Proceedings World Geothermal Congress 2026, Calgary, Canada, June 8—11, 2026

Transforming Geothermal Exploration: The Integration of Physics-Informed Machine Learning and Digital Twins

Aref Lashin[1]*, Mahmoud AlGaiar[2]
King Saud University[1], Robert Gordon University[2]
*arlashin@ksu.edu.sa

INTRODUCTION

The convergence of Physics-Informed Machine Learning (PIML) and Digital Twin (DT) technologies is transforming geothermal exploration and operations under the emerging paradigm of Geothermal 4.0. Traditional geothermal modeling depends on costly and time-consuming reservoir characterization, while recent advances in artificial intelligence now enable data-driven modeling of subsurface and operational processes. PIML blends physical laws, such as mass, momentum, and energy conservation, with adaptive learning to increase model accuracy and interpretability. Digital twins complement this by dynamically mirroring geothermal assets, integrating sensor data, and enabling predictive simulations for performance optimization and failure prevention.

Recent frameworks, including **GOOML** (Geothermal Operational Optimization with Machine Learning) and the **GeoThermalCloud** platform, have demonstrated success in combining physics-based models with machine learning to optimize field management and production forecasting. Similarly, digital twin deployments at sites such as the **Delft University of Technology** (TU Delft) geothermal doublet and Türkiye's Alaşehir hybrid plant illustrate how real-time data assimilation improves operational reliability and sustainability. Collectively, these innovations mark a shift from static modeling to adaptive geothermal systems that self-learn from historical and sensor data, enabling operators to reduce risk, extend equipment life, and enhance efficiency.

METHODS

This study adopts a structured literature synthesis on PIML–DT integration in geothermal systems, integrating experimental, computational, and field-scale case studies. The review draws from operational and research frameworks spanning GOOML, GeoThermalCloud, and TU Delft's geothermal digital twin initiative. In the PIML domain, unsupervised and supervised learning methods such as non-negative matrix factorization (NMFk), Long Short-Term Memory (LSTM) autoencoders, and hybrid regression–neural network models were identified as enabling tools for anomaly detection, reservoir mapping, and real-time forecasting. These algorithms are augmented through physics-based constraints to ensure compliance with thermodynamic relationships and mass balance.

Digital twin methodologies emphasize multi-stage data pipelines combining geological modeling (e.g., Rapid Reservoir Modeling), high-fidelity numerical simulators (DARTS), and data assimilation techniques like the Ensemble Smoother with Multiple Data Assimilation (ESMDA). Across reviewed applications, real-time sensor integration, cloud computing, and modular software architectures facilitate reciprocal flow between models and physical systems. The synthesis highlights recurring workflows centered on data curation, feature extraction, uncertainty quantification, and validation against field observations.

RESULTS

The integrated PIML–DT approaches have yielded significant improvements in geothermal prediction, optimization, and control. Case studies demonstrate operational benefits across multiple contexts. In the **GOOML** framework, machine learning models trained on multi-year datasets from sites such as Wairakei and Kawerau improved generation forecasts, enhanced flow control, and suggested operational adjustments resulting in a 1–2% power gain without additional infrastructure. The **GeoThermalCloud** platform efficiently identified hidden geothermal signatures and resource attributes using unsupervised learning on sparse regional datasets, improving mapping of low-, medium-, and high-enthalpy systems. At **TU Delft**, incorporation of real-time well-log and fiber-optic data into a digital twin environment constrained uncertainty in bottom-hole pressure and reservoir permeability, informing safe injection strategies. Additionally, at **Alaşehir GPP**, an AI-driven predictive maintenance DT based on an LSTM Autoencoder model detected pump faults nine days before failure with 0.88 precision and 0.78 F1-score. Collectively, these implementations illustrate how data-centric workflows combining physical and machine learning models improve model generalization, accelerate computation, and reduce field risk, leading to more resilient geothermal systems.

Table 1. Summary of Reviewed PIML–DT Frameworks

Framework	Focus Area	Methodology	Reported Outcome
GOOML	Plant optimization	ML + physical constraints	+2% generation gain
GeoThermalCloud	Hidden resource discovery	Unsupervised ML	Improved anomaly detection
TU Delft	Digital twin validation	DARTS + ESMDA	<5% prediction error
Alaşehir GPP	Predictive maintenance	LSTM Autoencoder	9-day fault lead time

DISCUSSION

The integration of PIML and DT technologies demonstrates that geothermal systems benefit most from hybridized, data-centric workflows. Unlike traditional model-centric approaches that prioritize algorithmic complexity, data-oriented strategies improve accuracy and adaptability by refining data quality and ensuring consistent real-world applicability. Lessons from GOOML and GeoThermalCloud underline that successful geothermal digitalization depends on rigorous data curation, standardized metadata, and consistent communication between engineers and data scientists.

Digital twins further enhance interpretability by embedding geophysical constraints and enabling near real-time feedback loops between model predictions and plant conditions. Nonetheless, challenges remain in standardizing data formats, protecting proprietary field data, and scaling models across facilities with varying complexity. Future advances should focus on open data exchange protocols, cloud scalability, and combining PIML with uncertainty-focused physics solvers to balance predictive fidelity and computational efficiency. Ultimately, the coevolution of PIML and DT frameworks is expected to establish intelligent, self-calibrating geothermal systems.

CONCLUSION

The combined use of physics-informed machine learning and digital twin frameworks represents a transformative pathway toward sustainable geothermal energy development. Case studies show that these methods enhance resource assessment accuracy, predict equipment performance, and enable proactive maintenance while reducing computational overhead and operational uncertainty.

A data-centric shift, where physical constraints and high-quality curated data govern model training, ensures interoperability and reproducibility across diverse geothermal contexts. Looking ahead, adoption of standardized PIML–DT platforms will accelerate innovation through cross-disciplinary collaboration, from real-time reservoir management to field-scale optimization. These integrated technologies are poised to reshape geothermal operations as part of the broader Industry 4.0 transition, promoting transparency, efficiency, and economic competitiveness in clean energy production.

Acknowledgements

This research was conducted through the collaborative efforts of **King Saud University** and **Robert Gordon University**. The authors express sincere appreciation to the Petroleum and Natural Gas Engineering Department at King Saud University and the School of Engineering at Robert Gordon University for providing access to computational resources and academic support. The authors especially acknowledge the contributions of colleagues and reviewers whose constructive feedback enhanced the quality of this work.

REFERENCES

(1) Blair, A., Siratovich, P., Buster, G., Taverna, N., Weers, J., Siega, C., Mannington, W., Urgel, A., Cen, J., Quinao, J., Watt, R. and Akerley, J. (2020) *GOOML: Geothermal operational optimization with machine learning. GRC Transactions*, 44, pp. 1–9.
(2) Konstantopoulos, I., Siratovich, P., Buster, G., Taverna, N., Weers, J., Blair, A., Huggins, J., Siega, C. and Akerley, J. (2023) *Deploying digital twins for geothermal operations with the GOOML framework. Proceedings, 48th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, CA, pp. 1–10.
(3) Muther, T., Syed, F.I., Lancaster, A.T., Salsabila, F.D., Kalantari-Dahaghi, A. and Negahban, S. (2022) *Geothermal 4.0: AI-enabled geothermal reservoir development — current status, potentials, limitations, and ways forward. Geothermics*, 100, 102348.
(4) Taverna, N., Buster, G., Huggins, J., Rossol, M., Siratovich, P., Weers, J., Blair, A., Siega, C., Mannington, W., Urgel, A., Cen, J., Quinao, J., Watt, R. and Akerley, J. (2022) *Data curation for machine learning applied to geothermal power plant operational data. Proceedings, NREL CP-6A20-81649.*
(5) Tut Haklidir, F.S., Özen, F., Halepmollas, R. and Haklidir, M. (2024) *Artificial intelligence approaches for sustainable geothermal energy systems with a case study. Proceedings, 49th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, CA, pp. 1–8.
(6) Vesselinov, V.V., Ahmmed, B., Frash, L. and Mudunuru, M.K. (2022) *GeoThermalCloud: Machine learning for discovery, exploration, and development of hidden geothermal resources. Proceedings, 47th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, CA, pp. 1–7.
(7) Voskov, D., Abels, H., Barnhoorn, A., Chen, Y., Daniilidis, A., Bruhn, D., Drijkoningen, G., Geiger, S., Laumann, S., Song, G., Vardon, P.J. and Verschuure, E. (2024) *Initial modeling and establishment of a geothermal digital twin at TU Delft. Proceedings, 49th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, CA, pp. 1–10.