

Generative Adversarial Network Applied to the Energy Efficiency of Buildings

Julio Barzola-Monteses, MSc.^{1,2}, Franklin Parrales-Bravo, PhD.^{1,2}, Gary Reyes, PhD.^{1,2}, Vicente Macas-Espinosa, MSc.³, Jean Merchan Merchan, Ing.², Carlos Maridueña Muñoz, Ing.², and A.H. Yanez, MSc.^{4,2},

¹Universidad Bolivariana del Ecuador, Durán, Ecuador, jjbarzola@ube.edu.ec,

²Artificial Intelligence research group, University of Guayaquil, Guayaquil, Ecuador, julio.barzola@ug.edu.ec, franklin.parralesb@ug.edu.ec, gary.reyesz@ug.edu.ec, jean.merchanm@ug.edu.ec, carlos.mariduenam@ug.edu.ec

³Facultad de Arquitectura y Urbanismo, Universidad de Guayaquil, Guayaquil, Ecuador, vicente.macase@ug.edu.ec

⁴Escuela Técnica Superior de Ingeniería Informática, Universidad de Málaga, Málaga, España, alcibar.yaneze@uma.es

Abstract.— *Energy consumption in buildings represents a significant proportion of the global energy consumption, which raises the need to develop strategies for its optimization. However, datasets can often be incomplete when analyzing energy variables, such as electricity consumption, due to missing measurements or equipment failures. Generative adversarial networks (GANs) can generate high-quality synthetic data that mimics actual data distribution. Through a literature review, this study examined how GANs have been applied to study building energy efficiency. In addition, as a case study, we consider a dataset generated from the historical data of the FCMF-UG building of the University of Guayaquil. The findings demonstrated that the variability of the original data influences the results of curve generation with GANs. These preliminary results can serve as a baseline for future analysis of GANs applied to building energy efficiency.*

Keywords. - *energy efficiency building, generative adversarial network, machine learning, time series.*

I. INTRODUCCIÓN

The building sector is responsible for significant global energy consumption, underscoring the need to implement energy efficiency strategies in this area. However, the analysis of energy and environmental variables, such as electricity consumption and temperature, is often hampered by incomplete data sets due to lost measurements or equipment failures. To address this challenge, generative adversarial networks (GANs) have emerged as a promising tool in generating high-quality synthetic data that mimics the distributions of actual data [1].

GANs, introduced by Goodfellow, consist of two competing neural networks: a generator and a discriminator. The generating network creates synthetic data while the discriminator evaluates its authenticity, thus refining the generator's ability to produce indistinguishable data from the real thing. This architecture has been widely adopted in a variety of applications, including generating images, music, and synthetic data to train other machine-learning models [1].

In the context of energy efficiency in buildings, GANs have been explored for various applications. For example, its use in predicting energy demand, generating missing data in time series of energy consumption, and simulating scenarios to

optimize the design and operation of energy systems in buildings has been investigated. These applications seek to improve the accuracy of predictive models and facilitate informed decision-making in building energy management [2].

Despite the advances, applying GANs in this domain faces significant challenges. The inherent variability in energy consumption data, influenced by factors such as weather conditions, occupancy patterns, and system efficiency, can affect the ability of GANs to generate accurate and representative data. In addition, the evaluation of the quality of the synthetic data generated and its impact on energy efficiency decisions requires robust methodologies and detailed empirical studies [3].

This article aims to review the existing literature on the application of GANs in the energy efficiency of buildings. In addition, a case study is presented that uses historical data from the FCMF-UG building of the University of Guayaquil to generate synthetic datasets using GANs, evaluating their effectiveness and accuracy in the representation of energy consumption patterns. The findings of this study will provide a basis for future analyses and applications of GANs in optimizing energy consumption in buildings.

The rest of the document is organized as follows: Section II summarizes the related works to the article's subject. Section III details the method used to apply GANs as a case study of the FCMF-UG building dataset. Section IV explains the analysis of the study's results, and Section V focuses on determining the conclusions and future work that can be derived from this research.

II. RELATED WORKS

Bibliographic research based on scientific articles and relevant studies has been conducted to analyze the energy curves of educational buildings using generative adversarial networks (GANs). This approach allows the identification and comparison of various techniques used to generate data and predict energy consumption, highlighting the application of GANs.

The study conducted by Montero et al. [4] presented an innovative approach by applying energy curves in educational buildings using GANs to model and predict the energy

consumption. This method allows the generation of synthetic data that reflects actual energy behavior, thus facilitating the planning and management of consumption in these spaces. To measure the effectiveness of the model, metrics such as the correlation coefficient (R) and mean square error (MSE) were used, achieving remarkable accuracy with $R = 0.9435$, which indicates a high effectiveness in predicting energy consumption and suggests a significant potential to optimize energy use in educational environments.

Similarly, in [5] They applied to model the energy curves of educational buildings, which allowed a more accurate representation of energy consumption by integrating historical data and environmental variables such as temperature and occupancy. This innovative approach focuses on generating synthetic data that complements existing information, thereby improving the model's predictive capability. To evaluate the model's effectiveness, metrics such as MAE and the coefficient of determination (R^2) were used, achieving an MAE of 29.5 kWh/d and an R^2 of 0.863. These results not only indicate good performance in the prediction of energy consumption but also highlight the potential of GANs to optimize energy management in educational environments, facilitating the identification of consumption patterns and the implementation of energy efficiency measures.

A studio in [6] explored the application of GANs to generate a solar irradiance time series for urban facades using fisheye images. By integrating a Variational Autoencoder (VAE) with GANs, the model effectively captured complex urban features and generated high-fidelity solar irradiance data. The results demonstrate that the model can produce realistic time series that align closely with ground truth data, significantly reducing the computation time compared to traditional simulation methods. This approach demonstrates the potential of GANs to enhance urban energy planning and design under varying climatic conditions.

In the same way, Baasch et al. [7] A conditional generative adversarial network (C-TimeGAN) is applied to generate energy load profiles in residential and commercial buildings with scarce data using time series and variables such as monthly average outdoor temperature. The architecture employs temporal convolutional networks (TCNs) with an autoencoder and supervised learning components to model temporal patterns. In terms of metrics, accuracy was evaluated using the Jensen-Shannon Divergence (JSD), achieving 0.012 for residential data and 0.037 for commercial data, in addition to mean absolute errors (MAE) close to the originals, demonstrating high fidelity in data generation and competitiveness against previous approaches with only 1-2% of the data size used in other studies.

The study conducted by Ortega-Diaz et al. [8] analyzed the application of GANs to model energy curves in educational buildings, highlighting their ability to generate synthetic data that mimics energy consumption patterns. This approach makes it possible to improve the accuracy of consumption predictions, which can optimize energy management and promote sustainability in the education

sector. This research highlights the potential of GANs as innovative tools in building planning and energy analysis. They used performance metrics, such as MSE, MAE, and R^2 , to evaluate the effectiveness of GANs in modeling energy curves in educational buildings. The results showed that the GANs achieved remarkable accuracy, with an R^2 value greater than 0.90, indicating a high correlation between the generated and actual data. This suggests that GANs are effective in predicting energy consumption and in improving management and sustainability in the educational field.

A study developed by Yu et al. [9] Implemented GANs to address the loss of energy data in educational buildings and to improve the accuracy of consumption predictions. These GANs learn the distribution of the original data to generate virtual data that, when combined with the actual data, significantly improves the model's performance. Energy consumption was predicted using a backpropagation neural network optimized with the Levenberg-Marquardt algorithm (BPNN optimized). The metrics used to evaluate the model were MAE.

The study conducted by Labiadh [10] addressed the use of GANs to model the energy curves of educational buildings, focusing on generating synthetic data for scenarios where historical data are limited or incomplete. These networks allow simulation of energy consumption patterns and improve predictions, thereby providing a robust approach for energy planning. The metrics used to evaluate the performance included MAE, MSE, and R^2 , achieving accuracies of more than 90% in the validation scenarios.

The results highlighted the effectiveness of GANs not only in generating realistic data but also in their ability to improve the performance of predictive models, representing a significant advance in the energy analysis of educational buildings.

In addition, Choi et al. [11] presented a hybrid framework based on Conditional Adversarial Generative Networks (CGNs) and Time GAN for the generation of synthetic PV power data integrated into educational buildings, addressing the data shortage in this area. The methodology includes the incorporation of temporal attributes as conditioning information, which ensures chronological order and improves the fidelity of the generated data. Metrics such as the discriminative score (D-score) and predictive score (P-score) were used to evaluate the quality and usefulness of the data generated. The model showed a 79.58% improvement in the D-score and a 13.46% improvement in the P-score compared with Time GAN. In addition, the integration of synthetic data into the prediction models resulted in an increase of up to 23.56% in the accuracy of the MAE error for power generation predictions.

Likewise, researchers Li et al. [12] applied a hybrid methodology for short-term prediction of energy consumption in industrial and commercial buildings, using temporal antagonistic generative networks (TimeGAN), convolutional neural networks (CNNs), and long-term and short-term memory networks (LSTMs). First, TimeGAN generated

Table I: Literature of GANs Applied to the Energy Efficiency of Buildings

Ref.	Application	GAN Method Used	Result obtained
[4]	Generation of synthetic data to model energy behaviour	GANs	$R = 0.9435$
[5]	Integration of historical data and environmental variables	GANs	MAE = 29.5 kWh/d; $R^2 = 0.863$
[6]	Data Generation for Residential and Commercial Buildings	C-TimeGAN	JSD: 0.012 (residential), 0.037 (commercial)
[7]	Generation of synthetic solar irradiance time series	VAE + GAN	$JSD \leq 0.1$, Fast Generation and Robust Generalization
[8]	Generating Synthetic Data for Energy Consumption Predictions	GANs	$R^2 > 0.90$, high correlation between generated and actual data
[9]	Generation of virtual data combined with real data for improved predictions	GANs	MAE reduced by 14%, $R^2 = 0.944$
[10]	Generating Synthetic Data for Limited Historical Data	GANs	Accuracy > 90% in validation scenarios
[11]	Generation of synthetic photovoltaic power data	CGAN + TimeGAN	79.58% (D-score) and 13.46% (P-score) improvement over TimeGAN
[12]	Data-limited energy prediction through knowledge transfer	DANN	40-90% improvement compared to non-optimized models
[13]	Fault Diagnosis (FD) in Building Energy Systems, Specifically HVAC Systems	FD+CNN+DANN+GANS	Accuracy 93% average improvement of 55%
[14]	Synthetic Data Generation and Analysis for Hybrid Models	TimeGAN + CNN + LSTM	MAPE < 5%, $R^2 = 0.812$
[15]	Generating Synthetic Data for Unbalanced Data	GANs	R^2 improvement of 17.13% (overall) and 14.82%

synthetic data to complement limited actual datasets. The CNNs then extracted relevant features, and the LSTMs made the prediction of energy consumption. The model was evaluated using metrics such as mean absolute percentage error (MAPE), RMSE, and R^2 . The results showed an average ASM of less than 5% and an average R^2 of 0.812, meeting high standards of accuracy and effectiveness in predictions.

In the work carried out by Gao et al. [13] developed a comparative study on deep transfer learning strategies for fault diagnosis in building energy systems under cross-conditions, this study addresses fault diagnosis (FD) in building energy systems, specifically in HVAC systems, using three deep transfer learning (DTL) strategies: fine-tuning (FT), domain adaptive neural network (DaNN), and adversarial neural network (DANN). These techniques improve diagnostic capability when using data from different systems and operating conditions. The evaluation showed that fine-tuning (FT) achieved 93% accuracy across all tasks, with an average improvement of 55% compared to the convolutional neural network (CNN) base model. The study also looked at the impact of source and target data volume on diagnostic performance.

Liu et al. [14] presented a comprehensive evaluation of the transferability of building energy prediction using deep adversarial network transfer learning. This study explores the use of deep adversarial networks (DANN) to improve the energy prediction of buildings with limited data and transfer knowledge from source buildings with sufficient information. Factors such as data similarity, structural characteristics, and the volume of data required were analyzed. Using data from the Building Data Genome Project, the methodology was validated for 36 buildings of six different types. The results showed significant improvements in accuracy, between 40% and 90%, compared to non-optimized LSTM models.

Finally, Zhang et al. [15] applied GANs to address the data imbalance between working days and holidays in the prediction of the energy consumption of educational buildings by combining global and independent modelling using CNN-LSTM. Metrics such as R^2 , normalized RMSE, and normalized MAE were used. The results showed significant improvements: R^2 increased from 0.7466 to 0.8745 (17.13%) in the global model and from 0.7659 to 0.8794 (14.82%) in the standalone model, whereas the NRMSE and NMAE decreased by 14.77% and 16.95%, respectively, demonstrating the

effectiveness of the approach in improving the accuracy of energy predictions. Table 1 summarizes the key information from the 14 studies previously mentioned on various applications of advanced GANs and ANNs methods. This contrasts with the different applications, models used, and error metrics considered.

III. METHODOLOGY

After the literature review presented in the previous section, an application was carried out in the GANs to generate synthetic energy data considering a dataset of actual data taken by the sensors installed in the Faculty of Mathematical and Physical Sciences of the University of Guayaquil building. More project details can be found in [16], [17]. Also, to analyze the further performance of the GANs, a 2nd dataset with temperature measurement data was considered.

Figure 1 presents the flow diagram of the five processes carried out to apply the GANs to the previously mentioned datasets.

Historical Data Collection

Data on the energy consumption of the FCMF-UG building were obtained using smart metres. All the data collected during the study period were from September 2021 to August 2023. The dataset consisted of 16 variables measured by the sensors at 15-minute intervals. Historical data were stored in a database (DB).

Data pre-processing:

To facilitate its organization and analysis, the data were classified according to the days of the week because of the users' energy consumption behavior. For this study, Monday through Friday was considered, while Saturdays and Sundays were omitted because of their low energy consumption because there was little academic activity in the building on those days. In this phase, cleaning, filtering, and standardization techniques were applied to ensure the quality and suitability of the model as an input.

Design and implementation of the GAN model:

The GAN was implemented based on a deep neural network architecture comprising a generator and a discriminator. The generating neural network is responsible for generating new energy-consumption curves from random noise. Its goal is to produce synthetic data that resemble actual data. The discriminating neural network receives both real and generator-generated data, and learns to distinguish between them. Its function is to improve the quality of the generated data progressively. Both networks were trained in a competitive environment, where the generator attempts to deceive the discriminator and the latter strives to improve its detection capacity, that is, both networks are simultaneously optimized to improve the accuracy of the simulations. According to Figure 1:

a) **Generative network:** A generative neural network introduces an input signal (represented as a wave). It processes this data and generates an output (shown as a graph with a linear trend).

b) **Discriminating network:** The generated images are compared with real images to evaluate the quality of the generation. A discriminator classifies data as "real" or "false".

c) **Integration of the GAN model:** This section describes the training process of a GAN network composed of a generator and a discriminator. Noise is generated as the input, the generator creates synthetic data, and the discriminator evaluates whether the data is real or false. Backpropagation was applied to improve GAN's performance.

Training and validation:

The GAN model was trained in a competitive scheme to progressively improve the quality of the simulations. The proportion of 80% for training data and 20% for validation data was considered. Performance evaluation:

In this phase, the datasets generated by the GANs were compared with the original dataset to evaluate their performance in the synthetic generation of energy curves. To analyze the performance of the GANs and ensure their validity and robustness, the following metrics were considered [18]:

a) *Coefficient of determination (R^2):* Assesses how well the model captures variability in the data.

b) *Mean Square Error (MSE):* quantifies the average difference between the model's predictions and the actual values.

c) *Accuracy (acc):* measures the model's ability to generate synthetic data that is indistinguishable from real data.

IV. RESULTS AND DISCUSSION

Once the methodology proposed in Figure 1 is considered, the GAN architecture is applied and synthetic data are generated. Figure 2a) shows a set of bell-like curves of the energy consumption (total active power, TAP) of a day or 24 h for the FCMF-UG building. The same graph shows the curves obtained with the GANs, whose routes are horizontal and bounded between 55 kW and 90 KW. From this figure, it is evident that good results were not obtained with the GANs; they were not able to learn and generalize the pattern and behavior of the TAP variable of the FCMF-UG building and its bell shape. The GANs were not good at generating synthetic data from actual measurements in the low KW intervals of the TAP curve.

On the other hand, Figure 2b) shows a set of temperature data in the same time window compared to those generated by the GANs, and there is evidence of a better trajectory and capture of patterns compared to the previous case. Table II corroborates these results through the R^2 , MSE, and Accuracy

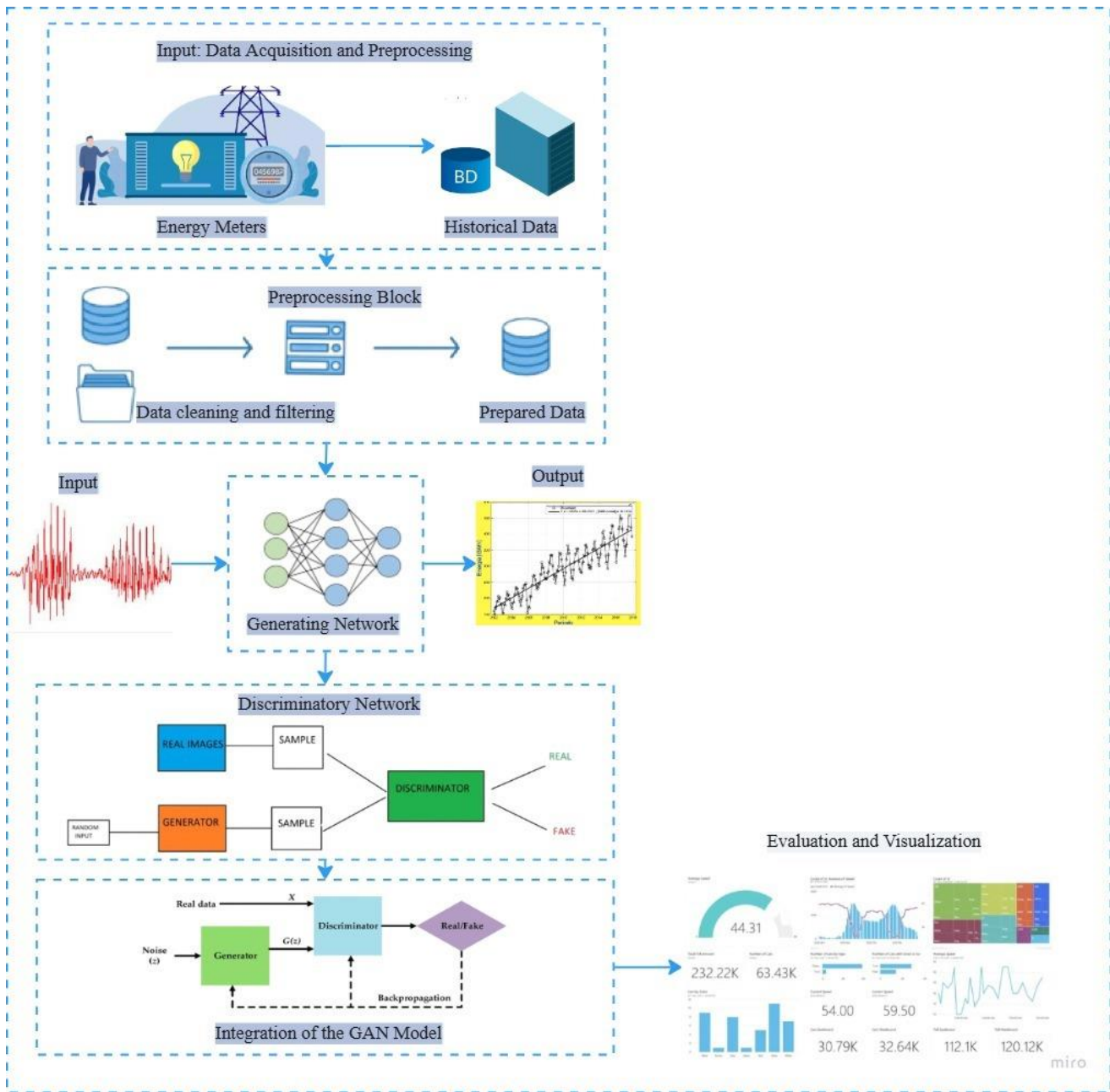


Fig. 1. Methodological scheme considering GANs

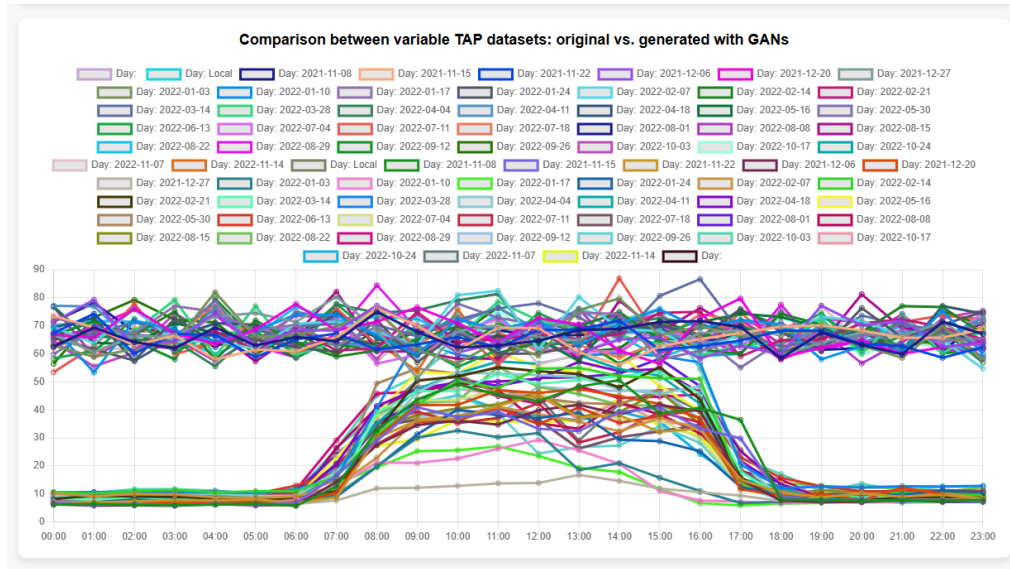
metrics, whose values can be further improved. That is, the GANs in this case study show some limitations of synthetic data generation in intervals with significant linearity changes, as observed with the TAP variable.

GANs have been widely applied to time-series data generation, but they often face challenges that lead to suboptimal results. These challenges include convergence, information loss, instability, and difficulty in capturing complex temporal dependencies. Despite advancements, there are notable cases where GANs have not achieved satisfactory outcomes in generating time-series data, particularly in

financial and industrial contexts [19]–[21]. This highlights the need for continued research and development to overcome these limitations and improve the efficacy of GANs for time-series data generation.

V. CONCLUSIONS

This study demonstrates the influence of original data variability on the results of curve generation using GANs. The case study, which involved generating a dataset from the



patterns. By overcoming these challenges, the full potential of GANs for advancing energy efficiency can be realized.

Table II: Results obtained through GAN applied to the dataset of the temperature variable

Day	Metric	
Monday	R ²	0.8097813646284659
	MSE	22.536538467180776
	ACC	0.588095238095238
Tuesday	R ²	0.6745009433962108
	MSE	20.478232016949406
	ACC	0.861904761904762
Wednesday	R ²	0.578017150186836
	MSE	23.10693078500037
	ACC	0.4190476190476190
Thursday	R ²	0.8862053413180912
	MSE	18.32590658291022
	ACC	0.5476190476190477
Friday	R ²	0.4688749219778885
	MSE	24.598955352769806
	ACC	0.5690476190476192

REFERENCES

- [1] I. Goodfellow *et al.*, “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
- [2] L. A. Cuenca Alava, “Aplicación de las redes inteligentes para el aumento de la eficiencia energética en las redes eléctricas,” *Rev. Investig. en Energía, Medio Ambient. y Tecnol. RIEMAT ISSN 2588-0721*, vol. 1, no. 1, p. 49, Jan. 2016, doi: 10.33936/riemat.v1i1.209.
- [3] L. Calcagni, “Redes Generativas Antagónicas y sus aplicaciones,” Universidad Nacional de La Plata, 2020.
- [4] R. Montero Laurencio, O. Torres Breff, S. Marrero Ramirez, and D. Jiménez jiménez, “Predicción de Consumo y Demanda de Electricidad Mediante Redes Neuronales Artificiales y Algoritmo Iterativo,” *Rev. Politécnica*, vol. 54, no. 3, pp. 45–58, Dec. 2024, doi: 10.33333/rp.vol54n3.05.
- [5] D. A. Kazacos Winter Pais de Brito, “Predicción de Consumo Energético en Edificios de Oficinas mediante Modelos de Tipo Caja Negra,” Universidad Politécnica de Madrid, 2017.
- [6] Y. Zhang, A. Schlueter, and C. Waibel, “SolarGAN: Synthetic annual solar irradiance time series on urban building facades via Deep Generative Networks,” *Energy AI*, vol. 12, no. October 2022, p. 100223, Apr. 2023, doi: 10.1016/j.egyai.2022.100223.
- [7] G. Baasch, G. Rousseau, and R. Evins, “A Conditional Generative adversarial Network for energy use in multiple buildings using scarce data,” *Energy AI*, vol. 5, p. 100087, Sep. 2021, doi: 10.1016/j.egyai.2021.100087.
- [8] L. Ortega-Díaz, J. Cárdenas-Rangel, and G. Osma-Pinto, “Estrategias de predicción de consumo energético en edificaciones: una revisión,” *Tecnológicas*, vol. 26, no. 58, p. e2650, Sep. 2023, doi: 10.22430/22565337.2650.
- [9] M. Yu, L. Li, and Z. Guo, “Model analysis of energy consumption data for green building using deep learning neural network,” *Int. J. Low-Carbon Technol.*, vol. 17, no. January, pp. 233–244, Feb. 2022, doi: 10.1093/ijlct/ctab100.
- [10] M. Labiadh, “Methodology for construction of adaptive models for the simulation of energy consumption in buildings,” Université de Lyon, 2021.
- [11] D. H. Choi, W. Li, and A. Y. Zomaya, “Enhancing Building-Integrated Photovoltaic Power Forecasting with a Hybrid Conditional Generative Adversarial Network Framework,” *Energies*, vol. 17, no. 23, p. 5877, Nov. 2024, doi: 10.3390/en17235877.
- [12] G. Li, Y. Wu, S. Yoon, and X. Fang, “Comprehensive transferability assessment of short-term cross-building-energy prediction using deep adversarial network transfer learning,” *Energy*, vol. 299, no. February, p. 131395, Jul. 2024, doi: 10.1016/j.energy.2024.131395.
- [13] Y. Gao, Z. Hu, W.-A. Chen, and M. Liu, “Solutions to the insufficiency of label data in renewable energy forecasting: A comparative and integrative analysis of domain adaptation and fine-tuning,” *Energy*, vol. 302, no. March, p. 131863, Sep. 2024, doi: 10.1016/j.energy.2024.131863.
- [14] Y. Liu, Z. Liang, and X. Li, “Enhancing Short-Term Power Load Forecasting for Industrial and Commercial Buildings: A Hybrid Approach Using TimeGAN, CNN, and LSTM,” *IEEE Open J. Ind. Electron. Soc.*, vol. 4, no. May, pp. 451–462, 2023, doi: 10.1109/OJIES.2023.3319040.
- [15] Y. Zhang, Z. Ma, J. Wang, Y. Shao, and J. Kang, “Local Data Augmentation of Building Electricity Load Prediction in Time-Series Generative Adversarial Network to Improve Computing Speed and Accuracy,” pp. 1–8, 2024, doi: 10.2139/ssrn.4715208.
- [16] V. Macas-Espinosa, L. Vera-Rodríguez, and J. Barzola-Monteses, “Analysis and Evaluation of Energy Efficiency in Buildings Based on Building Information Modeling,” vol. 7, no. 1, 2022, pp. 137–149.
- [17] J. Barzola-Monteses, F. Parrales-Bravo, V. Macas-Espinosa, G. Bórquez-Vargas, and M. Espinoza-Andaluz, “Energy Consumption of a Building Using the XGBoost Algorithm: A Forecasting Study,” in *2024 43rd International Conference of the Chilean Computer Science Society (SCCC)*, Oct. 2024, pp. 1–7, doi: 10.1109/SCCC63879.2024.10767630.
- [18] J. Barzola-Monteses, J. Gómez-Romero, M. Espinoza-Andaluz, and W. Fajardo, “Time series forecasting techniques applied to hydroelectric generation systems,” *Int. J. Electr. Power Energy Syst.*, vol. 164, p. 110424, Mar. 2025, doi: 10.1016/j.ijepes.2024.110424.
- [19] S. Kwon and Y. Lee, “Can GANs Learn the Stylized Facts of Financial Time Series?,” in *Proceedings of the 5th ACM International Conference on AI in Finance*, Nov. 2024, pp. 126–133, doi: 10.1145/3677052.3698661.
- [20] Y. Wang, X. Xu, L. Hu, J. Fan, and M. Han, “A time series continuous missing values imputation method based on generative adversarial networks,” *Knowledge-Based Syst.*, vol. 283, p. 111215, Jan. 2024, doi: 10.1016/j.knosys.2023.111215.
- [21] L. Ren, H. Wang, and Y. Laili, “Diff-MTS: Temporal-Augmented Conditional Diffusion-Based AIGC for Industrial Time Series Toward the Large Model Era,” *IEEE Trans. Cybern.*, vol. 54, no. 12, pp. 7187–7197, Dec. 2024, doi: 10.1109/TCYB.2024.3462500.