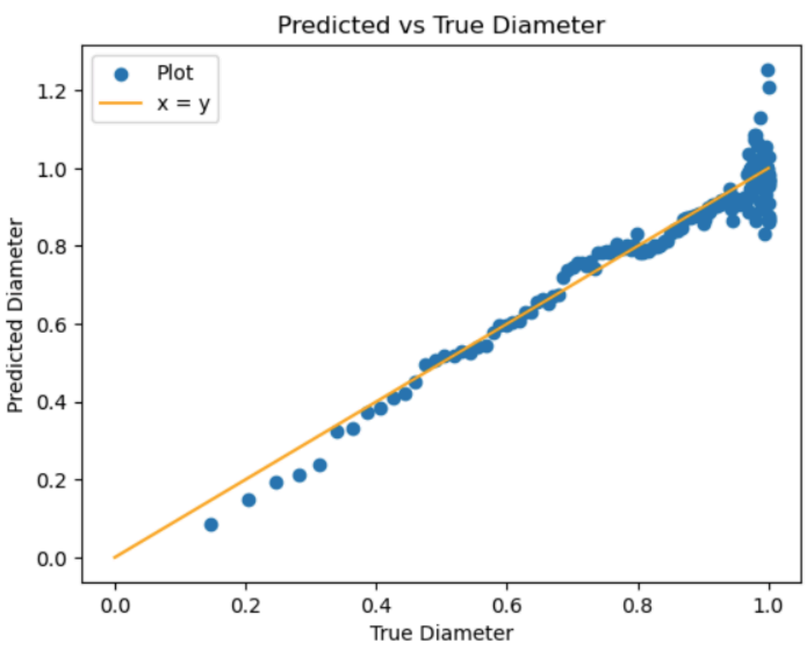
Condition 1 : pressure - 1.1 bar , heat flux - 173 kW/m^2 , mass flux - 495 kg/m^2s , subcooling - 6.5 , channel diameter - 13.33 mm
  ```


True and Predicted Diameters
Predicted Diameter
True Diameter
Diameter
Time
0.0
0.2
0.4
0.6
0.8
1.0
1.2

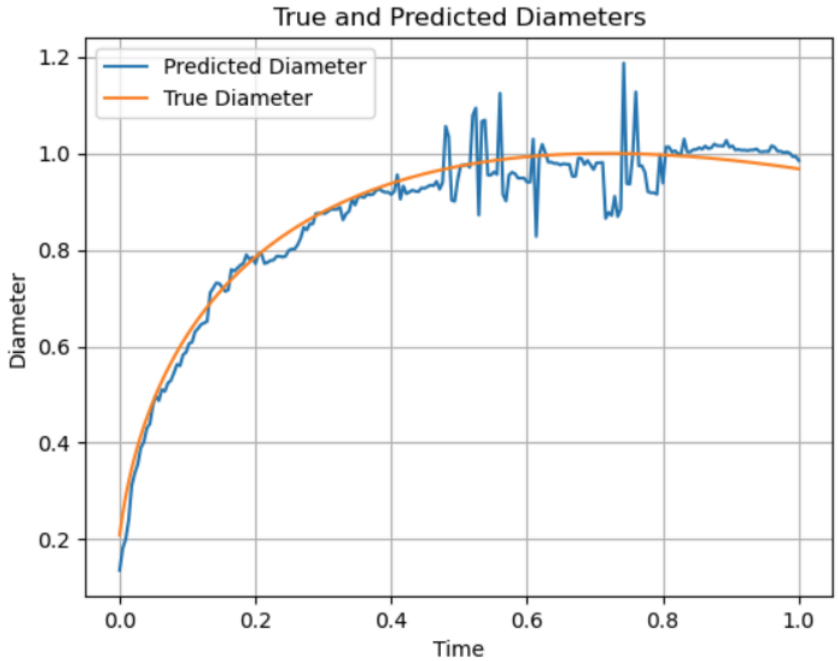

Predicted vs True Diameter
Plot
x = y
Predicted Diameter
True Diameter
0.0
0.2
0.4
0.6
0.8
1.0
1.2

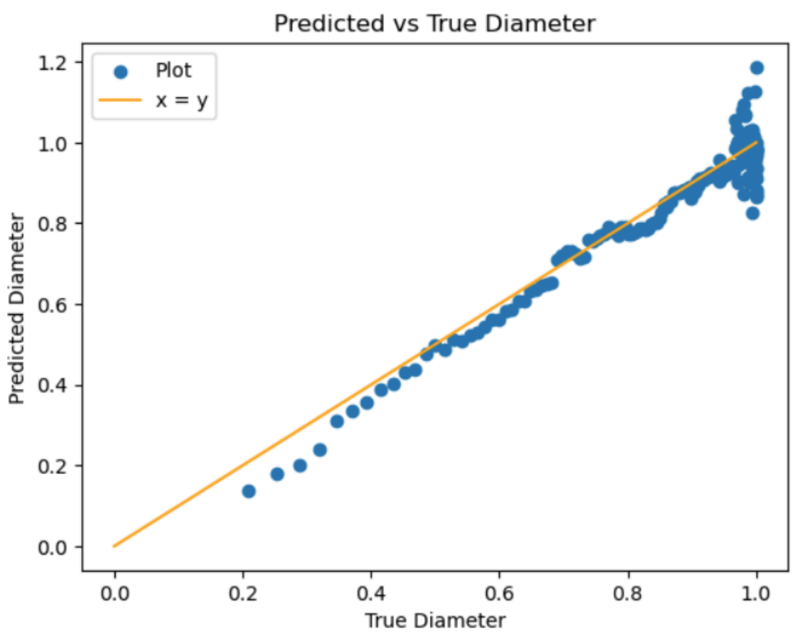
Condition 2 : pressure - 1.08 bar , heat flux - 207 kW/m^2 , mass flux - 497 kg/m^2s , subcooling - 10.33K , channel diameter - 13.33 mm





Condition 3 : pressure - 1.13 bar , heat flux - 269 kW/m^2 , mass flux - 489 kg/m^2s , subcooling - 16.4K , channel diameter - 13.33 mm





Evaluation matrix for the above three conditions :

```
[120]: evaluation(testingy1 , yp1)

the mean absolute error is  0.042640989206342635
the median absolute error is  0.0338384694486511
the mean poisson deviance is  0.005244872100167102
the r2_score is  0.9033060681909382
the explained_variance is  0.9158883248114164

[121]: evaluation(testingy2 , yp2)

the mean absolute error is  0.03322793639751075
the median absolute error is  0.02428819277018429
the mean poisson deviance is  0.0027042539374809837
the r2_score is  0.9287097394718593
the explained_variance is  0.9308258765773424

[122]: evaluation(testingy3 , yp3)

the mean absolute error is  0.03265381644022791
the median absolute error is  0.025876214137683096
the mean poisson deviance is  0.0025899994927438075
the r2_score is  0.9312001600278942
the explained_variance is  0.936237460355679

[123]: evaluation(testingy4 , yp4)

the mean absolute error is  0.03456695201300266
the median absolute error is  0.02164139072255855
the mean poisson deviance is  0.004485152275384383
the r2_score is  0.8918886288953248
the explained_variance is  0.8948556545856442
```

Conclusion :

The XGBoost model has proven effective in predicting bubble growth dynamics in flowing liquids, significantly advancing multiphase flow research and practical applications. Through thorough evaluation, the model exhibited good predictive capabilities, accurately capturing trends and patterns in experimental data.

To further enhance predictive accuracy and robustness, expanding the dataset is crucial. Particularly, incorporating more data points depicting diameter vs. time relationships will better capture the complexities of bubble growth dynamics. Additionally, as dataset size increases, computational demands escalate. Exploring advanced techniques, such as

LSTM models, is prudent for efficiently handling larger datasets. LSTM models, renowned for modeling sequential data and capturing long-term dependencies, offer promise for addressing challenges associated with large-scale datasets.

In conclusion, while the current XGBoost model demonstrates commendable performance, there is room for refinement. Augmenting the dataset and leveraging advanced techniques like LSTM will facilitate the development of more accurate and robust predictive models for bubble growth dynamics. These efforts not only advance scientific knowledge but also present opportunities for practical applications across various industries and domains.

References :

1. Rouhollah Ahmadi, Tatsuya Ueno, Tomio Okawa https://www.sciencedirect.com/science/article/pii/S0017931011005552
2. Dmitry Zaitsev,Egor Tkachenko https://www.researchgate.net/publication/329254680_Subcooled_flow_boiling_in_a_flat_mini-channel_under_local_heating
3. B. B. Mikic, W. M. Rohsenow and P. Griffith, On bubble growth rates, Inl. J. Heat Mass Transfer 13, 657-666 (1970).
4. Murallidharan, J. S., Prasad, B., & Patnaik, B. (2018, July). A universal wall-bubble growth model for water in component-scale high-pressure boiling systems. *International Journal of Heat and Mass Transfer*, *122*, 161–181. https://doi.org/10.1016/j.ijheatmasstransfer.2018.01.070
5. Nikhil Chitnavis, Harish Pothukuchi, B. S. V. Patnaik; Bubble growth and departure behavior in subcooled flow boiling regime. *Physics of Fluids* 1 May 2023; 35 (5): 053327. https://doi.org/10.1063/5.0145889