Generative Adversarial Network-Based Approach to Scientific Time Series Analysis

Akanksha Vijayvergiya[1,†], Alsayed Algergawy[1,†]

[1]Chair of Data and Knowledge Engineering, University of Passau, Germany

Abstract

Acquiring time-series data in life sciences, such as soil, hydrological, and climatic measurements, presents significant challenges because of the susceptibility of real sensors to damage and malfunction. These issues often lead to incomplete, inconsistent, or missing data, where traditional machine learning and deep learning models struggle to handle effectively. To address this, the paper introduces a robust approach leveraging generative adversarial networks (GANs) to improve data reliability. GANs are used to generate synthetic data that fill in gaps and correct inconsistencies, resulting in a more complete and accurate dataset. The method involves training the GAN on the existing dataset to learn its fundamental patterns and subsequently producing new data that align with these patterns. The effectiveness of the proposed pipeline is validated through extensive set of experiments across various life-science datasets. The results demonstrate significant improvements in error metrics, including reduced mean absolute error (MAE) and root mean square error (RMSE), alongside increased R^2 scores. These findings highlight the enhanced accuracy and reliability of the pipeline compared to conventional approaches.

Keywords

Time series, GAN, Deep Learning, Machine Learning, Data Modeling,

1. Introduction

Scientific data is crucial in advancing our understanding of natural phenomena and driving innovations across various fields. It encompasses a range of data types, from categorical data requiring straightforward statistical methods to complex time-series data necessitating sophisticated analytical approaches and advanced artificial intelligence (AI) techniques. A significant subset of scientific data is life science data, which often involves time series measurements such as soil moisture levels and temperature fluctuations. These measurements are highly sensitive to climatic variations and external factors. Accurate monitoring, analysis, and prediction of these parameters are essential for environmental preservation, agricultural management, and climate change mitigation. However, collecting and analyzing time series data in life sciences presents challenges due to sensor issues such as noise, errors, and sensor drift, which complicate data collection. Enhancing data quality involves addressing these issues to improve reliability [1]. In addition, deploying physical sensors can be cost-prohibitive, logistically challenging, and often produce limited data volumes. Uneven sensor distribution and environmental variability further exacerbate these challenges, leading to incomplete datasets that affect the performance of traditional machine learning and deep learning models.

To overcome these data collection and analysis challenges, advanced techniques such as Generative Adversarial Networks (GANs) can be utilized. GANs create synthetic datasets that simulate real-world conditions, reducing the need for extensive physical data collection. Data augmentation using GANs improves the spatial and temporal resolution of environmental research data, providing a more comprehensive view of the monitored environment and enhancing the performance of ML models with more diverse training samples [2, 3]. Moreover, these techniques can help in anomaly detection and enhance the robustness of predictive models by addressing data imbalance [4, 5].

To sum up, the followings are the main contributions of the paper:

1. Introducing an effective pipeline for analysis and processing sparse temporal life science data.
2. Investigating the performance of traditional machine learning and deep learning models in understanding climate-soil interactions, and
3. Applying GANs in life science data analysis.

2. Background

The field of soil-climate interactions and life science data analysis encompasses many complex concepts and methodologies. Establishing a robust foundation for this work, this section presents a thorough examination of the basic principles, ideas, and approaches that are pertinent to the argument.

2.1. Time Series

Time series analysis, as described in [6], is a statistical technique used to examine data points that are organized sequentially. The primary objective of time series analysis is to understand the underlying structure and process that produced the data. This approach is widely used for many reasons, such as economic forecasting, stock market analysis, weather prediction, and many more[7]. The selected dataset for our investigation comprises soil-climate data, which exemplifies the scientific data that may greatly benefit from time series analysis. Long-term collection of soil-climate data offers essential knowledge on the relationships between soil characteristics and climatic elements including temperature, precipitation, and humidity.

2.2. Tradition ML & DL Methods for Time Series

Imputation techniques are crucial for estimating and substituting missing values to facilitate comprehensive data analysis. There are several ways to deal with missing values,

Published in the Proceedings of the Workshops of the EDBT/ICDT 2025 Joint Conference (March 25-28, 2025), Barcelona, Spain

*Corresponding author.

vijayv01@ads.uni-passau.de (A. Vijayvergiya);

alsayed.algergawy@uni-passau.de (A. Algergawy)

such as cubic and linear imputation, Gaussian imputation, and K-Nearest Neighbors (KNN) imputation. Cubic and linear imputation methods introduced by [8] [9] are widely used for their simplicity and effectiveness in time series and continuous data. Gaussian imputation requires that the data conforms to a Gaussian (normal) distribution, as stated by Little in 1987 [10] [11]. K-closest Neighbors (KNN) imputation is a non-parametric technique that utilizes the k closest neighbors to approximate the missing values [12] [13] [14].

Machine learning utilizes a range of models to examine data and make predictions about future events. This section presents two often used models: Linear Regression and Random Forest [15]. Both models were selected based on their efficacy in managing time-varying data, which is essential for precise forecasting and analysis in dynamic settings.

Deep Neural Network (DNN) [16] is an extension of a simple neural network with multiple hidden layers between the input and output layers. The addition of these hidden layers allows the network to model more complex relationships in the data. The working of each layer in a DNN follows the same principles as in a simple neural network, but with repeated layers, the depth of the network increases.

Generative Adversarial Networks GAN are a class of machine learning frameworks invented by Ian Goodfellow and his colleagues in 2014 [17]. GANs consist of two competing neural networks, the Generator (G) and the Discriminator (D), which are trained simultaneously through a process known as adversarial training. GAN is designed to generate synthetic data that resembles real data. It achieves this through the interaction of two neural networks: **Generator (G):** Generates new data samples by taking an *Input Noise Vector* and producing *Generated Data*, and **Discriminator (D):** Evaluates whether the data samples are real or generated.

The training of GANs involves a minimax game between the generator and the discriminator:

1. **Discriminator Training:** The discriminator is updated to maximize the probability of correctly classifying real and fake data. The loss function for the discriminator is given by:

$$L_D = -\Big(\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]\Big) \text{ [17]}$$

2. **Generator Training:** The generator is updated to minimize the probability that the discriminator correctly classifies the generated data as fake. The loss function for the generator is given by:

$$L_G = -\mathbb{E}_{z\sim p_z(z)}[\log D(G(z))] \text{ [17]}$$

The overall objective function of the GAN is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \text{ [17]}$$

3. Related Work

Key advancements include the optimization of deep learning models using GANs and the Sailfish Optimization Algorithm (SOA) for soil moisture prediction [18, 19]. This method enhances the quality of synthetic data generation, addressing the challenge of incomplete and inconsistent soil moisture readings. Machine learning models such as Random Forest (RF) and Support Vector Machines (SVM) have been utilized for soil moisture and temperature prediction, demonstrating robustness and improved generalization capabilities. However, these models also faced limitations in handling missing data and variability in data quality [20]. Recent studies have leveraged GANs for data augmentation, which significantly enhances classifier performance by increasing the volume and diversity of the training data [21]. Additionally, the integration of GANs with Long Short-Term Memory (LSTM) networks (GAN-LSTM) has been explored to improve the accuracy of soil moisture predictions by generating high-quality synthetic time series data [22]. Moreover, GANs have been applied to refine seasonal weather predictions, demonstrating significant potential for high-resolution forecasting and capturing intricate spatial patterns among climate variables [23].

This paper builds on these advancements by incorporating advanced imputation techniques to preprocess datasets before applying GANs, ensuring the generation of high-quality synthetic data. By integrating various deep learning and machine learning models, including GANs, DNNs, SNNs, and CNNs. This work further aims to develop highly accurate and reliable prediction models for soil moisture and temperature. This comprehensive strategy results in enhanced prediction models that exhibit high levels of accuracy and reliability.

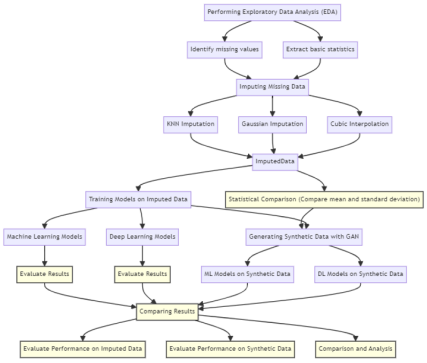

Figure 1: Flowchart of the Methodology

4. Methodology

In this section, we outline the proposed approach, which consists of the following main tasks, as shown in Figure 1: *Exploratory Data Analysis (EDA)*, *imputing missing data*, *training models on the completed datasets*, *generating synthetic data using GANs*, and *evaluating the effectiveness of these models*.

In the following we are going to provide more details for each task.

4.1. Performing Exploratory Data Analysis (EDA)

EDA is an essential process for understanding the organization, integrity, and attributes of the information. The first step involves identifying any missing values present in the dataset. This stage is crucial as it establishes the magnitude of the missing data and guides the approach for addressing

it. The number of missing values in each row and column is computed to assess the degree of data incompleteness. Hence, comprehending the dispersion and quantity of missing data aids in determining the appropriate imputation methods to guarantee the integrity and use of the dataset.

4.2. Imputing Missing Data

After identifying missing data, the subsequent action is to impute these absent values. Three different algorithms: KNN, Cubic and Gaussian which are explained in section (2.2) are employed and later compared to pick the best method for further model training.

A statistical study is conducted to evaluate the efficacy of each imputation strategy following the imputing of missing data. By comparing the imputed data with the original data, the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) offer insights on both the imputed values precision and fluctuation. Whereas RMSE is susceptible to outliers and assigns greater weight to higher errors, MAE calculates the average size of prediction errors. These measures aid in assessing how effectively the imputation techniques maintain the distribution and structure of the underlying data [14] [24]. This comparison facilitates the assessment of the efficacy of each imputation technique as indicated by the results in section (5.3).

4.3. Model Selection

When analyzing the relationship between soil and the environment over time, it is essential to choose suitable models that can accurately represent the intricate dynamics and inter-dependencies included in the data. The model selection procedure was guided by the need to achieve a balance between predicted accuracy, interpretability, and the capability to handle many types of data patterns. The following models—Random Forest, Linear Regression, Simple Neural Network, Deep Neural Network, and Convolutional Network—were chosen because of their suitability for training time-series data and are described in section (2.2).

The pre-processed data obtained through imputation in subsection (4.2) and EDA in subsection (4.1) is used as training and validation dataset for further training Models.

4.4. Machine And Deep Learning Model Evaluation on Real Data

The outcomes shown in subsections (5.4) analyze the efficacy of ML and DL models when utilized with real-world datasets. The findings suggest that both ML and DL models have subpar performance.

The inadequate performance highlights the need for other methods to improve the accuracy and resilience of the model. An encouraging strategy is the use of Generative Adversarial Networks as explained in subsection (2.2).

4.5. GAN Model Implementation and Generating Synthetic Data

The procedures for data processing, GAN model training, and assessment are outlined in subsection (2.2). The primary objective is to produce artificial data samples using GANs, using diverse input characteristics obtained from several sources as training data.

4.6. Evaluating ML and DL Models on Synthetic Data

After generating the synthetic data, ML and deep learning DL models are trained and assessed using this data to verify its fidelity to the original dataset and results are described in section (5.6). The technique entails a meticulous selection of synthetic data, guided by statistical measurements, to guarantee its resemblance to the original data.

5. Results

This section summarizes the results obtained from the analysis and modeling conducted in this research. The sections are arranged in a systematic manner to comprehensively address the following topics: exploration of the dataset, comparison of imputation methods, application and evaluation of machine learning and deep learning models on real data, generation and assessment of synthetic data using GANs, and comparative analysis of model performance on real versus synthetic data.

5.1. Dataset Description

To validate the performance of the proposed approach, we use three different datasets from the open data provided by biodiversity exploratory information system (BExIS)[1]) [25] [26], which serves as the data portal for biodiversity datasets collected within the framework of the Biodiversity Exploratories project[2]. The collection includes a variety of climate and soil factors that are consistently recorded and classified.The characteristics of selected datasets are illustrated in Table 1. As shown in the table, we selected datasets that represent two different domains, *soil* and *climatic*.

Dataset	Start Date	End Date	No. Features	No. tuples
ds_set1.csv	05-01-2008	06-09-2010	26	801
ds_set2.csv	03-01-2009	04-09-2011	26	801
ds_set3.csv	09-02-2020	04-09-2024	26	801

Table 1
Dataset Description

Climate parameter	Description
Ta_10	Air temperature in 10 cm
Ta_200	Air temperature in 2 m
Ta_200_min	Minimum temperature
Ta_200_max	Maximum temperature
Ta_200_ice_days	Eistage
Ta_200_cool_days	cold days
Ta_200_cold_days	Frost days
Ta_200_warm_days	warm days
Ta_200_summer_days	Summer days
Ta_200_tropical_days	talk
Ta_200_tropical_nights	Tropical nights
Ta_200_growing_degree_days_10	Growing degree days
precipitation_radolan	Precipitation
precipitation_radolan_rain_days	Rain days
WD	Wind direction
WV	Wind speed
WV_gust	Peak gusts
SD_Olivieri	Sunshine duration

Table 2
Climate Features

[1] https://www.bexis.uni-jena.de/
[2] https://www.biodiversity-exploratories.de/en/

The climatic features specified in Table (2) serve as input variables for machine learning (ML) and deep learning (DL) models. These features encompass a range of meteorological parameters, including temperature, precipitation, wind attributes, and duration of sunshine. They offer a thorough comprehension of the climate dynamics across various altitudes and temporal dimensions.

Soil parameter	Description
Ts_05	Soil temperature in 5 cm
Ts_10	Soil temperature in 10 cm
Ts_20	Soil temperature in 20 cm
Ts_50	Soil temperature in 50 cm
SM_10	Soil moisture in 10 cm
SM_20	Soil moisture in 20 cm

Table 3
Soil Features

The soil features listed in Table (3) are used as output features for the ML and DL models. These features include soil temperature and soil moisture at various depths. By understanding the relationship between climate inputs and soil outputs, the models can predict soil conditions based on climatic variations, which is essential for applications in agriculture, environmental monitoring, and land management.

5.2. Analysis Using Exploratory Data Analysis

In the domain of actual data, sparsity is a prevalent problem, as shown by the abundance of NaN (Not a Number) values. The presence of these gaps in the data may often be ascribed to sensor malfunctions or other difficulties related to data collecting. The analysis, shown in Figure 2, indicates that the columns SD Olivier, WD, WV, and WV gust do not have data and should be excluded. Figure 3 demonstrates that a considerable proportion of rows have over 50% missing data.

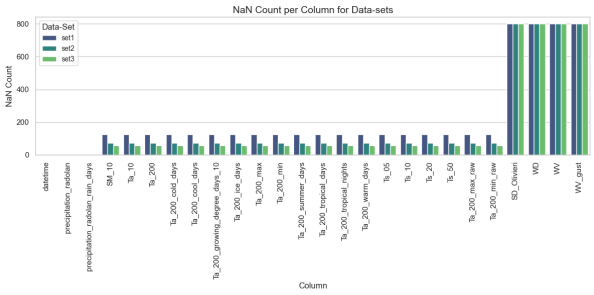

Figure 2: NaN Count Per Column for Datasets

5.3. Analysis of Different Imputation Methods

Various imputation approaches may be used to address missing data. To assess the efficacy of these techniques, we may compare the mean absolute error (MAE) and the root mean square error (RMSE) values obtained from the original and imputed datasets. The performance of three imputation approaches, namely Cubic, K-nearest neighbors (KNN), and Gaussian, was evaluated in three datasets (set1, set2, set3), as shown in Figure 4.

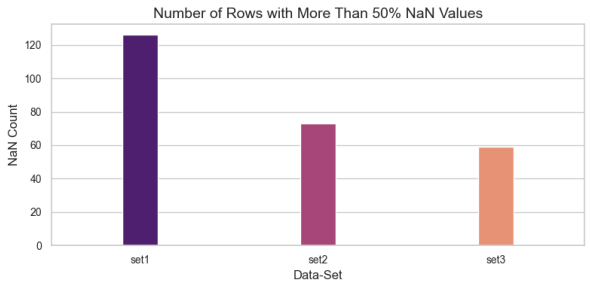

Figure 3: Number of Rows with More Than 50% NaN Values

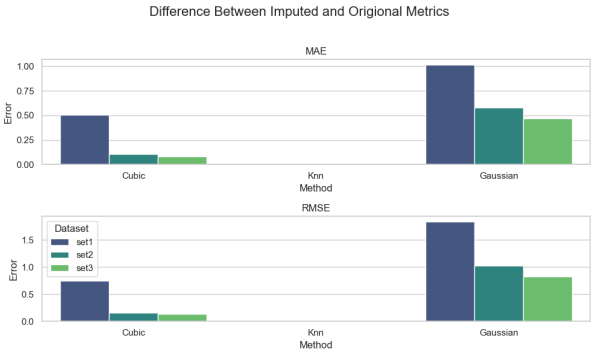

Figure 4: MAE and RMSE Errors for Different Imputation Methods

5.4. Analysis of Machine Learning and Deep Learning Models on Real Data

The examination of various machine learning models on the dataset Figure 5 uncovers a consistent trend where the output feature `SM_10` displays significantly higher MSE and MAE, as well as notably lower R^2 scores in comparison to other target variables (`Ts_05`, `Ts_10`, `Ts_20`, `Ts_50`). This suggests that the `SM_10` input feature has the most impact, leading to poor prediction performance and poor model fit. On the other hand, alternative target variables exhibit reduced error rates and increased R^2 scores, indicating superior model performance and predictive accuracy. Despite experimenting with several hyperparameters in the Random Forest model, no substantial improvements were seen. Similar trends were observed in all three datasets. Similar results were also visible with the deep learning models on all three datasets as described by Figure 6.

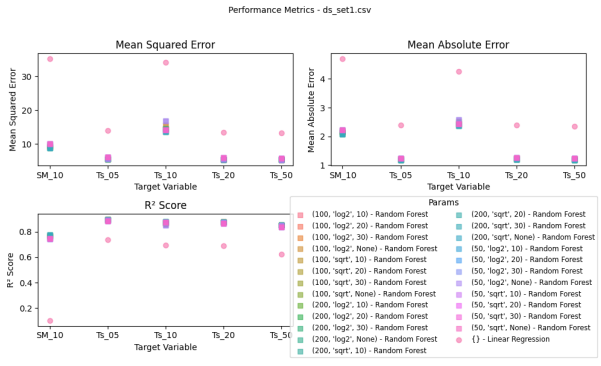

Figure 5: ML Performance Metrics for ds_set1.csv

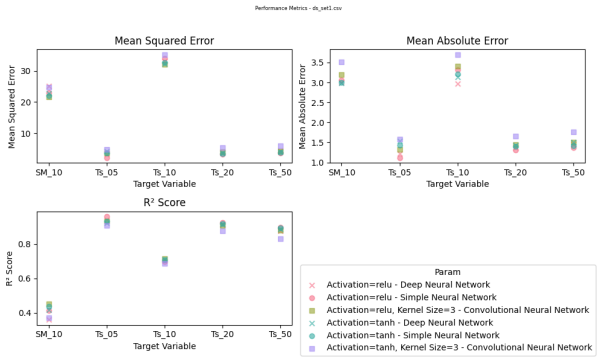

Figure 6: Deep Learning Performance Metrics for ds_set1.csv

5.5. Analysis of Real vs GAN Synthetic Data

The figures provided provide a comparison examination of performance metrics between the original data and the data produced by a GAN for three datasets (ds_set1, ds_set2, ds_set3). Each image consists of two subplots: one displaying the mean values and the other representing the standard deviations. When examining the subplots that compare the average values, it becomes evident that the produced data closely resembles the original data in virtually all aspects. This suggests that the GAN successfully captures the fundamental distribution of the original dataset. The strong agreement seen across several hyperparameter configurations, as shown in Figures 7, 8, and 9, highlights the GAN model's ability to faithfully reproduce the average values of the original dataset. Similarly, the analysis of the subplots comparing the standard deviations shows that the generated data closely resemble the variability of the original data. However, there are some deviations in certain features, indicating that while the GAN performs well overall, reproducing specific features accurately may be more challenging.

The evaluation also considers the effects of several hyperparameters, which are represented by various markers and colors. These combinations consist of learning rates (0.0002, 0.003, 0.001) and loss functions (mean squared error, binary cross-entropy).

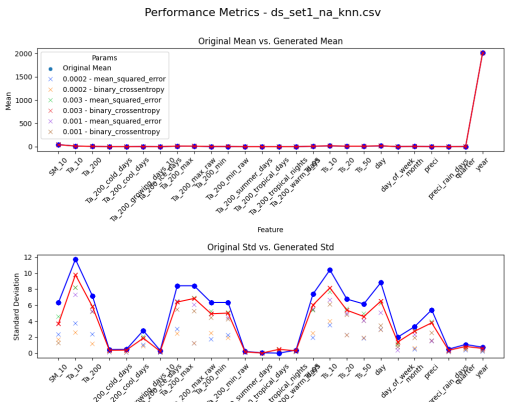

Figure 7: Mean and SD Performance Metrics for ds_set1.csv

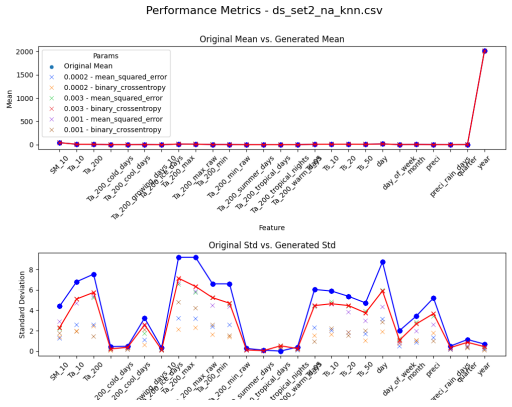

Figure 8: Mean and SD Performance Metrics for ds_set2.csv

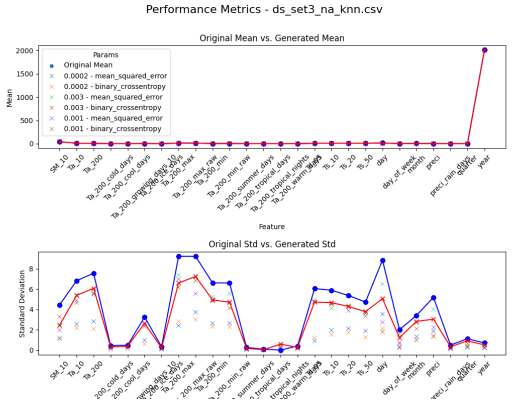

Figure 9: Mean and SD Performance Metrics for ds_set3.csv

5.6. Analysis of DL and DL Models on Real vs Synthetic Data

The assessment of deep learning models on three datasets (Figures 10, 11, and 12) reveals significant improvements when using GAN-generated synthetic data in comparison to the original data. Notable observations consist of:

- **Mean Squared Error**: Models trained on synthetic data consistently provide decreased MSE values for all target variables, with significant enhancements seen for `SM_10`.
- **Mean Absolute Error**: The use of GAN-generated data leads to a notable decrease in MAE, especially for the `SM_10` target variable.
- R^2: Higher R^2 scores show more explanatory power for models trained on synthetic data, particularly improving the prediction performance for `SM_10`.

The results emphasize the effectiveness of using synthetic data produced by GANs to improve the accuracy and reliability of models, especially in predicting the `SM_10` variable.

Similar observations were observed using ML models trained on the synthetic data. The assessment of machine learning models in three datasets (Figures 13, 14, and 15) reveals substantial improvements when using synthetic data generated by GAN compared to the original data. Notable observations consist of:

- **Mean Squared Error**: Models trained on synthetic data consistently exhibit decreased MSE values for

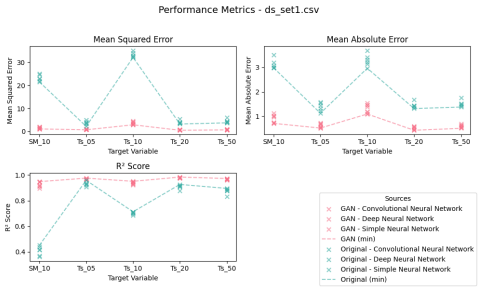

Figure 10: DL Metrics for ds_set1.csv: GAN Data vs. Original Data

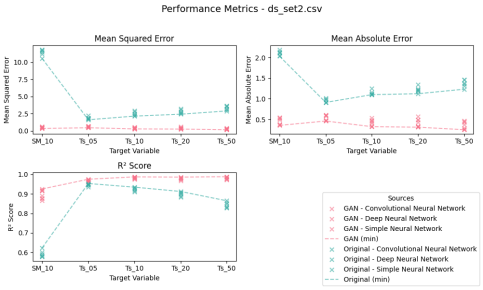

Figure 11: DL Metrics for ds_set2.csv: GAN Data vs. Original Data

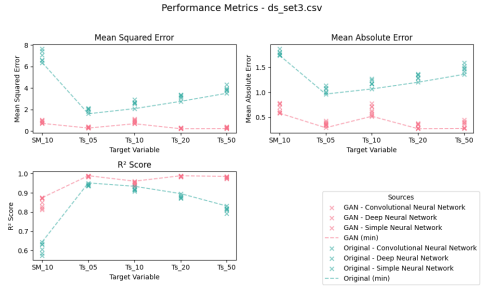

Figure 12: DL Metrics for ds_set3.csv: GAN Data vs. Original Data

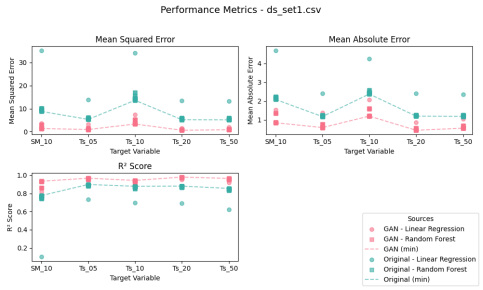

Figure 13: GAN Data vs. Original Data: ds_set1.csv

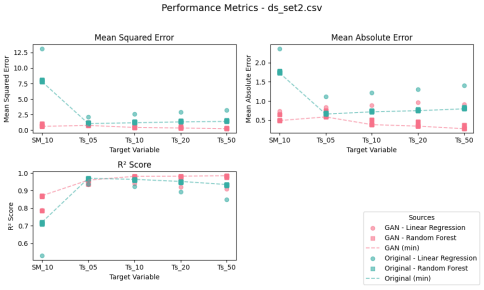

Figure 14: GAN Data vs. Original Data: ds_set2.csv

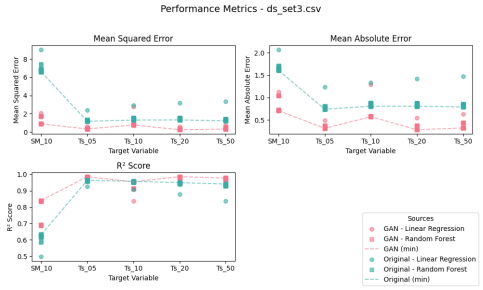

Figure 15: GAN Data vs. Original Data: ds_set3.csv

all target variables, with particularly noticeable enhancements for SM_10.

- **Mean Absolute Error**: The use of GAN-generated data significantly decreases the MAE, especially for the SM_10 target variable.
- R^2: Higher R^2 scores suggest better explanatory power for models trained on synthetic data, particularly improving predictive performance for SM_10.

The results demonstrate the efficacy of using synthetic data generated by GAN to improve the accuracy and resilience of the model, particularly for the SM_10 target variable.

6. Conclusion

In this paper we investigated the analysis of time series datasets collected with the life science domain. We demonstrate the effect of KNN-based imputation techniques and show how KNN imputation consistently outperforms other methods, making it the optimal choice for addressing missing data in this scenario. The data generated by GANs exhibits a high degree of similarity to the original data, demonstrating GANs' ability to accurately replicate the underlying distribution of the actual dataset (5.5). Models trained on GAN-generated data show superior performance compared to those trained on real data, as evidenced by significantly improved evaluation metrics, such as lower MSE and higher R2 scores. These improvements are observed in both machine learning and deep learning models across various datasets described in section (5.6). The findings of this paper have broad applicability in biological sciences and environmental research. This study enhances the resilience and precision of models predicting soil properties under different climatic conditions, facilitating more reliable agricultural planning and environmental monitoring.

Declaration on Generative AI

Either:
The author(s) have not employed any Generative AI tools.

During the preparation of this work, the author(s) used X-GPT-4 and Gramby in order to: Grammar and spelling check. Further, the author(s) used X-AI-IMG for figures 3 and 4 in order to: Generate images. After using these

tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, S. Gómez-Canaval, Data augmentation techniques in time series domain: a survey and taxonomy, Neural Computing and Applications 35 (2023) 10123–10145.

[2] M. Cekić, Anomaly detection in medical time series with generative adversarial networks: a selective review, Anomaly Detection-Recent Advances, AI and ML Perspectives and Applications (2023).

[3] T. Chakraborty, U. R. KS, S. M. Naik, M. Panja, B. Manvitha, Ten years of generative adversarial nets (gans): a survey of the state-of-the-art, Machine Learning: Science and Technology 5 (2024) 011001.

[4] B. K. Iwana, S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, Plos one 16 (2021) e0254841.

[5] H. Jeon, D. Lee, A new data augmentation method for time series wearable sensor data using a learning mode switching-based dcgan, IEEE Robotics and Automation Letters 6 (2021) 8671–8677.

[6] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[7] P. Diggle, E. Giorgi, Time series: a biostatistical introduction, Oxford University Press, 2024.

[8] M. G. Kendall, The advanced theory of statistics. (1946).

[9] K. Kornelsen, P. Coulibaly, Comparison of interpolation, statistical, and data-driven methods for imputation of missing values in a distributed soil moisture dataset, Journal of Hydrologic Engineering 19 (2014) 26–43.

[10] R. J. Little, D. B. Rubin, Statistical analysis with missing data, New York: Wiley (1987).

[11] X. Qiu, F. Wang, Q. Zhang, G. Tao, S. Zhou, An improved gaussian process for filling the missing data in gnss position time series considering the influence of adjacent stations, Scientific Reports 14 (2024) 19268.

[12] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R. B. Altman, Missing value estimation methods for dna microarrays, Bioinformatics 17 (2001) 520–525.

[13] P. Dixneuf, F. Errico, M. Glaus, A computational study on imputation methods for missing environmental data, arXiv preprint arXiv:2108.09500 (2021).

[14] N. A. Zainuri, A. A. Jemain, N. Muda, A comparison of various imputation methods for missing values in air quality data, Sains Malaysiana 44 (2015) 449–456.

[15] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.

[16] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, nature 323 (1986) 533–536.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27 (2014).

[18] D. M. N. Sivasankaran S1, Dr.K. Jagan Mohan2, Soil moisture quantity prediction using optimized deep learning model with sailfish optimization algorithm, Journal of Neural Networks 15 (2024) 4268–3515.

[19] E. Brophy, Z. Wang, Q. She, T. Ward, Generative adversarial networks in time series: A systematic literature review, ACM Computing Surveys 55 (2023) 1–31.

[20] S. Siddharth, R. Abhishek, S. Karthik, Machine learning applications for predicting soil moisture, Environmental Modelling & Software 134 (2020).

[21] V. Venkatesan, K. Nithya, B. Karthikeyan, A. Adilakshmi, A deep learning data augmentation experiment to classify agricultural soil moisture to conserve plants, International Journal of Intelligent Systems and Applications in Engineering 11 (2023) 114–119. URL: https://www.ijisae.org/index.php/IJISAE/article/view/2834.

[22] Y. Wang, L. Shi, Y. Hu, X. Hu, W. Song, L. Wang, A comprehensive study of deep learning for soil moisture prediction, Hydrology and Earth System Sciences Discussions 2023 (2023) 1–38.

[23] G. Gousios, T. Mamouka, P. Vourlioti, S. Kotsopoulos, Downscaling seasonal weather forecasting with generative adversarial networks, preprint (2023).

[24] N. M. Noor, M. M. Al Bakri Abdullah, A. S. Yahaya, N. A. Ramli, Comparison of linear interpolation method and mean method to replace the missing values in environmental data set, in: Materials science forum, volume 803, Trans Tech Publ, 2015, pp. 278–281.

[25] BEXIS 2, Tropical climate data: Exported sensor data, 2024. URL: https://www.biodiversity-exploratories.de/en/klimatool/.

[26] M. Fischer, T. Weithoener, W. W. Weisser, TubeDB: An on-demand processing database system for climate station data, Computers & Geosciences 145 (2020) 104641. doi:10.1016/j.cageo.2020.104641.