Evaluating the Efficiency of Machine Learning Algorithms in Identifying Geothermal Energy Potential Areas in Akita and Iwate Provinces, Japan

Majid Kiavarz [1], Mohammadreza Jelokhani-Niaraki[1], Avin Meysami[1], Yasaman Ghorbani[1], Najmeh Neysani Samany[1]
[1] Department of Remote Sensing and GIS, Faculty of Geography, University of Tehran, Iran - kiavarzmajid@ut.ac.ir, mrjelokhani@ut.ac.ir, avin.meysami@gmail.com, yasaman.qorbani@ut.ac.ir, nneysani@ut.ac.ir

Keywords: Geothermal Energy, Machine Learning, Random Forest, Generative Adversarial Networks (GANs), Spatial Analysis, Geothermal Exploration

Abstract

The growing demand for clean and renewable energy sources has intensified the need to identify and exploit geothermal resources as a key solution for sustainable energy development. However, geothermal exploration faces significant challenges including geological complexity, high drilling costs, economic risks, and spatial data limitations. This study evaluates the efficiency of advanced machine learning algorithms, specifically Random Forest and Generative Adversarial Networks (GANs), in identifying geothermal energy potential areas in Akita and Iwate provinces, Japan. Using a limited dataset of 152 geothermal well locations, seven key parameters were analysed: volcanic activity, fault and fracture density, hot springs, surface thermal indices, fumaroles, mud volcanoes, and surface alteration evidence. Data were collected from geological and remote sensing sources and pre-processed for modelling. Results demonstrate that both algorithms effectively identify high-potential areas despite data scarcity. Random Forest achieved 94.08% accuracy in well identification with a C/S(C) index of 10.93, demonstrating robust performance and spatial correlation. The Generative Adversarial Network showed superior performance with 96.71% accuracy and a C/S(C) index of 4.36, indicating exceptional capability in identifying geothermal potential areas and detecting complex spatial patterns. These findings confirm that hybrid approaches combining machine learning and deep learning, particularly GANs, possess high capability for accurate geothermal prospectivity mapping and can effectively overcome limitations posed by data scarcity, providing valuable tools for exploration prioritization and investment decision-making.

1. Introduction

Global concerns about climate change and the depletion of fossil fuel resources have elevated the importance of clean and renewable energy sources. Geothermal energy stands out as a sustainable, clean, and reliable resource with significant potential for electricity generation and direct heating applications. Beyond its environmental compatibility and operational stability, geothermal energy plays a crucial role in reducing greenhouse gas emissions and advancing sustainable development goals (Lund and Boyd, 2016). Despite these advantages, identifying geothermal energy potential areas remains a fundamental and costly challenge. The complexity arises from the intricate interplay of geological, geophysical, and geochemical factors that control geothermal systems. Traditional exploration methods, while proven, face significant limitations: they are expensive, time-consuming, and most critically, lack the capability to simultaneously analyse the complex nonlinear relationships between multiple geological parameters. These limitations create substantial barriers to efficient geothermal resource development, particularly in regions where exploration budgets are constrained.

The integration of spatial data analysis and machine learning algorithms offers a promising solution to these challenges. Machine learning methods can process complex, high-dimensional geospatial datasets to identify subtle patterns and relationships that may not be apparent through conventional analysis. Given the nonlinear and complex nature of geological, geophysical, and geochemical data, machine learning algorithms have become essential tools for modern geothermal exploration (Lund and Boyd, 2016).

The present research addresses the critical need for efficient and accurate geothermal prospectivity mapping by evaluating advanced machine learning techniques. The main objective is to identify geothermal energy potential areas in Akita and Iwate provinces, Japan, using two sophisticated algorithms: Random Forest and Generative Adversarial Networks (GANs). Random Forest was selected for its proven capability in handling complex spatial data, managing noise, and providing interpretable results through ensemble learning. GANs were chosen for their unique ability to learn complex patterns and generate realistic predictions even with limited training data, a critical advantage given the scarcity of geothermal well data in many regions. The study area in northern Japan was selected due to its active geothermal systems and well-documented geological characteristics, making it an ideal testbed for algorithm evaluation. Seven key surface evidence parameters were utilized: faults and fractures, hot springs, mud volcanoes, volcano density, thermal indices, fumaroles, and alteration zones. These parameters represent critical indicators of subsurface geothermal activity. Using data from 152 existing geothermal wells across the two provinces, this research evaluates and compares the performance of Random Forest and GANs in geothermal prospectivity mapping under conditions of limited data availability.

2. Related Works (Background)

In recent decades, the application of machine learning and advanced computational methods to geothermal exploration has evolved significantly over the past decade. This section reviews key studies chronologically to illustrate the progression of methodologies and the increasing sophistication of approaches used in geothermal prospectivity mapping.

(Kiavarz et al., 2014) used several spatial analysis methods including Fry analysis, well presence mapping, distance distribution, Weights of Evidence, and Evidential Belief Function to study geothermal resources in northern Japan (Akita and Iwate). They found geothermal energy is closely linked to volcanoes, faults, and hot springs. Geochemical and geophysical data provided stronger evidence than geological data, and the

Weights of Evidence model was supported by expert opinion. The results help improve geothermal potential mapping in similar regions. Building on this work, (Kiavarz et al., 2017) explored geothermal area identification as the first step in resource use. They introduced the concept of risk through the Ordered Weighted Averaging method in GIS based analysis to create geothermal prospectivity maps. Their model produced maps reflecting different pessimistic and optimistic strategies. The well percentage in highly favorable areas reached 85 percent under the most pessimistic strategy and 100 percent under the most optimistic. The prediction rates were 18.55 and 1.18 for the most pessimistic and most optimistic strategies, respectively. (Tut Haklidir et al., 2020) introduced deep learning for geothermal reservoir characterization by developing a model that predicts reservoir temperature using hydrogeochemical data. They compared it with linear regression and a linear Support Vector Machine. Their results showed that deep learning captures complex nonlinear patterns in the data and gives better temperature prediction performance.

(Faulds et al., 2020) used machine learning methods, including ensemble boosting and artificial neural networks, to support the Nevada Geothermal Play Fairway Analysis project. Working with about 85 active geothermal systems and more than 12 geological, geophysical, and geochemical features, they aimed to build an algorithmic approach to identify new geothermal systems. Machine learning was first applied to the original datasets and then to enhanced and additional datasets. (Shahdi et al., 2021) used bottom hole temperature data from oil and gas wells to produce heat flow and depth temperature maps for identifying active geothermal areas. They tested several machine learning models for predicting subsurface temperature and geothermal gradients, finding that Extreme Gradient Boosting and Random Forest performed best. They also validated Gradient Boosting and Deep Neural Network models using temperature measurements from 58 wells in West Virginia. The results show that machine learning models are comparable to, and sometimes better than, physics based thermal conduction models.

(Jiang et al., 2022) focused on improving prediction of geothermal reservoir responses to different production scenarios. Alongside traditional physics based models, they introduced a recurrent neural network that also uses convolutional layers to predict energy production. A labeling scheme was added to handle real field issues such as data gaps. They also presented a full workflow for applying the model, including data preprocessing, feature selection, and training strategies for short and long term prediction.

(Kshirsagar et al., 2022) noted that although geothermal energy is clean and renewable, its assessment in development projects is often overlooked. They highlighted that using artificial intelligence and machine learning with geothermal concepts can improve the sector, especially for identifying subsurface resources and predicting subsurface temperature and geothermal gradients. (Moraga et al., 2022) introduced a method that combines remote sensing, machine learning, and artificial intelligence to assess geothermal potential using indicators such as mineral signatures and surface temperature. They applied it to the Brady and Desert Peak areas, which differ in surface expression. Machine learning was used to identify patterns in surface manifestations, and an artificial intelligence model then predicted geothermal potential for each pixel. (Shah et al., 2022) attempted to predict surface temperature, as geothermal wells are highly dependent on surface temperature. The algorithms implemented for this article were Support Vector Machines (SVM) and Random Forest algorithms, which also have better accuracy compared to other regression models. (Mogaji et al., 2023) evaluated a GIS based machine learning approach for analysing remote sensing and geophysical data. Remote sensing datasets were processed to extract key surface factors such as land use, vegetation index, lineament density, land surface temperature, and slope. The GIS spatial analysis module produced thematic maps of geothermal conditioning factors. Using MATLAB, optimal weights for these factors were derived with Adaptive Neuro Fuzzy Inference System models and metaheuristic optimization methods including Genetic Algorithm, Invasive Weed Optimization, and Particle Swarm Optimization. The results were used to generate geothermal potential predictive index maps for the study area. (Mudunuru et al., 2023) showed that the Tularosa Basin in New Mexico has strong geothermal potential. They identified probable geothermal sites using an unsupervised machine learning method based on non negative matrix factorization with k means clustering, implemented in the open source GeoThermal Cloud framework. The method revealed likely geothermal locations and key controlling parameters, and the results matched findings from previous studies of the basin.

(Al-Fakih et al., 2024) provided a comprehensive review of machine learning and deep learning applications in geothermal resource development from 2000 to 2024, highlighting the exponential growth in AI applications for seismic interpretation, reservoir characterization, and thermal energy prediction. Their work emphasized the use of advanced techniques such as artificial neural networks (ANNs), convolutional neural networks (CNNs), and support vector machines (SVMs) in improving geothermal exploration efficiency.
Most recently, (Koray et al., 2025) applied machine learning techniques including clustering algorithms and support vector machines to calculate field permeability in the Williston Basin for optimizing geothermal reservoir quality. (Al-Fakih et al., 2025) used Bayesian-optimized machine learning models including XGBoost and random forest for forecasting geothermal temperatures in western Yemen's complex tectonic settings.

3. Methodology

This section presents the methodological framework employed to identify and map geothermal energy potential areas. The approach integrates spatial data processing, feature extraction, and advanced machine learning and deep learning algorithms to generate geothermal prospectivity maps. The workflow encompasses four main phases: (1) data collection, (2) spatial feature extraction and data preprocessing, (3) model training using Random Forest and GANs, and (4) prospectivity mapping and accuracy assessment. Figure 1 illustrates the complete workflow of this ML/DL-based framework, showing how geospatial data flows through preprocessing, model training, and validation stages to produce the final geothermal potential maps.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-699-2026 |

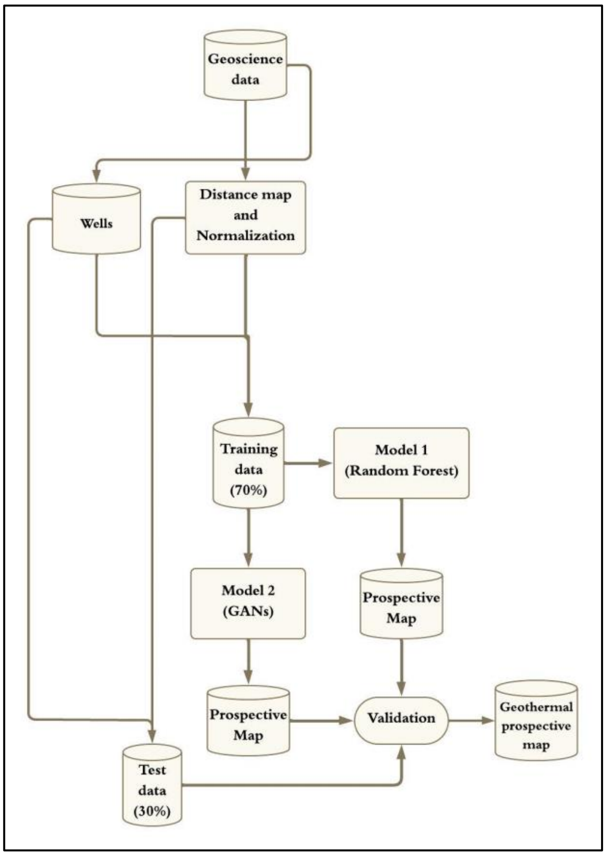

Figure 1. Overall workflow of the proposed AI-based framework.

3.1 Case Study

The study area of this research includes Akita and Iwate provinces in Japan (Figure 2), which are recognized as geothermal energy potential areas due to geothermal activities and specific geological characteristics. This region has faults, active and semi-active volcanic craters, hot springs, and extensive alteration zones, all of which are important indicators for identifying geothermal potential. Additionally, there are 152 drilled geothermal wells in the region that will be used for training and testing the algorithms. Spatial indices including distance from faults, volcanic activity density, alteration intensity, and hydrothermal indicators were calculated using Geographic Information System (GIS) tools to be used as input variables for the models. Due to scale differences in various data, all parameters were normalized and converted to a specific range.

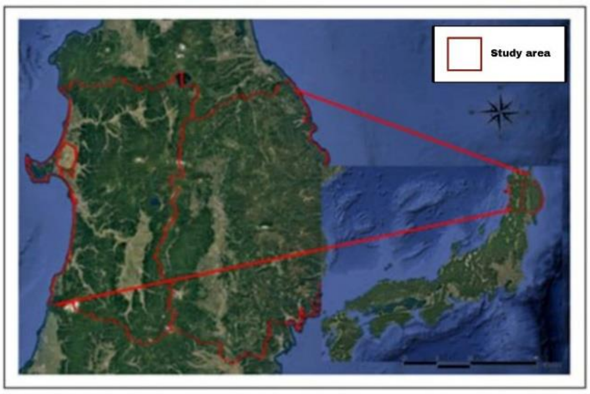

Figure 2. Study area map showing Akita and Iwate provinces

3.2. Data Preparation and Model Training

After data collection, pre-processing operations including normalization and preparation of layers in a common format were performed. All layers were prepared in raster format with uniform spatial resolution and were normalized using Min-Max normalization method to ensure all features have comparable scales for machine learning algorithms.
For spatial analysis, distance maps were prepared from all geological and geochemical layers including faults, volcanic craters, hot springs, fumaroles, mud volcanoes, and alteration zones. These distance maps quantified the proximity of each location to these key geothermal indicators, providing crucial spatial information for the machine learning models. The thermal indices layers (heat flux data) were used directly without distance transformation as they already represented continuous spatial distributions of thermal properties (Figure 3).

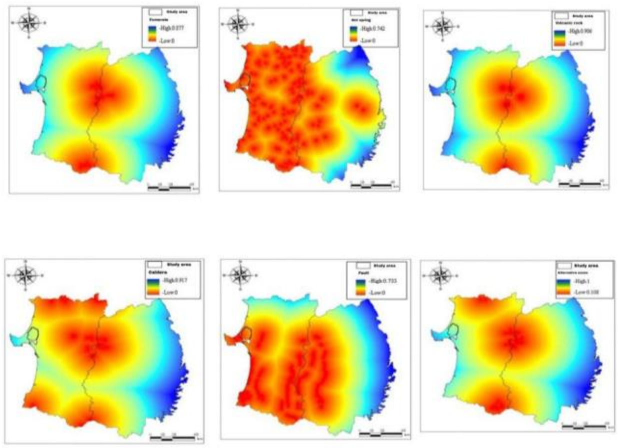

Figure 3. Distance map

For training the models and evaluating their performance, point data were divided into two sets: 70% of the data were used for training the models so that the algorithms could learn the relationships between criteria and geothermal potential, and the remaining 30% were used as test data to evaluate model performance on unknown data. This division method, in addition to preventing overfitting, enables examination of the models' ability to predict new data and helps the researcher better evaluate the actual performance and accuracy of the algorithms.

The Random Forest method, which is a combination of multiple decision trees, was employed to increase prediction accuracy and reduce overfitting. This method, in addition to managing noisy data, has high capability in working with complex and spatial data and typically provides better results in generating high-accuracy potential maps.

Additionally, Generative Adversarial Networks (GANs) were employed as a deep learning method. By utilizing two competing networks (generator and discriminator), they can extract hidden and complex patterns from data and even generate new and meaningful samples with limited data. This capability provides more accurate and realistic identification of geothermal energy potential areas compared to classical methods (Fig 3.). After training the models, they were applied to the entire region considering seven input criteria to generate geothermal energy potential maps. Then, each map was classified into three classes: high potential areas (grade above 0.7), medium potential (grade between 0.5 and 0.7), and low potential (grade below 0.5).

4. Experiments and Results

To evaluate the spatial correlation between model predictions and actual geothermal well locations, we employed the C (Contrast Index) and C/S(C) indices, which are specifically designed for spatial validation of prospectivity models. The Contrast Index (C) quantifies the concentration of known geothermal wells within areas predicted to have high potential, measuring how well the model identifies key geothermal locations. A higher C value indicates that actual wells are concentrated in high-potential areas predicted by the model. The C/S(C) index is a complementary metric that normalizes the Contrast Index by its standard deviation, measuring the statistical significance and reliability of the spatial correlation between actual wells and predicted potential areas. This index helps determine whether the concentration of wells in high-potential areas is statistically significant rather than occurring by chance. The spatial distribution of predicted geothermal potential for both models is presented in Figures 4 and 5.

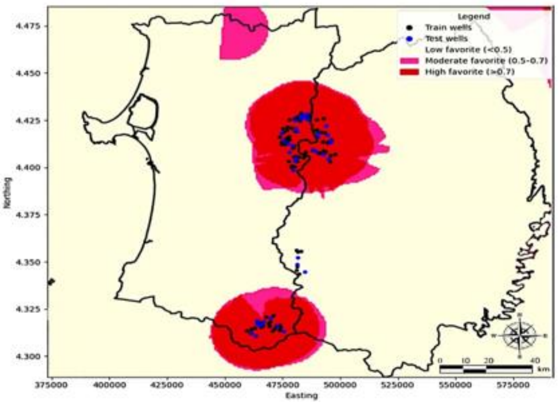

Figure 4. Geothermal potential map generated by Random Forest algorithm. The map displays three classes of geothermal potential: high (red), moderate (pink), and low (white).

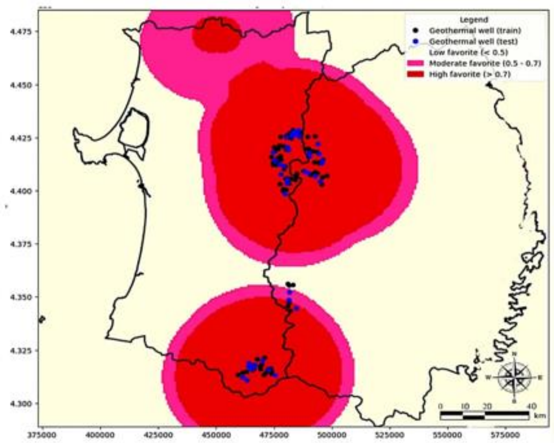

Figure 5. Geothermal potential map generated by Generative Adversarial Network (GANs). The classification scheme matches. The map displays three classes of geothermal potential: high (red), moderate (pink), and low (white).

Results showed that the Random Forest algorithm achieved a C/S(C) value of 10.93, successfully identifying 94.08% of geothermal wells in the high potential class. The Generative Adversarial Network demonstrated superior performance with a C/S(C) value of 4.36, placing 96.71% of wells in the high potential class. Both algorithms achieved C/S(C) values indicating statistically significant spatial correlations (both identifying over 90% of wells in high potential areas), confirming their effectiveness as reliable models for geothermal prospectivity mapping.

The comparison reveals an important distinction between the two models: while GANs achieved higher well prediction accuracy (96.71%), the Random Forest model demonstrated a higher C/S(C) ratio (10.93 vs 4.36). This suggests that Random Forest concentrated the wells in a more compact high-potential area, while GANs distributed high-potential predictions across a broader spatial extent while maintaining high accuracy. In practical terms, the higher C/S(C) ratio of Random Forest indicates that it can identify geothermal potential in more focused, smaller areas, which may be advantageous for targeted exploration efforts and cost-effective drilling campaigns. Conversely, GANs' broader prediction pattern, enabled by their ability to capture complex nonlinear patterns through generative modeling, may be beneficial for comprehensive regional assessments. The application of Generative AI through GANs demonstrates how advanced deep learning architectures can analyze geothermal data more comprehensively by learning the underlying data distribution and generating synthetic samples that capture complex geological relationships.

5. Discussion

To assess the spatial correlation between model predictions and actual geothermal well data, the C and C/S(C) indices were used. Results showed that the Random Forest algorithm was able to identify 94.08% of wells in the high potential class with a C/S(C) index value of 10.93, indicating reliable correlation and high concentration of wells in high potential areas. However, in comparing the two models, the GANs model demonstrated better performance and was able to place 96.71% of wells in the high potential class, with a C/S(C) index value of 4.36, which indicates the very high capability of this model in accurate and reliable identification of geothermal energy potential areas.

The accuracy assessment through C/S(C) indices revealed that both models achieved reliable performance levels exceeding 90% well prediction in high potential areas, confirming their suitability for geothermal prospectivity mapping. The Random Forest model's accuracy of 94.08% with high C/S(C) ratio demonstrates its strength in stable, interpretable predictions that concentrate potential areas effectively. The GANs model's 96.71% accuracy represents the highest prediction rate achieved, showcasing the power of deep learning approaches in capturing complex geological patterns. However, these high accuracy rates must be interpreted within the context of the spatial prediction task, where both models successfully identified the majority of known wells while minimizing false positives in low-potential areas.

Results showed that each algorithm has its own specific characteristics and limitations in identifying geothermal energy potential areas. Random Forest was able to provide a realistic and balanced pattern of potential areas by combining input criteria. The high correlation of output with faults and alteration zones demonstrates its power in identifying key geological structures. The main strength of Random Forest is stability and balance between accuracy and extent of high potential areas, but its weakness is a slight reduction in interpretability compared to simpler models. The GANs algorithm covered broader ranges of high potential and was able to place the highest percentage of actual wells in the high class. The main capability of this model is in identifying complex and potentially unknown patterns.

However, its weaknesses included instability during training and reduced generalization on test data.

The primary limitation affecting both models was the restricted dataset size of 152 geothermal well locations. This data scarcity constrains the models' ability to learn the full spectrum of geothermal indicators and reduces confidence in predictions for areas geologically different from the training data. With limited data, complex models like GANs face increased risk of overfitting, potentially memorizing training patterns rather than learning generalizable relationships. The spatial distribution of training wells also introduces bias, as the models learn predominantly from explored regions and may underperform in geologically distinct unexplored areas. Despite sophisticated algorithms, the fundamental challenge remains that machine learning models cannot extrapolate far beyond the characteristics of their training data, making comprehensive data collection essential for reliable geothermal exploration.

6. Conclusion and Future Work

The present research was conducted with the aim of identifying and zoning geothermal energy potential areas in Akita and Iwate provinces, Japan. For this purpose, seven key criteria were used including faults, volcanic craters, volcanic rocks, fumaroles, hot springs, alteration zones, and heat flux. The collected data, after preprocessing and normalization, were used as input in two machine learning algorithms: Random Forest and Generative Adversarial Networks. The main objective of this modeling was to compare the algorithms' ability to predict areas with high potential and analyze the strengths and weaknesses of each under conditions of limited data from 152 actual well points.

Analysis using the C (Contrast Index) and C/S(C) indices provided robust spatial validation of both models. The C index quantifies how effectively a model concentrates known geothermal wells in predicted high-potential areas, serving as a measure of spatial correlation strength. The C/S(C) index further assesses the statistical significance of this correlation by normalizing the contrast by its standard deviation, indicating whether the observed well clustering in high-potential zones is meaningful or merely coincidental. Random Forest achieved a C/S(C) of 10.93 with 94.08% well prediction accuracy, indicating highly focused identification of geothermal areas. GANs achieved a C/S(C) of 4.36 with 96.71% accuracy, demonstrating superior well prediction but across a broader spatial extent (See Table 1). Both models' C/S(C) values well above threshold levels confirm statistically significant spatial correlations, validating their reliability for geothermal prospectivity assessment.
The application of Generative AI through GANs represents a significant advancement in geothermal exploration methodologies. Unlike traditional discriminative models that learn to classify or predict based on existing patterns, generative models learn the underlying probability distribution of geothermal data, enabling them to capture subtle and complex relationships between geological features. The generator-discriminator architecture of GANs creates a competitive learning environment where the generator learns to produce realistic geothermal potential distributions while the discriminator learns to distinguish between generated and real patterns. This adversarial training process enables GANs to identify non-obvious geothermal signatures and explore broader areas of potential, explaining their superior well prediction accuracy of 96.71%. The success of GANs in this research demonstrates that generative approaches can complement traditional machine learning methods, particularly in scenarios with complex, nonlinear geological relationships characteristic of geothermal systems.

Model	Well Prediction Rate (%)	C/S(C)	Performance Interpretation
Random Forest	94.08	10.93	High spatial concentration;
GAN	96.71	4.36	Broader spatial coverage; highest well identification rate

Table 1. Comparative model characteristics

The comparative analysis highlights distinct performance characteristics of the two models (See Table 2). The Random Forest algorithm demonstrates stable and robust performance, generating more spatially compact high-potential zones with strong correlation to key structural geothermal indicators such as faults and alteration zones. This spatial concentration behavior makes it particularly suitable for targeted exploration and cost-efficient drilling campaigns. However, its well detection rate is slightly lower than that of the GAN model. In contrast, the Generative Adversarial Network (GAN) exhibits superior capability in capturing complex nonlinear spatial relationships, achieving the highest well prediction rate among the tested models. Nevertheless, GAN predictions tend to cover broader spatial extents, which may increase exploration uncertainty, and the model may be more susceptible to overfitting under limited data conditions. These differences suggest that Random Forest is advantageous for focused exploration strategies, whereas GAN is better suited for regional-scale reconnaissance assessments.

Model	Strengths	Limitations
Random Forest	Stable performance; compact prediction zones; strong correlation with structural features	Slightly lower well detection rate compared to GAN
GAN	Captures complex nonlinear patterns; highest well prediction rate	Broader prediction extent; potential overfitting under limited data

Table 2. Strengths and limitations of models

Based on the results obtained from this research and existing limitations, it is suggested that in future studies and practical applications, several measures be considered. First, increasing field and actual data is important, as the limited number of geothermal wells (152 wells) has reduced the models' generalization capability and increased the probability of overfitting in complex algorithms such as Generative Adversarial Networks.

Using hybrid methods combining machine learning algorithms and classical multi-criteria decision-making analysis methods, such as Fuzzy logic or Ordered Weighted Averaging (OWA), can integrate the best features of each method, leading to error reduction in boundary areas, improved prediction accuracy, and provision of more reliable potential maps.

A critical future direction for advancing geothermal resource assessment lies in improving the interpretability and transparency of the machine learning models utilized. While the current study successfully demonstrated the efficiency of spatial analysis in prospectivity mapping, a subsequent and highly valuable step is the application of Explainable Artificial Intelligence (XAI) techniques. XAI encompasses methods that reveal the internal decision-making processes of a machine learning model, thereby facilitating sensitivity analysis on the input parameters. By implementing XAI it is possible to quantitatively determine the contribution of each geological criterion to the model's predictive output. This kind of analysis is paramount as it directly quantifies the importance of a criterion based on its effect on the final prediction of geothermal potential. This rigorous feature importance assessment will not only validate the meaningful geological relationships learned by the models but will also provide a scientifically robust basis for prioritizing exploration efforts based on the criteria that exhibit the strongest predictive power. This research demonstrates that spatial analysis using machine learning enables efficient exploration of extensive geographic areas while significantly reducing costs compared to traditional field-intensive methods. By integrating multiple geological and geophysical data layers through automated machine learning algorithms, we can systematically evaluate geothermal potential across entire provinces, a task that would require years of fieldwork and substantial financial investment if conducted through conventional exploration approaches. Machine learning models process and analyse spatial relationships at scale, identifying prospective areas that can be prioritized for detailed ground surveys and drilling programs. This targeted approach minimizes unnecessary exploration expenditure in low-potential areas while focusing resources on locations with highest probability of success. Furthermore, the spatial analysis framework developed in this study can be readily adapted and applied to other geographically similar regions, enabling cost-effective reconnaissance-level assessments that guide subsequent detailed exploration campaigns.

Despite research limitations, particularly limited data, the results of this study show that machine learning algorithms can be a solid foundation for future studies and method optimization. Collecting more data, using advanced deep learning models, and generalizing methods to other regions can significantly increase prediction accuracy, generalization capability, and practical application of results. Finally, this research emphasizes that employing intelligent and flexible algorithms for identifying geothermal energy potential areas, in addition to improving scientific accuracy, can assist in management decision-making, investment planning, and sustainable energy resource development, and transform the proposed models into practical tools for designing geothermal energy development policies and identifying exploration priorities.

References

Al-Fakih, A., Al-Khudafi, A., Koeshidayatullah, A., Kaka, S., & Al-Gathe, A., 2025: Forecasting geothermal temperature in western Yemen with Bayesian-optimized machine learning regression models. *Geothermal Energy*, 13(1).

Al-Fakih, A., Abdulraheem, A., & Kaka, S., 2024: Application of machine learning and deep learning in geothermal resource development: Trends and perspectives. *Deep Underground Science and Engineering*, 3(3), 286-301.

Faulds, J. E., et al., 2020: Preliminary report on applications of machine learning techniques to the Nevada geothermal play fairway analysis. In the 45th workshop on geothermal reservoir engineering.

Jiang, A., et al., 2022: Recurrent neural networks for short-term and long-term prediction of geothermal reservoirs. *Geothermics*, 104, 102439.

Kiavarz, M., & Jelokhani-Niaraki, M., 2014: Geothermal prospectivity mapping using spatial analysis methods in Akita and Iwate provinces, Japan. *Geothermics*, 70, 295-304

Kiavarz, M., & Jelokhani-Niaraki, M., 2017: Geothermal prospectivity mapping using GIS-based Ordered Weighted Averaging approach: A case study in Japan's Akita and Iwate provinces. *Geothermics, 70*, 295-304.

Koray, I., et al., 2025: Leveraging machine learning for enhanced reservoir permeability estimation in geothermal hotspots: a case study of the Williston Basin. *Geothermal Energy*, 13(1).

Kshirsagar, A., & Sanghavi, P., 2022: Geothermal, oil, and gas well subsurface temperature prediction employing machine learning. In the 47th workshop on geothermal reservoir engineering.

Lund, J. W., & Boyd, T. L., 2016: Geothermal direct utilization: 2015 worldwide review. *Geothermics*, 60, 66–93.

Mogaji, K. A., Adeniyi, O. T., & Komolafe, A. A., 2023: Geothermal energy potentiality modeling using GIS-based machine learning algorithm concept in Southwestern Nigeria: Insights from geophysical and remote sensing data sets.

Moraga, J., et al., 2022: The geothermal artificial intelligence for geothermal exploration. *Renewable Energy*, 192, 134-149.

Mudunuru, M. K., et al., 2023: Machine learning for geothermal resource exploration in the Tularosa Basin, New Mexico. *Energies*, 16(7), 3098.

Shah, K., Mehta, A., & Shah, M., 2022: Prediction of geothermal wells location and hot water temperature using machine learning. *Petroleum Research*.

Shahdi, A., et al., 2021: Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of the Northeastern United States. *Geothermal Energy, 9*, 1-22.

Tut Haklidir, F. S., & Haklidir, M., 2020: Prediction of reservoir temperatures using hydrogeochemical data, Western Anatolia geothermal systems (Turkey): a machine learning approach. *Natural Resources Research*, 29(4), 2333-2346.