ALGAIAR, M., HOSSAIN, M., PETROVSKI, A., LASHIN, A. and FAISAL, N. 2024. Applications of artificial intelligence in geothermal resource exploration: a review. *Deep underground science and engineering* [online], 3(3): geothermal energy, pages 269-285. Available from: https://doi.org/10.1002/dug2.12122

Applications of artificial intelligence in geothermal resource exploration: a review.

ALGAIAR, M., HOSSAIN, M., PETROVSKI, A., LASHIN, A. and FAISAL, N.

2024

Received: 14 May 2024 | Revised: 28 July 2024 | Accepted: 4 August 2024

DOI: 10.1002/dug2.12122

REVIEW ARTICLE

DUSE WILEY

Applications of artificial intelligence in geothermal resource exploration: A review

Mahmoud AlGaiar[1] | Mamdud Hossain[1] | Andrei Petrovski[2] | Aref Lashin[3] | Nadimul Faisal[1]

[1]School of Engineering, Robert Gordon University, Aberdeen, UK

[2]National Subsea Centre, Aberdeen, UK

[3]Petroleum and Natural Gas Engineering Department, College of Engineering, King Saud University, Riyadh, Saudi Arabia

Correspondence
Mahmoud AlGaiar and Nadimul Faisal, School of Engineering, Robert Gordon University, Aberdeen AB10 7GJ, UK.
Email: m.algaiar@rgu.ac.uk and n.h.faisal@rgu.ac.uk

Funding information
None

Abstract

Artificial intelligence (AI) has become increasingly important in geothermal exploration, significantly improving the efficiency of resource identification. This review examines current AI applications, focusing on the algorithms used, the challenges addressed, and the opportunities created. In addition, the review highlights the growth of machine learning applications in geothermal exploration over the past decade, demonstrating how AI has improved the analysis of subsurface data to identify potential resources. AI techniques such as neural networks, support vector machines, and decision trees are used to estimate subsurface temperatures, predict rock and fluid properties, and identify optimal drilling locations. In particular, neural networks are the most widely used technique, further contributing to improved exploration efficiency. However, the widespread adoption of AI in geothermal exploration is hindered by challenges, such as data accessibility, data quality, and the need for tailored data science training for industry professionals. Furthermore, the review emphasizes the importance of data engineering methodologies, data scaling, and standardization to enable the development of accurate and generalizable AI models for geothermal exploration. It is concluded that the integration of AI into geothermal exploration holds great promise for accelerating the development of geothermal energy resources. By effectively addressing key challenges and leveraging AI technologies, the geothermal industry can unlock cost-effective and sustainable power generation opportunities.

KEYWORDS

artificial intelligence, geothermal energy, geothermal exploration, geothermometry, hidden/blind geothermal resources, machine learning

Highlights

- Progress in the use of Artificial intelligence (AI) methodologies is presented in detail.
- Geophysical data analysis is the most notable AI application.
- Neural networks are the most-used AI technique across geothermal exploration groups.
- Challenges and recommendations for future research using AI are provided.
- Large-scale AI applications are reasonably novel in geothermal exploration.

1 | INTRODUCTION

Geothermal energy is a renewable, sustainable, and low-emission energy source derived from the Earth's subsurface layers through natural heat sources, such as rock formation and radioactive decay. It is used for heating, cooling, and power generation due to its cost-effectiveness, stable supply, and high-capacity factors throughout the year. The development of power generation from hydrothermal reservoirs started in 1913 and has since expanded to include various technologies such as flash and dry steam plants for high-temperature resources and binary cycle technologies for medium-temperature resources. Global electricity generation from geothermal energy grew from 69.8 GW·h in 2011 to 95.3 GW·h in 2021, providing a significant share of electricity demand in countries, such as El Salvador, New Zealand, Kenya, and the Philippines, and more than 90% of heating demand in Iceland (International Renewable Energy Agency, 2023).

The development of geothermal energy takes place in successive stages, starting with surface surveys, followed by exploration drilling for resource realization. If the resource is proven, delineation drilling follows to confirm the extent of the reservoir's productivity and its development plan. Production drilling and power plant construction can commence once the resource has been confirmed and financial viability has been established. Preliminary studies, exploration, and delineation drilling require significant investment and involve high financial risk, which can hinder resource evaluation plans. For example, in a recent appraisal study for geothermal exploration in Indonesia, the World Bank estimated that the predevelopment program would cost approximately USD 30 million, assuming a minimum of three wells for greenfield development and at least two wells producing an acceptable level of steam for site exploration to provide satisfactory evidence or resource availability (The World Bank, 2012). Figure 1 illustrates the geothermal development project stages, the level of risk at each stage, and the associated percentage of cumulative project cost (Gehringer & Loksha, 2012; ©World Bank; The World Bank et al., 2012).

Many countries are exploring hidden or blind geothermal resources (hydrothermal resources without surface manifestations), which require detailed knowledge of subsurface features (including hydrological, geophysical, geological, geomechanical, geochemical, and thermal characteristics) to assess their commercial potential (Pandey et al., 2018). Traditional methods of subsurface feature analysis rely heavily on expert knowledge for resource evaluation and reserve estimation, leading to uncertainties in the discovery of hidden geothermal resources. Advances in data-driven models have led to the use of artificial intelligence (AI) to replace traditional expert-based and statistical methods, where AI can uncover hidden patterns and develop predictive models from large multivariate datasets, thus enhancing exploration outcomes by reducing uncertainty and improving prediction accuracy. With the rapid increase in the creation of data repositories for the preservation, processing, and management of subsurface data, data-driven models offer an efficient and cost-effective approach to identifying key features of hidden geothermal resources (He et al., 2019). This supports resource evaluation, problem solving, and decision-making while reducing predevelopment costs in the geothermal industry.

2 | REVIEW OF AI IMPLEMENTATION IN GEOTHERMAL RESOURCE EXPLORATION

Several reviews have been published on the use of AI in geothermal applications (Aljubran et al., 2022; Liu & Misra, 2022; Muther et al., 2022; Okoroafor et al., 2022; Wang et al., 2023). However, the current review focuses on the specific applications of AI in the resource exploration domain of the geothermal industry; such information has not been detailed in any other published review or article. Here, exploration is defined as any process employed to discover geothermal potential before confirmation as a commercial resource; this includes exploratory drilling for resource evaluation. In conducting the current study, published articles were critically analyzed to identify studies applying AI to geothermal exploration. Consequently, the specific applications of AI in geothermal exploration were summarized and the potential benefits and challenges of using AI in this field were highlighted.

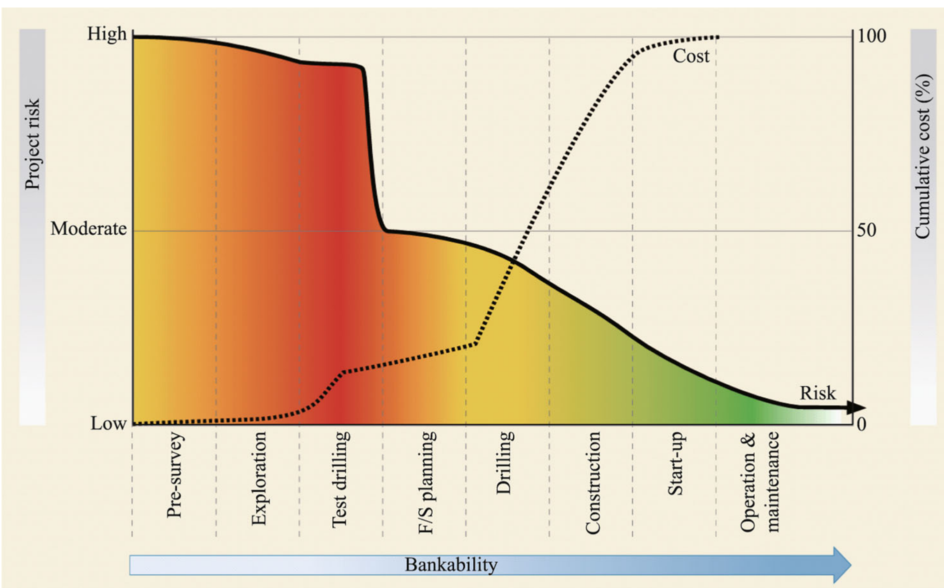

FIGURE 1 Geothermal development project cost and risk profile throughout various project stages (Gehringer & Loksha, 2012; The World Bank et al., 2012; reproduced under the terms of the CC BY 3.0 IGO copyright licenses, ©World Bank).

The reviewed publications on the application of AI were classified into six subsections: (1) play fairway analysis (PFA), (2) integrated subsurface data set applications, (3) specific geochemical data applications, (4) specific geophysical data applications, (5) thermal data applications, and (6) other data applications. These six subsections provided a systematic approach to understanding the different AI applications in geothermal exploration. The application of each was critically evaluated and is presented in detail in the following subsections.

2.1 | PFA applications

PFA is a regional evaluation approach used to define geothermal potential by integrating geological, geophysical, and geochemical parameters indicative of geothermal activity (Faulds et al., 2017). These parameters are divided into subsets and assigned specific weights to provide rankings that collectively constitute the geothermal play.

The original Nevada geothermal fairway defined by Faulds et al. (2015, 2018, 2019, 2021) included subsurface features associated with geothermal activity. The features were linked by multiplying each by a unique "weight" and then combining the weighted parameters in a linear sum to produce the fairway; a value scaled to represent the degree of geothermal potential. Due to recognized limitations in the data and the limited training sites, the study included expert opinion, and a machine learning (ML) approach including logistic regression (LR), weight of evidence, and other statistical metrics (Faulds et al., 2017). The fairway helped to locate at least two blind geothermal systems; however, it faced several limitations, including the limited number of training sites, certain data set limitations, and the need to determine the influence weights of features. Figure 2 illustrates the modeling workflow used to identify the Nevada play fairway. Relative weights are determined by the weights of the evidence method and are shown in red, while expert-driven weights are shown in black.

In 2020, Faulds et al. (2020) used advanced AI methods (such as artificial neural networks [ANNs] and training set augmentation) to address the fundamental problems associated with the Nevada play fairway project. In their ongoing project, ML techniques are applied to enhanced datasets of the original fairway, and a modified PFA workflow is used to incorporate newly added datasets. Their project are aimed to develop an algorithm-based approach that learns to estimate the influence weights for various parameters and minimizes the expert-driven inputs, to identify undeveloped geothermal sites in the Great Basin region and investigate the implementation of ML principles in geothermal

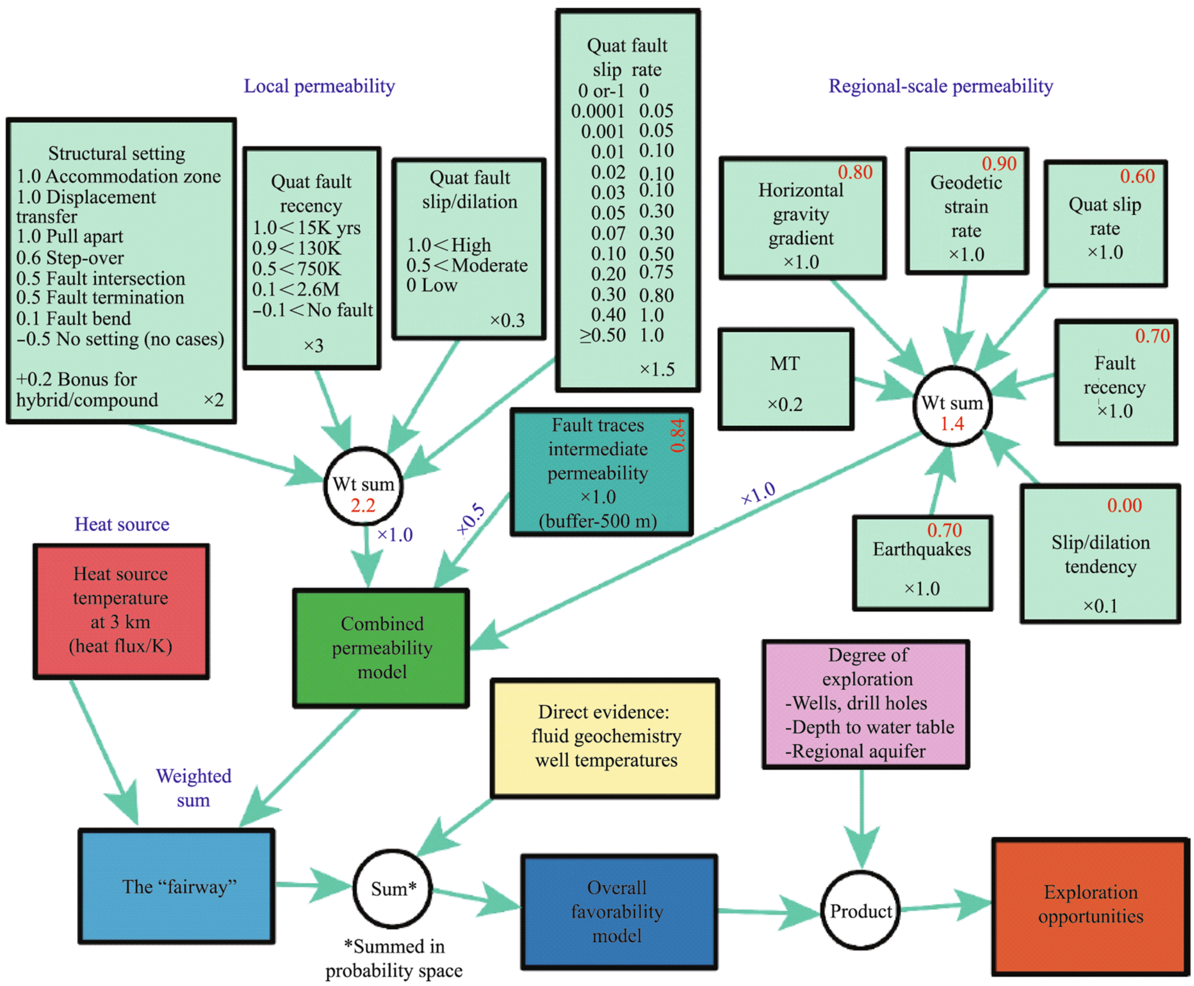

FIGURE 2 Nevada play fairway modeling workflow (Faulds et al., 2017; permission obtained© J.E. Faulds, all rights reserved).

potential assessments, in addition to introducing groundbreaking aspects to the process of identifying previously undiscovered blind geothermal systems by determining new signatures.

In the framework of Nevada PFA, Smith et al. (2021) used a combination of supervised and unsupervised ML approaches to assess the impact of specific geological and geophysical features in predicting geothermal potential and discovered new methods to empirically structure correlations between feature weights and labels in an improved manner. First, a permutation-supervised filter model was developed to discover feature dependencies in the positive and negative categorizations of training and test data. Second, the unsupervised principal component analysis (PCA) method was used to select the most relevant features for the favorability analysis. The reduced data set was then clustered using the semi-supervised k-means algorithm to detect geographic patterns in the data. The results indicated potential means of uncovering favorable sources of predictive information to locate blind geothermal systems and improve knowledge of complex geothermal feature-label relationships in the Great Basin area and beyond.

There are benchmarks (such as an existing power plant or conclusive positive or negative drilling results) in certain geographical areas where geothermal potential is known. Using the known benchmarks in Nevada and Nevada PFA data, Brown et al. (2020, 2022) interpreted the original and enhanced PFA datasets, added new datasets in a form suitable for ML algorithms, and explored a variety of ANN architectures to predict geothermal potential as a probability map. A supervised ML approach was applied to a series of maps based on 10 geological and geophysical features that were used to categorize geographical regions as either positive or negative resources. The training data set contained 83 positive sites from known geothermal systems (subsurface temperature ≥39°C) and 62 negative sites from wells with a negative geothermal potential, the majority of which were from deep oil and gas exploration wells. These positive and negative sites were combined to form the "labels" that were used to train and improve various ANN models including Bayesian neural networks (BNNs). The authors concluded that the Bayesian NN can predict geothermal resource potential and provide measures of confidence and reliability. The main challenges encountered throughout the study were the small size and possible imbalance of the training set, the diverse data types (a mixture of categorical and numerical data), and the complicated feature-label relationships.

Vesselinov et al. (2020) combined data from three previous PFA studies in Southwest New Mexico (SWNM), an area of known geothermal resources, to generate a data set containing 42 geological, geophysical, geochemical, and geothermal attributes at 207 locations. They then applied an unsupervised ML framework, the non-negative matrix factorization with a customized *k*-means clustering (NMF *k*) SmartTensors tool, developed by the Los Alamos National Laboratory (LANL), to the combined data set. Nonnegative matrix factorization splits the main data matrix into two smaller matrices representing hidden data structures, known as the signature and mixing matrices, while *k*-means clustering determines the optimal number of signatures. According to the authors, NMF *k* has important advantages over PCA, singular value decomposition, and independent component analysis (ICA) methods because it can deal with true and categorical variables as well as sparse datasets with large amounts of missing data. The analysis showed that NMF *k* found hidden structures in the data and revealed the optimal signal numbers. The study concluded that the main parameters characterizing SWNM geothermal systems include quartz water vapor temperature at 2 m depth; temperature at 250 m depth; silica (SiO_2); Calcium (Ca); Sulfate (SO_4); Sodium (Na); Na/Ca/K, Na/K, and Potassium/Magnesium (K/Mg) geothermometers; and bottom-hole temperature (BHT).

Holmes and Fournier (2022) suggested a new method for extending the use of ML in PFA predictions with uncertainty estimates. Normalized Shannon entropy was employed as the uncertainty metric to evaluate three sources of uncertainty: model representation, model parameterization, and feature interpolation. Four ML algorithms—LR, decision trees (DT), extreme gradient boosting (XGB), and ANN—were evaluated and trained to assess the potential of subsurface enthalpy resources in a research area with known geothermal resources in SWNM. The advanced XGB and ANN models outperformed the simpler LR and DT models for all geothermal gradient classes. This research identified continuous enthalpy trends hidden in a high-dimensional feature set and effectively generated geothermal gradient classification maps from four independent ML algorithms and a weighted ensemble model that showed a higher overall predictive ability. The ensemble also outperformed standard interpolation techniques that rely only on spatial patterns for prediction.

In summary, PFA seeks to increase the success rate of geothermal exploration by integrating the geological, geophysical, and geochemical parameters indicative of geothermal activity. When integrated with AI-based algorithms, PFA can identify signatures that lead to the location of hidden geothermal systems and improve the knowledge of complex geothermal feature-label relationships to predict resource potential and provide measures of confidence and reliability. However, the limitations of PFA include the need for extensive and costly diverse surface and subsurface data types that are not always available during the early stages of geothermal exploration. In addition, a significant amount of training data is required to replace expert knowledge in determining the influence weights of features for future applications.

2.2 | Integrated subsurface data set applications

Expert knowledge has been incorporated into the widely used PFA technique to analyze complex geological, geophysical, hydrochemical, and thermal datasets and the findings are combined to produce probabilistic predictions of geothermal resource prospectivity.

Researchers have used recent advances in ML on integrated subsurface datasets to provide an alternative approach to performing PFA by simultaneously processing integrated datasets to produce predictive results.

To analyze various datasets and effectively guide geothermal research and development, LANL created an open-source tool called GeoThermalCloud (GTC), which uses unsupervised and physics-informed ML algorithms. Vesselinov et al. (2022) used GTC and reviewed multiple geothermal datasets with promising results. This review included: (1) analyzing 18 data attributes at 44 SWNM sites, identifying low- and medium-temperature hydrothermal systems, determining prominent attributes and the spatial distribution of extracted hidden hydrothermal signatures, and proving blind predictions of regional physiographic regions; (2) analyzing 18 geochemical and thermal attributes with sparse datasets at 14 341 Great Basin locations and extracting hidden geothermal signatures associated with low-, medium-, and high-temperature geothermal systems, their main characterizing features, and their spatial distribution within the study area; (3) identifying key geological factors controlling geothermal production in the Brady Geothermal Field, Nevada; (4) investigating 19 geothermal attributes in 41 wells at Tohatchi Springs, New Mexico; (5) analyzing data attributes from four different Hawaiian islands separately and jointly and identifying low-, intermediate-, and high-temperature geothermal systems and their main characterizing features; and (6) performing prospectivity analysis to identify future drilling locations using 22 data attributes at 102 locations at the Utah FORGE site. The authors concluded that GTC ML techniques were validated with various datasets and were capable of producing prospectivity maps with favorable drilling areas for future geothermal exploration.

Mudunuru et al. (2022) also used GTC and proposed an enhanced PFA using ML. In their model, ML was applied throughout the PFA process, which helped to find hidden signatures from the filtered data and uncover physical information from site-specific workflows. In this model, GTC extracted critical geothermal signals, which were then analyzed by subject matter experts for effective blind geothermal discovery.

Meshalkin et al. (2020) and Shakirov et al. (2021) applied ML to an integrated subsurface data set from well log data to predict rock thermal conductivity (TC) and support geothermal exploration and enhanced oil recovery applications, which rely heavily on understanding the process of heat transfer within a reservoir. The data set featured the reservoir rock lithology, physical rock properties (porosity, mineralogy, and cementation degree), in situ pressure, and temperature. Meshalkin developed and evaluated eight regression models using well logs and core sample data from a heavy oil reservoir to characterize and predict vertical changes in rock TC. Among the ML methods, the random forest (RF) approach most accurately predicted rock TC from well logs. The research tabulated the coefficient of determination (R^2), root mean square error (RMSE), random error, and systematic error evaluation metrics of the eight regression models, and the results showed that when using RF, TC was determined with a total relative error of 12.54% compared to the experimental data.

Shakirov proposed and evaluated an improved approach for estimating TC and the volumetric heat capacity (VHC) from well data. This approach used three alternative theoretical models incorporating correction factors to determine rock TC, as well as advanced regression analysis to predict the thermal properties of subsurface rocks. The proposed method was evaluated using experimental data (thermal core log and well log data) from five drilled wells. The estimation of TC from well log data was achieved with uncertainty of less than 12% for TC parallel to the bedding plane and 15% for TC perpendicular to the bedding plane. In addition, rock VHC was estimated from well-log data with uncertainty under 5%. These applied case studies demonstrated that well-log-based estimates using the ML approach were more accurate than the theoretical model predictions.

Sadeghi and Khalajmasoumi (2015) assessed the geothermal potential of northwestern Iran (East Azerbaijan Province) by combining data on volcanic and intrusive rocks, volcanoes, geochemical parameters, faults, and fractures using fuzzy logic approaches with a binary index overlay in a Geographic Information Systems (GIS) framework. The integration of the information layers resulted in a conceptual model that identifies the relationship between information layers and targets. The authors compared all the approaches and the weighted index overlay proved to be the best approach, and therefore its use may be beneficial in future assessments, particularly when the number of information layers is insufficient. Their research indicated that the central part of the study area has the greatest potential for future geothermal exploration; however, an increase in the number of detailed information layers at larger scales is required to provide more information for exploration.

In summary, by simultaneously processing integrated geological, geophysical, geochemical, and thermal datasets, recent advances in ML have provided an alternative approach to conducting PFA and predicting subsurface thermal properties, and predictive results have been obtained that effectively guide geothermal research and development. However, this approach requires a significant amount of available data, and therefore significant upfront investment. In addition, this type of analysis requires expertise in data realization and interpretation to develop robust and efficient data-driven predictive models.

2.3 | Geochemical data applications

Groundwater chemistry (geochemistry) data are often used in the early stages of geothermal exploration because groundwater samples are typically less expensive to obtain and are more widely available than other geothermal data. The geochemical analysis provides useful information about hydrogeological processes, groundwater types, heat sources, recharge conditions, flow patterns, hydrogeochemical interactions between water and the host rock, and geothermal characteristics of the reservoir (such as reservoir temperature, heat flow, boundary conditions, and the geographical extent). Geochemical data include aqueous species, major cations, anions, isotopes, geothermometry, and tracer elements. Chemical, gas, and/or isotope geothermometers are used to calculate the subsurface reservoir temperature using empirical formulae based on the chemical composition of water/gas samples. However, geothermometers' use is limited because they assume an overall chemical equilibrium between the fluid and reservoir rock, whereas fluid re-equilibration can occur during migration to the surface. Other limitations include limited spatial coverage, microbial effects, and correction for total silica composition. These can lead to large uncertainties when determining the reservoir temperature. To this end, researchers have used AI to solve the current limitations of conventional geothermometers.

2.3.1 | Enhanced Na/K geothermometers

Ferhat Bayram (2001) developed new equations for the Na/K chemical geothermometer using a simple ANN with two input layers (Na, K), two hidden layers, and an output layer (reservoir temperature). The ANN was trained using multilayer perceptrons (MLP) and backpropagation, which are widely used in ANN classification, with 600 reservoir temperature training data derived from the outputs of six known geothermometers and synthetic Na/K data (Na given different values between 200 and 2000 mg/L, while K was kept constant at 20, 60, 80, 100, 120, 140, 180 and 200 mg/L for each calculation). However, the ANN model overestimated the subsurface temperature when compared with the actual temperature measurements from 20 geothermal fields in Turkey.

In a similar approach, Can (2002) used an ANN architecture with two input layers, a hidden layer, and an output layer to develop a new Na/K chemical geothermometer for estimating the subsurface temperature of geothermal reservoirs. The model was trained using the error backpropagation method to obtain the optimal network weight values, using 39 geothermal well data from worldwide locations with temperatures ranging from 94 to 345°C. The new geothermometer provided competent subsurface temperature predictions compared to recorded BHTs and had the lowest normalized mean square error value. However, as the developed geothermometer was derived on a purely empirical basis, the author advised that it should not be used outside the calibration range.

Similarly, Diaz-Gonzalez et al. (2008) developed three novel Na/K geothermometers, with two obtained using the ANN technique and the third obtained using the linear regression method. In the first geothermometer, the ANN design was trained using 212 actual reservoir temperature data collected from several geothermal fields worldwide as the input layer neurons. Backpropagation training was performed using a linear activation function, taking advantage of the linear nature of the Na/K geothermometer. The second geothermometer was trained via backpropagation using a time-tangent hyperbolic activation function. According to the results, the new geothermometers consistently provided deep equilibrium temperature estimates (for temperatures over 160°C) that were more accurate and consistent than those derived from traditional geothermometer equations mentioned in the literature.

Serpen et al. (2009) introduced a novel ANN-based Na/K geothermometer, adapted from the work of Ferhat Bayram (2001), Can (2002), and Diaz-Gonzalez et al. (2008), using the same 212 actual reservoir temperature data as training data, with Na and K measurements used as inputs and geothermometer temperatures used as outputs. Instead of backpropagation, a multilayer feedforward NN (FNN) was trained using a genetic algorithm to optimize the weights of the hidden layer neurons and a linear regression algorithm to optimize the weights of the output neurons. The model was evaluated successfully and demonstrated a deviation of approximately 10% when compared with actual subsurface temperature measurements. The results also suggested that additional reliable data are required to obtain a Na/K geothermometer with a good temperature prediction capability below 160°C, as the available training data set was limited in size.

2.3.2 | Enhanced gas geothermometers

To predict subsurface reservoir temperatures, Perez-Zarate et al. (2019) used a three-layer FNN, with four variables in the input layer, eight neurons for the hidden layer, and one neuron for the output layer, to perform a multivariate analysis of the gas phase composition of geothermal fluids. They collected gas phase fluid compositions and downhole temperature measurements from published literature and created a new database consisting of 591 samples from 149 production wells worldwide. The principal gas phase compositions of CO_2, H_2S, CH_4, and H_2 were specified as inputs, and the recorded downhole temperatures were used as output variables. Thus, 455 ANN architectures were successfully tested, and a sensitivity analysis was performed to identify the relationship between input factors and results. A robust ANN modeling technique involving three functional geochemical sub-databases was successfully implemented, and six ANN architectures emerged as the best tools for estimating subsurface temperatures based on statistical parameter estimation (with error percentage differences ranging from 2% to 11%). The best ANN architecture had three variables in the input layer, eight neurons in the hidden layer, and one neuron in the output layer. According to the authors, the results support the possibility of using an ANN for gas geothermometry to estimate subsurface temperatures.

Acevedo-Anicasio et al. (2021) developed eight new gas geothermometers using an ANN and an analytical computer program (GaS_GeoT) to evaluate the reliability of predicting subsurface temperatures. For the first time, a geochemometric study based on multi-criteria decision analysis was used to determine the best ANN for geothermal reservoir temperature prediction. The database of 591 geochemical samples of Perez-Zarate et al. (2019) was used to evaluate the predictive effectiveness of new gas geothermometers in geothermal fields. The prediction accuracies of these geothermometers were then compared with those of 25 other gas geothermometers currently in use. Figure 3 illustrates the workflow of the proposed model. The eight new gas geothermometers provided accurate reservoir temperature predictions for fluid-dominated reservoirs, and two of the eight provided accurate predictions for steam-dominated reservoirs. The results suggest that the developed gas geothermometers and the GaS_GeoT program can be used as geothermometric tools for accurate reservoir temperature estimation.

2.3.3 | Novel chemical geothermometers

Haklidir and Haklidir (2019) generated a geochemical data set from up to 83 thermal springs in Western Anatolia, Turkey, which represented relatively high-, medium-, and low-temperature geothermal systems. Critical hydrogeochemical indicators were selected, including the reservoir temperature, pH, electrical conductivity (EC), and Na, K, SiO_2, chloride (Cl), and boron (B) concentrations for each spring. Using the geochemical data set, the authors predicted subsurface reservoir temperatures using various ML methods, including support vector machines (SVMs), *k*-nearest neighbors (KNNs), and deep NNs (DNNs). They aimed to identify the algorithm with the lowest RMSE and mean absolute error (MAE), and the DNN approach produced the least errors and the most accurate reservoir temperature predictions when numerically compared with the actual temperature data. However, owing to limited hydrogeochemical data, they assumed that all the data provided were valid for conducting temperature estimations. The study also showed

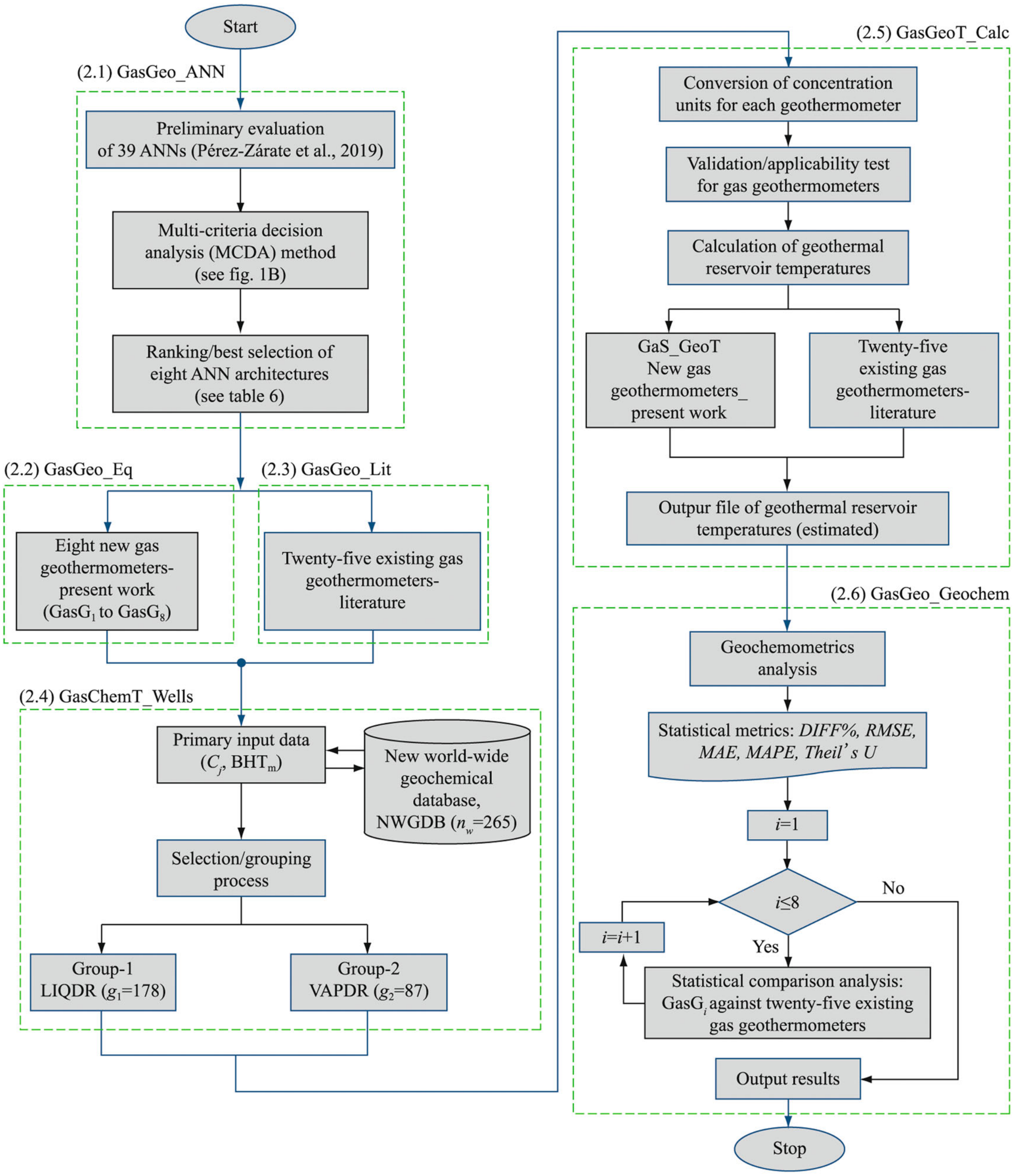

FIGURE 3 Schematic flow diagram showing the work methodology used in the study of Acevedo-Anicasio et al. (2021). (Reproduced under the terms of the Creative Commons Attribution (CC BY 4.0) license).

that when the data quality and quantity are increased, the accuracy and reliability of the estimates are also increased.

When applying AI to estimate subsurface reservoir temperatures, large amounts of geochemical data are required as training samples. However, data are limited for some geothermal sites. In this respect, Yang et al. (2022) proposed a new approach to train five ANN architectures with varied hidden layers to estimate reservoir temperature in the absence of training data. A numerical simulation of water–rock interaction was used to generate a geochemical and thermal data set for the Lindian geothermal field in China, which was used as training data, while 29 hot water samples and temperature log data from 11 geothermal wells in the same field were used as test data. The results were then correlated with conventional geothermometric calculations, and the prediction error of the ANN was found to be the lowest, followed by that of the Na/K geothermometer. In contrast, the prediction errors of the chalcedony geothermometer and integrated multicomponent geothermometry methods were the highest. The authors suggest that ANNs provide accurate reservoir temperature estimates, but only when there is a high degree of water-rock interface and no other complicated activities are involved, otherwise, the ANN-based technique is not valid and needs to be improved.

The Great Basin is the largest contiguous endorheic watershed region in the western United States, and it contains many geothermal reservoirs ranging from low- to high-temperature resources. Ahmmed et al. (2020) characterized the geochemical features of geothermal resources in the Great Basin by analyzing 15 geochemical attributes at 14 341 locations. Three temperature-constrained datasets were generated, and an unsupervised NMF k approach was applied to each data set; hidden signals were discovered that were indicative of hidden geothermal resources. The results showed that most of the major cations/anions were related to low-temperature geothermal resources, whereas fewer were related to medium-temperature geothermal resources. This research also suggested that tracer elements were key features to consider when characterizing high-temperature geothermal resources.

In summary, geochemical analysis is a cost-effective and noninvasive method for estimating subsurface reservoir temperatures in the initial phases of geothermal resource evaluation. However, geochemical analysis can limit the accuracy of temperature prediction because traditional geothermometers rely on certain assumptions about the composition of the geothermal fluid and reservoir conditions, which may not always be accurate. The main challenge in integrating AI models with geochemical datasets for reservoir temperature prediction is the lack of appropriate training data, which can lead to prediction uncertainty. Therefore, to improve the accuracy of subsurface temperature predictions, combining geochemical analysis with other methods, such as geophysical analysis, may be necessary.

2.4 | Geophysical data applications

Geophysical data is widely used in geothermal exploration to help scientists characterize the geology and structure of the subsurface and identify hidden geothermal resources. Electromagnetic (EM), magnetotelluric (MT), and seismic surveys, as well as remote sensing (RS), are the most commonly used geophysical techniques in geothermal exploration. Both EM and MT are used to investigate rock resistivity; however, the main difference between MT and EM surveys is that MT uses naturally occurring electromagnetic fields whereas EM uses artificial, controlled sources to probe the Earth's subsurface and map its electrical conductivity structure.

Rock resistivity and subsurface reservoir temperature have a strong relationship because they are influenced by the same parameters, such as porosity, permeability, and fluid salinity. The resistivity of rocks containing water as a pore fluid often decreases with increasing temperature because temperature tends to reduce the viscosity of pore fluids. Researchers used the pattern of resistivity change with temperature to train data-driven algorithms for estimating the subsurface temperature.

2.4.1 | Electromagnetic survey applications

Spichak and Zakharova (2009a) developed a novel method to estimate subsurface temperatures at depth using an indirect EM geothermometer. This method relies on an ANN architecture, which does not require prior knowledge of electrical conductivity mechanisms and provides temperature predictions based on an analysis of indirect conductivity–temperature relationships. Their method has been applied in different geological environments, such as in Tien Shan, Kyrgyzstan (Spichak et al., 2011); Soultz-sous-Forêts, France (Spichak et al., 2010); and Hengill, Iceland (Spichak & Zakharova, 2009b). Using the indirect EM geothermometer, four factors were found to influence temperature estimation errors: faulting, distance between the EM position and the predicted temperature location, meteoric and groundwater flow, and horizontal geological inhomogeneity. The results showed that the distance between the EM site and the wellbore and the ratio of the wellbore length to the extrapolation depth determined the temperature extrapolation accuracy. The application showed that the use of an indirect EM geothermometer allows correct temperature estimates to be made at depths beyond those of the boreholes for which temperature data are available. In particular, the relative error of extrapolation to a depth of twice the total borehole depth did not exceed 5% on average, and the error at three times the depth was approximately 20%. The indirect EM geothermometer was used to reconstruct the two- and three-dimensional (3D) temperature models of the study areas using EM sounding data, allowing important conclusions to be drawn about the dominant heat transfer processes, fluid circulation pathways, and new drilling locations.

Ishitsuka et al. (2018) observed that despite the successful application of the indirect EM geothermometer developed by Spichak and Zakharova, the estimation error increased with increasing distance from the recorded temperature logs. In an attempt to reduce the estimation error in their study, the authors used a neural kriging (NK) method to predict subsurface temperature patterns from the resistivity and temperature data of seven drilled wells in the Kakkonda geothermal field in northern Japan. Kriging was incorporated into the NN, and a variogram was used to improve the NN design. The study showed that the use of NK reduced both estimation and variogram errors better than the standard NN method, due to the fact that NK takes into account the spatial correlation of temperature. The temperature distribution predicted by NK is consistent with previous findings, and the results show that NK is effective in estimating temperatures from resistivity data.

Namaswa et al. (2021) developed a data-driven ML model to predict reservoir temperatures using 297 data samples of recorded EM resistivity and associated subsurface temperature measurements from the Olkaria Domes geothermal field in Kenya. The regression techniques used were decision tree regression (DTR), adaptive boosting, RF, and support vector regression (SVR). The R^2 and MAE metrics were used to assess the performance of the models, and the DTR algorithm proved to be the optimal model for predicting subsurface temperatures ($R^2 = 0.81$ and MAE = 29.8) based on the performance scores. The authors concluded that the DTR algorithm could be used to determine the subsurface temperature from the resistivity in high-temperature hydrothermal fields.

2.4.2 | Magnetotellurics survey applications

Akpan et al. (2014) used a shallow NN consisting of five inputs, two hidden layers, and one output layer, with limited volume MT-derived electrical resistivity records and 203 borehole temperature logs from the Tattapani geothermal field in central India as inputs to estimate subsurface temperatures. The data set was divided into a training and a test set (61% training and 39% test). Conventional statistical techniques (including the adjusted coefficient of determination, relative error, absolute average deviation, RMSE, and regression analysis) were used to evaluate the performance of the network. The study concluded that the developed network was structurally flexible and could be used to estimate subsurface temperatures despite the modest data size used to train the network.

Maryadia and Mizunaga (2022) used the ANN methods originally proposed by Spichak and Zakharova (2009a) and Akpan et al. (2014) in a geothermal field in Japan to outline the subsurface temperature distribution in the area and analyze the capacity of the ANN method in a new location with different geological conditions. In contrast with previous studies, the subsurface resistivity profile used in their investigation was generated using audio-MT (AMT) soundings rather than classical MT, because AMT is better at reproducing the shallow subsurface resistivity structure. The NN temperature profiles matched those measured from the surrounding boreholes, thereby providing a satisfactory assessment of the predictive capabilities of the model. Both resistivity and temperature anomalies showed good agreement, indicating the presence of subsurface anomalies, such as altered layers, reservoir zones, and probable faults. The authors concluded that the use of indirect geothermometers based on calibrated AMT survey data can help to reduce the cost of subsurface temperature estimation when a suitable training method is used. However, because of the complexity of the subsurface structure, the optimal NN model may differ from that used in their study. This method therefore requires testing in more locations to confirm its predictive accuracy. In addition, further data attributes, such as alteration intensity and/or hydrothermal minerals, could improve the accuracy of this estimation technique.

Sutarmin and Daud (2021) investigated the correlation between borehole temperature data and 3D inversion MT resistivity values to obtain reliable interpretation results. Borehole temperature records linked to resistivity values were used to predict temperature spread using an NN. The coordinates of the borehole location, resistivity gradient, and resistivity values of the target temperature at that location were used to create the vector, and the weights from NN training were employed to determine the temperature of the upcoming drill location (temperature vs. depth analysis). During training, the best MT readings close to the borehole were used so that the weight value represented the geothermal field. The weight of the NN was employed to estimate the temperature in 3D to provide the temperature distribution across the geothermal field. Together with other geophysical data, the results of NN 3D temperature modeling can be employed as a drilling guideline because they help to determine the flow area (the target area) and the outflow, thereby minimizing risks associated with geothermal development. Figure 4 shows the 200°C isothermal line from the NN predictions and the best of conductor (BOC) line from the 3D MT inversion data; the BOC and 200°C isothermal trend lines are similar. The architecture of the NN design has a significant influence on the prediction results; therefore, to select the optimal network for temperature prediction, the number of layers and neurons with significant variations must be determined.

Bayesian frameworks have been instrumental in quantifying and reducing uncertainty when modeling multiple data sets for predicting shallow features in geothermal fields. Most Bayesian frameworks incorporate a geophysical/thermal model to connect geophysical measurements to temperatures; they also allow the inclusion of different geological properties. Using the simulation results of a numerical reservoir model incorporating MT-derived resistivity, temperature records, and the geological boundary of the Kakkonda geothermal field in Japan, Ishitsuka et al. (2021) evaluated an NN technique with a Bayesian estimation method to estimate the subsurface temperature distribution. The study determined that both Bayesian estimation and NN methods provided subsurface temperature predictions compatible with temperature distributions produced by numerical models. Still, when considering the number of wells (25, 50, and 100), the Bayesian estimation method was more robust than the NN method. The Bayesian method also assesses parameters and estimation uncertainty. However, the NN method provides superior estimations of complicated temperature patterns with fewer/simpler assumptions.

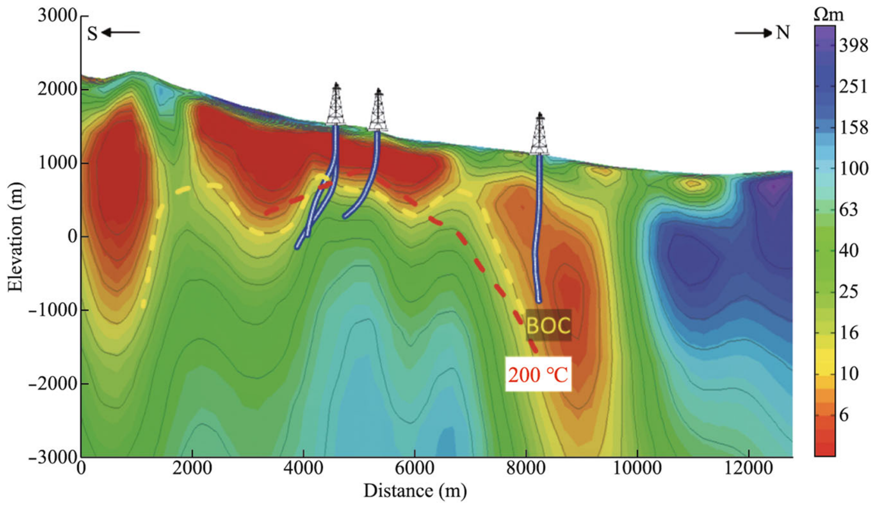

FIGURE 4 Isothermal at 200°C and Best of Conductor (BOC) lines (Sutarmin & Daud, 2021; reproduced with permission).

In late 2021, Ishitsuka et al. (2022) improved the previously proposed Bayesian rock physics model to include temperature, effective porosity, and salinity at depth. In particular, they investigated the spatial continuity of temperature and explored a novel scenario where pore-fluid salinity was considered as a variable. When applying the model to the Kakkonda geothermal field data, the estimated temperatures were consistent with the subsurface temperature records, supporting the geological hypothesis that Kakkonda granite is a heat source rock. The two examples analyzed showed remaining uncertainties depending on the geological context, namely effective porosity and the salinity distribution. However, the results indicated the presence of a magmatic-hydrothermal setting at depth. The authors found that Bayesian rock physics modeling was useful for predicting temperatures and constraining effective porosity and salinity distributions, even with limited data.

Hokstad and Tanavsuu-Milkeviciene (2017) used multi-geophysical inversion methods followed by a Bayesian framework to estimate the subsurface temperature of the Iceland Deep Drilling Project in Reykjanes. In this project, the formation temperature of the drilling target was estimated using electrical resistivity from MT inversion and density from gravity inversion. The temperature estimate was then updated, and a geological model was built using resistivity logs and core samples collected during drilling. The Bayesian network represented the temperature dependence of the geophysical model parameters. Various rock physics relationships provided the first set of dependencies, whereas differential equations (such as Maxwell's equations of electromagnetism, Newton's law of gravity, and the elastic wave equation) provided the second set of dependencies. Resistivity log updates during drilling indicated that the formation temperature was most likely above the model temperature estimate. The well was then drilled to a depth of approximately 4500 m, and core samples were collected at various depths. The model temperature estimates were found to be within the temperature ranges suggested by the alteration of minerals in the cores and the changes in rock properties. This proposed approach was only calibrated and validated for mid-ocean-ridge basalts; it is thus necessary to recalibrate the rock physics models used in inversion when applying it in other geological settings.

2.4.3 | Seismic survey applications

Seismic images provide important geological information on subsurface structure, stratigraphy, and reservoir properties. This information, including the fault/fracture zones, is critical to understanding the stress field and potential fluid flow paths for geothermal permeability evaluation. Researchers are using ML techniques to interpret seismic data in complex geological environments where traditional methods struggle to detect patterns and indicators of geothermal activity.

The conventional double-beam fracture characterization method only provides information on the fracture characteristics around a target, such as the fracture direction, density, and compliance. The discrete fracture network, which is almost invisible in conventional seismic migration images, was mapped onto the data by Zheng et al. (2021) using the double-beam NN (DBNN) model, an image-to-image learning technique. The authors used a synthetic seismic data set to construct a subsurface model that included small-scale fractures within the Soda Lake geothermal field in the United States. The study showed that the DBNN model accurately predicted the number of fracture sets and the orientation of the fractures. Errors in the DBNN fracture location were within 5 m using 60 Hz P-wave data for targets at a depth of approximately 300 m.

Gao et al. (2021) also produced high-resolution fault maps for the Soda Lake geothermal field using seismic imagery and a novel ML-based fault detection method based on a multiscale connection-fusion U-shaped (MCFU) convolutional NNs (CNN). The MCFU approach uses skip connections to connect feature maps of variable spatial resolution, and a fusion operation to produce the final fault map. According to the authors, the MCFU also provides sharper and explainable fault maps for complicated seismic images compared with the commonly used ant-tracking approach and the typical U-shaped CNN, potentially leading to improved geological interpretability of the detected faults, which is critical for optimizing well placement to maximize geothermal energy recovery. The authors' future work aims to train an NN for 3D fault detection.

Perozzi et al. (2021) concentrated on the quantitative interpretation of existing seismic lines in the Canton of Geneva, Switzerland, using an unsupervised *k*-means clustering algorithm, which allows the identification of similar groups of clusters within a seismic image. The *k*-means method classifies seismic data by dividing the samples into n clusters of equal variances, where the number of clusters must be specified. The study aimed to demonstrate how quantitative approaches can be used to automatically detect faults, analyze seismic facies, and identify lithofacies from geophysical borehole logs. The results showed that when used in conjunction with expert knowledge (geologist or geophysicist), these techniques are an additional tool for gaining knowledge about and characterizing geological reservoirs, thereby reducing uncertainty in the subsurface.

Matzel et al. (2021) used an unsupervised ML approach to analyze and define high-permeability productive zones as potential drilling locations within the Raft River geothermal resource area, in the United States of America, by combining 3D seismic and MT datasets. For the combined data set, *k*-means clustering was used to differentiate the lithologies and validate them against known lithological boundaries. The study focused on 3D seismic reflection and the 3D resistivity volume, which were supported by additional datasets. The 3D seismic reflection data incorporated several seismic features in addition to the basic physical aspects of density and velocity. The MT data were reprocessed to produce the spatial gradients of resistivity, and three lithological units were clearly distinguished: the Raft River Formation (low resistivity), Salt Lake Formation (medium resistivity), and basement (very high resistivity). The results were correlated with known formations based on well-log records and mapped seismic horizons, and a strong correlation was observed between the geophysical measurements and the recognized geology. Cluster analysis showed that the boundary between the basement and overlying geology was clearly defined, and this revealed the main production zone of the geothermal reservoir.

2.4.4 | Remote sensing applications

RS data provide information on surface deformation caused by geothermal activity and gas emissions, which can be used for mineral mapping, structural analysis, and temperature and heat flow measurements. However, combining all the surface information to gain a full understanding of the distribution of geothermal sites requires a high level of expertise. Researchers used AI techniques to process and analyze satellite imagery and geospatial remote sensing data to accurately map geothermal potential and identify areas with hidden geothermal resources.

Using RS satellite imagery, Abubakar et al. (2019) evaluated the efficiency of using ICA and mixture-tuned matched filtering clustering algorithms to detect hydrothermal alteration and thermal anomaly zones, to conduct an initial assessment of the geothermal potential in the aseismic geological setting of Yankari Park, Nigeria. When mixture-tuned matched filtering was applied, the use of the verified image endmember spectra yielded much better results than the use of library spectra. The geographical association between anomalous regions and the major fault-fracture systems on the geological map of the study area showed a strong relationship with the previously observed thermal gradients within thermally preserved sedimentary formations. The study also indicated that further research should be conducted to identify hidden geothermal areas in unexplored regions by examining the spatial correlation between geological structures and thermal anomalies.

Moraga et al. (2022) developed a method incorporating RS and AI to provide a preliminary assessment of the geothermal potential and to analyze multiple satellite images and geospatial data to determine mineral markers, surface temperature, faults, and deformation. The research was conducted at the Brady and Desert Peak geothermal sites in the United States, which are adjacent but have different resource properties. In addition, the former has visible surface manifestations, and the latter is within a blind/hidden location. To generate patterns of the surface manifestations of the former and markers of the latter, they used unsupervised ML methods and conducted spatial and temporal analyses of geothermal indicators and automatic labeling. Subsequently, they created "Geothermal AI," a deep learning (DL) AI algorithm that used the patterns to predict geothermal sites. Its use was tested by employing independent datasets from each site; an accuracy of 92%–95% was achieved, with the lowest accuracy relating to the blind site. Geothermal AI trained on one location was also tested by running it in the other site to predict geothermal/nongeothermal classification, and it showed good performance with a prediction accuracy of 72%–76%. However, the model was designed for early exploration, and it only identified the outline of a geothermal reserve and not its prospect. A subsurface model was required to determine the total capacity of the geothermal resources, and the subsurface prediction model was therefore extended to estimate and assess the total capacity.

In summary, geophysical analysis is a noninvasive and economical method that can be used to improve the knowledge of subsurface features and rock properties. However, they also have limitations, such as the limited resolution, which means that they may not be capable of detecting small-scale features or variations in the subsurface. In addition, the data obtained can be complex and require specialized knowledge to interpret. AI algorithms can be used to find patterns and anomalies in the geophysical data collected, such as those obtained using seismic, electrical, and magnetotelluric surveys. These patterns can reveal information about the location, dimension, and quality of potential geothermal reservoirs. The geophysical analysis combined with ML algorithms can help identify hidden geothermal reserves, reduce the cost and time of exploration, and diminish the environmental impact of the exploration phase.

2.5 | Thermal data applications

Thermal gradient, BHT, shallow and deep heat flows, and temperature measurements at different depths are examples of subsurface thermal data. Researchers used data-driven models integrated with existing thermal data to extend the knowledge of thermal properties in new regions.

Koike et al. (2001) evaluated the interpolation accuracy of an NN-based interpolator, NK, and applied it to the distribution analysis of subsurface temperatures in the Hohi geothermal area, southwest Japan, using limited log data from 20 wellbores with an average depth range of 1260 m. The values estimated by NK were similar to those of sample data, and the spatial correlation associated with regionalized variables was recreated with interpolation errors lower than those of the ordinary multilayer NN (MNN) and ordinary kriging. Despite the limited number and distribution of temperature survey data, the NK extrapolated sample values and the interpolation results indicated the existence of high-temperature zones and convection patterns of hydrothermal fluids. In this respect, the NK method can deal with any isotropic or anisotropic semivariogram, even if it cannot be represented by any experimental model.

Spichak (2006) also used a shallow NN approach and downhole temperature logs to map the subsurface temperature distribution of a hypothetical geothermal reservoir. The NN, also known as the neuronet approach, was calibrated against an analytical model result, and results showed that errors in neuronet temperature prediction were influenced by the "training level" of the neuronet and the distance between the estimation point and the location for which data were available. These findings were validated by predicting the temperatures in eight wells using 50 temperature records from other geothermal wells. Spichak found that increasing the training datasets for the NN reduced the average estimation error for all cases studied to 16.9%, illustrating the importance of obtaining sufficient training data to reduce errors in the NN model. The impact of the training data pattern (conductive vs. convective temperature profiles) was investigated, and a methodology for NN training based on the available data was established.

Shahdi et al. (2021) investigated the applicability of four different ML algorithms (deep NN, ridge regression model, decision tree-based XGB, and RF model) to predict the subsurface temperature and geothermal gradient parameters from BHT data and geological information from over 20 750 oil and gas wells in the northeastern United States. In terms of predicting the subsurface temperature, XGB and RF outperformed all other models. The authors used the XGB model to generate continuous two-dimensional temperature maps at three different depths, which were subsequently used to identify potential geothermal resources, as shown in Figure 5. They also compared physics-based and ML models using an additional data set of vertical temperature profiles from 58 wells in West Virginia and concluded that ML models are highly comparable to physics-based models in predicting subsurface temperatures according to the evaluation metrics.

BHT is the temperature measured at a certain depth in the well during logging operations. Owing to transient disturbances associated with mud circulation, BHT is generally lower than the actual stabilized or static formation temperature (SFT). SFT is used to identify the original temperature of the surrounding formation rock and is valuable in many geothermal applications, such as locating heat flow or lost circulation zones, estimating heat reserves in a geothermal reservoir, evaluating geothermal gradients, interpreting electrical logs, and evaluating in-situ thermophysical rock properties. However, estimating SFT at any depth is time-consuming and expensive because drilling needs to be paused temporarily and specialized logging equipment is required. SFT is, therefore, often estimated using analytical and numerical simulation approaches that use BHT and shut-in time data as inputs and linear or nonlinear regression models as solutions. However, significant errors in predicting SFT occur, and these are related to a variety of factors, including impractical models offered to characterize the drilling process, heat transfer models with simple assumptions, and errors in BHT measurements.

Bassam et al. (2010) used an ANN method to generate a novel estimation model for calculating the SFT in geothermal wells. They trained a three-layer ANN structure with BHT records and shut-in times as primary inputs and transient temperature gradients as secondary inputs using an experimental geothermal well database containing "statistically normalized" SFT estimates. The best training data set was achieved using an ANN design consisting of five neurons in the hidden layer; this allowed for the prediction of SFT with good efficiency ($R^2 > 0.95$). A statistical comparison between recorded and predicted values showed accurate predictions with errors <5%, confirming that the new ANN model is a reliable tool for SFT estimations using only BHT and shut-in time as input data (mainly when BHT data are in the training interval of 45–263°C), making it a more advantageous tool than the existing analytical methods relying heavily on other complicated input variables.

Espinoza-Ojeda and Santoyo (2016) also attempted to estimate the SFT using a new empirical technique based on logarithmic transformation regressions, where several linear and polynomial regression models were applied to the BHT and logarithmic transformation shut-in times. The best regression models were selected using four statistical factors, and the optimum model was used to reproduce the thermal recovery process of the wells and calculate the SFT. The original BHT and shut-in-time data were then used to illustrate the effectiveness of the new technique. When the new log-transformation regression approach was applied to a full range of geothermal, oil, and permafrost wells, the polynomial models were found to be the best regression models for describing the thermal recovery processes. The outcomes showed that this new approach can be used for an accurate estimation of SFT. The approach involves BHT and shut-in time measurements as the primary input data, and it therefore offers significant advantages over conventional analytical methods that require many measurements. Notably, the quantity and quality of the recorded data are crucial for the accurate

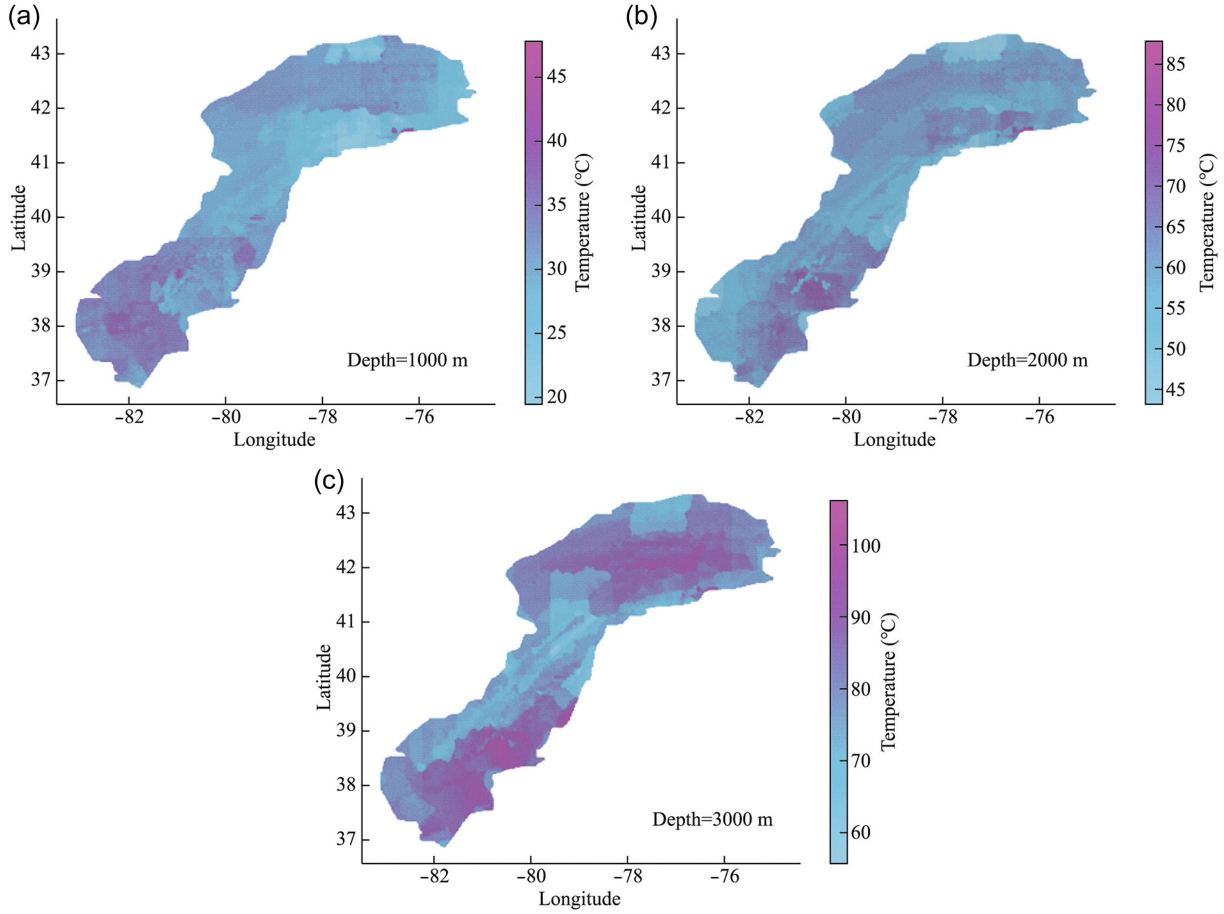

FIGURE 5 Temperature map at three different depths obtained using an extreme gradient boosting (XGB) model. (a) 1000 m, (b) 2000 m, and (c) 3000 m. (Shahdi et al., 2021; reproduced under the terms of the Creative Commons Attribution (CC BY 4.0) license).

characterization, prediction, testing, and estimation of the potential of geothermal wells.

In summary, geothermal exploration is a complex process that requires a thorough understanding of subsurface temperature and heat flow parameters. The application of AI to geothermal exploration can provide an efficient and accurate method for estimating subsurface thermal properties, including subsurface temperatures, temperature gradients, TC, and heat flows. AI algorithms can analyze large amounts of data to build accurate subsurface temperature and heat transport models, which can then be used to identify hidden geothermal resources.

2.6 | Other data applications

Researchers have used other available datasets, including well surface locations, borehole angles, site photographs, and AI, to assist in the exploration of geothermal resources. Porkhial et al. (2015) used a data set extracted from six exploration wells at the Sabalan geothermal site in Iran with five variables as input data (well northing and easting, total depth, hole inclination, and hole azimuth) and one output (subsurface temperature at total depth) to train a group method of data handling (GMDH)-type NN model for predicting geothermal reservoir temperature behavior. To determine the best set of acceptable quadratic expression coefficients, the authors examined the number of neurons in each hidden layer and their connectivity configuration in conjunction with singular value decomposition. A comparison between the actual results and the developed GMDH model demonstrated the predictive capabilities of the proposed model and showed that two elements (well northing and azimuth direction) did not affect the borehole subsurface temperature.

Xiong et al. (2022) investigated the possibility of using the GoogLeNet deep learning approach with photographs and image interpretation to detect geothermal surface manifestations (GSMs) and compared their results with existing ML models to aid geothermal resource exploration. They generated a new image data set containing seven GSM types (warm springs, hot springs, geysers, fumaroles, mud pots, hydrothermal alteration, and crater lakes) and one non-GSM type, as demonstrated in Figure 6. The results of the GoogLeNet model were correlated with those of three ML approaches (SVMs, DTs, and KNN) using the evaluation metrics of overall accuracy (OA), overall F1 score (OF), and computational time (CT) to train and test the models via cross-validation. This study demonstrated that the GoogLeNet model, retrained using transfer learning, significantly outperformed the SVM, DT, and KNN models and provided the highest OA, OF, and CT for both validation and testing. The results showed that deep

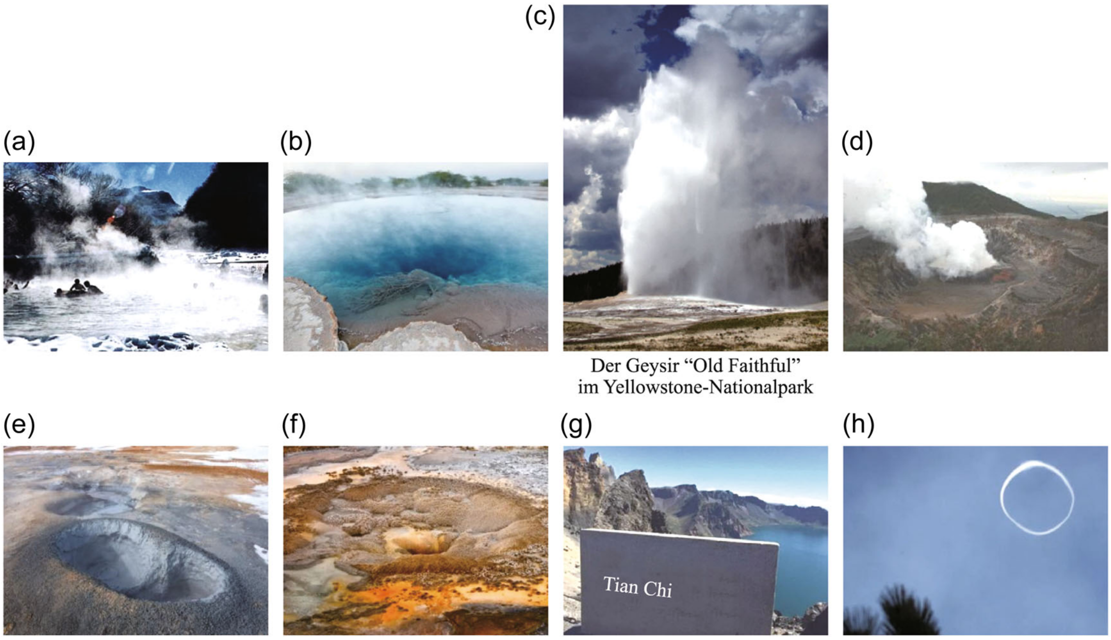

FIGURE 6 Example images of the types of geothermal surface manifestations (a) warm spring, (b) hot spring, (c) geyser, (d) fumarole, (e) mud pot, (f) hydrothermal alteration, (g) crater lake, and (h) a non-GSM type. (Xiong et al., 2022; reproduced under the terms of the Creative Commons Attribution (CC BY 4.0) license).

transfer learning using a pre-trained network may be a viable option for GSM recognition.

3 | CONSIDERATIONS AND OPPORTUNITIES

Based on a literature review of recent advances in AI applications, this study provides a perspective on the use of AI in geothermal exploration, identifying the AI algorithms used, the challenges being addressed, and the opportunities for further application. The review highlights that the most important subsurface features in geothermal exploration decision-making are reservoir temperatures, fluid types, and flow rates. It also highlights that most research has focused on the application of geophysical data, while NNs are the most widely used AI techniques among all geothermal exploration groups, both in ML and DL.

The exploration of geothermal resources involves the management of subsurface, technical, and economic uncertainties. Subsurface uncertainties arise from limited knowledge of geological, geophysical, geochemical, and thermal properties. Technical uncertainties arise from the limitations of tools and data acquisition, which create challenges in the processing and interpretation of subsurface features. Affected by the first two uncertainties, economic uncertainties can influence decision-making on field development strategies. The use of AI in geothermal exploration applications can reduce uncertainties and minimize costs, but the scarcity of large, high-quality subsurface datasets and the challenges of accessing them present a critical barrier to the widespread adoption of data-driven models in geothermal exploration.

Organized data repositories and efficient data processing, preparation, and transformation are critical to enabling generalizable AI algorithms. Facilitating access to data, optimizing it for AI applications, and providing specialized AI training to professionals in the geothermal industry can make geothermal energy a cost-effective, widely accessible, and geographically diverse source of power generation. One example is the OpenEI Geothermal Data Repository (Open, 2023), an open-source data collection node funded by the US Department of Energy's Geothermal Technologies Office and used by many researchers.

The authors suggest collaborative efforts between governments, industry organizations, and data owners to establish comprehensive databases of subsurface hydrothermal resources. The Geothermal Operational Optimization using ML (GOOML) project exemplifies industry collaboration to enhance the capacity and capabilities of the geothermal industry by analyzing historical production data. Siratovich et al. (2020) suggest that the use of datasets from New Zealand and the United States can lead to the development of generalized data-driven geothermal algorithms. These algorithms can predict market conditions, optimize maintenance operations, and recommend system component setpoints for optimal power generation. These insights will enable the prediction and scheduling of maintenance tasks, resulting in improved operational efficiency, higher capacity factors, and minimized levelized energy costs, ensuring the competitiveness of geothermal power generation (Craig et al., 2021; Smith et al., 2023).

4 | CONCLUSIONS

The rapid development of AI algorithms offers a significant opportunity for their integration into the geothermal industry. Transfer learning enables AI models trained in one domain to be applied to related domains with limited data, such as using techniques from the oil and gas industry in geothermal exploration. This study investigates advanced AI approaches in geothermal resource exploration, emphasizing innovative geoscience and subsurface engineering concepts resulting from AI applications.

The combination of hydrothermal data with NNs (both ML and DL) has been primarily used to predict subsurface hydrothermal reservoir temperatures. This review showcases the effectiveness of AI in determining reservoir properties, predicting lithology and fluid content, detecting mineral constituents, and identifying optimal locations. In comparison to basic physics-based and statistical methods, AI has the potential to improve the efficiency of geothermal exploration. However, there are challenges to overcome, such as reduced prediction accuracy when models trained in one location are applied elsewhere. Future studies should focus on improving AI accuracy across different regions to increase exploration efficiency. There is a significant opportunity to leverage AI in geochemical data applications to infer reservoir characteristics and reduce development costs and risks associated with geothermal projects.

The growing use of AI in geothermal exploration suggests its continued growth. Although DL techniques provide additional research opportunities, the lack of significant geothermal data remains a crucial challenge that must be addressed for AI to have a transformative impact on geothermal exploration. It is believed that combining all accessible geothermal information (raw and moderated) from various open-source data repositories can revolutionize geothermal resource exploitation and hidden geothermal system discovery.

Shared initiatives across the geothermal industry are essential to make multiple datasets and insights accessible for the rapid development of AI approaches. Encouraging interdisciplinary collaboration and promoting AI use through various channels, such as conferences and training activities, can enhance exploration performance and reduce costs. Partnerships between academic and professional organizations play a crucial role in accelerating the development of AI approaches on a larger scale. The application of AI in geothermal exploration has significant potential to improve efficiency, effectiveness, and productivity in geothermal energy development, ensuring competitiveness in the broader energy industry. Our findings suggest that AI will be pivotal in the future of geothermal energy exploration and development.

AUTHOR CONTRIBUTIONS

Mahmoud AlGaiar: Conceptualization; methodology; formal analysis; investigation; writing—original draft. **Mamdud Hossain**: Writing—review and editing. **Andrei Petrovski**: Writing—review and editing. **Aref Lashin**: Writing—review and editing. **Nadimul Faisal**: Writing—review and editing; supervision; project administration.

ACKNOWLEDGMENTS

The authors would like to acknowledge Robert Gordon University for providing resources and support for this review paper and their invaluable contributions to the completion of this research. The authors have no funding to report.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

We would like to declare that no specific software or script is associated with the work presented in this review. We acknowledge the importance of code availability for reproducibility and transparency in research and assure readers that all methodologies and results are thoroughly described in the review paper.

ORCID

Nadimul Faisal http://orcid.org/0000-0001-5033-6336

REFERENCES

Abubakar AJA, Hashim M, Pour AB. Remote sensing satellite imagery for prospecting geothermal systems in an aseismic geologic setting: Yankari Park, Nigeria. *Int J Appl Earth Obs Geoinf*. 2019;80: 157-172. doi:10.1016/j.jag.2019.04.005

Acevedo-Anicasio A, Santoyo E, Pérez-Zárate D, Pandarinath K, Guevara M, Díaz-González L. GaS_GeoT: a computer program for an effective use of newly improved gas geothermometers in predicting reliable geothermal reservoir temperatures. *Geothermal Energy*. 2021;9:1-41. doi:10.1186/s40517-020-00182-9

Ahmmed B, Vesselinov VV, Mudunuru MM, Middleton RS, Karra S. *Geochemical Characteristics of Low-, Medium-, and Hot-Temperature Geothermal Resources of the Great Basin, USA (No. LA-UR-20-28011)*. Los Alamos National Laboratory; 2020.

Akpan AE, Narayanan M, Harinarayana T. Estimation of subsurface temperatures in the Tattapani geothermal field, Central India, from limited volume of magnetotelluric data and borehole thermograms using a constructive back-propagation neural network. *Earth Interact*. 2014;18:1-26. doi:10.1175/2013EI000539.1

Aljubran J, Nwosu C, Okoroafor E, Smith C, Gudmundsdottir H. Recent trends in artificial intelligence for subsurface geothermal applications. In *47th Workshop on Geothermal Reservoir Engineering*. Stanford University of California; 2022:8. https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2022/Aljubran1.pdf

Bassam A, Santoyo E, Andaverde J, Hernández JA, Espinoza-Ojeda OM. Estimation of static formation temperatures in geothermal wells by using an artificial neural network approach. *Comput Geosci*. 2010;36(9):1191-1199. doi:10.1016/j.cageo.2010.01.006

Brown S, Coolbaugh MF, Deangelo J, et al. Machine learning for natural resource assessment: an application to the blind geothermal systems of Nevada. *Trans Geotherm Resour Counc*. 2020; 44:920-932.

Brown S, Rodi WL, Gu C, et al. Bayesian neural networks for geothermal resource assessment: prediction with uncertainty. *arXiv preprint arXiv*. 2022;2209:15543.

Can I. A new improved Na/K geothermometer by artificial neural networks. *Geothermics*. 2002;31(6):751-760. doi:10.1016/S0375-6505(02)00044-5

Craig JW, Faulds JE, Hinz NH, et al. Discovery and analysis of a blind geothermal system in southeastern Gabbs Valley, western Nevada, USA. *Geothermics*. 2021;97:102177. doi:10.1016/j.geothermics.2021.102177

Diaz-Gonzalez L, Santoyo E, Reyes-Reyes J. Three new improved Na/K geothermometers using computational and geochemiometrical tools: application to the temperature prediction of geothermal systems. *Rev Mex Cienc Geológicas*. 2008;25:465-482.

Espinoza-Ojeda OM, Santoyo E. A new empirical method based on log-transformation regressions for the estimation of static formation temperatures of geothermal, petroleum and permafrost boreholes. *J Geophys Eng*. 2016;13:559-596. doi:10.1088/1742-2132/13/4/559

Faulds J, Craig J, Coolbaugh M, Hinz N, Glen J, Deoreo S. Searching for blind geothermal systems utilizing play fairway analysis, western Nevada. *GRC Bullet*. 2018;47:10. https://www.osti.gov/servlets/purl/1724102

Faulds J, Hinz N, Coolbaugh M, et al. Integrated geologic and geophysical approach for establishing geothermal play fairways and discovering blind geothermal systems in the Great Basin region, western USA: a progress report. *GRC Trans*. 2015;39: 691-700. https://www.osti.gov/servlets/purl/1724089

Faulds J, Hinz N, Coolbaugh M, et al. Progress report on the Nevada play fairway project: integrated geological, geochemical, and geophysical analyses of possible new geothermal systems in the Great Basin region. In *Proceedings, 42nd Workshop on Geothermal Reservoir Engineering*. Stanford University, Nevada Bureau of Mines and Geology, University of Nevada; 2017.

Faulds J, Hinz N, Coolbaugh M, et al. Vectoring into potential blind geothermal systems in the Granite Springs Valley Area, western Nevada: application of the play fairway analysis at multiple scales. In *Proceedings, 44th Workshop on Geothermal Reservoir Engineering*. Stanford University, Nevada Bureau of Mines and Geology, University of Nevada; 2019:11. https://www.osti.gov/servlets/purl/1724085

Faulds J, Brown S, Coolbaugh M, et al. Preliminary report on applications of machine learning techniques to the Nevada geothermal play fairway analysis. In *45th Workshop on Geothermal Reservoir Engineering*. Stanford University, Nevada Bureau of Mines and Geology, University of Nevada; 2020:229-234. https://pubs.usgs.gov/publication/70211874

Faulds JE, Hinz NH, Coolbaugh M, et al. *Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways: All Phases*. University of Nevada; ATLAS Geosciences, Inc.; United States Geological Survey; Hi-Q Geophysical, Inc.; Innovate Geothermal Ltd; Utah Geological Survey; 2021. https://doi.org/10.2172/1724080

Ferhat Bayram A. Application of an artificial neural network model to a Na–K geothermometer. *J Volcanol Geother Res*. 2001;112:75-81. doi:10.1016/S0377-0273(01)00235-9

Gao K, Huang L, Lin R, Hu H, Zheng Y, Cladohous T. Delineating faults at the Soda Lake geothermal field using machine learning. In *Proceedings of the 46th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2021:8. https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2021/Gao2.pdf

Gehringer M, Loksha V. *Geothermal Handbook: Planning and Financing Power Generation: ESMAP Technical Report No. 002/12*. World Bank; 2012.

Haklidir FST, Haklidir M. The fluid temperature prediction with hydro-geochemical indicators using a deep learning model: a case study western Anatolia (Turkey). In *43rd Workshop on Geothermal Reservoir Engineering*, 2019:6. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2019/Tut.pdf

He Y, Bai Y, Tian D, Yao L, Fan R, Chen P. A review of geoanalytical databases. *Acta Geochimica*. 2019;38(5):718-733. doi:10.1007/s11631-019-00323-3

Hokstad K, Tanavasuu-Milkeviciene K. Temperature prediction by multigeophysical inversion: application to the iddp-2 well at Reykjanes. *Iceland. Trans Geotherm Resour Counc*. 2017;41:1141-1152.

Holmes RC, Fournier A. Machine learning-enhanced play fairway analysis for uncertainty characterization and decision support in geothermal exploration. *Energies*. 2022;15:1929. doi:10.3390/en15051929

International Renewable Energy Agency. *Geothermal, international renewable energy agency*. 2023. Accessed March 18, 2023. https://www.irena.org/Energy-Transition/Technology/Geothermal-energy

Ishitsuka K, Kobayashi Y, Watanabe N, et al. Bayesian and neural network approaches to estimate deep temperature distribution for assessing a supercritical geothermal system: evaluation using a numerical model. *Nat Resour Res*. 2021;30:3289-3314. doi:10.1007/s11053-021-09874-w

Ishitsuka K, Mogi T, Sugano K, Yamaya Y, Uchida T, Kajiwara T. Resistivity-based temperature estimation of the Kakkonda Geothermal Field, Japan, using a neural network and neural kriging. *IEEE Geosci Remote Sens Lett*. 2018;15:1154-1158. doi:10.1109/LGRS.2018.2832647

Ishitsuka K, Yamaya Y, Watanabe N, et al. Constraining temperature at depth of the Kakkonda Geothermal Field, Japan, using Bayesian rock-physics modelling of resistivity: implications to the deep hydrothermal system. *Geothermics*. 2022;100:102316. doi:10.1016/j.geothermics.2021.102316

Koike K, Matsuda S, Gu B. Evaluation of interpolation accuracy of neural kriging with application to temperature-distribution analysis. *Math Geol*. 2001;33:421-448. doi:10.1023/A:1011084812324

Liu R, Misra S. *Machine Learning Assisted Recovery of Subsurface Energy: A Review*. ESS Open Archive; 2022.

Maryadi M, Mizunaga H. Subsurface temperature estimation in a geothermal field based on audio-frequency magnetotelluric data. *Explor Geophys*. 2022;53:275-288. doi:10.1080/08123985.2021.1949945

Matzel E, Magana-Zook S, Mellors RJ, Pullimmanappallil S, Gasperikova E. Looking for permeability on combined 3d seismic and magnetotelluric datasets with machine learning. In *Proceedings of the 46th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2021:5. https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2021/Matzel.pdf

Meshalkin Y, Shakirov A, Popov E, Koroteev D, Gurbatova I. Robust well-log based determination of rock thermal conductivity through machine learning. *Geophys J Int*. 2020;222:978-988. doi:10.1093/gji/ggaa209

Moraga J, Duzgun HS, Cavur M, Soydan H. The geothermal artificial intelligence for geothermal exploration. *Renew Energy*. 2022;192: 134-149. doi:10.1016/j.renene.2022.04.113

Mudunuru MK, Vesselinov VV, Ahmmed B. GeoThermalCloud: machine learning for geothermal resource exploration. *J Mach Learn Model Comp*. 2022;3:57-72. doi:10.1615/JMachLearnModelComput.2022046445

Muther T, Syed FI, Lancaster AT, Salsabila FD, Dahaghi AK, Negahban S. Geothermal 4.0: AI-enabled geothermal reservoir development- current status, potentials, limitations, and ways forward. *Geothermics*. 2022;100:102348. doi:10.1016/j.geothermics.2022.102348

Namaswa S, Githiri J, Mariita N, K'Orowe M, Njiru N. Application of data-driven discovery machine learning algorithms in predicting geothermal reservoir temperature from geophysical resistivity method. *Trans Mach Learn Artif Intell*. 2021;9:1-12. doi:10.14738/tmlai.94.10281

Okoroafor ER, Smith CM, Ochie KI, Nwosu CJ, Gudmundsdottir H, (Jabs) Aljubran M. Machine learning in subsurface geothermal energy: two decades in review. *Geothermics*. 2022;102:102401. doi:10.1016/j.geothermics.2022.102401

Open EI. Linked energy information on hundreds of topics crowdsourced from industry and government agencies; 2023. Accessed March 18, 2023. https://openei.org/wiki/Main_Page

Pandey SN, Vishal V, Chaudhuri A. Geothermal reservoir modeling in a coupled thermo-hydro-mechanical-chemical approach: a review. *Earth Sci Rev*. 2018;185:1157-1169. doi:10.1016/j.earscirev.2018.09.004

Pérez-Zárate D, Santoyo E, Acevedo-Anicasio A, Díaz-González L, García-López C. Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. *Comput Geosci*. 2019;129:49-68. doi:10.1016/j.cageo.2019.05.004

Perozzi L, Guglielmetti L, Moscariello A. Geothermal reservoir characterization using seismic and machine learning: a case study from the Geneva Basin. In *Proceedings of the World Geothermal Congress*, Reykjavik, Iceland. 2021:8. https://pangea.stanford.edu/ERE/db/WGC/papers/WGC/2020/32007.pdf

Porkhial S, Salehpour M, Ashraf H, Jamali A. Modeling and prediction of geothermal reservoir temperature behavior using evolutionary design of neural networks. *Geothermics*. 2015;53:320-327. doi:10.1016/j.geothermics.2014.07.003

Sadeghi B, Khalajmasoumi M. A futuristic review for evaluation of geothermal potentials using fuzzy logic and binary index overlay in GIS environment. *Renew Sustain Energy Rev*. 2015;43:818-831. doi:10.1016/j.rser.2014.11.079

Serpen G, Palabiyik Y, Serpen U. An artificial neural network model for Na/K geothermometer. In *Proceedings of the 34th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2009:12. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2009/serpen.pdf

Shahdi A, Lee S, Karpatne A, Nojabaei B. Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of Northeastern United States. *Geothermal Energy*. 2021;9:18. doi:10.1186/s40517-021-00200-4

Shakirov A, Chekhonin E, Popov Y, et al. Rock thermal properties from well-logging data accounting for thermal anisotropy. *Geothermics*. 2021;92:102059. doi:10.1016/j.geothermics.2021.102059

Siratovich PA, Blair A, Gooml WJ. Geothermal operational optimization with machine learning. *GRC Trans*. 2020;44:955-964. https://cdn.prod.website-files.com/63db257f39a0ee46cb665e15/64582b4633d41e090d50dfc2_1034265%20GRC%202020%20Siratovich.pdf

Smith CM, Faulds JE, Brown S, et al. Characterizing signatures of geothermal exploration data with machine learning techniques: an application to the Nevada play fairway analysis. In *Proceedings of the 46th Workshop on Geothermal Reservoir Engineering*. Stanford University of California; 2021:13. https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2021/Smith1.pdf

Smith CM, Faulds JE, Brown S, et al. Exploratory analysis of machine learning techniques in the Nevada geothermal play fairway analysis. *Geothermics*. 2023;111:102693. doi:10.1016/j.geothermics.2023.102693

Spichak VV. Estimating temperature distributions in geothermal areas using a neuronet approach. *Geothermics*. 2006;35:181-197. doi:10.1016/j.geothermics.2006.01.002

Spichak VV, Geiermann J, Zakharova OK, Calcagno P, Genter A, Schill E. Deep temperature extrapolation in the Soultz-sous-Forêts geothermal area using magnetotelluric data. In *Expanded Abstracts, 35th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2010:9. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2010/spichak.pdf

Spichak VV, Zakharova OK. The application of an indirect electromagnetic geothermometer to temperature extrapolation in depth. *Geophys Prospect*. 2009a;57:653-664. doi:10.1111/j.1365-2478.2008.00778.x

Spichak VV, Zakharova OK. Electromagnetic temperature extrapolation in depth in the Hengill geothermal area, Iceland. In *Expanded Abstracts, 34th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2009b.

Spichak VV, Zakharova OK, Rybin AK. Methodology of the indirect temperature estimation basing on magnetotelluric data: Northern Tien shan case study. *J Appl Geophys*. 2011;73:164-173.

Sutarmin S, Daud Y. Subsurface temperature prediction in geothermal field with neural network using 3D MT data inversion and borehole temperature data. In *AIP Conference of Proceedings*. AIP Publishing; 2021:2320. doi:10.1063/5.0039616

The World Bank. International bank for reconstruction and development. Project appraisal document. Report no: pad2932. Energy & Extractives Global Practice, East Asia and Pacific Region. Accessed March 10, 2023. https://documents1.worldbank.org/curated/pt/972941569608786496/pdf/Indonesia-Geothermal-Resource-Risk-Mitigation-Project.pdf

Vesselinov VV, Ahmmed B, Frash L, Mudunuru MK. GeoThermalCloud: machine learning for discovery, exploration, and development of hidden geothermal resources. In *Proceedings, 47th Workshop on Geothermal Reservoir Engineering*. 2022; 2022. Stanford University.

Vesselinov VV, Ahmmed B, Mudunuru MM, Karra S, Middleton RS. Hidden Geothermal Signatures of the Southwest New Mexico. In *World Geothermal Congress, Los Alamos National Laboratory*; 2020. https://www.osti.gov/servlets/purl/1673337

Wang L, Yu Z, Zhang Y, Yao P. Review of machine learning methods applied to enhanced geothermal systems. *Environ Earth Sci*. 2023;82:69. doi:10.1007/s12665-023-10749-x

Xiong Y, Zhu M, Li Y, Huang K, Chen Y, Liao J. Recognition of geothermal surface manifestations: a comparison of machine learning and deep learning. *Energies*. 2022;15:2913. doi:10.3390/en15082913

Yang F, Zhu R, Zhou X, et al. Artificial neural network based prediction of reservoir temperature: a case study of Lindian geothermal field, Songliao Basin, NE China. *Geothermics*. 2022;106:102547. doi:10.1016/j.geothermics.2022.102547

Zheng Y, Li J, Lin R, Hu H, Gao K, Huang L. Physics-guided machine learning approach to characterizing small-scale fractures in geothermal fields. In *Proceedings, 46th Workshop on Geothermal Reservoir Engineering*. Stanford University; 2021:9. https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2021/Zheng.pdf

AUTHOR BIOGRAPHIES

Mahmoud AlGaiar is a senior well engineer at Baker Hughes in the Kingdom of Saudi Arabia. He has over 19 years of experience working in the Oil & Gas industry. He specializes in the design and execution of Oil & Gas wells and has worked in North and West Africa, South America, and Asia. His skills include well engineering, conventional and directional drilling, coring, downhole tools, turnkey project management, and business development. Mahmoud holds a BSc and PGDip in Mechanical Engineering, an MBA in Finance and Budgeting, and an MSc in Drilling and Well Engineering. He is currently a PhD student at Robert Gordon University, Aberdeen, UK, specializing in geothermal resource exploration and evaluation.

Nadimul Faisal (PhD, CEng, MIMechE, MIMMM, FHEA) is a Professor of Surface Engineering & Micromechanics at Robert Gordon University. His interest includes micromechanical behaviour analysis of materials and structures, sensor based instrumented mechanical testing, and acoustic emission (AE) sensor-based condition monitoring. He has over 80 peer-reviewed journals, 1 US Patent, and 4 book chapters, and over 50 conference publications. He is member of Royal Society of Edinburgh's Young Academy of Scotland, EPSRC peer review College member, and member of UK's Metamaterials Network.

How to cite this article: AlGaiar M, Hossain M, Petrovski A, Lashin A, Faisal N. Applications of artificial intelligence in geothermal resource exploration: a review. *Deep Undergr Sci Eng*. 2024;3(3):269-285. doi:10.1002/dug2.12122