



RIME-RF-RIME: A novel machine learning approach with SHAP analysis for predicting macroscopic permeability of porous media

Viet-Hung Phan¹, Hai-Bang Ly^{2,*}

¹University of Transport and Communications, Hanoi 100000, Vietnam

²University of Transport Technology, Hanoi 100000, Vietnam

Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2024.en.4.1.58-71>

*Corresponding author:

E-mail address:

banglh@utt.edu.vn

Received: 10/03/2024

Revised: 29/03/2024

Accepted: 30/03/2024

Abstract: Predicting the macroscopic permeability of porous media is critical in various scientific and engineering applications. This study proposes a novel model that combines Random Forest (RF) and rime-ice (RIME) optimization algorithm, denoted RIME-RF-RIME, to predict permeability based on six key features covering fluid phase dimensions, geometric characteristics, surrounding phase permeability, and media porosity. After the input space simplification process using RIME, the RF model achieves high predictive accuracy with a coefficient of determination (R^2) of 0.980. Furthermore, SHapley Additive exPlanations (SHAP) values are employed to decipher these features' importance and interaction effects on the model's predictions. The analysis reveals that porosity, permeability of the porous phase, and the size of the fluid phase perpendicular to the flow direction exert the most significant individual influences. This study not only unveils crucial insights into the underlying mechanisms governing permeability in porous media but also contributes to developing interpretable and reliable predictive models for related applications.

Keywords: Fluid flow; Permeability; Machine Learning; Random forest; SHAP values.

1. Introduction

In civil engineering, understanding fluid flow within porous media holds immense significance, impacting the design and performance of structures like dams, foundations, and underground reservoirs and informing strategies for geotechnical engineering and environmental remediation. Infiltration systems, ranging from clean water purification to life-saving drug manufacturing, precise separation hinges on optimal permeability [1]. Accurately predicting permeability empowers the design of efficient filters, minimizing energy consumption and maximizing throughput [2,3]. Beyond filtration, the energy sector critically relies on the intricacies of

permeability within porous rock formations [4]. Vast underground reservoirs hold valuable resources like oil and gas, and their extraction efficiency is directly tied to the flow characteristics governed by permeability [5]. Precise permeability predictions enable researchers to inform extraction strategies, maximizing resource utilization while minimizing environmental impact [6]. Furthermore, groundwater contamination, a significant threat to global water security, necessitates a deep understanding of flow dynamics within porous media. Predicting permeability allows researchers to map and manage groundwater flow, enabling proactive measures to prevent pollutant spread and safeguard this precious resource [7,8]. This

knowledge facilitates targeted interventions for containment and remediation in contaminated areas, ensuring the long-term sustainability of clean water supplies for future generations [9].

Empirical correlations, like those explored in Bachmat and Bear [10,11], Ergun [12], and Scheidegger [13,14], offer an established and efficient approach to predict fluid flow in porous media. Their broad applicability across diverse fields and established presence make them familiar and convenient tools. However, these correlations can be susceptible to inaccuracies in complex geometries, diverse flow regimes, and situations with inherent uncertainties in porous media properties. Thus, moving beyond the limitations of empirical correlations often requires alternative methods.

Traditional simplified models often struggle to capture the full complexity of multiphase interactions, turbulence, and diverse pore geometries, leading to inaccurate predictions. This is where the power of numerical simulations emerges. Advanced methods like the Finite Element Method (FEM) [15], Fast Fourier Transform (FFT) [16,17], and Lattice Boltzmann (LB) [18] methods offer a paradigm shift in permeability prediction. These methods reconstruct intricate pore geometries and diverse flow phenomena, transcending the limitations of simplified models. Their inherent flexibility allows them to adapt to diverse porous media and flow conditions, paving the way for highly accurate permeability predictions [19]. For instance, FEM simulations have been successfully employed to investigate the impact of pore geometry on permeability in fractured rock, providing insights for optimizing oil and gas extraction [20,21]. Similarly, LB simulations have shed light on multiphase flow phenomena in porous media, crucial for understanding processes like CO₂ sequestration [22–24].

However, achieving accurate predictions is not without its hurdles. The computational

demands of these methods can be substantial, requiring access to high-performance computing resources, especially for large or complex systems [25]. Furthermore, obtaining accurate representations of pore geometry often hinges on detailed experimental data, which can be scarce or challenging to acquire [26]. Additionally, choosing the appropriate numerical parameters is critical, demanding expertise and careful calibration, as even minor variations can significantly impact the results [27]. While offering greater interpretability than simpler models, understanding the intricate numerical computations within these methods can be challenging, requiring specialized knowledge. Finally, scaling these simulations to encompass vast porous systems remains an ongoing challenge, demanding advanced techniques and computational resources [28].

The traditional toolbox of civil engineering is getting a digital upgrade with artificial intelligence-powered machine learning (AI-ML). This transformative technology tackles complex challenges like earthquake simulations [29,30] material property prediction [31,32], and structural engineering [33,34], leading to optimized designs and informed decision-making. Indeed, advancements in AI-powered structural analysis hold promise for the design of bridges capable of withstanding extreme loads [35]. Similarly, AI contributes to the development of sustainable construction materials through its remarkable design accuracy [36,37]. Even project management benefits from AI's predictive capabilities, leading to optimized schedules and budgets [38]. Notably, the integration of advanced ML algorithms offers the ability to learn from vast datasets encompassing fluid-porous media interactions. This enables remarkably accurate predictions of fluid behavior within complex structures [39,40]. In geotechnical engineering, AIML-driven simulations can predict soil-water interactions with unprecedented accuracy, leading to safer foundation design and slope stability

analyses [41]. Similarly, environmental remediation efforts can benefit from targeted interventions guided by AI-modeled contaminant transport, minimizing environmental impact and optimizing cleanup strategies [42]. Despite the promise of ML in predicting fluid flow, limitations persist. Scarce and biased data hinders training, while complex models lack interpretability and require significant computational resources. Additionally, capturing intricate geometries remains a challenge. Overcoming these limitations requires further research, paving the way for more reliable, efficient, and generalized ML-driven fluid flow predictions across diverse applications.

This work addresses the limitations of previous studies by employing ML and readily available data to predict macroscopic permeability. While the dataset consists of 1728 data points

derived from FEM simulations, the innovative approach utilizes XGBoost to achieve accurate predictions. The model leverages various inputs, including porous phase geometry, flow behavior, and porosity. Notably, it employs a novel ML algorithm to identify the most impactful features before model development, leading to a more efficient and accurate prediction process. Following rigorous verification and evaluation, the optimal model is selected for further analysis. Finally, a comprehensive sensitivity analysis is conducted to assess the influence of each input on the prediction outcome, providing valuable insights into the key factors governing permeability and enhancing our understanding of the underlying physical processes.

2. Materials and Methods

2.1. Database acquisition

Table 1. The parameters used in the development of the ML model

	Min	Median	Average	Max
X ₁	0.05	0.25	0.25	0.45
X ₂	0.05	0.25	0.25	0.45
X ₃	0.3	1.5	13.3	50
X ₄	0.3	1.5	13.3	50
X ₅	1e-6	1e-5	3.7e-5	1e-4
X ₆	0.022	8.827	16.959	80.956
Y	0.59	25.00	27.13	79.10

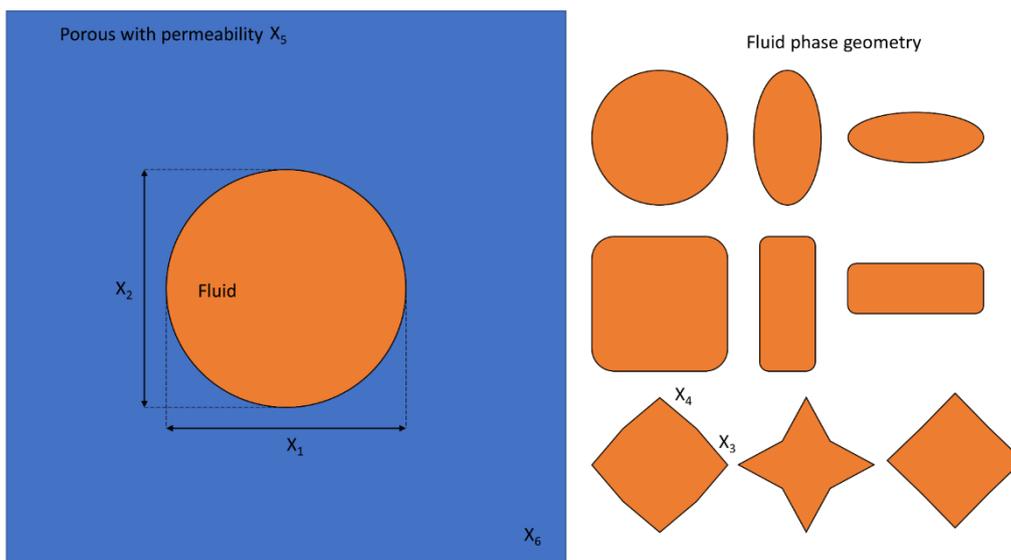


Fig. 1. Representation of the fluid flow problem in the unit cell

This study utilizes a dataset sourced from the literature [43]. The dataset encompasses 1728 results obtained from finite element method (FEM) simulations of fluid flow through a unit cell. Periodic boundary conditions are imposed on the unit cell to calculate the macroscopic permeability. The simulated flow scenario involves two distinct flow directions, aligned with the O_x and O_y axes. The dataset comprises six input variables (Table 1):

- X_1 and X_2 : Dimensions of the fluid phase within the unit cell, corresponding to the O_x and O_y axes, respectively.
- X_3 and X_4 : Geometric characteristics of the fluid phase.
- X_5 : Permeability of the porous phase surrounding the fluid phase.
- X_6 : Porosity of the unit cell or the porous media.

Notably, the unit cell employed in this study is unidimensional (size 1 x 1). Consequently, all input variables are unitless. While the simulations generate macroscopic permeability values for both O_x and O_y directions, only a single permeability value (denoted Y) is retained for analysis due to the considered symmetric flow. Readers seeking a more detailed description of the dataset and the underlying flow problem (Figure 1) are encouraged to refer to [43].

2.2. ML methods

Introduced by Leo Breiman in 2001, Random Forest (RF) stands as a powerful ensemble learning technique within the ML domain [44]. Its core principle lies in combining the predictions of multiple decision trees, ultimately leading to enhanced accuracy and robustness compared to individual trees. Regarding its operational mechanism, RF employs bootstrap aggregation (bagging) to generate numerous training subsets by drawing random samples (with replacement) from the original dataset. Each of these subsets then serves as the training ground for a distinct decision tree. At each node within these trees, a random subset of features is chosen from the

entire pool, and the optimal split amongst these features is identified to partition the data into child nodes. This inherent randomness introduced at every step contributes to reducing the correlation between trees, thereby mitigating the risk of overfitting. In the context of regression problems, after all trees have completed training, new data points are evaluated by averaging the predictions made by each individual tree. Notably, RF has garnered significant success in various civil engineering applications, encompassing Structural health monitoring, Material property estimation, Landslide susceptibility mapping and Traffic flow prediction. In conclusion, RF's ensemble approach, user-friendliness, and interpretability position it as a valuable ML tool capable of tackling complex prediction problems across diverse fields, including but not limited to the aforementioned civil engineering applications.

The physical phenomenon of the rime-ice (RIME) optimization algorithm, introduced by Su et al. in 2023 [45], is a novel optimization technique drawing inspiration from the captivating phenomenon of rime ice formation. Rime ice, characterized by its intricate crystalline structures, forms when supercooled water droplets solidify upon contact with a cold object. Regarding working mechanism, RIME mimics the two-stage growth process of rime ice, offering a compelling approach to optimization. **Soft-Rime Search:** This initial phase, akin to the feathery layer of young rime ice, utilizes a population of candidate solutions and evaluates their fitness. The algorithm leverages random exploration and information exchange between individuals to progressively refine these solutions. **Hard-Rime Puncture:** Drawing inspiration from the denser, crystalline structure of mature rime ice, this phase focuses on exploitation. The best solutions identified in the previous stage undergo mutation and recombination, aiming to discover improved solutions and escape potential local optima. Overall, RIME offers a fresh perspective on problem-solving through

optimization. Its balanced exploration and exploitation capabilities make it well-suited for tackling complex and challenging search landscapes. As research in this area continues to evolve, RIME's potential contributions to various fields, including civil engineering, are likely to become more significant.

2.3. Sensitivity analysis

SHAP (SHapley Additive exPlanations) values, introduced by Lundberg and Lee in 2017 [46], have become a fundamental tool for enhancing interpretability and transparency in ML. This innovative approach, grounded in game theory, offers a quantified and principled understanding of how individual features contribute to a specific prediction made by any ML model. Drawing inspiration from Shapley values within cooperative game theory, SHAP values estimate the marginal contribution of each feature to the model's prediction. This is achieved by considering all possible feature combinations, thereby revealing not only the individual influence of each feature but also its collective impact when interacting with other features. SHAP values can identify features that exert the most significant influence on model predictions. This crucial information enables engineers to prioritize resources more effectively and make informed decisions based on a deeper understanding of the factors governing the prediction task. By highlighting features that warrant further refinement or removal, SHAP values contribute to an iterative improvement process, ultimately

enhancing the performance and generalizability of ML model. By providing a rigorous and interpretable framework for understanding feature contributions within ML models, SHAP values empower researchers across various fields to confidently utilize and explain complex predictions. Their versatility and effectiveness have firmly established them as a responsible and trustworthy AI cornerstone.

2.4. Model metrics

R^2 (coefficient of determination) is a widely used metric for evaluating the proportion of variance in a dependent variable (target) explained by an independent variable (model prediction). R^2 values range from 0 to 1, with a value of 0 explaining no variance in the target variable (poor performance) and a value of 1 explaining all the variance in the target variable (perfect fit). RMSE (root mean squared error) measures the average magnitude of the difference between predicted and actual values. Lower RMSE values indicate better agreement between predictions and actual observations. Unlike R^2 , RMSE has a precise unit of measurement corresponding to the scale of the data, allowing for a more straightforward interpretation of the error magnitude. In conclusion, R^2 and RMSE are used in this study because they provide helpful information about model performance. These metrics can be calculated in the literature [47,48].

3. Results and discussions

3.1. Feature reduction analysis

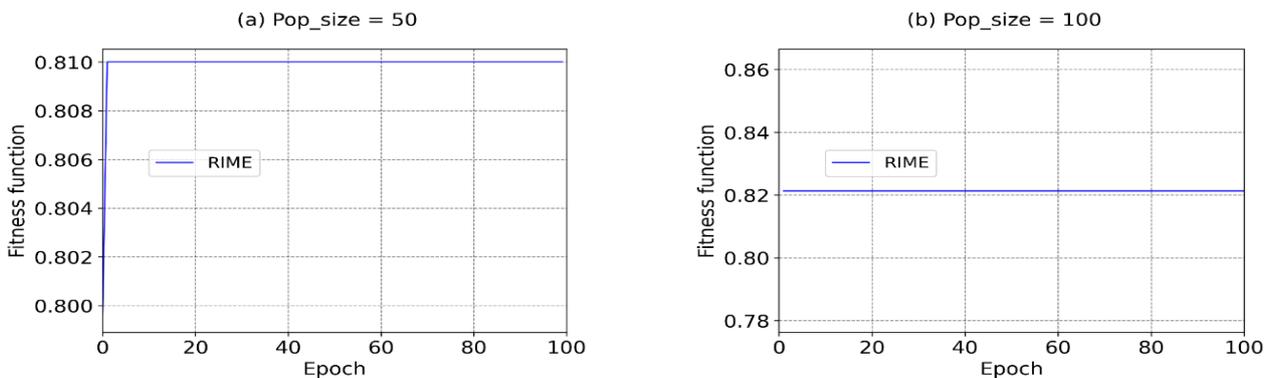


Fig. 2. Feature selection using RIME for the fluid flow problem: (a) pop=50, (b) pop=100

Table 2. Summarized results of feature selection process using RIME

Model	Pop_size	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Fitness function	Objective function
RIME	50		X			X	X	0.810	0.956
RIME	100		X			X	X	0.821	0.968

This section describes the process of analyzing and reducing features used for regression tasks. A metaheuristic optimization algorithm called RIME is employed for this purpose. Two population sizes (pop) of 50 and 100 are investigated, while the epoch is set to 100 for simplicity. The prediction model used for evaluation is an RF with default hyperparameters from the scikit-learn library. The fitness function considers two aspects: R^2 of the prediction and the reduction in the number of features compared to the original set. Both criteria are weighted equally (1-1) in the fitness function. The results of this analysis are presented in Figure 2.

For pop = 50, the fitness function, which combines R^2 and the reduction in input features, reaches a value of 0.81 after only 3 epochs. However, this value is surpassed by the case of pop = 100, where the fitness function reaches 0.821. This suggests that a larger population size improves the search process and leads to better solutions. Furthermore, the results indicate that no significant improvement in the fitness function is observed after a certain number of epochs. This suggests that the algorithm has converged and found a near-optimal solution. Finally, the analysis identifies a reduced input space of 3 features, compared to the original set of 6. This represents a significant reduction in dimensionality, achieved while maintaining high prediction accuracy. The fitness function value at convergence (0.821) indicates a good balance between feature reduction and prediction performance. Additionally, the objective function (R^2) reaches 0.968, demonstrating the excellent prediction capability of the RF model using only 3 selected features. Table

2 summarizes the detailed results of the feature reduction process, including the identified features and corresponding parameters.

The feature reduction process using RIME identified three key features (X_2 , X_5 , X_6) from the original set of six for predicting the target variable. This selection is supported by both ML and physical considerations. Firstly, X_2 , representing the size of the porous fluid perpendicular to the flow direction, is a relevant choice from a physical perspective. This feature likely captures important information related to the flow dynamics and contributes significantly to the permeability. Secondly, X_5 , corresponding to the permeability of the porous media, is well-known for its crucial role in governing the overall system behavior. Including this feature ensures that the model captures the fundamental physical characteristics relevant to the target variable, the macroscopic permeability. Lastly, X_6 , representing the porosity of the media, is considered, along with X_2 , to be a representative group of input features. While not directly related to X_1 , X_3 , and X_4 , their combined influence likely captures the overall pore structure and indirectly affects the target variable. This selection strategy demonstrates an understanding of the physical relationships between the features, leading to a more interpretable and meaningful ML model.

3.2. Hyperparameter selection

This section details the hyperparameter selection process for the RF model used in conjunction with the RIME feature reduction technique. The optimized feature space identified by RIME is employed for this stage. A 5-fold CV of R^2 is utilized as the baseline metric for fine-tuning the hyperparameters. RIME, a metaheuristic

optimization algorithm, is employed to optimize the hyperparameters. The number of iterations is set to 300, while the pop varies across different values: 25, 50, 75, and 100. This approach allows for exploring the impact of population size on the optimization process. It is important to note that the

model resulting from this process is denoted as "RIME-RF." The placement of "RIME" before "RF" emphasizes that RIME serves as a pre-processing step for the problem, responsible for selecting the optimal feature space before the RF model performs the prediction task.

Table 3. Hyperparameters used in the development of RF model

Hyperparameter	Explanation	Ranges	Best found
n_estimators	Number of trees in the forest	1-500	240
max_depth	Maximum depth of the tree	1-10	6
min_samples_split	Minimum number of samples required to split a node	2-8	3
min_samples_leaf	Minimum number of samples required to be at a leaf node	1-8	2

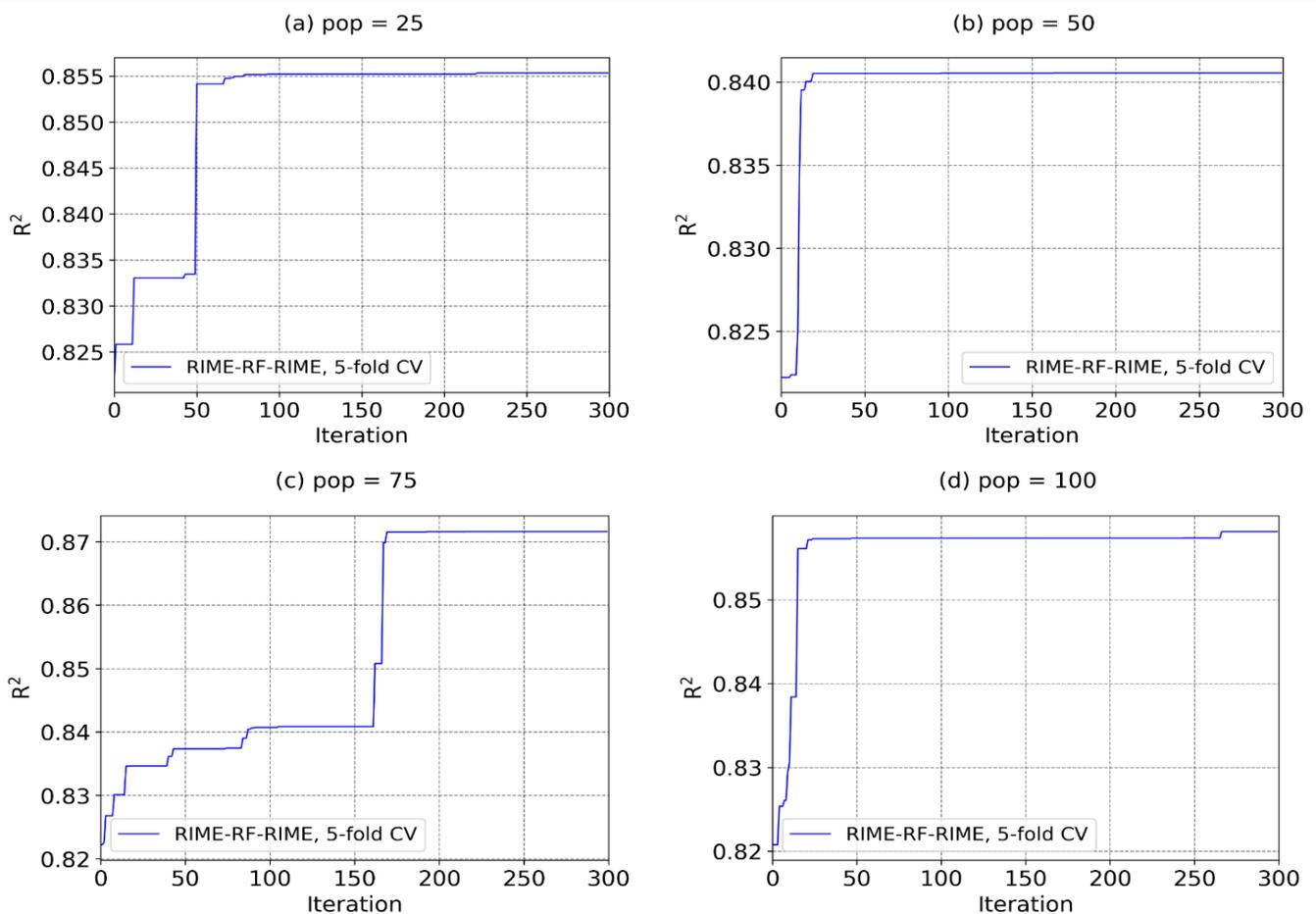


Fig. 3. Hyperparameters tuning of RIME-RF model using RIME with different pop sizes: (a) pop = 25, (b) pop = 50, (c) pop = 75, (d) pop = 100

The results, presented in Figure 3, reveal an interesting observation. Contrary to a potential

linear trend, the R^2 does not exhibit a clear relationship with increasing pop values. The

converged R^2 values after 300 iterations for pop sizes 25, 50, 75, and 100 are 0.855, 0.841, 0.872, and 0.857, respectively. This finding highlights that larger pop sizes do not necessarily guarantee superior prediction accuracy. Following the fine-tuning process, pop 75 emerges as the configuration leading to the best prediction accuracy ($R^2 = 0.872$). Consequently, the hyperparameters identified under this pop size are adopted for the final RF model. These hyperparameters are presented in detail in Table 3. This analysis underscores the importance of investigating the impact of hyperparameter tuning parameters, such as population size, on RIME's optimization performance. It demonstrates that an optimal configuration might not always be achieved by simply increasing a specific parameter value.

In conclusion, the hyperparameter selection process for the RIME-RF model has been successfully completed. Optimal hyperparameters are identified by utilizing RIME and employing a 5-fold CV of R^2 as the evaluation metric. To reflect the complete optimization process, the model is now denoted as "RIME-RF-RIME." This nomenclature highlights the sequential application of RIME for feature reduction and hyperparameter selection, culminating in the optimized RIME-RF model.

3.3. Model performance

This section presents the prediction results obtained using the RIME-RF-RIME model for estimating the macroscopic permeability on both the training and testing datasets. The model's performance is evaluated by comparing its predictions with the established FEM results.

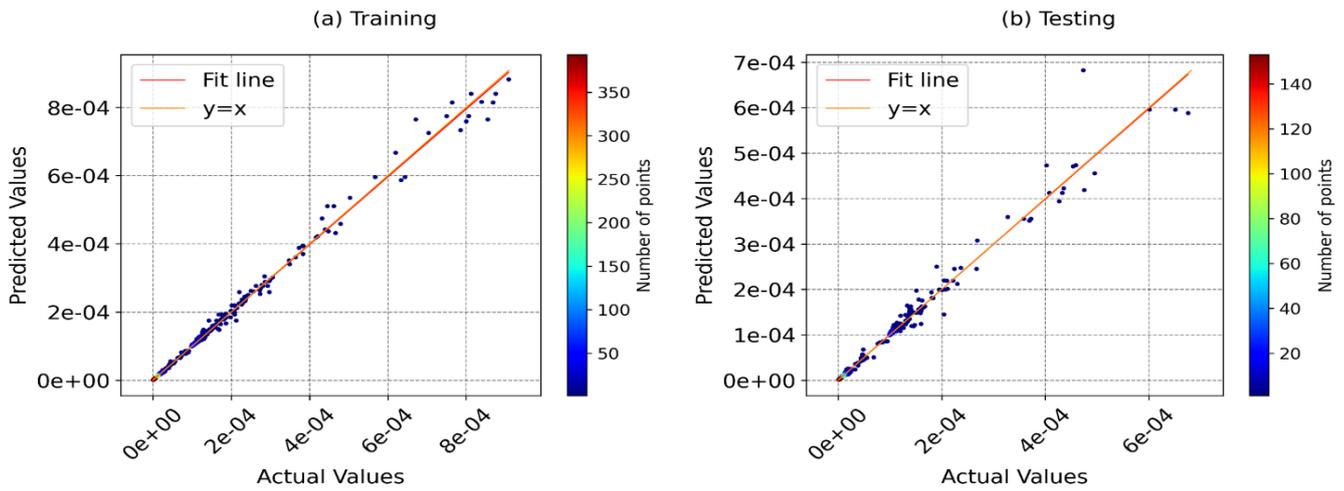


Fig. 4. Regression analysis of RIME-RF-RIME model: (a) training, (b) testing

The RIME-RF-RIME model exhibits promising results. Analyzing Figure 4 reveals a close agreement between the model's predictions and the FEM reference values for both the training and testing sets. This observation suggests that the feature reduction process, guided by RIME, has successfully identified a concise yet informative set of features crucial for accurate permeability prediction. Additionally, the optimized hyperparameters, also determined by RIME, effectively leverage the chosen features to achieve high prediction accuracy.

RIME-RF-RIME demonstrates exceptional

performance, achieving high accuracy on both training and testing datasets. The model exhibits an R^2 of 0.996 and 0.980 for the training and testing sets, respectively, and 0.992 for the entire dataset. The RMSE values are also remarkably low ($7.345e-6$ and $1.331e-5$ for training and testing, respectively), indicating minimal deviations between the predicted and actual permeability values. To provide context, a comparison is incorporated with relevant research. Of note, Ly and Nguyen [43] employed a Gradient Boosting model and achieved an R^2 of 0.998. This study demonstrates comparable predictive performance

while utilizing a more concise set of input variables. This finding implies that the proposed RIME-RF-RIME model successfully identifies the most influential factors driving the target outcome.

This successful outcome demonstrates the potential of RIME-RF-RIME as an efficient and reliable tool for estimating macroscopic permeability in similar applications. By combining feature selection and hyperparameter optimization, the model achieves comparable accuracy while utilizing fewer features, leading to a more efficient and interpretable solution than the complete feature set.

3.4. Sensitivity analysis

Understanding the rationale behind a

model's predictions is paramount for fostering trust and interpreting its results effectively. This section delves into a sensitivity analysis employing SHAP values to elucidate the inner working mechanism of the RIME-RF-RIME model and its predictions regarding macroscopic permeability in porous media. Figure 5 presents a SHAP bee-swarm plot visualization of the analysis. This plot depicts the distribution of SHAP values for each feature across various data points. An intriguing observation emerges from this analysis: the ordering method employed to estimate SHAP values can influence the perceived importance of specific features in predicting macroscopic permeability. Two scenarios are investigated.

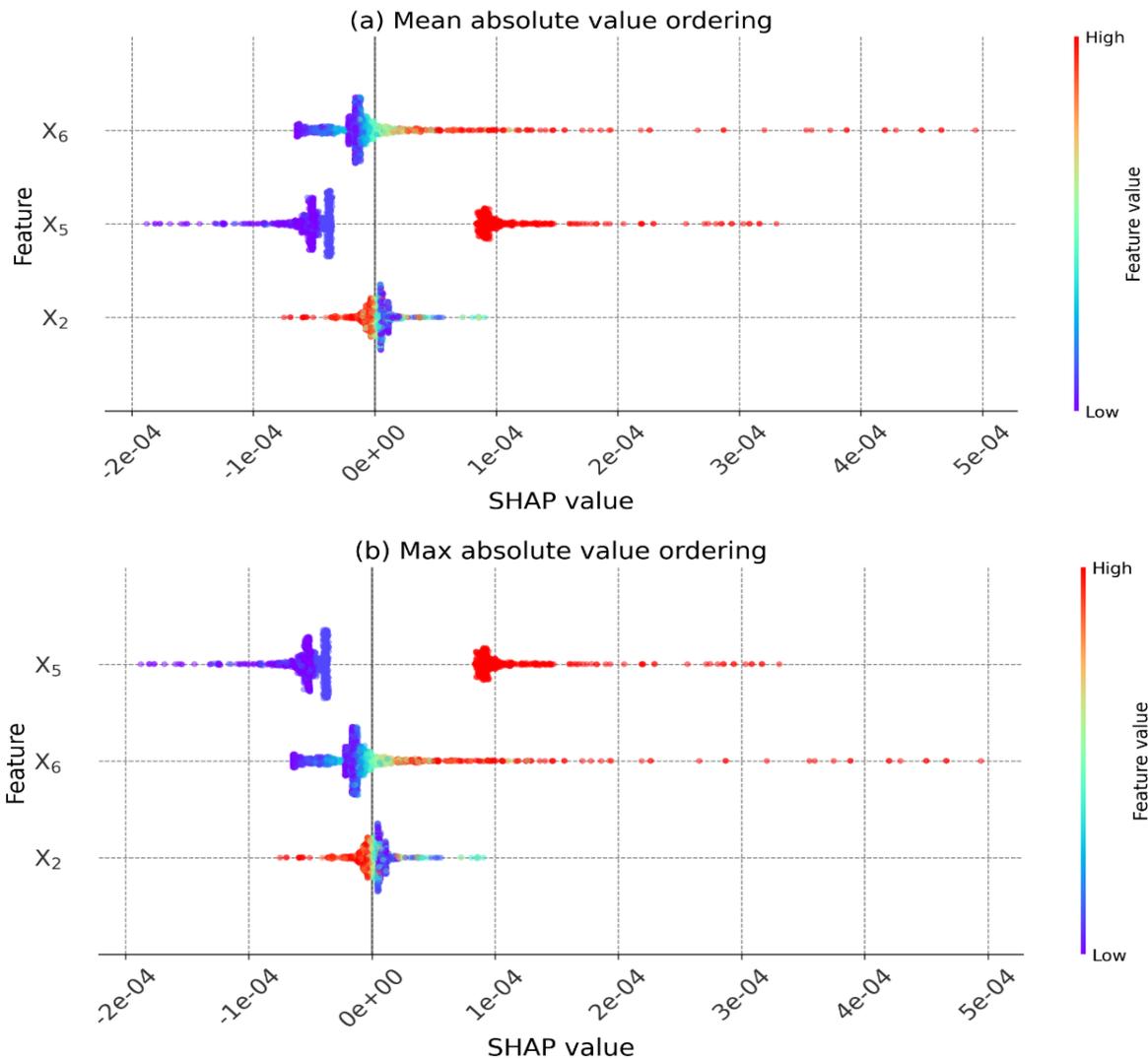


Fig. 5. Shap values analysis using bee-swarm plots: (a) mean absolute value ordering, (b) max absolute value ordering

In scenario 1, features are ranked based on the average of the absolute values of their SHAP values across all data points. This approach emphasizes the overall influence of each feature, regardless of the direction of its effect, whether positive or negative, on the permeability. This analysis reveals that porosity emerges as the most influential, suggesting a consistent impact on the model's predictions. The permeability of the porous media ranks second, highlighting its inherent significance in governing permeability. The size of the fluid phase perpendicular to the flow direction occupies the third position, indicating its non-negligible contribution to the model's output. Regarding scenario 2, the features are ranked based on the largest absolute SHAP value observed for each feature across all data points. This approach prioritizes features that can exert the most substantial influence, either positively or negatively, on specific predictions. The analysis reveals that the permeability of the porous media takes the first important position, underlining its potential for significant individual impact on predictions. The porosity of the media ranks second, showcasing its continued relevance. X_2 (size of the fluid phase perpendicular to flow direction) maintains its position in the top three.

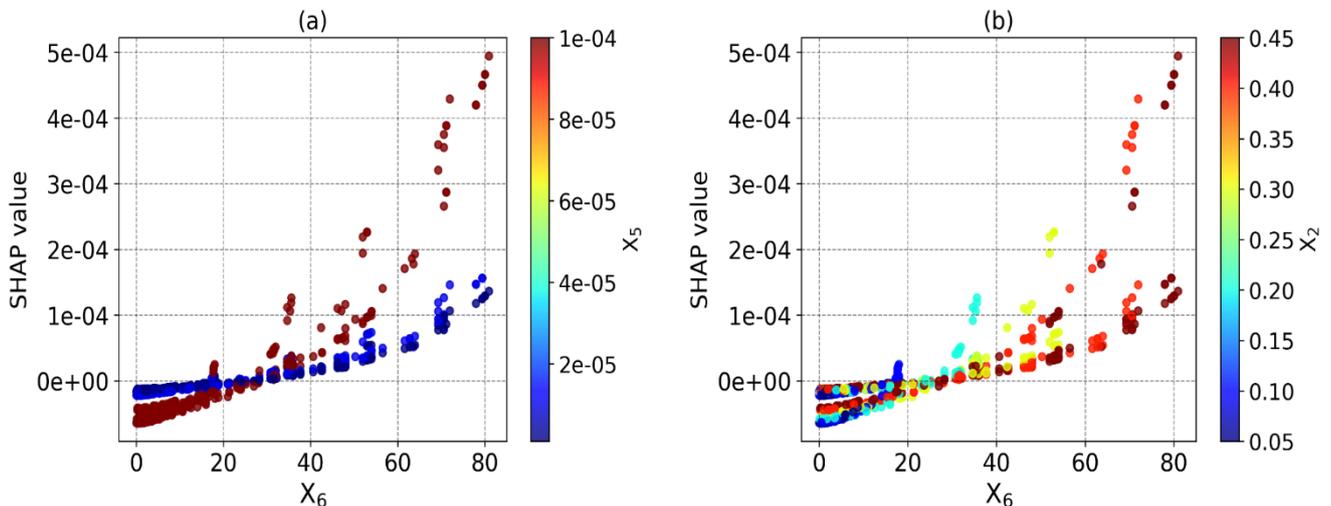
Furthermore, it is noteworthy that in both scenarios, X_5 (permeability of the porous media) and X_6 (porosity of the media) consistently exhibit

positive SHAP values, implying that an increase in their values leads to an increase in the model's predicted permeability. Conversely, X_2 consistently exhibits negative SHAP values, suggesting that a larger X_2 value corresponds to a decrease in predicted permeability.

These findings signify the complexity of interpreting model predictions and underscore the importance of employing various techniques to comprehensively understand the underlying mechanisms. The nuanced influence of SHAP ordering necessitates a rigorous and multifaceted approach to fully comprehend the model's decision-making processes and feature interactions, ultimately leading to a deeper scientific understanding of the problem domain.

A dependence plot SHAP analysis is conducted to further investigate the interaction between features and their impact on the predicted permeability (Figure 6).

Observations from the dependence plots reveal that increasing the porosity X_6 can lead to increased permeability, and this trend is true regardless of the remaining features, such as the porous phase X_5 (Figure 6a) and X_2 (Figure 6b). In contrast, increasing the permeability of the porous phase (X_5) could also increase the overall permeability. However, the coupled effect with X_2 is not clearly shown in this analysis, suggesting a complex behavior with X_2 .



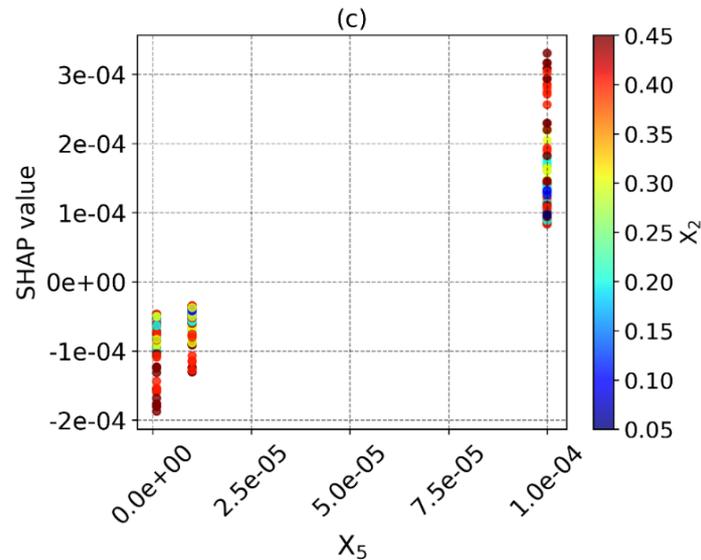


Fig. 6. Shap values dependence analysis: (a) X_6 and X_5 , (b) X_6 and X_2 , (c) X_5 and X_2

Overall, SHAP analysis demonstrates its value in investigating the RIME-RF-RIME model's underlying mechanisms and feature interactions. A comprehensive understanding of the model's decision-making processes is revealed by utilizing both single-feature SHAP analysis and dependence plot analysis, leading to a deeper scientific understanding of fluid flow in porous media problems.

4. Conclusions and Perspectives

This study introduced the RIME-RF-RIME model, a novel framework that leverages the Random Forest ML algorithm optimized by the RIME optimization algorithm to predict the macroscopic permeability of porous media. The model achieved exceptional predictive accuracy, with R^2 of 0.980, demonstrating its potential for simplification of FEM calculations. Furthermore, SHAP analysis identified porosity, surrounding phase permeability, and the size of the fluid phase perpendicular to the flow direction as the features exerting the most significant influence on the model's predictions. These findings not only offer insights into the mechanisms governing permeability in porous media but also contribute to the development of more interpretable and reliable predictive models for various scientific and engineering disciplines.

While the proposed model exhibits promising results, certain limitations warrant further investigation for future research. First, the study's reliance on a single dataset may limit its generalizability to broader configurations of porous media. Additionally, the focus on six specific features excludes the potential impact of other influential factors not considered in the current analysis. Future studies could address these limitations by enhancing the model's generalizability through validation on diverse datasets, employing sensitivity analysis, or incorporating domain knowledge to identify and explore the impact of additional features potentially relevant to permeability prediction.

Acknowledgments: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 107.03-2019.23.

Conflict of Interest: No.

Availability of data and material: No.

References

- [1] J.G. Herterich, D. Vella, R.W. Field, N.P. Hankins, I.M. Griffiths. (2015). Tailoring wall permeabilities for enhanced filtration. *Physics of Fluids*, 27, 053102.
- [2] F. Zhao, H.R. Landis, S.J. Skerlos. (2005). Modeling of Porous Filter Permeability via

- Image-Based Stochastic Reconstruction of Spatial Porosity Correlations. *Environmental Science & Technology*, 39(1), 239-247.
- [3] A. Zhu, P.D. Christofides, Y. Cohen. (2009). On RO membrane and energy costs and associated incentives for future enhancements of membrane permeability. *Journal of Membrane Science*, 344(1-2), 1-5.
- [4] D.J. Soeder. (1988). Porosity and permeability of eastern Devonian gas shale. *SPE Formation Evaluation*, 3(1), 116-124.
- [5] S. Baziar, M. Tadayoni, M. Nabi-Bidhendi, M. Khalili. (2014). Prediction of permeability in a tight gas reservoir by using three soft computing approaches: A comparative study. *Journal of Natural Gas Science and Engineering*, 21, 718-724.
- [6] A. Ghassemi. (2012). A Review of Some Rock Mechanics Issues in Geothermal Reservoir Development. *Geotechnical and Geological Engineering*, 30(3), 647-664.
- [7] S.B. Patil, H.S. Chore. (2014). Contaminant transport through porous media: An overview of experimental and numerical studies. *Advances in Environmental Research*, 3(1), 45-69.
- [8] M.L. Brusseau. (1994). Transport of reactive contaminants in heterogeneous porous media. *Reviews of Geophysics*, 32(3), 285-313.
- [9] D. Cohen-Tanugi, R.K. McGovern, S.H. Dave, J.H. Lienhard, J.C. Grossman. (2014). Quantifying the potential of ultra-permeable membranes for water desalination. *Energy & Environmental Science*, 7(3), 1134-1141.
- [10] Y. Bachmat, J. Bear. (1986). Macroscopic modelling of transport phenomena in porous media. 1: The continuum approach. *Transport in Porous Media*, 1, 213-240.
- [11] J. Bear, Y. Bachmat. (1986). Macroscopic modelling of transport phenomena in porous media. 2: Applications to mass, momentum and energy transport. *Transport in Porous Media*, 1, 241-269.
- [12] S. Ergun, A.A. Orning. (1949). Fluid Flow through Randomly Packed Columns and Fluidized Beds. *Industrial & Engineering Chemistry*, 41(6), 1179-1184.
- [13] A.E. Scheidegger. (1954). Statistical hydrodynamics in porous media. *Journal of Applied Physics*, 25, 994-1001.
- [14] A.E. Scheidegger. (1961). General theory of dispersion in porous media. *Journal of Geophysical Research*, 66(10), 3273-3278.
- [15] H.-B. Ly, H.-L. Nguyen, M.-N. Do. (2020). Finite element modeling of fluid flow in fractured porous media using unified approach. *Vietnam Journal of Earth Sciences*, 43(1), 13-22.
- [16] H.-B. Ly, V.-H. Phan, V. Monchiet, H.-L. Nguyen, L.N. Nguyen. (2022). Numerical investigation of macroscopic permeability of biporous solids with elliptic vugs. *Theoretical and Computational Fluid Dynamics*, 36(3), 689-704.
- [17] H.B. Ly, V. Monchiet, D. Grande. (2016). Computation of permeability with Fast Fourier Transform from 3-D digital images of porous microstructures. *International Journal of Numerical Methods for Heat and Fluid Flow*, 26(5), 1328-1345.
- [18] D. Bauer, L. Talon, Y. Peysson, H.B. Ly, G. Batot, T. Chevalier, M. Fleury. (2019). Experimental and numerical determination of Darcy's law for yield stress fluids in porous media. *Physical Review Fluids*, 4(6), 063301.
- [19] O. Dardis, J. McCloskey. (1998). Permeability porosity relationships from numerical simulations of fluid flow. *Geophysical Research Letters*, 25(9), 1471-1474.
- [20] S.E. Saberhosseini, K. Ahangari, H. Mohammadrezaei. (2019). Optimization of the horizontal-well multiple hydraulic fracturing operation in a low-permeability carbonate reservoir using fully coupled XFEM model. *International Journal of Rock Mechanics and Mining Sciences*, 114, 33-45.
- [21] Y. Peng, J. Liu, G. Zhang, Z. Pan, Z. Ma, Y. Wang, Y. Hou. (2021). A pore geometry-based permeability model for tight rocks and new sight of impact of stress on permeability. *Journal of*

- Natural Gas Science and Engineering*, 91(9), 103958.
- [22] R. Sadeghi, M.S. Shadloo, M. Hopp-Hirschler, A. Hadjadj, U. Nieken. (2018). Three-dimensional lattice Boltzmann simulations of high density ratio two-phase flows in porous media. *Computers & Mathematics with Applications*, 75(7), 2445-2465.
- [23] A.K. Gunstensen, D.H. Rothman. (1993). Lattice-Boltzmann studies of immiscible two-phase flow through porous media. *Journal of Geophysical Research*, 98(B4), 6431-6441.
- [24] S. Bakhshian, S.A. Hosseini, N. Shokri. (2019). Pore-scale characteristics of multiphase flow in heterogeneous porous media using the lattice Boltzmann method. *Scientific Reports*, 9, 3377.
- [25] S. Cant. (2002). High-performance computing in computational fluid dynamics: progress and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 360(1795), 1211-1225.
- [26] Y. Zaretskiy, S. Geiger, K. Sorbie, M. Förster. (2010). Efficient flow and transport simulations in reconstructed 3D pore geometries. *Advances in Water Resources*, 33(12), 1508-1516.
- [27] Z. Zhao, L. Jing, I. Neretnieks, L. Moreno. (2011). Numerical modeling of stress effects on solute transport in fractured rocks. *Computers and Geotechnics*, 38(2), 113-126.
- [28] S. Cai, Z. Mao, Z. Wang, M. Yin, G.E. Karniadakis. (2021). Physics-informed neural networks (PINNs) for fluid mechanics: a review. *Acta Mechanica Sinica*, 37, 1727-1738.
- [29] P.A. Johnson, B. Rouet-Leduc, L.J. Pyrak-Nolte, G.C. Beroza, C.J. Marone, C. Hulbert, A. Howard, P. Singer, D. Gordeev, D. Karaflos, C.J. Levinson, P. Pfeiffer, K.M. Puk, W. Reade. (2021). Laboratory earthquake forecasting: A machine learning competition. *Proceedings of the National Academy of Sciences USA*, 118(5), e2011362118.
- [30] D. Thaler, M. Stoffel, B. Markert, F. Bamer. (2021). Machine-learning-enhanced tail end prediction of structural response statistics in earthquake engineering. *Earthquake Engineering Structural Dynamics*, 50(8), 2098-2114.
- [31] B.-N. Phung, T.-H. Le, M.-K. Nguyen, T.-A. Nguyen, H.-B. Ly. (2023). Practical Numerical Tool for Marshall Stability Prediction Based On Machine Learning: An Application for Asphalt Concrete Containing Basalt Fiber. *Journal of Science and Transport Technology*, 3(3), 27-45.
- [32] M. Hasanipanah, R.A. Abdullah, M. Iqbal, H.-B. Ly. (2023). Predicting Rubberized Concrete Compressive Strength Using Machine Learning: A Feature Importance and Partial Dependence Analysis. *Journal of Science and Transport Technology*, 3(1), 27-44.
- [33] T.-A. Nguyen, K.N. Le, H.-B. Ly. (2024). Universal boosting ML approaches to predict the ultimate load capacity of CFST columns, *The Structural Design of Tall and Special Buildings*, 33(2), e2071.
- [34] T.-A. Nguyen, H.-B. Ly. (2024). Predicting axial compression capacity of CFDST columns and design optimization using advanced machine learning techniques. *Structures, Elsevier*, 59, 105724.
- [35] M. Alipour, D.K. Harris, L.E. Barnes, O.E. Ozbulut, J. Carroll. (2017). Load-Capacity Rating of Bridge Populations through Machine Learning: Application of Decision Trees and Random Forests. *Journal of Bridge Engineering*, 22(10), 04017076.
- [36] H. Adel, M.I. Ghazaan, A.H. Korayem. (2022). Machine learning applications for developing sustainable construction materials. *Artificial Intelligence and Data Science in Environmental Sensing, Elsevier*, pp. 179-210.
- [37] M. Khan, M.F. Javed. (2023). Towards sustainable construction: machine learning based predictive models for strength and durability characteristics of blended cement concrete. *Materials Today Communications*,

- 37, 107428.
- [38] P. Pospieszny, B. Czarnacka-Chrobot, A. Kobylinski. (2018). An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*, 137, 184-196.
- [39] K.R. Lennon, G.H. McKinley, J.W. Swan. (2023). Scientific machine learning for modeling and simulating complex fluids. *Proceedings of the National Academy of Sciences USA*, 120(27), e2304669120.
- [40] S.L. Brunton, B.R. Noack, P. Koumoutsakos. (2020). Machine Learning for Fluid Mechanics. *Annual Review of Fluid Mechanics*, 52, 477-508.
- [41] F. Karandish, J. Šimůnek. (2016). A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology*, 543, Part B, 892-909.
- [42] J. Luo, X. Ma, Y. Ji, X. Li, Z. Song, W. Lu. (2023). Review of machine learning-based surrogate models of groundwater contaminant modeling. *Environmental Research*, 238, Part 2, 117268.
- [43] H.-B. Ly, T.-A. Nguyen. (2023). Accelerating fluid flow simulations through doubly porous media using a FEM-assisted machine learning approach. *Results in Physics*, 54, 107036.
- [44] L. Breiman. (2001). Random forests. *Machine Learning*, 45, 5-32.
- [45] H. Su, D. Zhao, A.A. Heidari, L. Liu, X. Zhang, M. Mafarja, H. Chen. (2023). RIME: A physics-based optimization. *Neurocomputing*, 532, 183-214.
- [46] S.M. Lundberg, S.-I. Lee. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp 4768-4777.
- [47] B.-N. Phung, T.-H. Le, T.-A. Nguyen, H.-G.T. Hoang, H.-B. Ly. (2023). Novel approaches to predict the Marshall parameters of basalt fiber asphalt concrete. *Construction and Building Materials*, 400, 132847.
- [48] S.H. Trinh, H.-B. Ly. (2023). Enhancing Compressive Strength Prediction of Roller Compacted Concrete Using Machine Learning Techniques. *Measurement*, 218, 113196.