PROCEEDINGS, 50th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 10-12, 2025
SGP-TR-229

Enhancing Reservoir Temperature Prediction in Western Anatolia Geothermal Systems by Generating Synthetic Hydrogeochemical Data Using Generative Adversarial Networks

Füsun S. Tut Haklidir[1], Mehmet Haklidir[2]

[1]Zorlu Enerji Company, Istanbul, Turkiye, [2]TUBITAK BILGEM, Kocaeli, Türkiye

fusun.haklidir@zorlu.com, mehmet.haklidir@tubitak.gov.tr

Keywords: Synthetic Data Generation, Generative Adversarial Networks, Reservoir Temperature Prediction

ABSTRACT

Prior to geothermal investments, it is critical to accurately determine the potential of the geothermal resource and reservoir temperatures. For financial reasons, it may often be necessary to make these evaluations with limited data before the investment starts, and the estimation of the reservoir temperature, which is one of the most basic criteria, is very important at this stage. While hydrogeochemical data are used in geothermometer calculations, there may be significant differences between calculations due to assumptions. With the expansion of the usage areas of artificial intelligence algorithms, these algorithms have started to find a place in the estimation of reservoir temperatures in geothermal. For financial reasons, it may often be necessary to make these evaluations with limited data before the investment starts, and the estimation of the reservoir temperature, which is one of the most basic criteria, is very important at this stage. In our previous work, we employed machine learning models, including deep neural networks (DNNs), to predict reservoir temperatures based on hydrogeochemical data from Western Anatolia's geothermal systems. While the DNN model outperformed traditional regression approaches, the limited size and variability of the dataset constrained the model's predictive capabilities. In this study, we integrate Generative Adversarial Networks (GANs) to generate synthetic hydrogeochemical data, effectively augmenting the existing dataset. By enhancing the training data with realistic synthetic samples, we aim to improve the performance and generalization of machine learning models for reservoir temperature prediction by this new approach. Our results demonstrate that the GAN-augmented models achieve higher accuracy and lower error rates compared to models trained on the original dataset, with the best-performing model achieving an RMSE of 6.44 and an MAE of 4.72 using an augmented dataset with a 3:1 synthetic-to-real data ratio, offering a novel approach to address data scarcity in geothermal exploration.

1. INTRODUCTION

Accurate prediction of geothermal reservoir temperatures is essential for optimizing geothermal exploration and development. Reservoir temperature directly influences the feasibility of different geothermal energy applications, from electricity generation to district heating and industrial processes. Considering the criticality of near-realistic temperature predictions in geothermal exploration studies, it can support the reduction of risk and costs in drilling operations (Tut Haklidir and Haklidir 2020, Ibrahim et al., 2023, Shi et al., 2025).

Conventional geothermometers, which estimate reservoir temperatures based on the chemical and isotopic composition of geothermal fluids, have been the primary tool for geothermal exploration. However, their accuracy can often show significant deviations from actual values due to fluid mixing, pressure variations or different fluid interferences, as well as equilibrium assumptions that may not be universally valid. These limitations emphasize the need for improved prediction models that account for the complexities and non-linearities inherent in geothermal systems.

The Western Anatolia (Türkiye), a region of significant geothermal potential, features high-temperature reservoirs along the Büyük Menderes and Gediz Grabens (Tut Haklidir and Şengün, 2020). These areas, shaped by extensional tectonics, show high heat flow and thin crusts, making them ideal for geothermal energy production. Geothermal reservoir temperatures often exceed 200°C, supporting advanced geothermal power cycles and diverse energy applications in this region. Since 1984, 66 geothermal power plants with different technologies and with a total installed capacity of 1,734 MWe have been established in the region, and most of this power is supplied along these two large grabens. There are also a large number of low and medium temperature sources and 1338 MWt of residential heating is provided in the region (Figure 1).

In recent years, machine learning (ML) approaches have shown promise in addressing some of the challenges associated with geothermal exploration. For instance, studies by Rezvanbehbahani et al. (2017) used ML models to predict geothermal heat flux in Greenland, while Holtzman et al. (2018) applied ML to analyze seismic source spectra in the Geysers geothermal field. Closer to geothermal resource prediction, our 2020 study utilized ML models, including linear regression, support vector machines (SVM), and deep neural networks (DNN), to predict reservoir temperatures based on hydrogeochemical data from Western Anatolia’s geothermal systems. Among these models, the DNN provided the most accurate predictions, outperforming traditional regression approaches. However, the limited size and variability of the dataset constrained the model’s generalization capabilities, emphasizing the need for more comprehensive datasets. Subsequently, Ibrahim et al. 2023 showed that the model can be further improved with natural gradient boosting (NGB) using the data set in this publication. This estimation has also created a perspective that reservoir temperature estimation can be improved by using different algorithms.

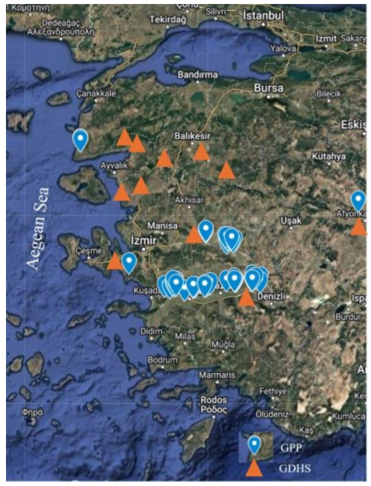

Figure 1: Geothermal power plants and district heating systems in the Western Anatolia (Türkiye)

Data scarcity is a pervasive issue in geothermal exploration, primarily due to the high costs and logistical challenges of acquiring comprehensive datasets. Several studies have addressed this challenge by proposing strategies to expand and enhance geothermal datasets. For example, Kelly and Mares. (1993) emphasized the importance of combining geophysical surveys with hydrogeological sampling to improve dataset comprehensiveness. In more recent work, Cladouhos et al. (2018) demonstrated how integrated datasets from diverse sources could mitigate data scarcity and enhance predictive modeling. These efforts highlight the critical role of innovative approaches in overcoming the limitations of sparse and incomplete datasets, enabling more accurate geothermal resource assessments.

Generative Adversarial Networks (GANs) offer a transformative solution to the challenge of data scarcity. GANs consist of two neural networks—a generator and a discriminator—that work in tandem to create highly realistic synthetic data. The generator produces data samples, while the discriminator evaluates their authenticity, iteratively refining the generator's outputs. This capability allows GANs to model complex data distributions and generate synthetic datasets that closely resemble real-world data.

GANs have been successfully employed in various resource exploration case studies to improve predictive modeling. For instance, Stabler et al. (2024) utilized GANs to generate synthetic seismic data, enhancing the accuracy of subsurface imaging in oil and gas exploration. Similarly, Ferreira et al. (2022) utilized the StyleGAN2-based PetroGAN model to generate synthetic petrographic datasets, improving the accuracy of mineral exploration and resource estimation models. These examples underscore the versatility of GANs in addressing data limitations across different domains, paving the way for their application in geothermal research.

In geothermal applications, Generative Adversarial Networks (GANs) address data scarcity by augmenting limited datasets with synthetic samples, enabling machine learning models to achieve better performance and generalization. This study focuses on integrating GANs to generate synthetic hydrogeochemical data tailored to geothermal systems in Western Anatolia, aiming to enhance reservoir temperature prediction accuracy. By addressing the limitations of small datasets, the study seeks to improve machine learning models' predictive capabilities and validate the enhanced models by comparing their predictions with actual reservoir temperatures. Through this approach, it highlights the potential of GAN-augmented datasets in advancing geothermal exploration and improving reservoir temperature estimation, paving the way for more accurate and efficient geothermal resource assessments.

2. METHODOLOGY

2.1 Study Area and Data Collection

The Western Anatolia is mainly renowned for its extensive geothermal resources, primarily concentrated along the Büyük Menderes and Gediz Grabens. These grabens, formed by extensional tectonics, exhibit high heat flow, making them some of the most promising geothermal areas such as; Kızıldere, Germencik, Alaşehir, Salihli geothermal fields in Türkiye. The geothermal fields in this region are characterized by a variety of manifestations, including thermal springs, fumaroles, and wells, with reservoir temperatures often exceeding 200°C. The diversity of hydrogeochemical conditions in these fields provides an excellent basis for understanding geothermal systems.

The amount of dissolved ions in geothermal fluids in general depends on the outcome of the reservoir temperature and geology. Due to water-rock interaction, fluids in the deep reservoir, especially under high temperature and pressure conditions, incorporate ions from the adjacent rock as they rise towards the surface, providing important information about the environment from which they came when they reach the surface especially Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl- ions and SiO_2, boron concentrations are important hydrogeochemical parameters to reveal this deep water-rock interaction. Therefore, it is predicted that these analytes can be used to estimate the reservoir temperature.

The hydrogeochemical dataset used in this study originates from our previous Tut Haklidir and Haklidir 2020 study, comprising measurements from 83 thermal springs and geothermal wells in the region. The dataset includes critical parameters essential for reservoir characterization: temperature, pH, electrical conductivity (EC), and the concentrations of Na^+, K^+, Cl^-, SiO_2, and $B_{total.}$ of the 83 samples, 66 were utilized for training machine learning models, while 17 were reserved for testing purposes. Reservoir temperatures were either directly measured from geothermal wells or estimated using established geothermometer equations for thermal springs.

2.2 Generating Synthetic Hydrogeochemical Data Using Generative Adversarial Networks

In this study, a Conditional Tabular Generative Adversarial Network (CTGAN) framework was implemented to generate synthetic hydrogeochemical data, enhancing the training dataset for reservoir temperature prediction in Western Anatolia's geothermal systems. (Xu et al., 2019) The CTGAN architecture comprises two neural networks—a generator and a discriminator—that work in tandem to learn the distribution of hydrogeochemical features from the real dataset. Below, we describe the design, hyperparameter settings, and workflow for synthetic data generation and validation with CTGAN (Xu et al., 2019).

The Generative Adversarial Network (GAN) training workflow for generating synthetic hydrogeochemical data involves a series of well-defined steps, as illustrated in Figure 2. The figure visually represents these stages, highlighting the critical transformations and interactions between different components of the workflow. By following this structured process, we ensured the generation of high-quality synthetic data, which significantly improved the predictive capabilities of the reservoir temperature models. This process is designed to ensure a seamless and effective augmentation of the original dataset.

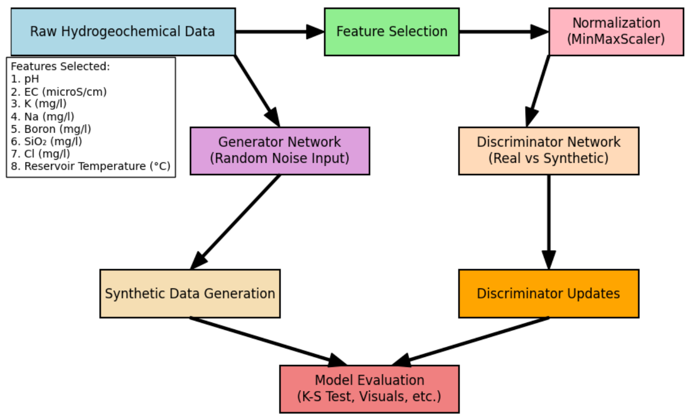

Figure 2: GAN Training Workflow

As illustrated in Figure 2, the CTGAN-based workflow for generating synthetic hydrogeochemical data involves several stages:

- Raw Hydrogeochemical Data Collection: The workflow begins with the collection of raw hydrogeochemical data from Western Anatolia's geothermal systems. This dataset contains critical features such as pH, electrical conductivity (EC), and ion concentrations (e.g., Na^+, K^+, Cl^-, boron), along with the reservoir temperatures. These features form the foundation for building a robust machine learning model.

- Feature Selection: From the raw dataset, eight key features were selected for their relevance to reservoir temperature prediction: pH, EC (microS/cm), K^+ (mg/l), Na^+ (mg/l), Boron (mg/l), SiO_2 (mg/l), Cl (mg/l), reservoir temperature (°C). These features were chosen based on domain knowledge and their statistical correlation with reservoir temperatures.

- Normalization: First, we applied a log transform (log1p) to wide-ranging features (EC, K, Na, Boron, SiO_2, Cl) to reduce the skew and stabilize training. To standardize the dataset and ensure compatibility with the neural network models, all numerical features were normalized to a range of 0 to 1 using the MinMaxScaler technique. Normalization prevents scale differences among features from affecting model performance.

- GAN Design: The architecture of the Generative Adversarial Network (GAN) utilized in this study was meticulously crafted to address the unique characteristics of hydrogeochemical data. We implemented a Conditional Tabular GAN (CTGAN) framework, selected for its proven capability in generating high-fidelity synthetic tabular data. (Xu et al., 2019) To facilitate accurate modeling, a comprehensive metadata schema was established from the log-transformed dataset, ensuring that each feature's data type and distribution were precisely recognized by the model.

The CTGAN framework consists of two primary neural networks: the generator and the discriminator. The generator network is designed to produce synthetic hydrogeochemical samples from random noise vectors, effectively learning to replicate the underlying distribution of the real data. Conversely, the discriminator network is tasked with distinguishing between genuine and synthetic data samples, thereby providing feedback to the generator. This adversarial interplay compels the generator to continuously improve its output quality, striving to produce data that the discriminator cannot easily differentiate from real samples.

Training of the GAN involves iterative optimization of both networks. (Figure 3) The generator aims to maximize the discriminator's error in identifying synthetic data, while the discriminator endeavors to minimize this error, enhancing its ability to accurately classify data samples. To ensure stability during training and prevent phenomena such as mode collapse, a gradient penalty was incorporated into the loss function. This regularization technique enforces Lipschitz continuity within the discriminator by penalizing large gradients, thereby promoting smoother convergence and more reliable training dynamics. The following hyperparameters were optimized based on experiments to achieve realistic synthetic data generation:

Table 1: Hyperparameters

Parameter	Value
Number of Samples	300 (generated)
Batch Size	64
Learning Rate	0.0001
Epochs	1000

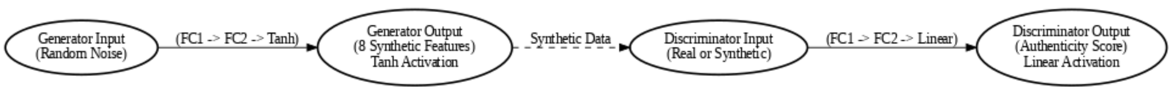

Figure 3: Architecture of the Conditional Tabular Generative Adversarial Network (CTGAN)

- Adversarial Training: The generator and discriminator are trained iteratively in an adversarial manner. The generator attempts to produce hydrogeochemical data indistinguishable from real samples, and the discriminator refines its ability to detect synthetic data. Training converges when both networks reach an equilibrium, leading to high-quality synthetic data.

- Synthetic Data Generation: Once trained, the Generator produces realistic synthetic hydrogeochemical data samples that augment the original dataset. These samples are validated to ensure they align with the statistical properties of the real data.

- Model Evaluation: The quality of synthetic data was evaluated through statistical and visual comparisons:
  - Kolmogorov-Smirnov (K-S) Test: Statistical similarity between real and synthetic feature distributions was assessed. Features such as pH, EC, and Boron showed strong alignment between real and synthetic data.
  - Visual Comparisons: Histograms and kernel density estimates (KDEs) were plotted for each feature, demonstrating close overlaps between real and synthetic distributions.
  - Model Performance: The GAN-augmented dataset improved the predictive performance of machine learning models for reservoir temperature estimation. Models trained on synthetic and real data outperformed those trained on the original dataset alone.

2.3 Dataset Augmentation and Machine Learning for Reservoir Temperature Prediction

To enhance the reservoir temperature prediction capabilities, the original dataset (Tut Haklidir and Haklidir, 2020) was augmented with synthetic hydrogeochemical data generated by the CTGAN model. The augmentation process involved combining the synthetic samples with the original dataset to create hybrid datasets with varying proportions of synthetic data. These augmented datasets were then used for downstream machine learning tasks. By varying the synthetic-to-real data ratios, the impact of synthetic data on model performance was systematically evaluated.

A Deep Neural Network (DNN) model was employed to predict reservoir temperatures using the augmented datasets. The workflow involved the following steps:

- Re-training with Augmented Data: The DNN model was re-trained using the hybrid datasets that combined real and synthetic data. This approach allowed the evaluation of the synthetic data's contribution to the model's predictive accuracy.

- Hyperparameter Tuning: To optimize the DNN's performance, hyperparameters such as the number of layers, neurons per layer, learning rate, and batch size were systematically tuned. Grid search and random search techniques were applied to identify the best-performing configuration.

- Performance Comparison: The augmented datasets were compared against the model trained solely on the original dataset. Evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to assess the impact of dataset augmentation.

3. RESULTS

The results of this study, focusing on synthetic hydrogeochemical data generation for the Western Anatolia geothermal systems, demonstrate the effectiveness of Generative Adversarial Networks (GANs) in addressing data scarcity and improving reservoir temperature predictions. Below, we summarize the key findings, supported by statistical analyses, visualizations, and predictive model evaluations.

3.1 Comparison of Real and Synthetic Data

To validate the quality of the generated synthetic hydrogeochemical data, we compared it with the original hydrogeochemical dataset via descriptive statistics, Kolmogorov-Smirnov tests, and histogram overlays:

- **Kolmogorov-Smirnov (K-S) Test:** The Kolmogorov-Smirnov (K-S) test results indicate the degree of alignment between synthetic and real data distributions. A lower K-S statistic and a higher p-value suggest a better match. Below is a summary table of the results (Table 2):

Table 2: Kolmogorov-Smirnov (K-S) Test Results

Feature	K-S Statistic	P-Value	Alignment
pH	0.0847	0.7122	Strong alignment
EC (microS/cm)	0.0830	0.7352	Strong alignment
K^+ (mg/l)	0.1263	0.2369	Good alignment
Na^+ (mg/l)	0.1706	0.0424	Marginal alignment
Boron (mg/l)	0.1359	0.1704	Good alignment
SiO_2 (mg/l)	0.0865	0.6877	Strong alignment
Cl^- (mg/l)	0.1001	0.5075	Good alignment
Reservoir Temperature (°C)	0.1586	0.0711	Marginal alignment

Overall, these K-S results suggest that most features exhibit strong or good alignment with the real distributions. Features such as pH, EC, boron, and SiO_2 demonstrate strong statistical alignment, while Na and Reservoir Temperature show marginal alignment. These results indicate that the synthetic data retains key statistical characteristics of the original dataset, making it a viable augmentation for hydrogeochemical analyses.

- **Statistical Comparison:** Table 3 shows summary statistics (mean, standard deviation, min, max) for both real and synthetic datasets. Table 3 presents summary statistics (mean, standard deviation, min, max) for both real and synthetic datasets. Most features exhibit aligned ranges and central tendencies. Features such as pH, Boron, SiO_2, and Cl^- show good agreement in both descriptive statistics and K-S test results. Meanwhile, Na^+ and reservoir Temperature reveal slightly larger deviations, consistent with their marginal alignment scores in the K-S test.

Table 3: Statistical Comparison

Feature	Real Mean	Synth Mean	Real Std	Synth Std	Real Min	Synth Min	Real Max	Synth Max
pH	7.53	7.46	0.78	0.70	5.80	6.00	9.10	9.10
EC (microS/cm)	2940.32	3095.75	1731.24	1800.65	350.00	403.56	5890.00	5890.00
K (mg/l)	64.94	73.18	56.87	61.38	0.80	1.21	191.00	191.00
Na (mg/l)	622.67	694.23	503.65	524.59	2.60	7.30	1513.00	1513.00
Boron (mg/l)	10.44	9.49	10.12	7.94	0.00	0.00	38.00	32.98

SiO_2 (mg/l)	168.69	170.61	154.65	159.88	11.00	12.25	650.00	650.00
Cl (mg/l)	81.14	78.62	64.07	51.43	3.00	3.00	326.00	279.00
Reservoir Temperature (°C)	145.67	153.74	56.82	53.98	50.00	60.00	245.00	245.00

- **Visualization**: Figures 4 illustrate histogram overlays with KDE (Kernel Density Estimate) curves for each feature, comparing real data to synthetic data distributions generated using CTGAN

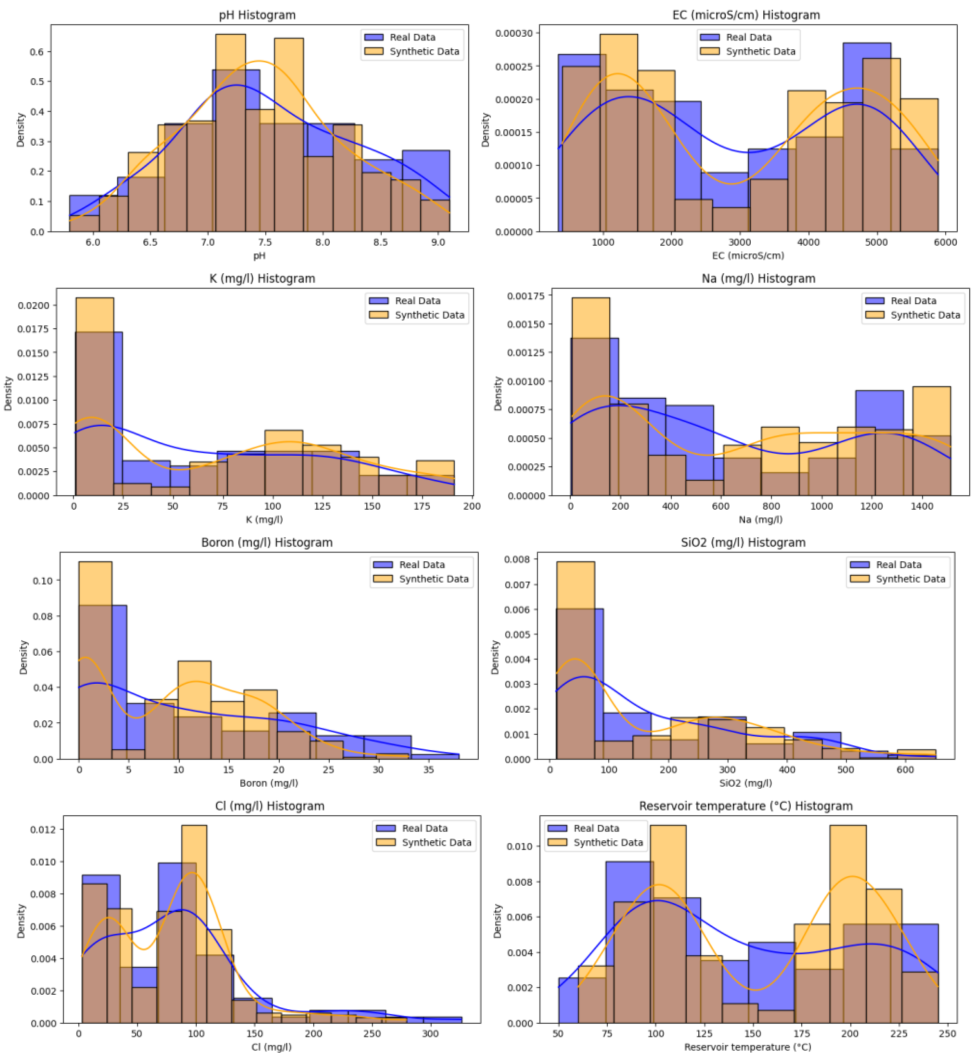

Figure 4: Visualization

The comparison reveals that CTGAN successfully captures the primary distribution characteristics of the real data for most features, as evidenced by substantial overlap in the histograms and KDE curves. This indicates that the synthetic data effectively models the main modes of the real data. However, some minor discrepancies were observed, particularly in the tails and secondary modes of certain features. These differences suggest that while the synthetic data aligns well with the real data in general, there are areas where the model could be further refined to improve accuracy.

Features such as pH and reservoir temperature show excellent alignment, with synthetic data closely mirroring the real data's central trends and capturing key distribution structures, including bimodal patterns where applicable. Other features, such as EC, K, and Na, exhibit slight overestimations or underestimations in specific ranges, particularly in higher concentration levels or mid-range densities. Additionally, features like boron, SiO2, and Cl– reveal good overall alignment but display minor deviations in density at the distribution tails.

3.2 Reservoir Temperature Prediction Evaluation

To evaluate the impact of synthetic data on reservoir temperature prediction, a Deep Neural Network (DNN) model was re-trained using datasets augmented with varying proportions of synthetic hydrogeochemical data. The analysis demonstrated significant improvements in prediction accuracy and reliability:

- Performance Improvements: Models trained on augmented datasets achieved consistently lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to models trained solely on the original dataset.
- Optimal Synthetic-to-Real Data Ratio: Among the tested ratios (1:1, 2:1, 3:1), the 3:1 ratio provided the best results, achieving an RMSE of 6.44 and an MAE of 4.72.
- Enhanced Generalization: The inclusion of synthetic data improved the model's ability to generalize, particularly in regions of the feature space that were underrepresented in the original dataset.

Table 4: Performance Metrics for DNN Models Trained on Different Datasets

Dataset	**MAE**	**RMSE**
Original Dataset (Tut Haklidir and Haklidir, 2020)	6.43	8.25
Augmented (1:1 ratio)	5.12	6.94
Augmented (2:1 ratio)	4.87	6.57
Augmented (3:1 ratio)	4.72	6.44

These results underscore the potential of synthetic data in improving the accuracy and robustness of reservoir temperature prediction models, particularly when original datasets are constrained in size or exhibit feature imbalances.

4. DISCUSSION AND CONCLUSION

This study demonstrates the effectiveness of Conditional Tabular Generative Adversarial Networks (CTGANs) in generating synthetic hydrogeochemical data and enhancing reservoir temperature prediction in geothermal systems. By integrating synthetic data with the original dataset, the research addressed challenges associated with data scarcity and feature imbalance, leading to significant improvements in machine learning model performance and generalization.

The synthetic data generated by the CTGAN closely captured the primary distribution characteristics of the original hydrogeochemical dataset. Kolmogorov-Smirnov (K-S) test results indicated strong or good alignment for most features, such as pH, EC, boron, and SiO_2, showcasing the high fidelity of the synthetic data. However, marginal alignment was observed for features like Na and reservoir temperature, particularly in the tails or secondary modes of their distributions. While the synthetic data successfully reflected central trends, these discrepancies suggest that extreme values and rare patterns could be better represented with further refinements to the model.

The integration of synthetic data into the machine learning pipeline for reservoir temperature prediction yielded significant benefits. Models trained with augmented datasets consistently outperformed those trained solely on the original dataset, achieving reductions in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Among the tested synthetic-to-real data ratios, a 3:1 ratio provided the best results, demonstrating the importance of balancing synthetic and real data for optimal performance. Moreover, the synthetic data enhanced the model's ability to generalize across the feature space, particularly in regions underrepresented in the original dataset. This improvement underscores the value of data augmentation in addressing limited data diversity, a common challenge in geothermal research.

In summary, the findings confirm that CTGAN-generated synthetic data can augment real datasets effectively, improving both model accuracy and reliability in reservoir temperature prediction. The ability of synthetic data to enhance model generalization further highlights its potential as a practical solution for data-constrained applications. Future efforts should focus on refining the CTGAN model to better capture extreme values and rare patterns, as well as exploring alternative generative models and hybrid approaches to enhance the quality of synthetic data. These advancements could lead to more robust and accurate predictions, ultimately supporting better exploration and management strategies in geothermal energy systems.

REFERENCES

Cladouhos, T.T., Petty, S., Bonneville, A., Schultz, A., and Sorlie, C.F. "Super Hot EGS and the Newberry Deep drilling Project." Proceedings: 43rd Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California (2018).

Ferreira I, Ochoa L, Koeshidayatullah A. On the generation of realistic synthetic petrographic datasets using a style-based GAN. Scientific Reports., 12(1):12845, (2022).

Holtzman, B.K., Paté, A., Paisley, J., Waldhauser, F., and Repetto, D.: Machine Learning Reveals Cyclic Changes in Seismic Source Spectra in Geysers Geothermal Field, Science Advances, 4(5), (2018), eaao2929.

Ibrahim, B., Konduah, J.O., and Ahenkorah, I.: Predicting Reservoir Temperature of Geothermal Systems in West Anatolia, Turkey: A Focus on Predictive Performance and Explainability of Machine Learning Models, SSRN Electronic Journal, (2022).

Kelly E.W., Mares, Staimislav. "Geophysical Surveys for Hydrogeological Purposes", Developments in Water Science, Elsevier, Volume 44, Pages 31-99, (1993). SSN 0167-5648.

Rezvanbehbahani, S., Stearns, L.A., Kadivar, A., Walker, J.D., and van der Veen, C.J.: Predicting the Geothermal Heat Flux in Greenland: A Machine Learning Approach, Geophysical Research Letters, 44(24), (2017), 12,271–12, 279.

Shi, Y., Robinson, J.H., and Garg, S.K.: The fluid temperature prediction with hydro-geochemical indicators using a deep learning model: A case study Western Anatolia (Turkey), Proceedings, 43rd Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2025).

Stabler, B., Zhang, H., and Lin, L.: SeisGAN: Improving Seismic Image Resolution and Reducing Random Noise Using a Generative Adversarial Network, Mathematical Geosciences, 56, (2024), 723–749.

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

Tut Haklidir, F.S., and Şengün, R.: Hydrogeochemical similarities and differences between high temperature geothermal systems with similar geologic settings in the Büyük Menderes and Gediz Grabens of Turkey, Geothermics, 83, (2020), 101717.

Xu, L., Skoularidou, M., Cuesta-Infante, A. and Veeramachaneni, K.: Modeling tabular data using conditional gan. Advances in neural information processing systems, 32, (2019).