INL/MIS-20-58225-Revision-0

Computationally Efficient CFD Prediction of Two-phase Flow using Deep Learning and Validation Data

May 2020

Changing the World's Energy Future

Han Bao, Jinyong Feng, Nam Dinh

INL/MIS-20-58225-Revision-0

Computationally Efficient CFD Prediction of Two-phase Flow using Deep Learning and Validation Data

Han Bao, Jinyong Feng, Nam Dinh

May 2020

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Computationally Efficient CFD Prediction of Two-phase Flow using Deep Learning and Validation Data

Han Bao, Hongbin Zhang
Idaho National Laboratory

Jinyong Feng
Massachusetts Institute of Technology

Nam Dinh
North Carolina State University

May 1st, 2020

www.inl.gov

Ideal Simulation Tool

An ideal simulation tool for system-level thermal-hydraulic analysis should be…

- **Fast-running**
- **Sufficiently accurate**
- **Scalable for extrapolation**

- **Coarse-mesh CFD**
- **Error estimation using machine learning**
- **Explore similarity in local features instead of global characteristics**

Experience-driven Expert Judgement

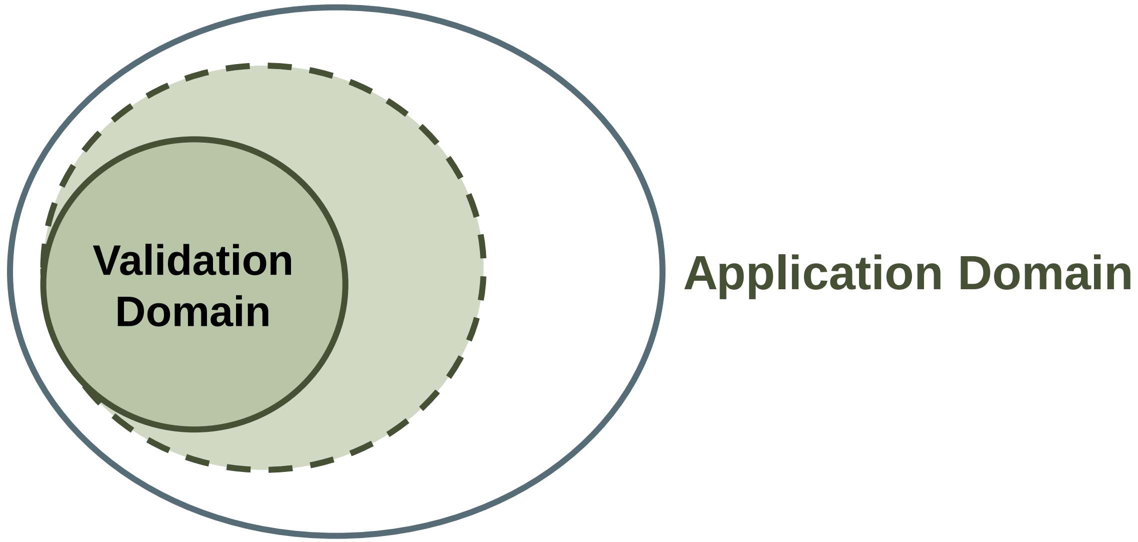
Data-driven Machine Learning

**Courtesy to The Matrix*

Some Questions about Data in this work…

- **What is “validation data”, “High-fidelity data” and “Low-fidelity data”?**
  - Validation data: Relevant experimental data and validated high-resolution numerical simulation results.
  - High-fidelity data: Validation data, or any data that satisfies requirements of “users”. (“Accurate”!)
  - Low-fidelity data: Simulation results using coarse meshes or simplified closures. (“Easy-to-get”!)

- **How to use existing numerical/experimental data to expand the validation domain?**
  - There is a lack of validation data due to scaling issues and costs…
  - To be continued…

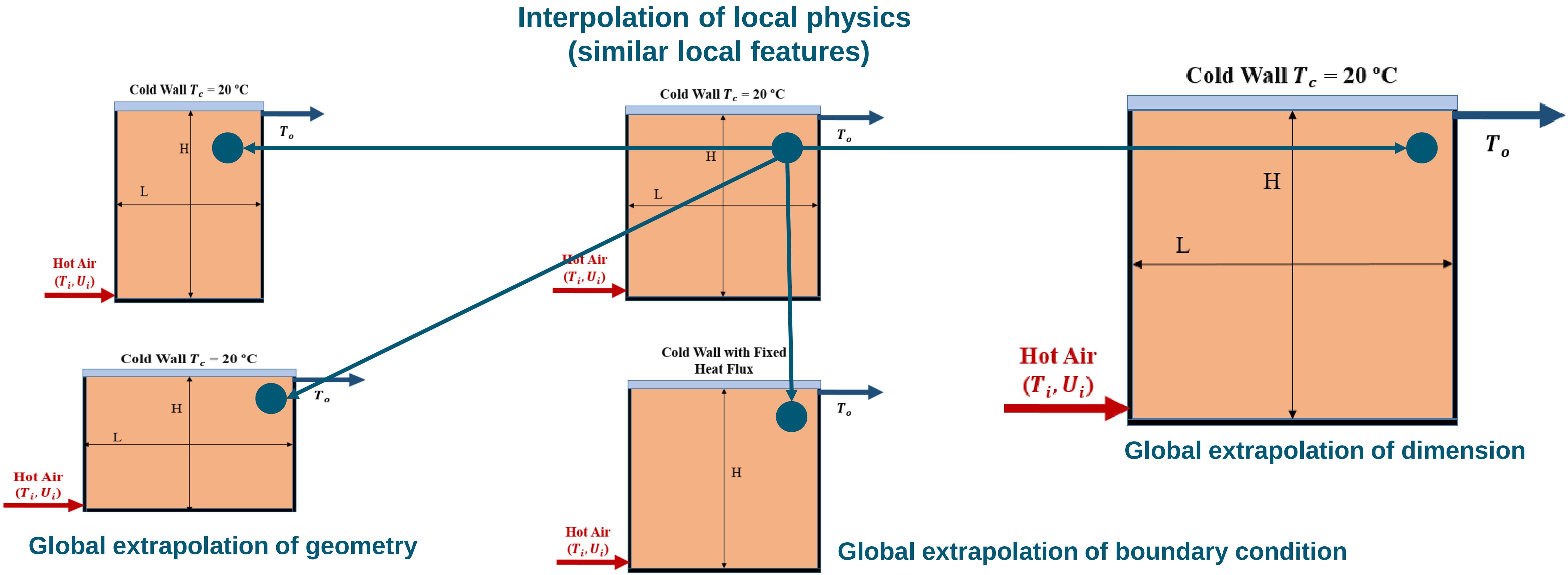


Some Questions about Data in this work…

- **Are both high-fidelity data and low-fidelity data useful?**
  - YES and NO!
  - Low-fidelity data ensures efficiency, high-fidelity data ensures accuracy.
  - The construction of training database is quite Target-oriented! Not every data point is useful!
- **How to fully extract information from data to efficiently improve CFD predictions for a target condition?**
  - Clarify goals. (Fast-running! Sufficiently accurate! Scalable!)
  - Analyze targets. (Phenomena, structure, geometry, IC/BCs….)
  - Build training database. (Find the most similar ones!)
  - Identify local similarities and global differences between targets and existing cases.

Global Characteristics vs. Local Features

- **Global characteristics** indicate the global state and observation of target system, such as the **dimension, geometry or boundary condition**;
- **Local physical features** refer to the local state and observation of target system.



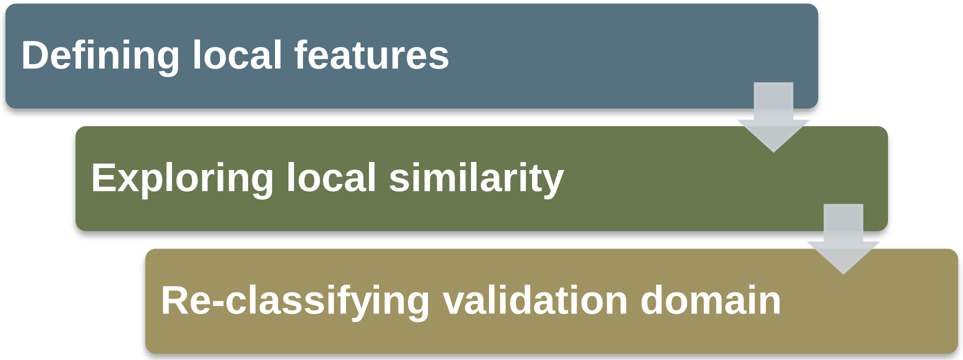
* **Bao H**., Dinh N., Lin L., Youngblood R., Lane J., Zhang H. "Using Deep Learning to Explore Local Physical Similarity for Global Scale Bridging in Thermal Hydraulic Simulation", (Under review). *Annals of Nuclear Energy* (2020).

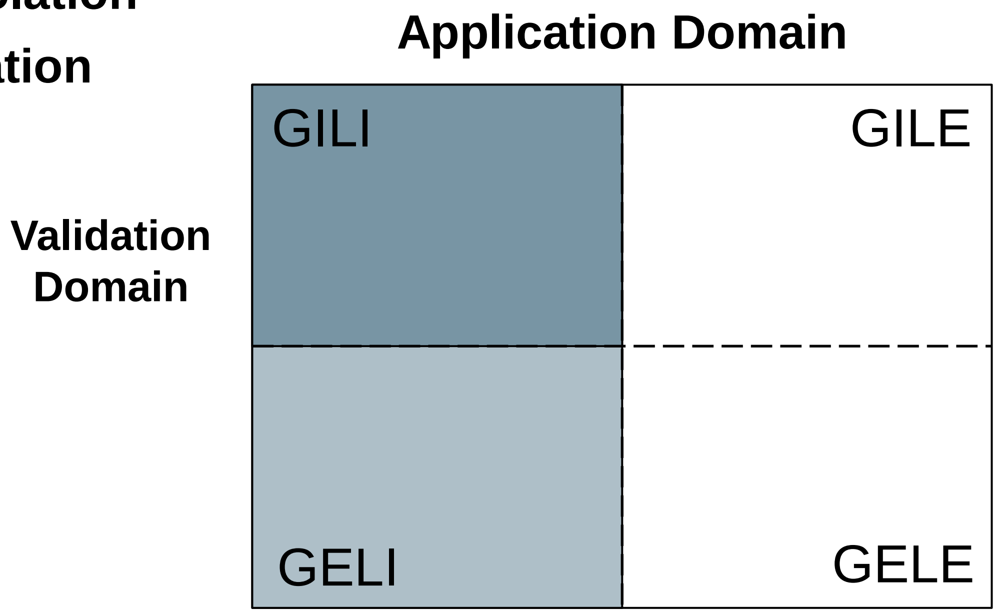
Q: How to Bridge Global Scale Gap?

---- Explore local similarity

Four Physics Coverage Conditions:

- **(GELI) Global Extrapolation but Local Interpolation**
  - The local interpolation (or local similarity) can be defined as:
    - Qualitatively, similar physics;
    - Quantitatively, similar data of physical features.
- **(GILI) Global Interpolation and Local Interpolation**
- **(GELE) Global Extrapolation and Local Extrapolation**
- **(GILE) Global Interpolation but Local Extrapolation**





Some Terms in "GELI Universe"

- **Feature Similarity Measurement (FSM):** A data-driven approach that was developed to
  - Identify local physical features
  - Measure data similarity of defined physical features
  - Enhance local similarity to improve the predictive performance of machine learning models
  - Estimate simulation errors

__H. Bao__, N. Dinh, L, Lin, R. Youngblood, J. Lane, H. Zhang. "Using Deep Learning to Explore Local Physical Similarity for Global Scale Bridging in Thermal Hydraulic Simulation", (Under review). Annals of Nuclear Energy (2020).
__H. Bao__, J. Feng, N. Dinh, H. Zhang. "Deep Learning Interfacial Momentum Closures in Coarse-Mesh CFD Two-Phase Flow Simulation Using Validation Data", (Under review) International Journal of Multi-phase Flow (2020).

- **Optimal Mesh/Model Information System (OMIS):** A data-driven framework that is formalized to
  - Estimate simulation error
  - Suggest optimal selection of computational mesh size and closure models

__H. Bao__, N. Dinh, J. Lane, R. Youngblood. "A Data-driven Framework for Error Estimation and Mesh-model Optimization in System-level Thermal-hydraulic Simulation", Nuclear Engineering and Design, 349, pp. 27-45 (2019).

- **In this work, FSM is applied to realize computationally efficient CFD prediction (CECFD).**

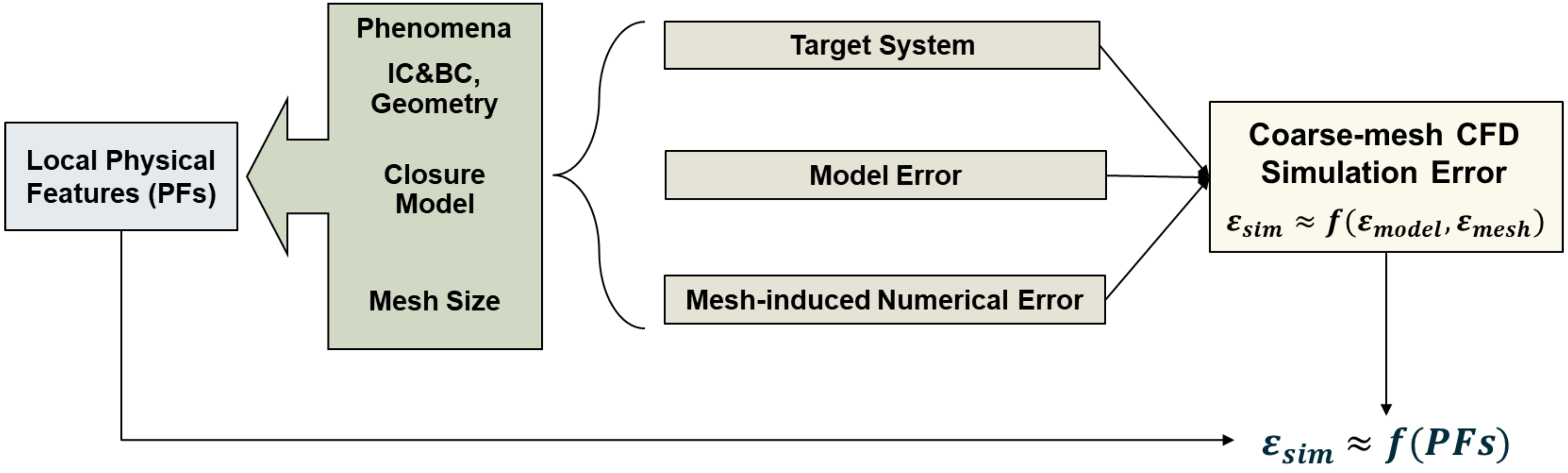
__H. Bao__, J. Feng, N. Dinh, H. Zhang. "Computationally Efficient CFD Prediction of Bubbly Flow using Physics-Guided Deep Learning", (Under review) International Journal of Multi-phase Flow (2020).

Case studies…

- **Case Study Part I**
  - **Goal:** Investigate how much **FSM** can improve the coarse-mesh CFD simulations for two-phase flow.
  - **Low-fidelity data:**
    - Coarse-mesh CFD results (Star CCM+)
  - **High-fidelity data**:
    - (Test 1) Fine-mesh CFD results (Star CCM+)
    - (Test 2) Experimental data

- **Case Study Part II**
  - **Goal:** Demonstrate **CECFD framework** for two-phase flow.
  - **Low-fidelity data:**
    - Coarse-mesh CFD results (Star CCM+)
  - **High-fidelity data**:
    - Fine-mesh CFD results (Star CCM+)

Data-driven Approach: Feature Similarity Measurement (FSM)



Goal

- Support simulation tools to build up efficient and scalable predictive capability

Application

- Different phenomena (e.g., mixed convection, two-phase flow)

Identification of Physical Features

- **Local physical features**
  - Derivatives of variables (← Regional information)
    - 1-order derivatives
    - 2-order derivatives
    - High-order derivatives
  - Local physical parameters (← Scalability and additivity of phenomena)
    - Non-dimensional groups
    - Parameters relevant to closure models
    - Parameters relevant to IC/BC, geometry

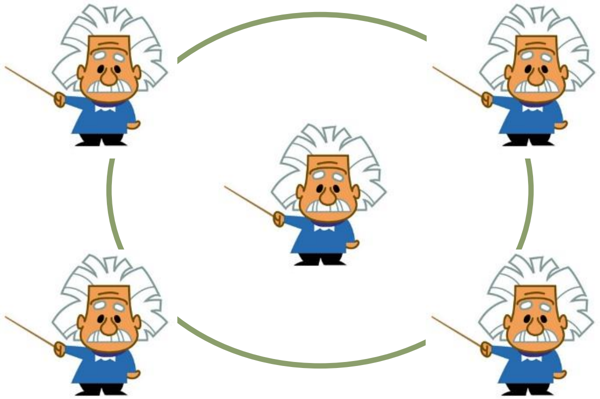
Phenomena

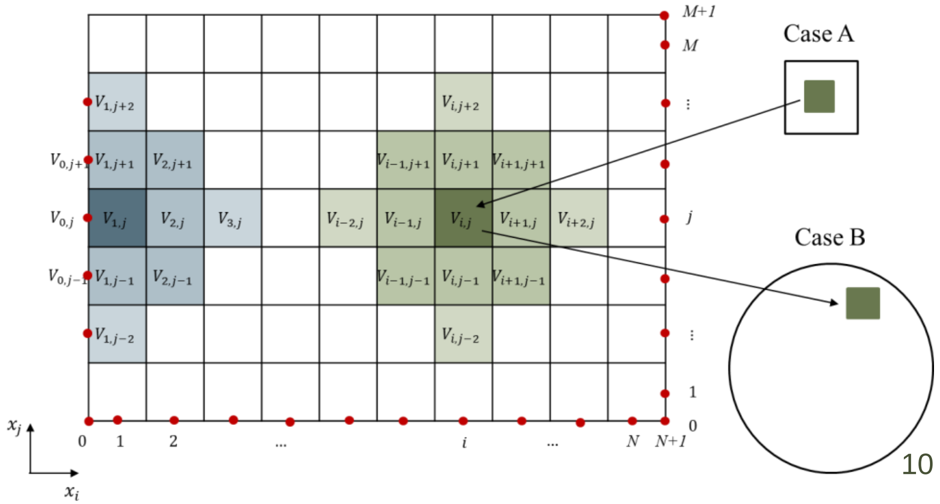
IC&BC, Geometry

Closure Model

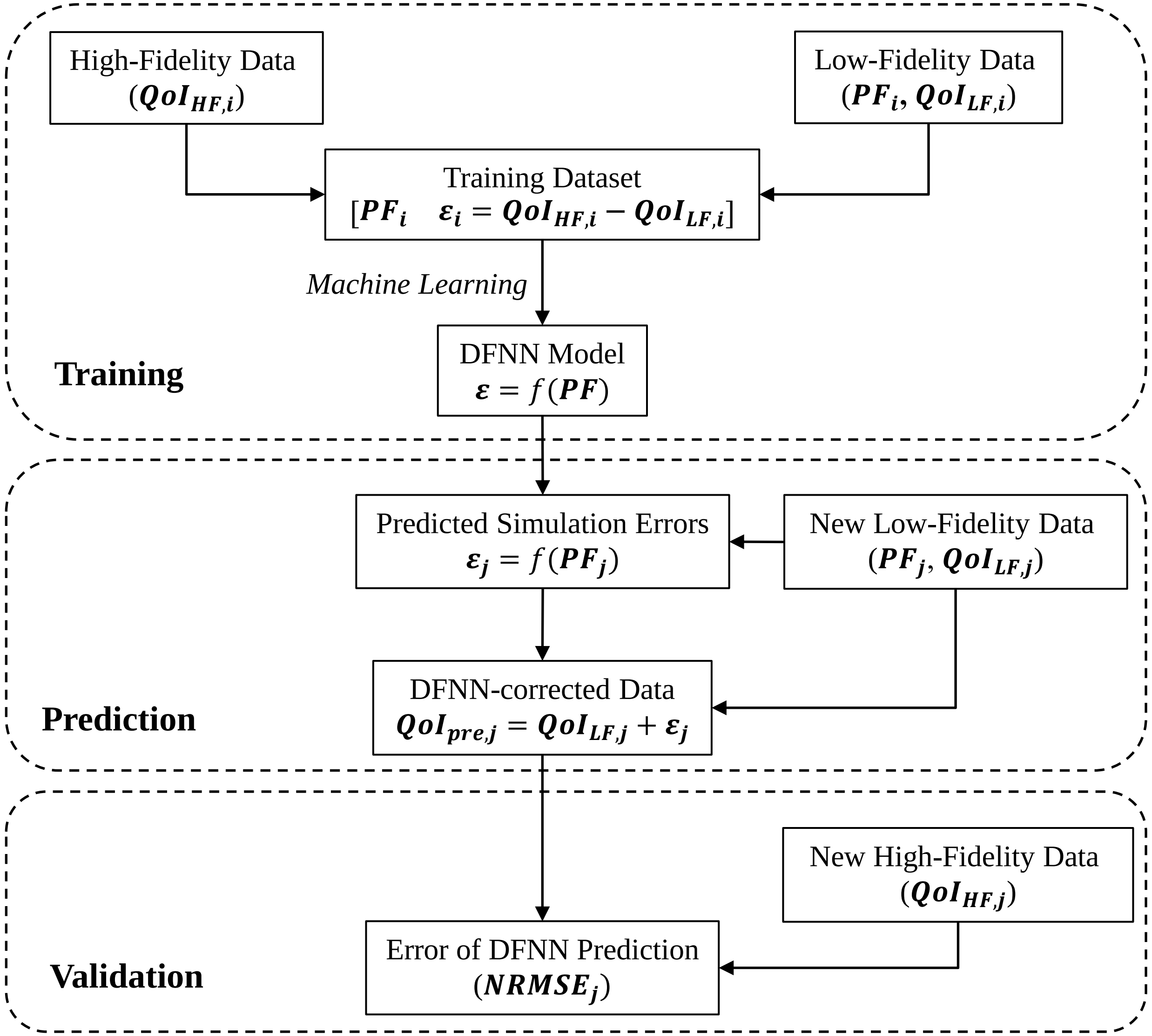
Mesh Size

Friction	PF_n
Turbulence	PF_n+PF_m
Convection	$PF_n+PF_m+PF_l$
…	…





Training Flow and Testing Flow of Feature Similarity Measurement (FSM)



$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum\left(QoI_{pre} - QoI_{HF}\right)^2}}{\frac{1}{n}\sum QoI_{HF}}$$

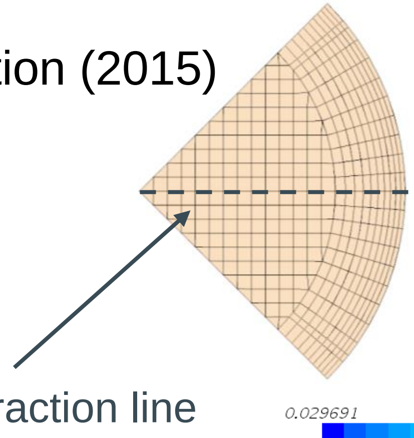
Case Study Part I

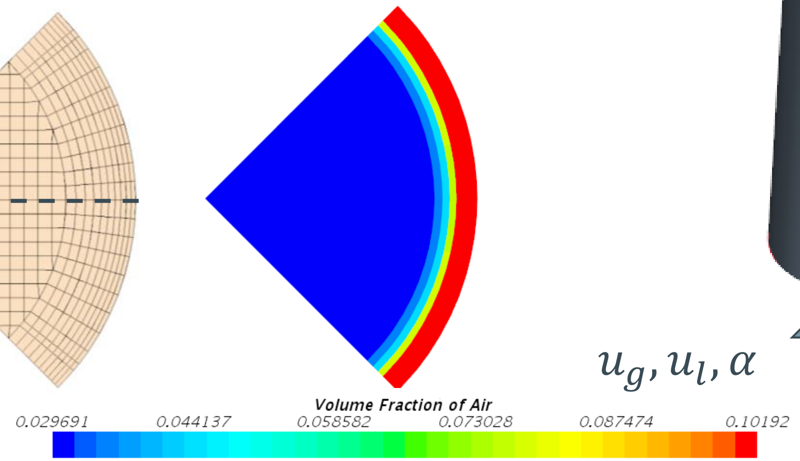
Goal: Investigate how much FSM can improve the coarse-mesh CFD simulations for two-phase flow.

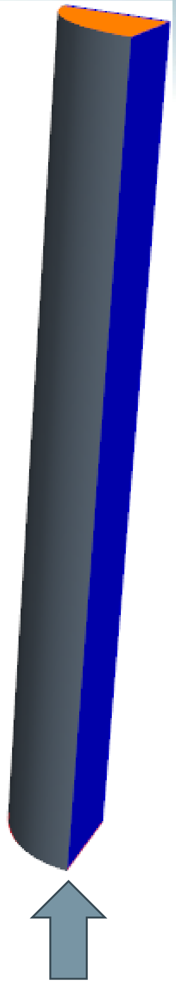
- **Machine learning:** Deep Feedforward Neural Network (DFNN)
- **Low-fidelity data:**
  - (Test 1) Coarse-mesh CFD results with validated interfacial momentum closures (BAMF model)
  - (Test 2) Coarse-mesh CFD results with simplified interfacial momentum closures
- **High-fidelity data**:
  - (Test 1) Fine-mesh CFD results with validated interfacial momentum closures (BAMF model)
  - (Test 2) Experimental data
- **Evaluation metrics:**
  - Accuracy: NRMSE
  - Efficiency: computational time

Problem Statement

- **Simulation of two-phase bubbly flow**
  - Reference experimental dataset: Liu & Bankoff (1989, 1993)
  - Interfacial momentum closures adopted from BAMF model (Sugrue, 2017):
    - Drag force model: Tomiyama (1998)
    - Lift force model (constant): 0.025
    - Wall lubrication force: Podowski's correction (2015)
    - Turbulent dispersion force: Burns (2004)
    - Geometry: L- 1.6 m; R- 0.019 m



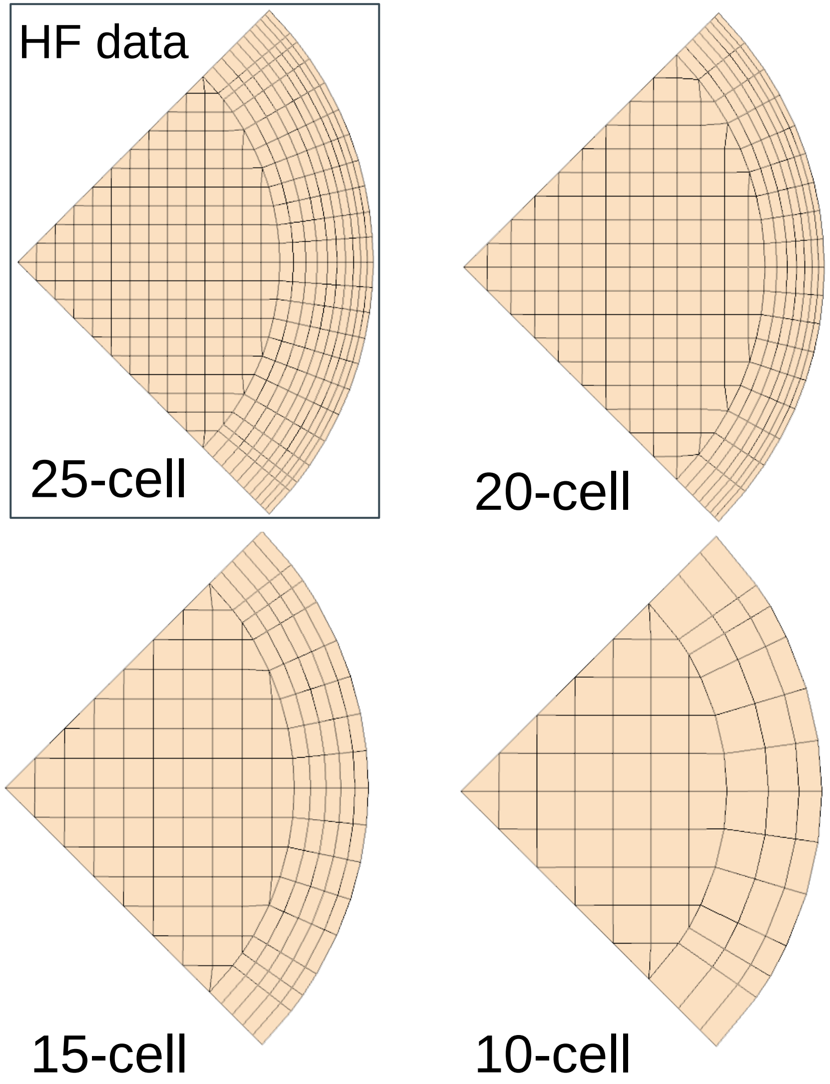




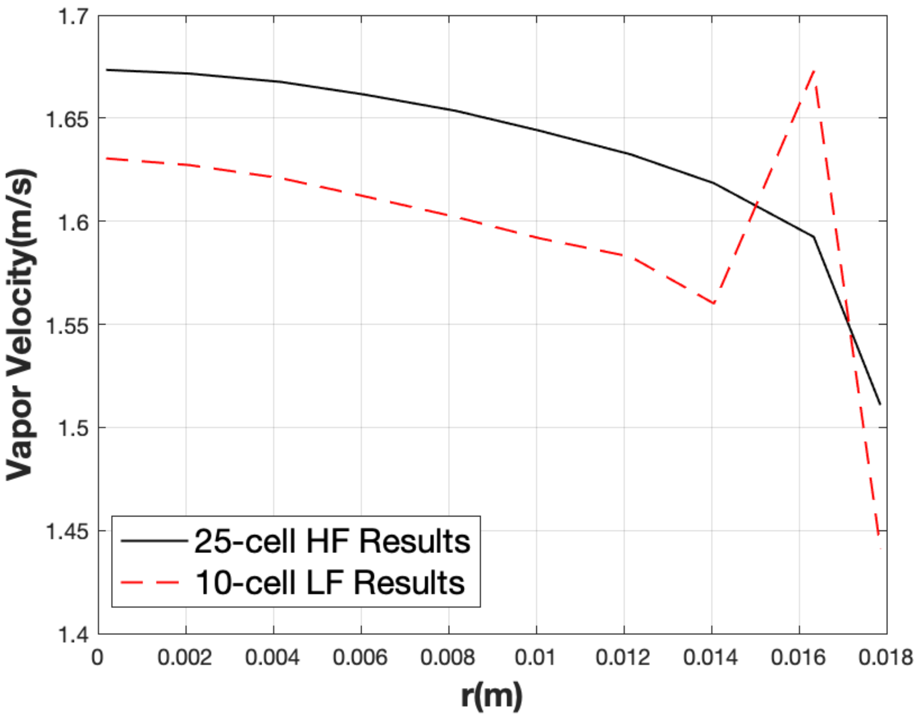
* BAMF: Bubbly And Moderate Void Fraction model

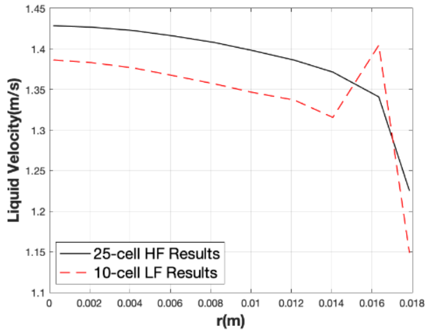
Liu, T.J., 1989. Experimental Investigation of Turbulence Structure in Two-Phase Bubbly Flow. Northwestern University, Evanston, Illinois. Ph.D. Thesis.
Liu, T.J., Bankoff, S.G., 1993a. Structure of air-water bubbly flow in a vertical pipe—I. liquid mean velocity and turbulence measurements. Int. J. Heat Mass Transf. 36 (4), 1049–1060.
Liu, T.J., Bankoff, S.G., 1993b. Structure of air-water bubbly flow in a vertical pipe—II. Void fraction, bubble velocity and bubble size distribution. Int. J. Heat Mass Transf. 36 (4), 1061–1072.
Sugrue, R., Magolan, B., Lubchenko, N., & Baglietto, E. (2017). Assessment of a simplified set of momentum closure relations for low volume fraction regimes in STAR-CCM+ and OpenFOAM. Annals of Nuclear Energy, 110, 79-87.

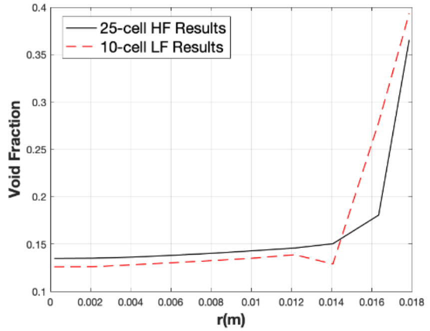
Test 1: Using Fine-mesh CFD results as High-fidelity Data



- Number of cells (million):
  - 10-cell: 0.07
  - 20-cell: 0.35
  - 15-cell: 0.18
  - 25-cell: 0.63
- BAMF model works well for 25-cell simulations, but not for coarse-mesh simulations.







Test 1: Analyze Target Cases and Identify Physical Features

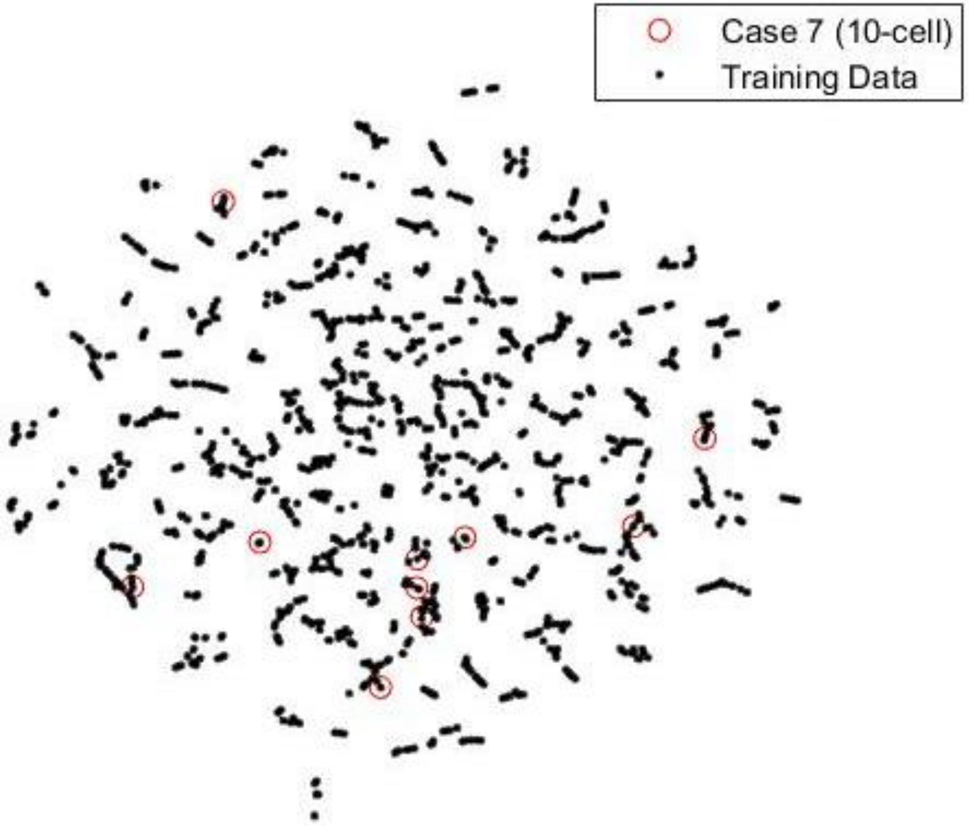
Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction	Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction
1	0.376	0.027	0.0407	22	0.974	0.027	0.0204
2	0.376	0.067	0.1167	23	0.974	0.067	0.0514
3	0.376	0.112	0.1843	24	0.974	0.112	0.0791
4	0.376	0.18	0.2449	25	0.974	0.18	0.1242
5	0.376	0.23	0.3079	26	0.974	0.23	0.1512
6	0.376	0.293	0.3657	27	0.974	0.293	0.1869
7	0.376	0.347	0.4168	28	0.974	0.347	0.2108
8	0.535	0.027	0.0312	29	1.087	0.027	0.0176
9	0.535	0.067	0.0877	30	1.087	0.067	0.0473
10	0.535	0.112	0.1406	31	1.087	0.112	0.0737
11	0.535	0.18	0.2016	32	1.087	0.18	0.1096
12	0.535	0.23	0.2344	33	1.087	0.23	0.1497
13	0.535	0.293	0.3102	34	1.087	0.293	0.1777
14	0.535	0.347	0.3398	35	1.087	0.347	0.1976
15	0.753	0.027	0.0235	36	1.391	0.027	0.0148
16	0.753	0.067	0.0622	37	1.391	0.067	0.0387
17	0.753	0.112	0.1091	38	1.391	0.112	0.0581
18	0.753	0.18	0.1554	39	1.391	0.18	0.0964
19	0.753	0.23	0.1816	40	1.391	0.23	0.1176
20	0.753	0.293	0.2381	41	1.391	0.293	0.1504
21	0.753	0.347	0.2692	42	1.391	0.347	0.1724

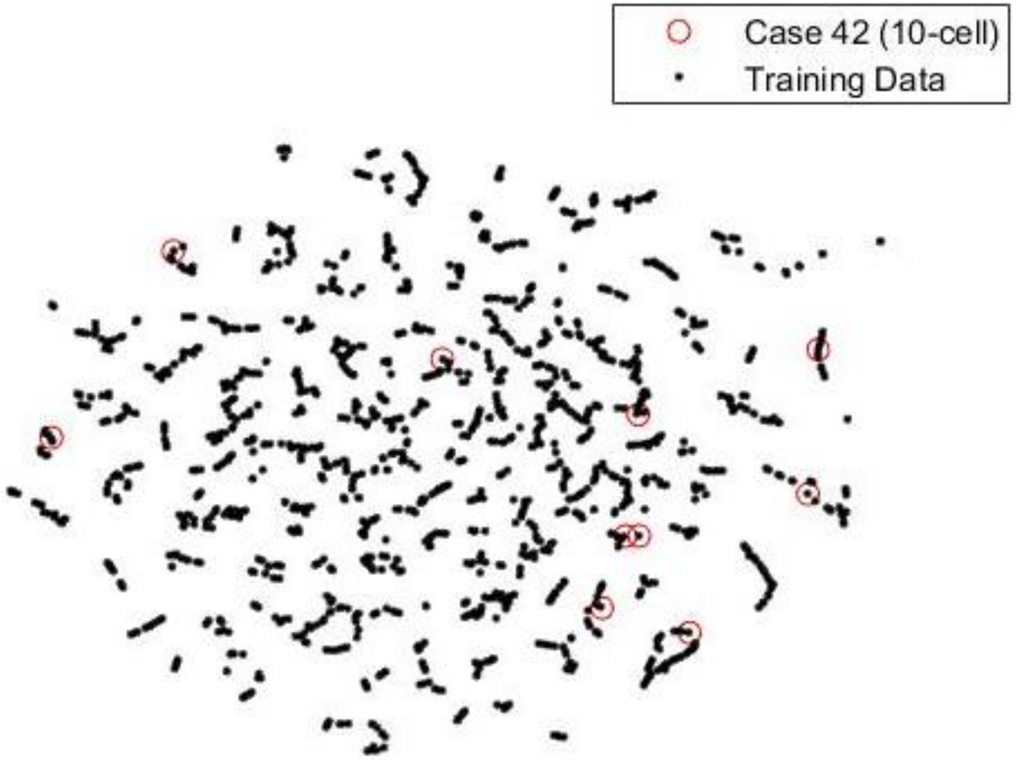
Test 1	Testing Case	Description
1.1	7 (10-cell)	**Global extrapolation** of high void fraction, vapor rate and low liquid rate
1.2	42 (10-cell)	**Global extrapolation** of high vapor rate and liquid rate

Local Physical Features							
Derivatives of variable				Local physical parameters			
1-order		2-order		Non-dimensional groups		Parameters relevant to closure models, IC/BC, geometry	
$\frac{du_l}{dx}$	$\frac{du_g}{dx}$	$\frac{d^2u_l}{dx^2}$	$\frac{d^2u_g}{dx^2}$	$Re_\Delta = \frac{\rho_l \Delta \cdot \Delta u}{\mu_l}$	$I_l = \frac{k_l}{{u_l}^2}$	$R_l = \frac{{k_l}^{\frac{3}{2}}}{\varepsilon_l D_b}$	$R_\mu = \frac{\mu_g^t}{\mu_L^t}$
$\frac{d\alpha}{dx}$	$\frac{dP}{dx}$	$\frac{d^2\alpha}{dx^2}$	$\frac{d^2P}{dx^2}$	$Re_b = \frac{\rho_l D_b \Delta u}{\mu_l}$	$I_g = \frac{k_g}{{u_g}^2}$	$R_g = \frac{{k_g}^{\frac{3}{2}}}{\varepsilon_g D_b}$	$r_l = \frac{\mu_l^t}{\mu_l}$
$\frac{dk_l}{dx}$	$\frac{dk_g}{dx}$	$\frac{d^2k_l}{dx^2}$	$\frac{d^2k_g}{dx^2}$	$We = \frac{\rho D_b \Delta u^2}{\sigma}$		$Re_y = \frac{\rho_l y \Delta u}{\mu_l}$	$R_b = \frac{D_b}{\Delta}$
$\frac{d\varepsilon_l}{dx}$	$\frac{d\varepsilon_g}{dx}$	$\frac{d^2\varepsilon_l}{dx^2}$	$\frac{d^2\varepsilon_g}{dx^2}$				

Visualization of Local Interpolation

- Dimensionality reduction technique (t-SNE) was used for visualization of the distances between testing data and training data.
- 27-D → 2-D.



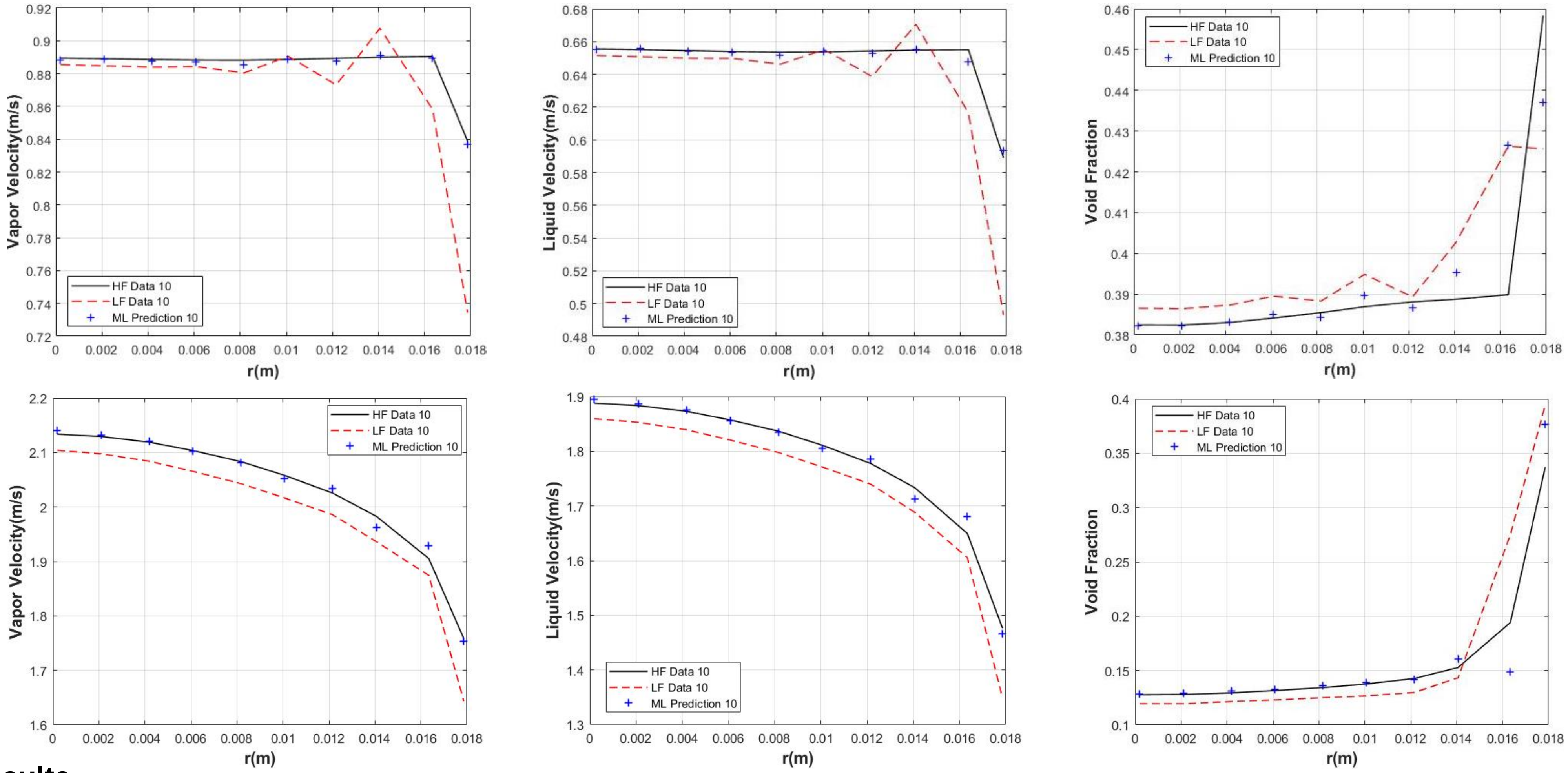


*t-SNE: t-Distributed Stochastic Neighbor Embedding.
Maaten, L., Hinton, G., 2012. Visualizing non-metric similarities in multiple maps. Mach. Learn. 87 (1), 33–55.

Test 1.1 and 1.2 : DFNN Predictions for GELI

**Testing case:
Case 7**

**Testing case:
Case 42**

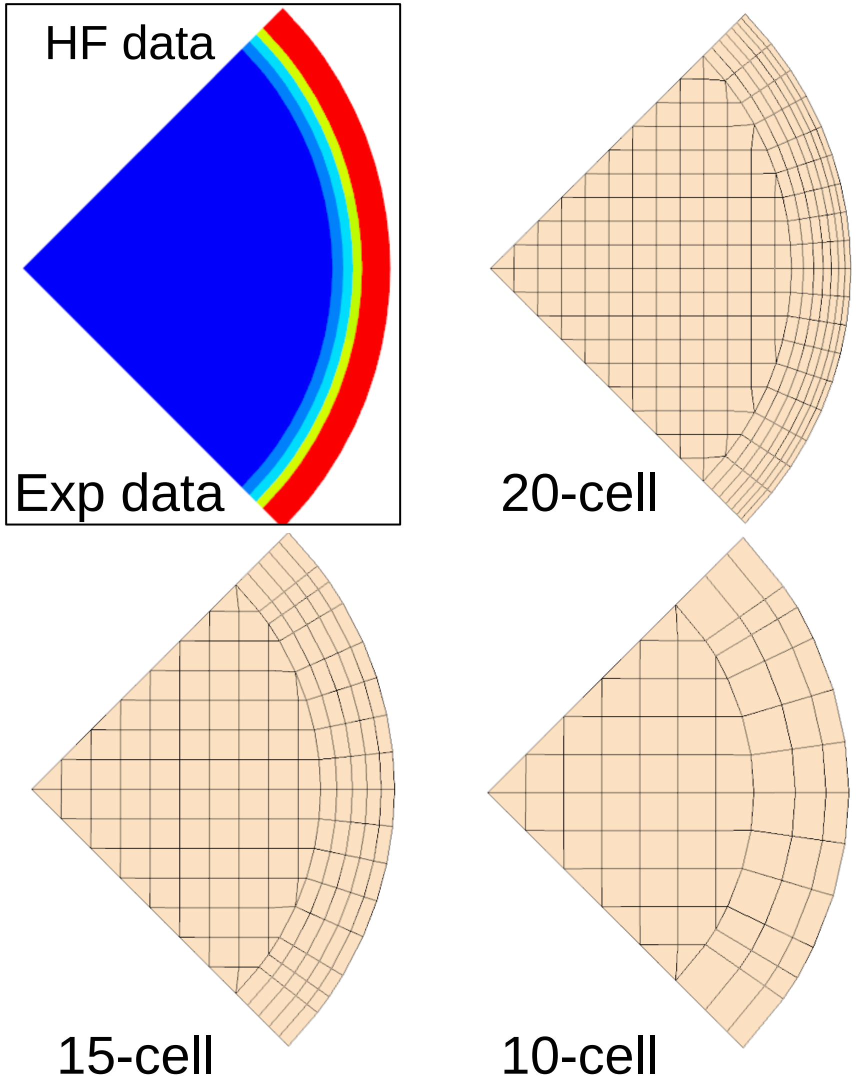


Vapor Velocity **Liquid Velocity** **Void Fraction**

Black Lines: HF 25-cell Results
Red Dashes: LF 10-cell Results
Blue Points: ML Prediction

Test 2: Using Experimental Data as High-fidelity Data

Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction	Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction
1	0.376	0.027	0.0407	22	0.974	0.027	0.0204
2	0.376	0.067	0.1167	23	0.974	0.067	0.0514
3	0.376	0.112	0.1843	24	0.974	0.112	0.0791
4	0.376	0.18	0.2449	25	0.974	0.18	0.1242
5	0.376	0.23	0.3079	26	0.974	0.23	0.1512
6	0.376	0.293	0.3657	27	0.974	0.293	0.1869
7	0.376	0.347	0.4168	28	0.974	0.347	0.2108
8	0.535	0.027	0.0312	29	1.087	0.027	0.0176
9	0.535	0.067	0.0877	30	1.087	0.067	0.0473
10	0.535	0.112	0.1406	31	1.087	0.112	0.0737
11	0.535	0.18	0.2016	32	1.087	0.18	0.1096
12	0.535	0.23	0.2344	33	1.087	0.23	0.1497
13	0.535	0.293	0.3102	34	1.087	0.293	0.1777
14	0.535	0.347	0.3398	35	1.087	0.347	0.1976
15	0.753	0.027	0.0235	36	1.391	0.027	0.0148
16	0.753	0.067	0.0622	37	1.391	0.067	0.0387
17	0.753	0.112	0.1091	38	1.391	0.112	0.0581
18	0.753	0.18	0.1554	39	1.391	0.18	0.0964
19	0.753	0.23	0.1816	40	1.391	0.23	0.1176
20	0.753	0.293	0.2381	41	1.391	0.293	0.1504
21	0.753	0.347	0.2692	42	1.391	0.347	0.1724



Test 2: New High-fidelity and Low-fidelity Data

Data Type	Low-fidelity Data	High-fidelity Data
Test 1.1, 1.2	• Coarse meshes • BAMF model	• Fine meshes • BAMF model
Test 2	• Coarse meshes • Simplified closures	• Experimental data

Model		Simplified Interfacial momentum closures	BAMF model
Turbulence model		Standard $k - \varepsilon$ linear	
Interfacial momentum forces	Drag coefficient	Tomiyama (Tomiyama et al., 1998)	
	Lift coefficient	N/A	Shaver and Podowski
	Turbulent dispersion force	N/A	Burns
	Wall lubrication force	N/A	Shaver and Podowski’s correction

Shaver, D.R., Podowski, M.Z., 2015. Modeling of interfacial forces for bubbly flows in subcooled boiling conditions. Trans. Am. Nucl. Soc. 113, 1368–1371.
Burns, A.D., Frank, T., Hamill, I., Shi, J.-M., 2004. The Favre averaged drag model for turbulent dispersion in Eulerian multi-phase flows. 5th Int. Conf. Multiph. flow, ICMF 4, 1–17.

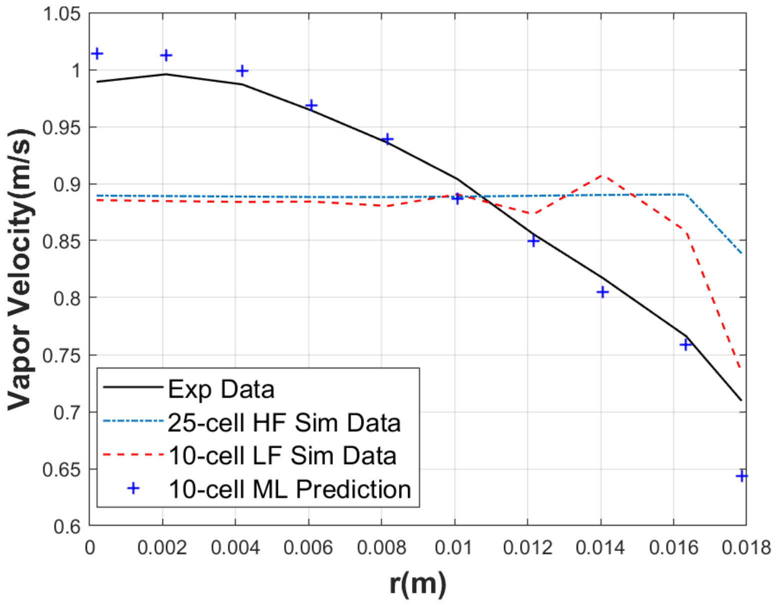
Test 2: DFNN Predictions

Motivations of using experimental data as high-fidelity data:

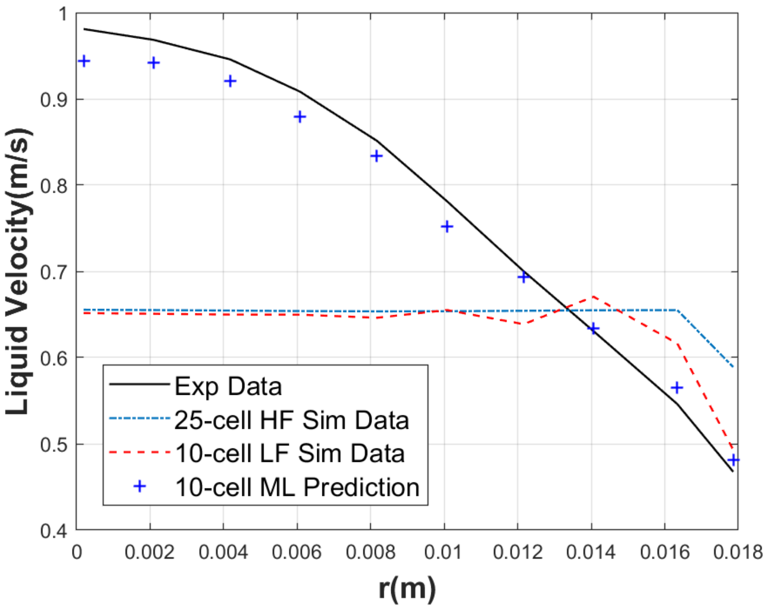
- Fine-mesh CFD needs validation, but experimental data does not.
- Even validated models have their applicable ranges.

Testing case: Case 7 [10 data points]
Training data: other 41 cases [1845=(41 × (25+20+15) data points]

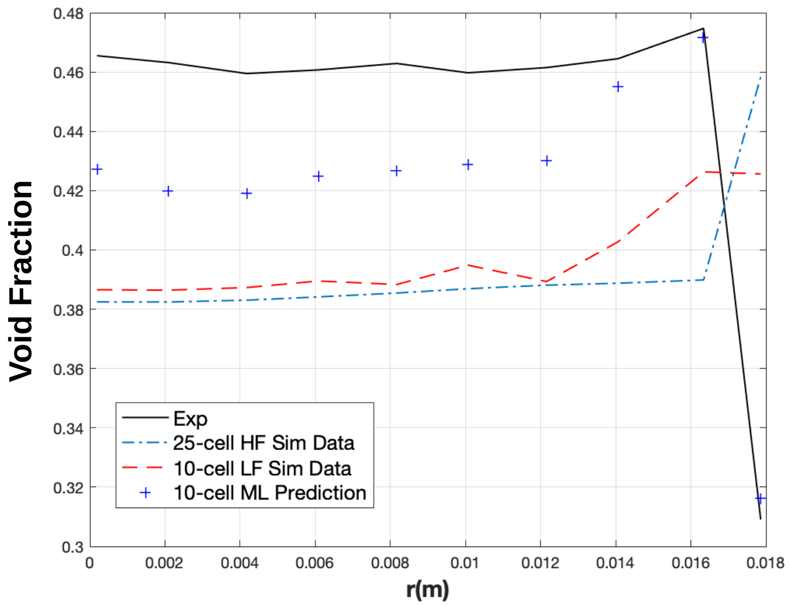
Black Lines: Experimental Data
Green Dashes: 25-cell Simulation Results
Red Dashes: 10-cell Simulation Results
Blue Points: ML Prediction



Vapor Velocity



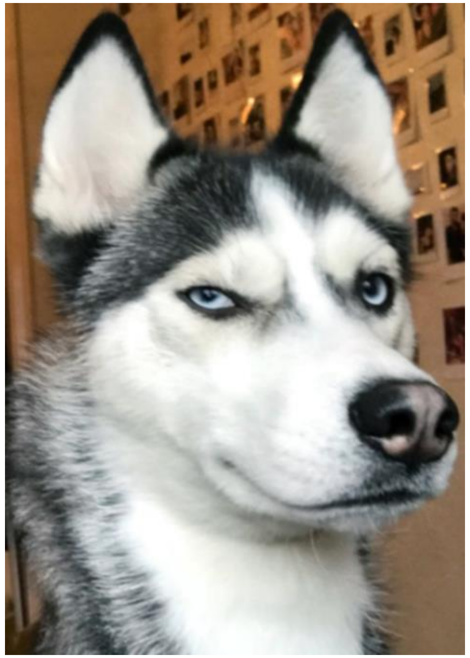
Liquid Velocity



Void Fraction

Test 2: Local Data Similarity Measurement and Enhancement (I)

- **One of the key hypotheses of FSM:** Predictive performance of FSM will be improved with the increase of data similarity between training and testing data.
- Several “most similar” data points from original training dataset will be selected to construct **an optimal training dataset with a smaller size, but higher similarity**.



* All photos from Google.

Test 2: Local Data Similarity Measurement and Enhancement (II)

- By measuring the data distances between each data point in testing case ($Q = 10$) and training cases ($M = 1845$), P data points with small values of data distance are selected for each target data point.
- There are totally $Q * P$ data points selected to build a new training dataset.

$$D_{m,q} = d(PF_m, PF_q) = \sqrt{\sum_{k=1}^{N} (x_{tr,k} - x_{ta,k})^2}$$

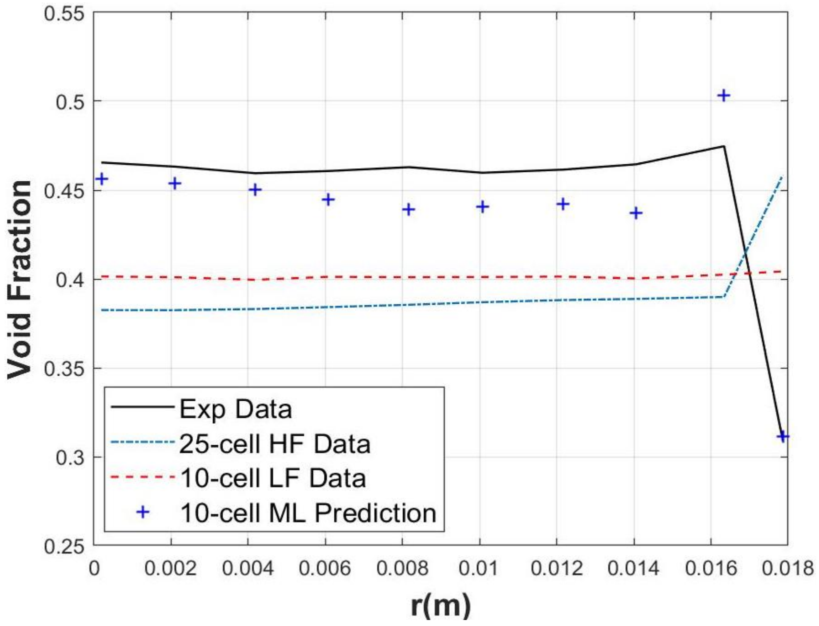
- $D_{m,q}$ is defined as the Euclidean distance between training data point PF_m and target data point PF_q, $(1 \leq m \leq M \; and \; 1 \leq q \leq Q)$.
- N is the number of physical features.
- $x_{tr,k}$ and $x_{ta,k}$ are respectively the values of physical feature number k of PF_m and PF_q.
- Values of physical features should be firstly standardized into [-1, 1].

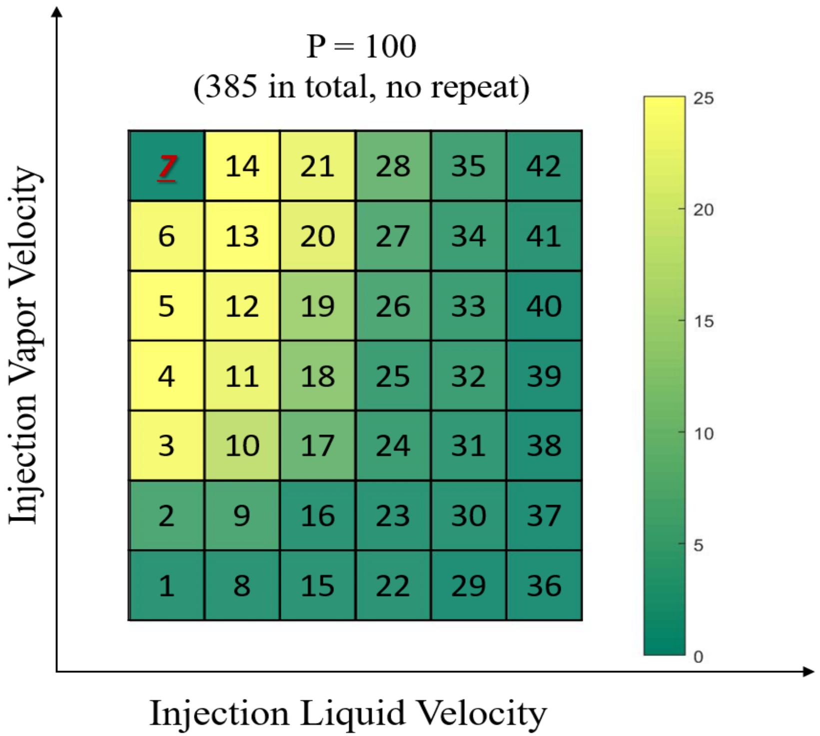
Test 2: Local Data Similarity Measurement and Enhancement (III)

Target Case: Case 7 10-cell mesh configuration						
# of target-similar data points for each target data point (P)	# of training data points from other 41 cases (no repeat)				Data similarity (S_{KDE})	NRMSE of prediction
	10-cell	15-cell	20-cell	Total		
50	143	75	8	226	0.4228	0.0578
100	196	165	20	385	0.3867	0.0409
200	247	290	107	664	0.3638	0.0493
400	314	431	293	1038	0.3255	0.0485
600	331	551	473	1315	0.2980	0.0539
800	367	569	557	1493	0.2820	0.0553
All	410	615	820	1845	0.2390	0.0600
Only 15-cell and 20-cell	0	615	820	1435	0.2147	0.0622
Original LF simulation	NA				NA	0.1487

Test 2: Local Data Similarity Measurement and Enhancement (IV)

P = 100 (385 data points from 1845)

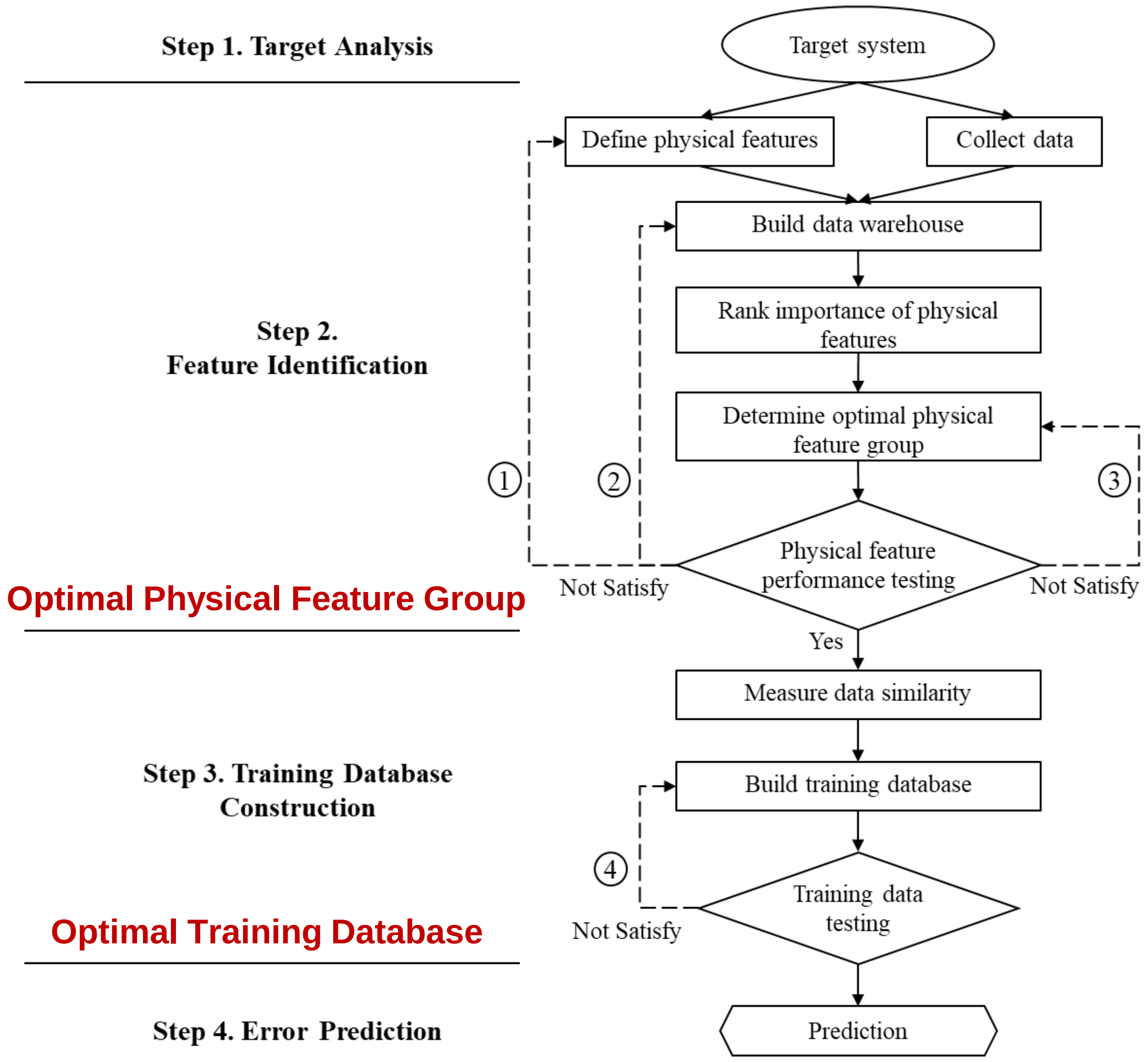




- All 41 cases are involved in the training-data selection.
- Even though global conditions of some cases (e.g., Case 36) are quite different from those of Case 7, they still have some local data points which are more similar than some of globally similar cases (e.g., Case 6).
- It denotes that even globally some cases are not similar to target case, but some locally similar data points of these cases can still be used to inform the prediction of target case.
- **Instead of choosing training data based on global similarity, local similarity is used as the metric to select "optimal" training data.**
- In this way, all existing data can be sufficiently utilized for specific targets.

Case Study part II: applying FSM for computationally efficient CFD prediction

- A target case
  - *Two-phase pipe flow*
- Validation data
  - *Fine-mesh simulation results*
- A CFD code
  - *Star CCM+*



Goal: Efficiently and Accurately Predict Behaviors of Case 35

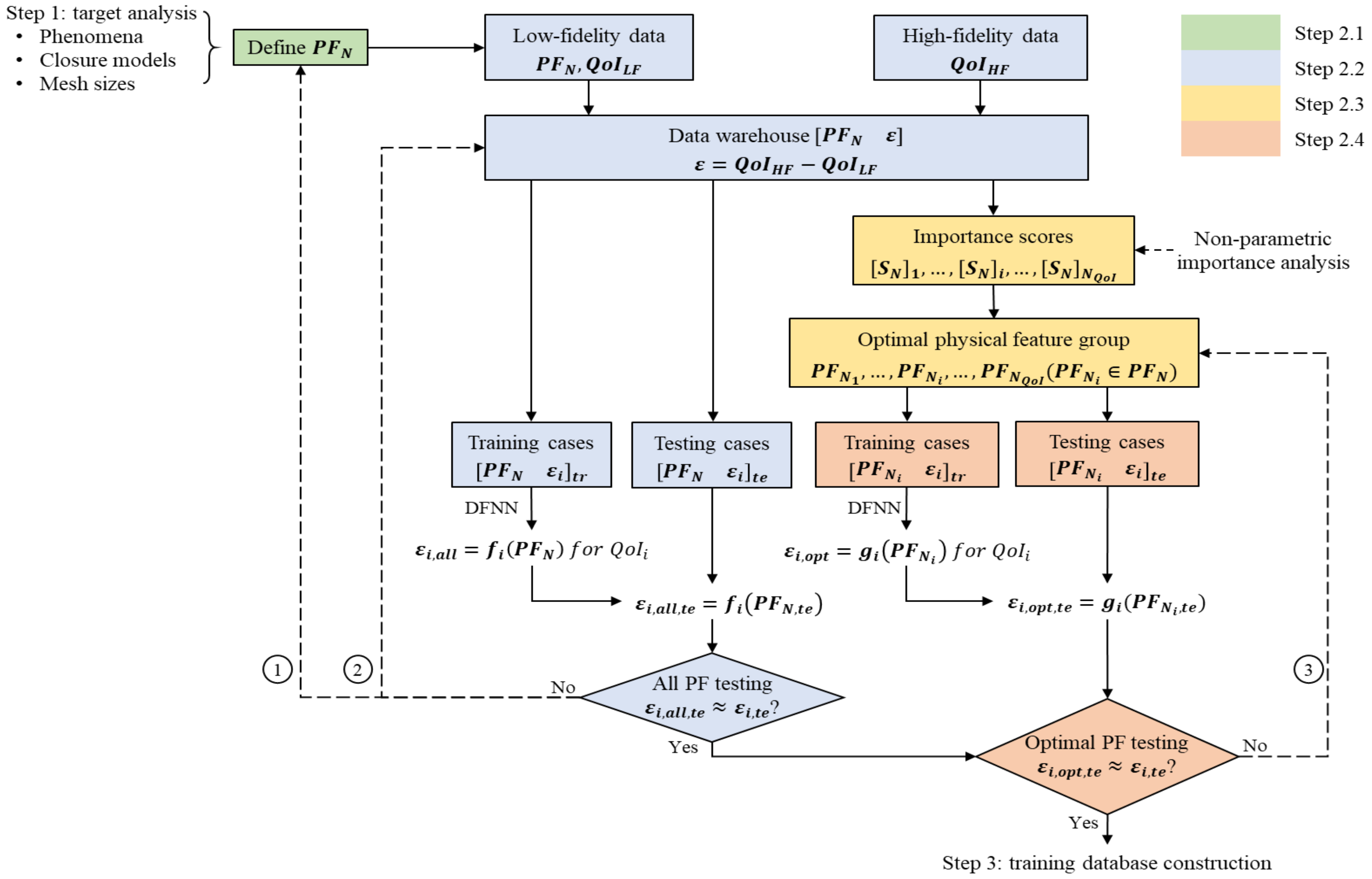
Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction	Set	Liquid rate (m/s)	Vapor rate (m/s)	Void fraction
1	0.376	0.027	0.0407	22	0.974	0.027	0.0204
2	0.376	0.067	0.1167	23	0.974	0.067	0.0514
3	0.376	0.112	0.1843	24	0.974	0.112	0.0791
4	0.376	0.18	0.2449	25	0.974	0.18	0.1242
5	0.376	0.23	0.3079	26	0.974	0.23	0.1512
6	0.376	0.293	0.3657	27	0.974	0.293	0.1869
7	0.376	0.347	0.4168	28	0.974	0.347	0.2108
8	0.535	0.027	0.0312	29	1.087	0.027	0.0176
9	0.535	0.067	0.0877	30	1.087	0.067	0.0473
10	0.535	0.112	0.1406	31	1.087	0.112	0.0737
11	0.535	0.18	0.2016	32	1.087	0.18	0.1096
12	0.535	0.23	0.2344	33	1.087	0.23	0.1497
13	0.535	0.293	0.3102	34	1.087	0.293	0.1777
14	0.535	0.347	0.3398	35	1.087	0.347	0.1976
15	0.753	0.027	0.0235	36	1.391	0.027	0.0148
16	0.753	0.067	0.0622	37	1.391	0.067	0.0387
17	0.753	0.112	0.1091	38	1.391	0.112	0.0581
18	0.753	0.18	0.1554	39	1.391	0.18	0.0964
19	0.753	0.23	0.1816	40	1.391	0.23	0.1176
20	0.753	0.293	0.2381	41	1.391	0.293	0.1504
21	0.753	0.347	0.2692	42	1.391	0.347	0.1724

- Case 35 was selected as target case because,
  - It has the highest injection vapor velocities and a second highest injection liquid velocity which makes it become **one of the extrapolative cases**.
  - It has the **highest simulation errors** between high-fidelity simulation results and low-fidelity simulation results.
- Core hours of 10-cell simulation are 8 while core hours of 25-cell simulation are 54.
- **Strategy:**
  - Identify optimal physical features and construct optimal training database.
  - Train a DFNN model.
  - Predict simulation errors of 10-cell simulation for Case 35.

Step 1: Target Analysis

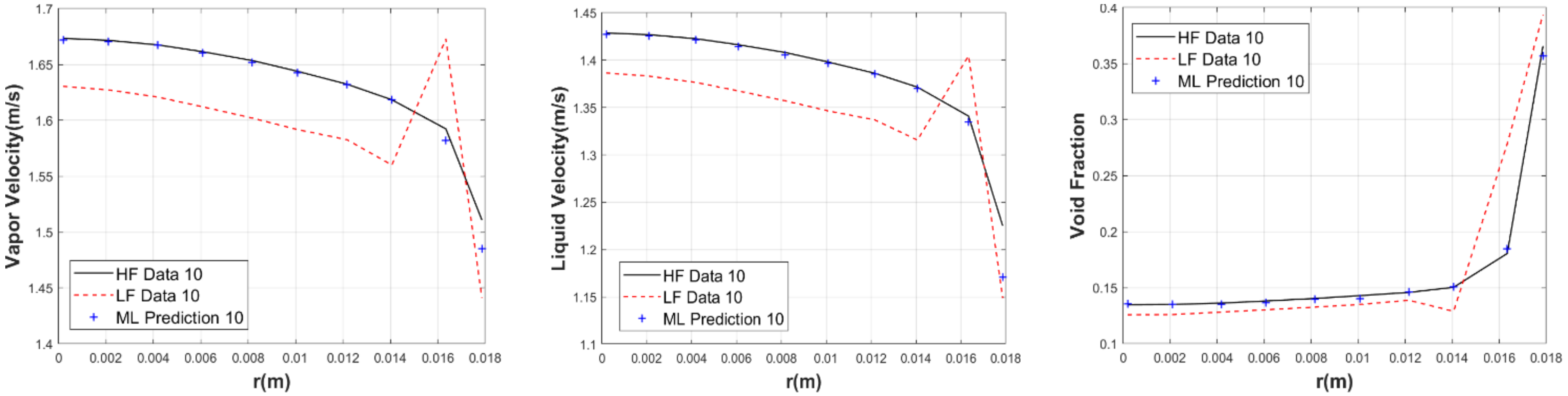
- **Key phenomena:** Two-phase pipe flow.
- **Closure models:**
  - Two-phase interfacial forces closures including drag force model and lift correction (BAMF model).
  - Turbulence models including turbulent dispersion force and the standard k-ε turbulence model.
- **QoIs:** liquid velocity, vapor velocity and void fraction.
- **Target case:** Case 35 10-cell simulation (10 points)
- **Training cases:** other 41 cases (1845 data points)

Step 2: Feature Identification (I)



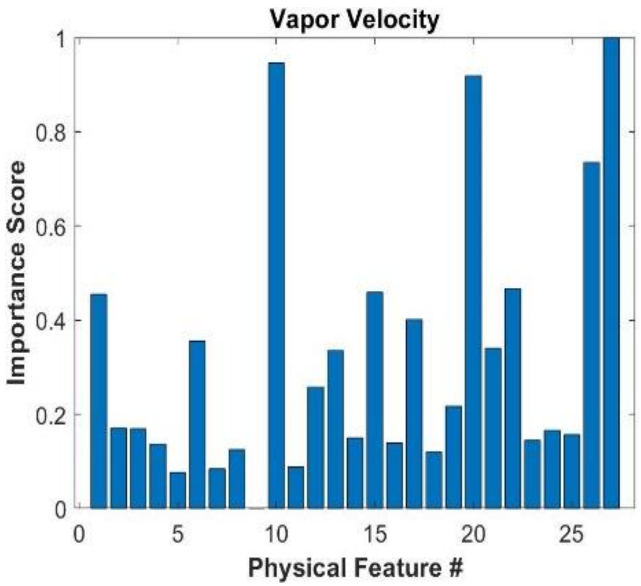
Step 2: Feature Identification (II)

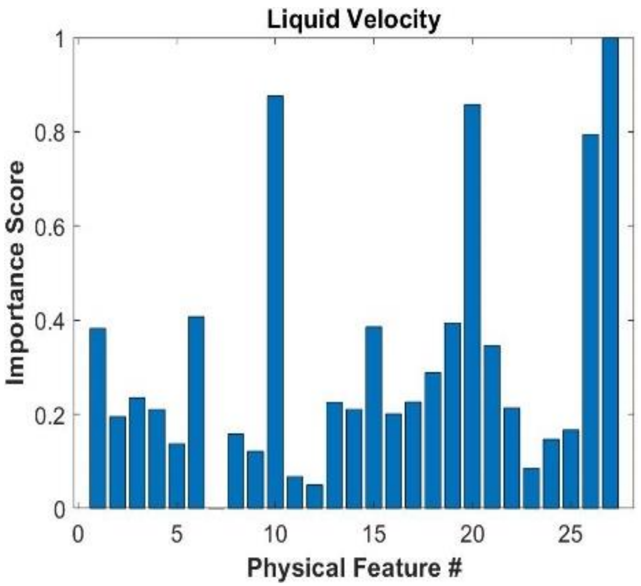
- **Step 2.1: define potential physical features**
- **Step 2.2: collect data and build data warehouse**
  - Only 41 cases are available
  - Test: Case 34 as the testing case and other 40 cases as training cases
  - 20-20-20 DFNN with 27 inputs and 3 outputs
  - The results indicate that these defined local physical features can represent local physics and provide sufficiently accurate prediction on simulation error of QoIs.
  - FSM represents good predictive capability on estimating the local simulation error even for the extrapolation of global physics (vapor injection rate in this test).

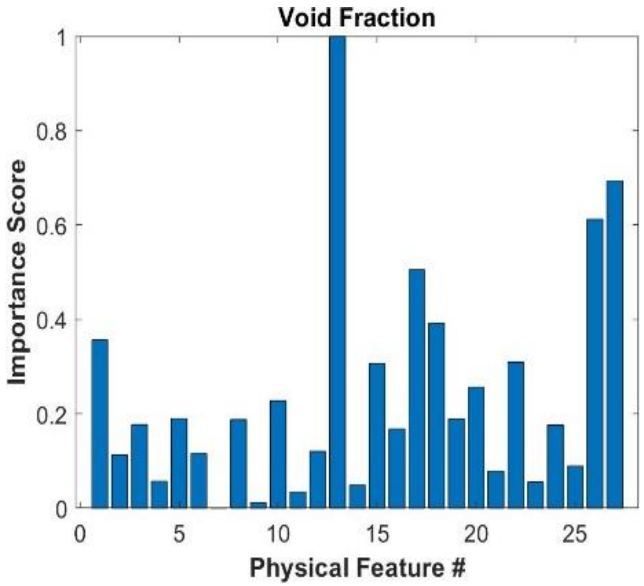


Step 2: Feature Identification (III)

- **Step 2.3: rank importance of physical features**
  - By applying RFR algorithm, the importance scores of all defined potential physical features for each QoI are quantified and ranked.
  - A greater score implies higher level of importance.
  - According to their importance scores for each QoI, physical features are classified in four levels:
    - Level 1 (score ≥ 0.2)
    - Level 2 (0.2 > score ≥ 0.15)
    - Level 3 (0.15 > score ≥ 0.1)
    - Level 4 (0.1 > score ≥ 0).







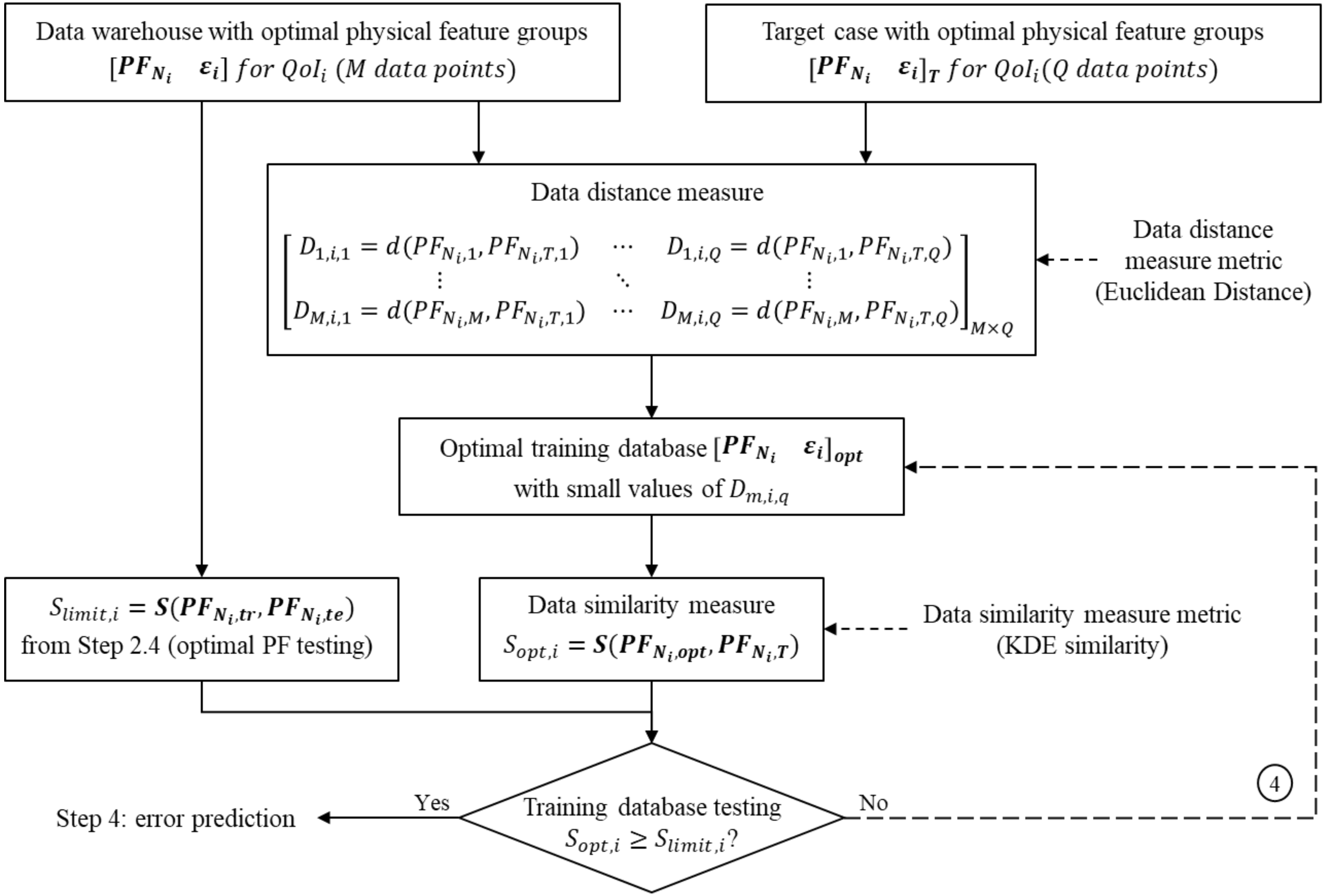
Step 2: Feature Identification (IV)

- **Step 2.4: determine optimal physical feature group**

QoI	Group #	Number of physical features	NRMSE (10 cells)	Training Time (core-hours)	Optimal physical feature	Number
$\mathbf{u_g}$	G1 (Level 1)	13	0.0108	0.35	$\frac{d\alpha}{dx}, \frac{du_g}{dx}, \frac{dk_g}{dx}, \frac{d^2\alpha}{dx^2}, \frac{d^2P}{dx^2}, \frac{d^2k_g}{dx^2}, \frac{d^2k_l}{dx^2}, \frac{d^2\varepsilon_l}{dx^2}$ $I_l, I_g, , We, Re_\Delta, R_b, R_g, R_l, R_\mu, r_l$	17
	G2 (Level 1~2)	17	0.0046	0.45		
	G3 (Level 1~3)	22	0.0076	0.85		
	G4 (Level 1~4, all)	27	0.0054	1.50		
	Original low-fidelity simulation		0.0341	8		
	25-cell simulation		-	54		
$\mathbf{u_l}$	G1 (Level 1)	17	0.0281	0.45	$\frac{d\alpha}{dx}, \frac{du_g}{dx}, \frac{du_l}{dx}, \frac{dP}{dx}, \frac{d\varepsilon_g}{dx}, \frac{d^2\alpha}{dx^2}, \frac{d^2u_g}{dx^2}, \frac{d^2k_g}{dx^2}, \frac{d^2k_l}{dx^2}, \frac{d^2\varepsilon_l}{dx^2}, \frac{d^2P}{dx^2}$ $I_l, We, Re_b, Re_\Delta, R_b, R_g, R_l, R_\mu, r_l$	20
	G2 (Level 1~2)	20	0.0050	0.50		
	G3 (Level 1~3)	23	0.0107	0.70		
	G4 (Level 1~4, all)	27	0.0079	1.50		
	Original low-fidelity simulation		0.0389	8		
$\boldsymbol{\alpha}$	G1 (Level 1)	10	0.1405	0.60	$\frac{d\alpha}{dx}, \frac{du_g}{dx}, \frac{du_l}{dx}, \frac{dk_g}{dx}, \frac{dk_l}{dx}, \frac{d^2\alpha}{dx^2}, \frac{d^2u_g}{dx^2}, \frac{d^2k_g}{dx^2}, \frac{d^2k_l}{dx^2}, \frac{d^2\varepsilon_l}{dx^2}, \frac{d^2P}{dx^2}$ $I_l, I_g, We, Re_b, Re_\Delta, R_b, R_l, R_\mu$	19
	G2 (Level 1~2)	16	0.0744	0.85		
	G3 (Level 1~3)	19	0.0301	0.90		
	G4 (Level 1~4, all)	27	0.0257	1.50		
	Original low-fidelity simulation		0.2013	8		

Step 3: Training Database Construction (I)

- This step aims to answer the question: which kind of data in the data warehouse should be used as the training database?



Step 3: Training Database Construction (II)

QoI	Data similarity (S_{KDE})	Data source (1000 points in total)			Actual data quantity (no repeat)			Number of involved cases
		10-cell	15-cell	20-cell	Total	10-cell	15-cell	
u_g	**0.4499**	733	267	0	442	267	146	36
u_l	**0.3902**	666	318	16	495	282	197	35
α	**0.5355**	884	116	0	372	281	91	33

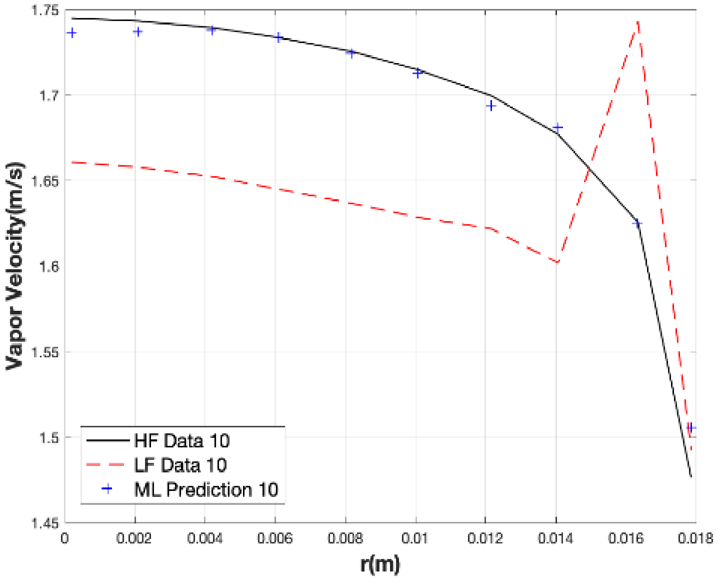
Data selection based on local similarity

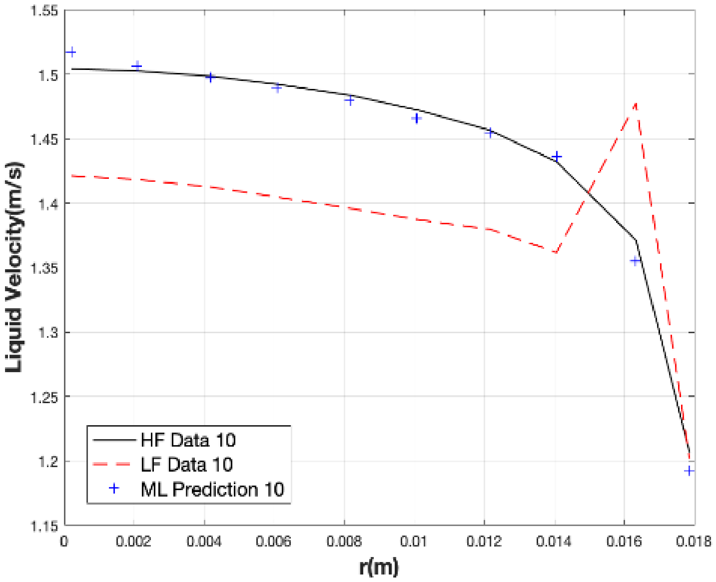
Backup training database #	Number of cases	Physical feature data similarity with Case 35 for QoIs (S_{KDE})			Description of cases included
		u_g	u_l	α	
1 (a), (c), (e)	18	0.2828	0.3758	0.4198	(a) Same u_g (0.347); (b) Similar u_g (0.293~0.347); (c) Same u_l (1.087); (d) Similar u_l (0.974~1.391); (e) Similar α (0.17~0.25) (f) Similar α (0.12~0.35)
2 (a), (c), (f)	29	0.2708	0.3473	0.3978	
3 (a), (c), (f)	25	0.2864	0.3830	0.4291	
4 (a), (d), (f)	33	0.2727	0.3548	0.4058	
5 (b), (c), (e)	21	0.2833	0.3776	0.4240	
6 (b), (d), (e)	31	0.2718	0.3512	0.4019	
7 (b), (c), (f)	26	0.2859	0.3835	0.4300	
8 (b), (d), (f)	34	0.2726	0.3560	0.4072	
9 All cases	41	0.2695	0.3488	0.3988	
Case study in Step 2.4 ($S_{limit,i}$)		0.2722	0.2292	0.3217	

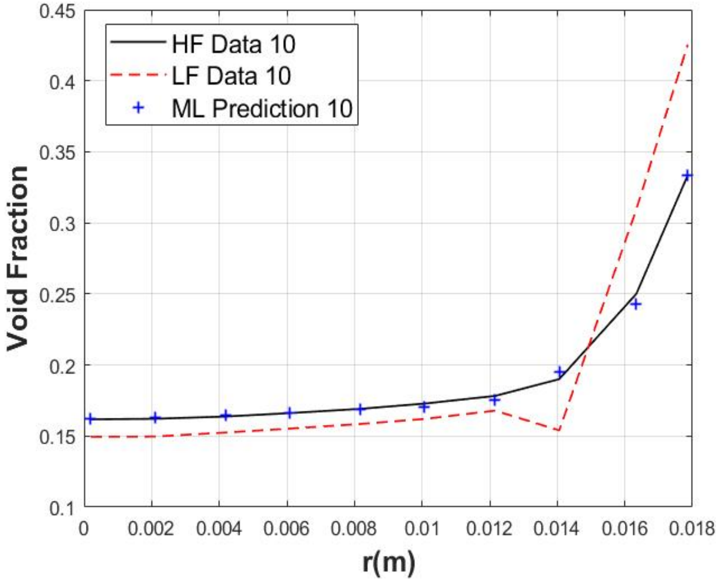
Data selection based on global similarity

Step 4: Error Prediction for Coarse-mesh CFD

- By using optimal physical feature group obtained in Step 2 and optimal training database obtained in Step 3, error prediction of local QoIs can be performed for Case 35.



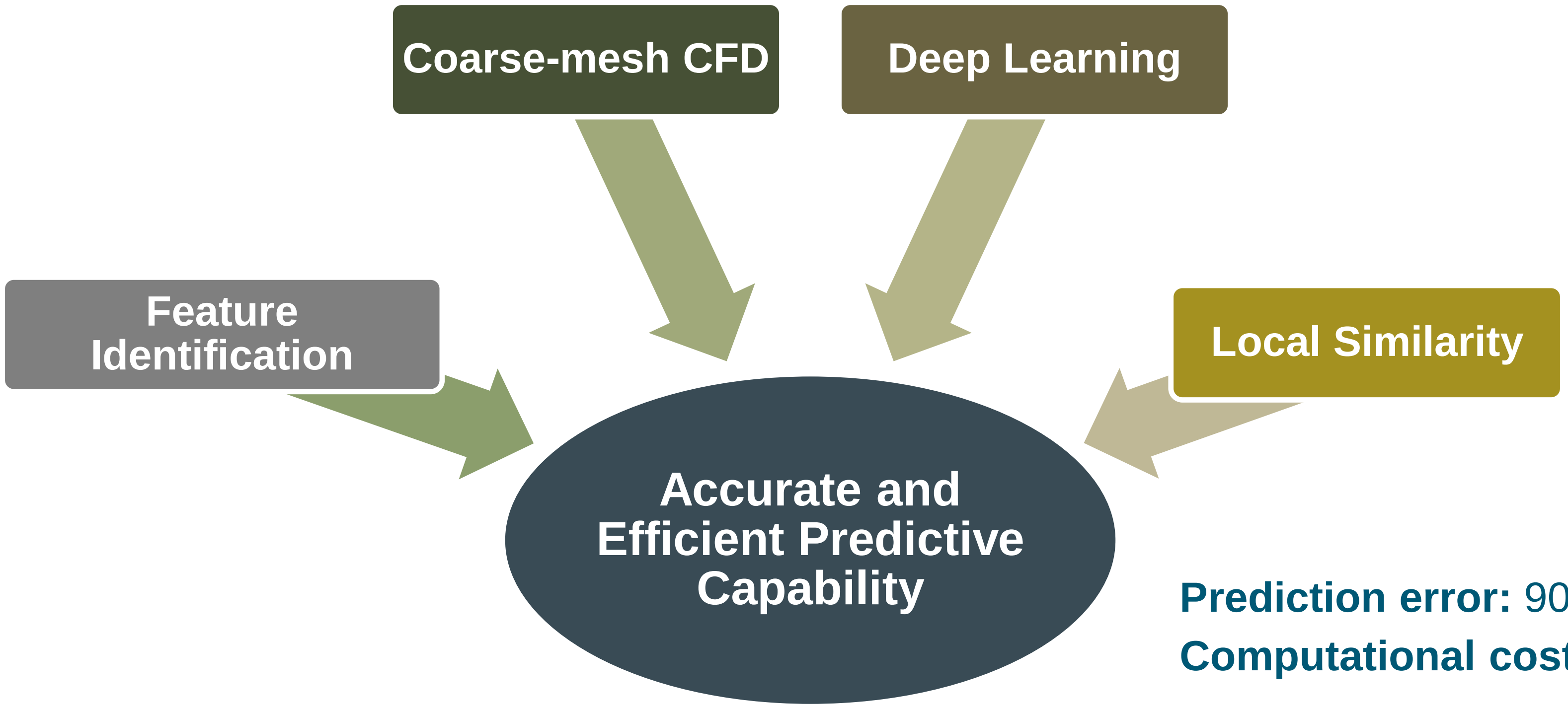




	$NRMSE_{u_g}$	$NRMSE_{u_l}$	$NRMSE_{\alpha}$	Mesh configuration
DFNN prediction	0.0057	0.0059	0.0158	10 cells
Low-fidelity simulation	0.0499	0.0564	0.1934	

Conclusions

- This paper demonstrated a data-driven approach **Feature Similarity Measurement (FSM)** to estimate simulation errors using coarse-mesh CFD to achieve a comparable accuracy as **fine-mesh CFD simulations or experimental data**.
- The **predictive performance** of the FSM approach has been investigated and applied to realize **computationally efficient CFD predictions** based on a **two-phase bubbly flow** case study.



Prediction error: 90% reduced from low-fidelity simulation

Computational cost:~7% of High-fidelity Simulation

Acknowledgement

- This work is supported by the **U.S. Department of Energy**, under Department of Energy Idaho Operations Office Contract DE-AC07-05ID14517. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.
- This work was originated from **INL** National University Consortium (**NUC**) program, INL Laboratory Directed Research & Development (**LDRD**) program under DOE Idaho Operations Office Contract DE-AC07-05ID14517.
- The preliminary work was also supported by U.S. Department of Energy’s Consolidated Innovative Nuclear Research program via the Integrated Research Project (**IRP**) on “Development and Application of a Data-Driven Methodology for Validation of Risk-Informed Safety Margin Characterization Models” under the grant DE-NE0008530.
- I am very grateful for the continuous supports from **Prof. Nam Dinh at NC State and Dr. Hongbin Zhang at INL.**
- I also would like to acknowledge the kindly support and computational resources from **Baglietto CFD Research Group at MIT**.

Thank you!