geoconvention
Calgary • Canada • May 11-13 2026

Unlocking Geothermal Resources: A Multi-Algorithm Machine Learning Framework with SHAP-Driven Insights from Western Canada

Gang Hui[1], Dan Wu[1], Yunli Lu[1], Feng Ni[1], Xiao Yang[2], Zhiyang Pi[1], Ye Li[1]

[1]State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, 102249, China;

[2]Department of Chemical and Petroleum Engineering, University of Calgary, Calgary T2N 1N4, Canada

Theory / Method / Workflow

This study develops a comprehensive machine learning (ML) framework to accurately predict geothermal potential in sedimentary basins. Using data from 301 wells in Alberta's Lower Cambrian sandstone, six advanced algorithms—SVR, DT, RF, GBDT, XGBoost, and LightGBM—were systematically benchmarked to predict Total Heat Capacity from integrated geological and engineering parameters (Fig.1). A rigorous protocol including a blind test set ensured robust evaluation. The optimal model was then interpreted using SHAP analysis to quantify and rank the influence of input features.

Results, Observations, Conclusions

The systematic comparison of six machine learning algorithms revealed a clear performance hierarchy. Among all models, XGBoost emerged as the optimal predictive engine for this geothermal regression task, delivering superior and robust performance. On the hold-out test set, it achieved a coefficient of determination (R^2) of 0.88 and a root mean square error (RMSE) of 0.12. Critically, this high level of accuracy was confirmed on a completely independent blind dataset, where the model maintained an R^2 of 0.86 and a mean absolute percentage error (MAPE) of less than 8% (Fig.2). This demonstrates exceptional generalization capability, confirming the model learned the underlying physical relationships rather than memorizing training data. The other ensemble methods (Random Forest and GBDT) also showed strong results, while simpler models like Decision Tree and SVR exhibited significantly lower predictive power for this complex, non-linear problem.

Beyond prediction accuracy, the application of SHAP (SHapley Additive exPlanations) interpretability analysis provided transformative insights into the model's decision-making process. The analysis quantitatively deconstructed the influence of all input features, identifying a dominant control triad: Extraction Depth, Reservoir Temperature, and Effective Geothermal Reservoir Thickness (Fig.3). These three parameters collectively accounted for over 75% of the feature importance, establishing their paramount role in controlling geothermal productivity in this setting. SHAP dependency plots further elucidated the specific nature of these relationships, revealing a strong positive monotonic relationship for depth and temperature, and a non-linear, "diminishing returns" effect for reservoir thickness beyond approximately 120 meters.

In conclusion, this study demonstrates that a carefully constructed ML workflow, combining rigorous algorithm benchmarking with advanced interpretability techniques, can serve as a powerful tool for sedimentary geothermal assessment. The final model provides not only a high-fidelity predictive tool for Total Heat Capacity but also delivers quantifiable, actionable geological intelligence. The identified key controls and their functional relationships offer a direct scientific basis for prioritizing data acquisition, defining economic cut-offs, and de-risking exploration strategies in analogous plays.

Novel/Additive Information

This work moves beyond conventional ML application by integrating rigorous multi-algorithm benchmarking with advanced model interpretability (SHAP) for geothermal prediction. It delivers a validated, end-to-end workflow that not only identifies the optimal predictive model (XGBoost) but also quantitatively deconstructs the "black box" to rank and elucidate the key subsurface controls. This framework provides a practical tool for rapid, insight-driven sweet-spot identification and risk reduction in sedimentary geothermal plays.

Acknowledgements

This research was funded by the Science Foundation of China University of Petroleum, Beijing (No. 2462023BJRC001).

References

Reference Style (use Arial 9pt normal)

[1] Awoleke, O., Lane, R, 2011. Analysis of data from the Barnett shale using conventional statistical and virtual intelligence techniques. SPE Reservoir Eval. Eng. 14 (5),544–556. https://doi.org/10.2118/127919-PA

[2] Bachu, S., Burwash, R. 1991. Regional scale analysis of the geothermal regime in the Western Canada Sedimentary Basin. Geothermics. 20(5):387-407. https://doi.org/10.1016/0375-6505(91)90028-T

[3] Bao, P., Hui, G., Hu, Y., Song, R., Chen, Z., Zhang, K., et al. (2025a). Comprehensive characterization of hydraulic fracture propagations and prevention of pre-existing fault failure in Duvernay shale reservoirs. Engineering Failure Analysis, 173, 109461. https://doi.org/10.1016/j.engfailanal.2025.109461

[4] Bao, P., Hui, G., Zhang, J., Wang, M., Liang, H., Zhang, R., et al. (2025b). Production-Increase Potential Evaluations after Refracturing Low-Shale-Oil-Producing Wells via Machine-Learning-Driven Multisource Data Mining. ACS Omega, 10(39), 45840–45854. https://doi.org/10.1021/acsomega.5c06404

[5] Chen, G., Jiao, J. J., Wang, Z., Dai, Q. Multi-fidelity machine learning with knowledge transfer enhances geothermal energy system. https://doi.org/10.46690/ager.2025.10.07

[6] design and optimization. Advances in Geo-Energy Research, 2025, 16(3): 244-259. https://doi.org/10.46690/ager.2025.06.05

[7] Hui G., Ren Y., Bi J., Wang M., Liu C. Artificial intelligence applications and challenges in oil and gas exploration and development. Advances in Geo-Energy Research, 2025, 17(3): 179-183. https://doi.org/10.46690/ager.2025.09.01

[8] Hui, G., Chen, S., He, Y., Wang, H., & Gu, F. (2021). Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors. Journal of Natural Gas Science and Engineering, 94, 104045. https://doi.org/10.1016/j.jngse.2021.104045

[9] Hui, G., Chen, Z., Wang, Y., Zhang, D., & Gu, F. (2023b). An integrated machine learning-based approach to identifying controlling factors of unconventional shale productivity. Energy, 266, 126512. https://doi.org/10.1016/j.energy.2022.126512

[10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

[11] Liu, Z., Gudala, M., Yan, B. Efficient optimization of coupled geothermal reservoir modeling and power plant off-design based on deep learning. Advances in Geo-Energy Research, 2025, 18(1): 84-98.
[12] Rashid, H. U., Olorode, O. Use of controlled fractures in enhanced geothermal systems. Advances in Geo-Energy Research, 2024, 12(1): 35-51. https://doi.org/10.46690/ager.2024.04.04
[13] Shi, G., Wang, J., & Gates, I. D. (2024a). Combined geothermal and CO2 sequestration in the Basal Cambrian Sandstone Unit (BCSU) in Alberta, Canada. Journal of CO2 Utilization, 80, 102685. https://doi.org/10.1016/j.jcou.2024.102685
[14] Shi, G., Wang, J., & Gates, I. D. (2024b). Investigation of enhanced geothermal system in the Basal Cambrian Sandstone Unit, Alberta, Canada. Heliyon, 10(2), e24763. https://doi.org/10.1016/j.heliyon.2024.e24763
[15] Wang, K., Yuan, B., Ji, G., & Wu, X. (2018). A comprehensive review of geothermal energy extraction and utilization in oilfields. Journal of Petroleum Science and Engineering, 168, 465–477. https://doi.org/10.1016/j.petrol.2018.05.012
[16] Zhang, K., Jiang, S., Chen, Z., Li, H., & Liu, S. (2023). Geothermal development associated with enhanced hydrocarbon recovery and geological CO2 storage in oil and gas fields in Canada. Energy Conversion and Management, 288, 117146. https://doi.org/10.1016/j.enconman.2023.117146

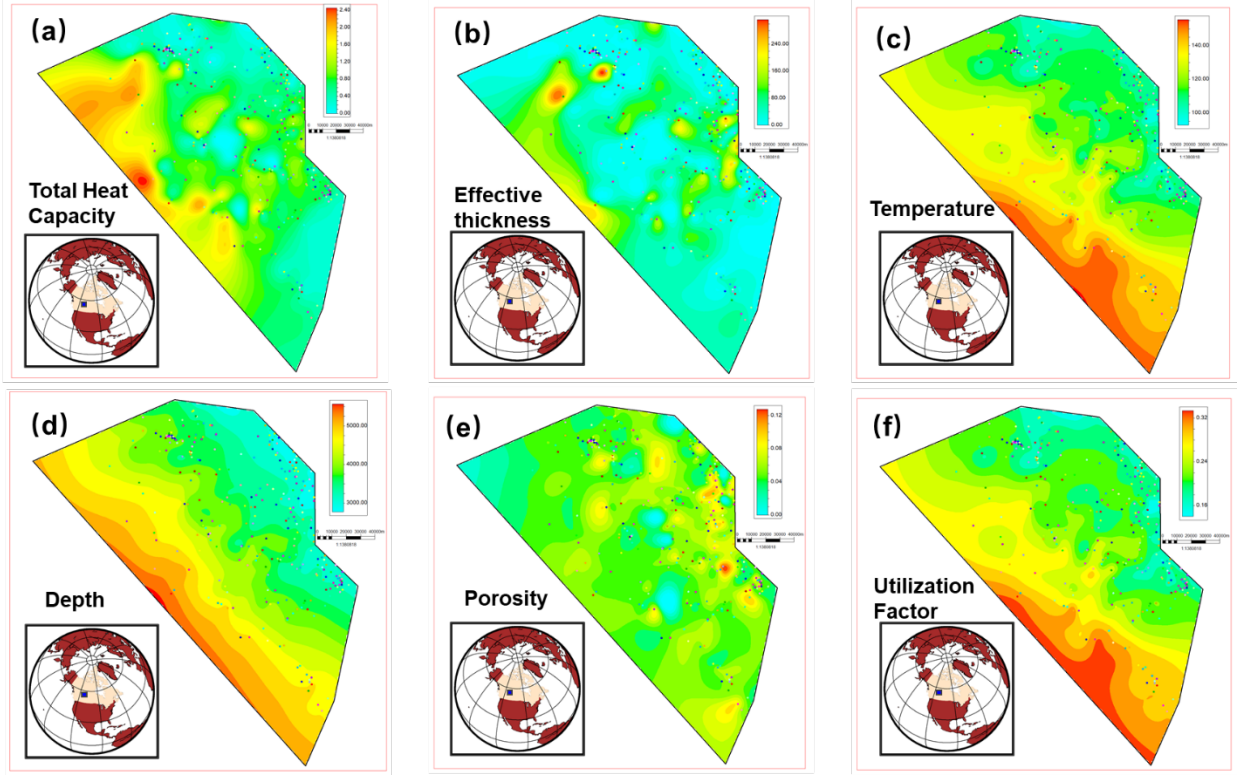


Fig. 1—Geological information in the studied region. The base map represents (a)Total Heat Capacity; (b) Effective thickness; (c) Temperature; (d) Depth; (e) Porosity; (f) Utilization Factor. The inner plot shows the location of the studied region in Canada.

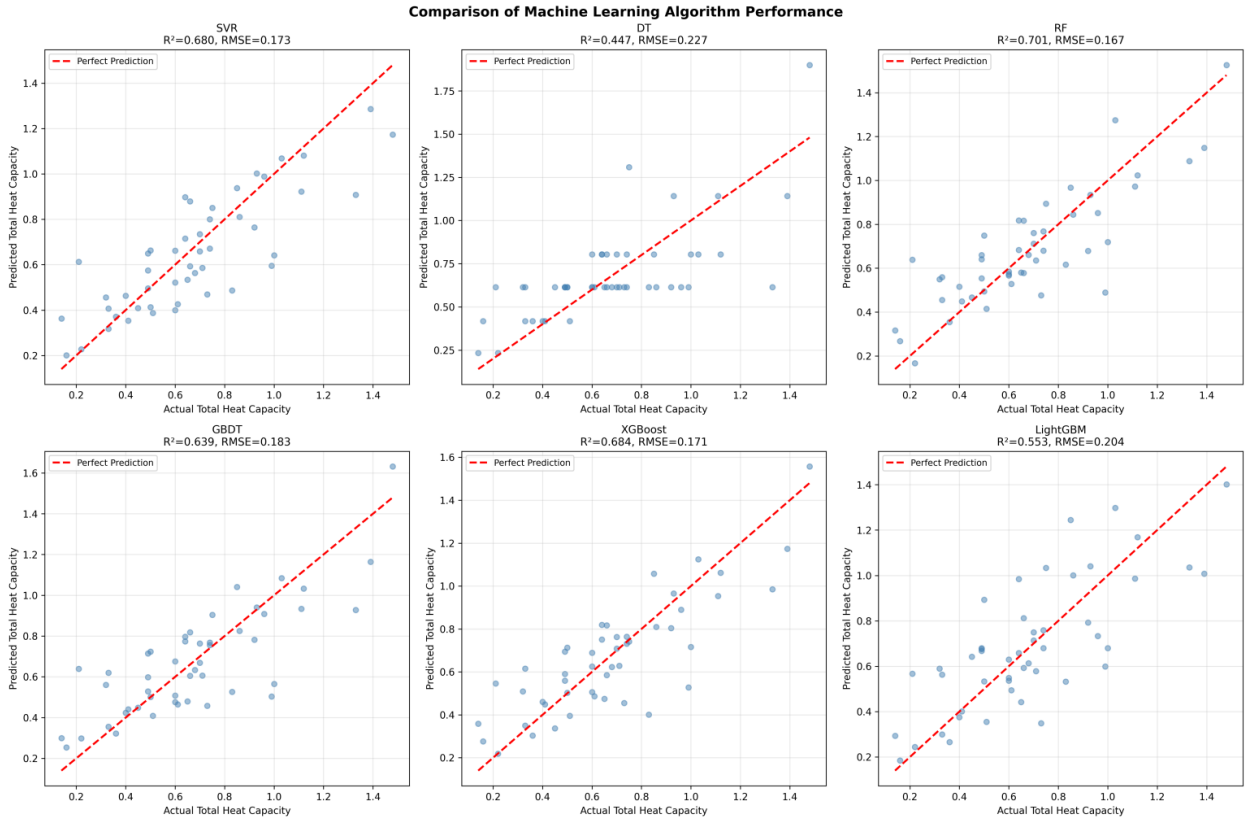


Fig. 2—Scatter Plot Comparison of Predicted vs. Actual Values for Each Algorithm

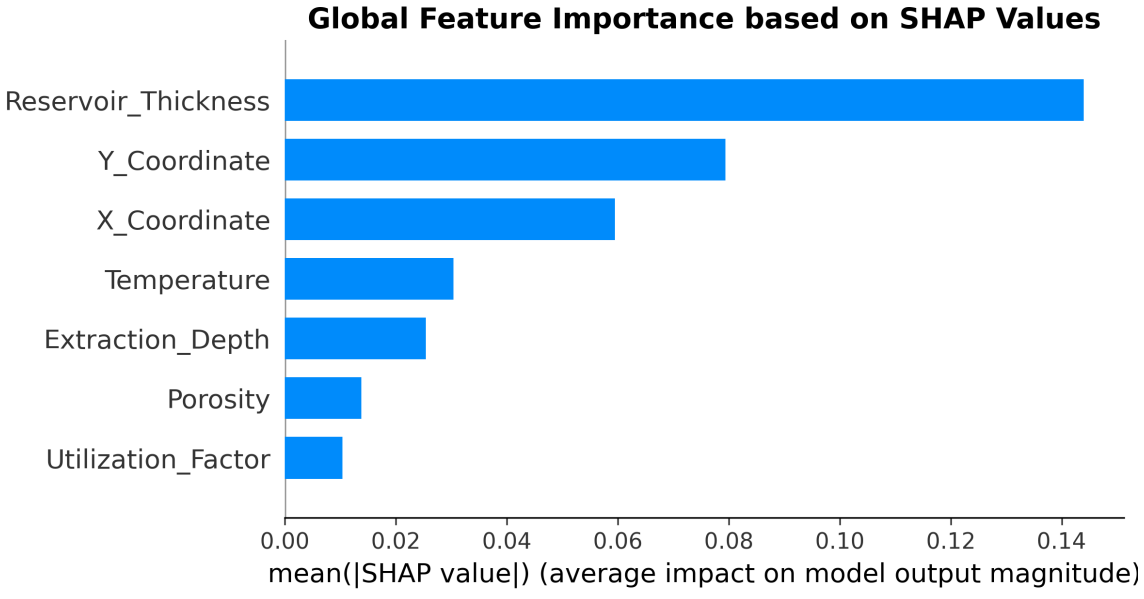


Fig. 3—Global Feature Importance Bar Chart Based on SHAP Values