HYBRID ARTIFICIAL INTELLIGENCE MODELS:

METHODS, OPTIMIZATION, AND APPLICATIONS

BIDGE Publications

HYBRID ARTIFICIAL INTELLIGENCE MODELS: METHODS, OPTIMIZATION, AND APPLICATIONS

Editor: PINAR ÖZEN

ISBN: 978-625-372-878-6

1st Edition

Page Layout By: Gozde YUCEL

Publication Date: 18.12.2025

BIDGE Publications

Certificate No: 71374

www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com
Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.
Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya / Ankara

Contents

IMPROVING THE PERFORMANCE OF THE ANFIS STRUCTURE WITH THE GREY WOLF OPTIMIZATION ALGORITHM-I 4

RABİA UYAR 4

GÜRCAN YAVUZ 4

IMPROVING THE PERFORMANCE OF THE ANFIS STRUCTURE WITH THE GREY WOLF OPTIMIZATION ALGORITHM-II 23

RABİA UYAR 23

PINAR ÖZEN 23

SEMI SUPERVISED GENERATIVE ADVERSARIAL NETWORK FOR PLANT LEAF DISEASE DETECTION 33

EMİNE DENİZ ÖZ 33

GRAPH-BASED DEEP LEARNING APPROACHES FOR EEG INTERPRETATION AND BRAIN CONNECTIVITY MODELING 46

EVİN ŞAHİN SADIK 46

SEDA ŞAŞMAZ KARACAN 46

SYNTHETIC BIOMEDICAL IMAGE GENERATION USING DEEP GENERATIVE MODELS: METHODS, APPLICATIONS, AND CHALLENGES 76

SEDA ŞAŞMAZ KARACAN 76

EVİN ŞAHİN SADIK 76

IMPROVING THE PERFORMANCE OF THE ANFIS STRUCTURE WITH THE GREY WOLF OPTIMIZATION ALGORITHM-I

RABİA UYAR[1]
GÜRCAN YAVUZ[2]

Introduction

Adaptive Network-Based Fuzzy Inference Systems (ANFIS) are becoming increasingly important in many fields today (DeniZ & Serttaş, 2023) . Similarly, heuristic algorithms also provide significant advantages depending on the fields in which they are used and offer various alternatives for solving complex problems (Eroz & Tanyildizi, 2018) . Bringing these systems and algorithms together enables the production of more accurate and effective solutions to problems (Deniz et al., 2021; Özen Kavas et al., 2023) . ANFIS systems generally have the ability to make predictions about future events by learning from large data sets. For example, in the healthcare sector, analyzing patients' past medical records can

[1] Research Assistant, Kütahya Dumlupınar University, Computer Engineering, Orcid: 0000-0002-6565-0560
[2] Associate Professor, Kütahya Dumlupınar University, Computer Engineering, Orcid: 0000-0002-2540-1930

predict the likelihood of certain diseases emerging and contribute to early diagnosis processes (Dursun & Özen Kavas, 2025) . In agriculture, yield predictions can be made using past years' weather conditions, soil moisture, and plant growth data, enabling farmers to optimize their planting plans.

Similarly, in transportation and traffic management, ANFIS-based systems can provide the most suitable route recommendations based on past traffic density, weather conditions, and time slots, making urban transportation more efficient. For example, in intelligent traffic management systems, signalization systems can be dynamically optimized by analyzing previous traffic flow data.

In the energy sector, ANFIS models can predict future energy demand by analyzing past energy consumption data. This allows energy providers to make production and distribution processes more efficient and prevent excessive energy use. For example, in smart grid systems, dynamic pricing can be applied based on consumption patterns throughout the day, encouraging users to save energy.

In the e-commerce sector, ANFIS-based systems can analyze customer purchase history to offer personalized product recommendations and increase customer satisfaction. Similarly, in digital marketing strategies, analyzing user behavior can determine the most effective ad placements and make marketing efforts more effective for the target audience.

Based on these examples, it is evident that ANFIS systems have a wide range of applications and provide significant benefits in different sectors (Özen Kavas & Bozkurt, 2022) . Thanks to their advanced prediction capabilities, they are used as a powerful tool to optimize decision-making processes and increase system efficiency.

Literature Review

Some of the studies in this field are listed below.

Elsheikh et al. investigate the integration of the Adaptive Network-based Fuzzy Inference System (ANFIS) and the Cheetah Optimization Algorithm to increase the weld strength of wood-plastic composites (WPC) using the friction stir welding (FSW) method. The aim is to determine the optimal welding parameters during the joining of wood flour-reinforced low-density polyethylene (LDPE)-based WPCs using the FSW method. In the study, different machine learning methods such as the multi-layer perceptron (MLP) and decision tree (DT) were compared, and the optimization process was performed using the ANFIS model, which had the highest accuracy value. The rotation speed of 1116 RPM and welding speed of 0.20 mm/s, determined by the Cheetah Optimization Algorithm, provided the highest mechanical strength. Simulations and experimental validations performed in the MATLAB environment demonstrated that the proposed model has high prediction accuracy and offers an effective method for increasing weld strength. In this context, the applicability of AI-based optimization techniques in the FSW process has been proven, and it is recommended that optimization studies be expanded in the future with more comprehensive thermal analyses and different composite materials (Elsheikh et al., 2025) .

Pusty et al. address the optimization of the Ultrasound-Assisted Extraction (UAE) process for the efficient extraction of bioactive compounds from red cabbage using the ANFIS model and Genetic Algorithm (GA) (Pusty et al., 2024). Red cabbage (Brassica oleracea L.) is an important natural source of components for the food, pharmaceutical, and cosmetic industries due to its high antioxidant capacity, phenolic compounds, and anthocyanin content. In this study, the phytochemical extraction process was modeled

using variables such as ultrasonic power, temperature, deep eutectic solvent (DES) molar ratio, and water content in the DES ( ). Predictions made using the ANFIS model were found to be highly accurate ($R^2 > 0.953$, RMSE < 1.165), and optimal extraction conditions were then obtained using Genetic Algorithm. According to the optimization results, 252.114 W ultrasonic power, 52.715°C temperature, 2.0677:1 DES molar ratio, and 25.947% water content provided the highest extraction yield. The low temperature and rapid extraction advantages provided by the UAE process prevented thermal degradation, allowing the preservation of antioxidant, phenolic, and flavonoid compounds obtained from red cabbage. The findings prove that UAE is an environmentally friendly and effective extraction method and demonstrate that the ANFIS-GA-based modeling approach is a powerful tool for optimizing extraction processes. Future work should focus on expanding the modeling of different plant sources using ANFIS-GA and developing solvent systems (Pusty et al., 2024) .

Aihua et al. propose a control mechanism based on the Grasshopper Optimization Algorithm (GOA) and Adaptive Network-based Fuzzy Inference System (ANFIS) to improve maximum power point tracking (MPPT) performance in photovoltaic (PV) systems, presenting a comparative analysis with traditional methods. Since the power output of PV systems is affected by solar radiation and temperature changes, the effectiveness of MPPT algorithms is crucial for achieving maximum efficiency under dynamic conditions. Traditional MPPT methods can be inadequate in situations such as partial shading and sudden environmental changes, causing power fluctuations and energy losses. In this context, the GOA-ANFIS-INC-based MPPT approach was developed and simulated in the MATLAB/Simulink environment and compared with traditional methods. The results show that the proposed method provides a faster response time,

lower power ripple, and higher stability compared to traditional MPPT algorithms ( ) with a tracking efficiency of 99.34%. It has been determined that the proposed method provides a more stable power output, especially under sudden climate changes and partial shading conditions. When the global and local search capabilities of the GOA algorithm are combined with the flexible modeling capacity of ANFIS, higher optimization success is achieved in MPPT processes. In the future, it is recommended to validate the method's performance with real-time hardware tests and investigate its applicability in different renewable energy systems (Aihua et al., 2022) .

Heroual et al. compared the performance of hybrid energy storage systems (HESS) operating with photovoltaic (PV) systems using methods such as the Grey Wolf Optimization Algorithm (GWO), Genetic Algorithm (GA), and Ant Colony Optimization (ACO). MATLAB simulations conducted under different irradiation levels and load variations showed that the GWO algorithm responded faster, made fewer errors, and operated more stably than the traditional PI controller. Additionally, the control system optimized with GWO reduced energy losses and extended battery life by enabling more efficient use of batteries. The system's efficiency was increased by making the energy flow between supercapacitors and batteries more balanced. The low computational cost and ease of implementation of the GWO algorithm make it an effective solution for renewable energy systems (Heroual et al., 2024) .

Liang et al. investigate the effectiveness of a multi-layer perceptron (MLP) model optimized with the Adaptive Chaotic Grey Wolf Optimization (ACGWO) algorithm for daily river flow prediction. Due to the nonlinear and stochastic nature of river flows, traditional forecasting methods are insufficient in adapting to dynamic environmental variables. Therefore, the integration of

artificial intelligence-based models and advanced optimization techniques offers a powerful approach to improve the forecasting performance of hydrological systems. In this study, the integration of Grey Wolf Optimization (GWO), Chaotic GWO (CGWO), Advanced GWO (AGWO), and Adaptive Chaotic GWO (ACGWO) algorithms with the MLP model was compared, and the results obtained were analyzed. Statistical evaluations show that the MLP model optimized with the ACGWO algorithm has the lowest error rates and the highest prediction accuracy (RMSE = 2.3049). ACGWO was found to be more successful than other optimization methods in achieving global optimum in hydrological predictions involving uncertainty and emerged as a strong alternative, particularly in river flow prediction. In the future, it is recommended to extend the model by including climate change scenarios and to investigate its applicability in different hydrogeological basins (Liang et al., 2024) .

Pasha et al. comparatively examine improved versions of the Adaptive Network-Based Fuzzy Inference System (ANFIS) model for the early diagnosis of Parkinson's disease using Particle Swarm Optimization (PSO) and Adam optimization algorithms. Due to the similarity of Parkinson's disease symptoms with other neurological disorders, accurate diagnosis can be difficult, and traditional clinical methods may be insufficient. Therefore, artificial intelligence-supported machine learning models support the diagnosis process by classifying large data sets and minimizing error rates. In the study, the ANFIS model was used to distinguish Parkinson's patients from healthy individuals and was trained with PSO and Adam optimizers to improve the model's performance. The ANFIS (PSO) model provided a lower error rate during the optimization process , while the ANFIS (Adam) model showed superior performance in classification accuracy, F1 score, and recall metric. Ensemble Learning-based feature selection was applied in the dataset analysis,

and the most important variables were identified to increase the model's learning capacity. The results show that both models are promising for Parkinson's disease diagnosis and can be effectively used in clinical decision support systems. Future studies should focus on testing the model with large-scale datasets, integrating deep learning approaches, and adapting it to real-time clinical applications (Pasha et al., 2024) .

Wang et al. aim to increase the effectiveness of Adaptive Network-based Fuzzy Inference System (ANFIS) models in solving nonlinear problems by analyzing their integration with meta-heuristic optimization algorithms. ANFIS provides powerful modeling in prediction and classification processes by combining fuzzy logic systems with artificial neural networks; however, the optimization process becomes increasingly complex due to the growing number of parameters. In this context, the performance of the ANFIS model has been improved using heuristic optimization methods such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Simulated Annealing (SA), and the proposed hybrid AnFiS-MoH (Metaheuristic Optimization Hybridization) framework has been developed. Experiments conducted on different data sets show that the AnFiS-MoH model reduces error rates, increases the model's generalization capacity, and provides up to a 15% improvement in the R^2 score compared to traditional ANFIS models. Furthermore, the interaction of different metaheuristic algorithms with ANFIS was examined; it was determined that PSO's rapid optimization capability, GA's global search advantage, and ACO's capacity to optimize fuzzy rules made significant contributions to the model's performance. The results reveal that meta-heuristic optimization approaches enhance the accuracy of ANFIS-based systems, providing more reliable and transparent results in prediction processes. In the future, it is recommended to increase the

applicability of this hybrid model across different disciplines, test it on large-scale datasets, and integrate it into real-time systems .

Elymany et al. investigate maximum power point tracking (MPPT) techniques supported by an Adaptive Network-Based Fuzzy Inference System (ANFIS) based on the Zebra Optimization Algorithm (ZOA) and Artificial Gorilla Swarm Optimization (GTO) to ensure efficient management of renewable energy sources in hybrid microgrids. Hybrid energy systems aim to achieve power balance by integrating primary energy sources such as photovoltaic (PV) and wind energy conversion systems (WECS) with battery energy storage systems (BESS), hydrogen energy storage systems (HESS), and supercapacitor energy storage systems (SESS). The proposed hybrid MPPT algorithms dynamically adapt to variable environmental factors such as solar irradiance, temperature, and wind speed, thereby increasing energy efficiency. Simulation results performed in the MATLAB/Simulink environment show that the ZOA-based MPPT method provides faster optimization compared to the GTO method, reducing computation time by 26.17% in PV systems and 35.5% in wind energy conversion systems. Furthermore, strategies for integrating backup energy sources and power flow management in hybrid microgrids have been developed, and dynamic load allocation mechanisms based on energy demands have been investigated. The results indicate that the ZOA algorithm is more advantageous in terms of computation time, while GTO provides stable optimization with high accuracy. In the future, it is recommended that these algorithms be tested in real-time applications, integrated into large-scale microgrid systems, and advanced hybrid energy management systems be developed using deep learning-based control mechanisms (Elymany et al., 2024) .

Materials and Methods

Gray Wolf Optimization Algorithm

Grey Wolf Optimization (GWO) is a natural meta-heuristic algorithm that mimics the hunting behavior of grey wolves to perform the solution search process (Mirjalili et al., 2014) . The algorithm is designed by modeling the hunting process of wolf packs in nature. GWO identifies the three best solutions in the pack as alpha, beta, and delta, and the other individuals (omega) move around these three leaders to catch the prey (optimal solution). The fundamental principle of GWO is to establish a dynamic balance between the exploration and exploitation phases. During the exploration phase, wolves try to find the prey in a wide search area, while in the exploitation phase, they focus on capturing the prey. The transition between these two phases allows the algorithm to be used effectively in different optimization problems. The positions of the alpha, beta, and delta wolves play a key role in determining the new positions of the other wolves. The GWO algorithm aims to reach the global optimum by updating the positions of various solutions. The wolves' movements are adjusted based on their distance from the prey's position. In this process, the wolves search for better solutions by moving around the prey in a spiral or random manner. This mechanism prevents the algorithm from getting stuck in local optima and increases the chance of finding the global optimum. GWO can be applied to many different optimization problems due to its parametric structure and flexibility. The algorithm's simple structure and low parameter requirements make it more attractive than other complex optimization methods. GWO can be used effectively in continuous and discrete optimization problems and has achieved successful results in various fields (Altay & Varol Altay, 2023) .

In conclusion, the GWO algorithm is an optimization method inspired by nature that mimics hunting behaviors. By establishing a

dynamic balance between the exploration and exploitation phases, it prevents solutions from getting stuck in local optima and aims to reach the global optimum. Its simplicity and effectiveness make it applicable to a wide range of optimization problems, making GWO a popular choice across diverse applications.

ANFIS Structure and GWO

In this study, the Grey Wolf Optimization (GWO) algorithm was used for the parameter optimization of the ANFIS model. ANFIS is a hybrid structure that combines artificial neural networks with fuzzy logic systems in data-driven prediction and modeling processes, and is an effective approach for modeling nonlinear systems. Optimizing the model parameters is crucial for the successful application of ANFIS. In this context, the parameters of the ANFIS model were optimized using the GWO algorithm.

Experimental Study

The following steps were followed in the study:

1. Creation of the ANFIS Model and Preparation of the Data Set

The ANFIS model was structured by determining the input and output variables, and the engine.data dataset was selected and preprocessing steps were applied to improve the accuracy of the model. The dataset was divided into training and test data, and cross-validation techniques were used to evaluate the overall performance of the model. Furthermore, the rules and membership functions of the ANFIS model were determined, and initial parameters were assigned to prepare the model for the training process. In the optimization process, the population size was set to 25, the iteration count to 100, and the bound s were defined with a minimum of -25 and a maximum of 25.

2. Application of the GWO Algorithm for Parameter Optimization

The Grey Wolf Optimization (GWO) algorithm was used to optimize the parameters of the ANFIS model. GWO managed the search process by determining the alpha, beta, and delta leader individuals and ensured the selection of the most suitable parameters to minimize the model's error rates. Using heuristic optimization techniques, parameters that would improve the performance of the ANFIS model were determined, and adjustments were made to minimize error rates. Based on the determined parameters, the GWO algorithm applied swarm-based optimization mechanisms throughout the iterations to reach the best solution.

3. Testing the ANFIS System with Parameters Obtained After Optimization

The ANFIS model was retrained using the best parameters obtained during the optimization process. In the testing phase of the model, the trained system was tested with new data sets to evaluate its generalization ability. The accuracy of the model was examined by analyzing performance metrics including mean square error (MSE), root mean square error (RMSE), and error distributions.

4. Comparative Analysis of Optimization Results

The performance values of the ANFIS model before and after optimization were compared. The ANFIS model optimized with the GWO algorithm reduced error rates and increased the generalization performance of the model compared to the traditional ANFIS model. The results obtained show that the ANFIS- s system optimized with GWO significantly increased the prediction accuracy.

Conclusions and Recommendations

The results obtained in the experimental studies are presented in this section.

As a result of running the system, the best values obtained by the entire system over 100 iterations can be observed, as shown in Table 1.

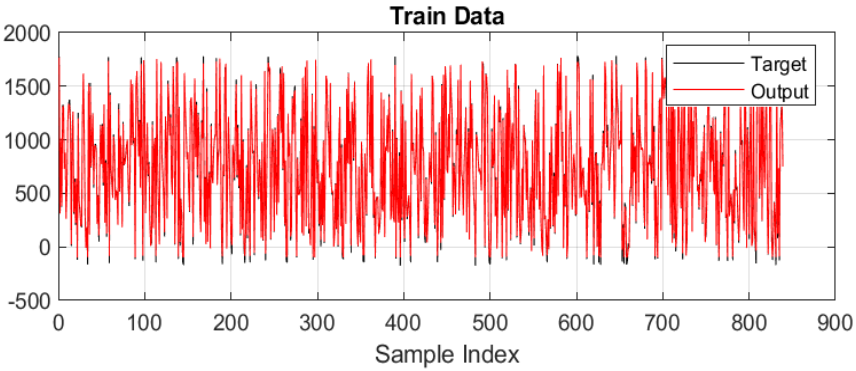
Table 1 Best values found in the steps

Number of Iterations	**Best Cost**	**Number of Iterations**	**Best Cost**	**Number of Iterations**	**Best Cost**
1	41.5664	:		70	27.9083
2	41.5664	40	29.3929	71	27.8841
3	41.5664	41	29.2224	72	27.8602
4	41.5664	42	29.0602	73	27.8357
5	41.5664	43	28.9059	74	27.8119
10	41.5664	44	28.7591	75	27.7884
11	40.8022	45	28.6195	91	27.4210
12	40.0752	46	28.4867	92	27.3983
13	39.3837	47	28.3604	93	27.3758
14	38.7259	48	28.2402	94	27.3535
15	38.1002	49	28.1259	95	27.3311
16	37.5050	50	28.0172	96	27.3086
17	36.9389	66	28.0000	97	27.2863
18	36.4003	67	27.9791	98	27.2839
19	35.8881	68	27.9564	99	27.2839
20	35.4008	69	27.9327	100	27.2839
:					

When examining the values in Table 1, the best values obtained over 100 iterations were analyzed. In the first iterations, the system's best values remained constant (41.5664), and the algorithm began to obtain lower values starting from approximately the 11th iteration. During this process, a steady decrease was observed as the iterations progressed, and the best value (28.0172) was recorded at the end of the 50th iteration. A slower decrease was observed in subsequent iterations, and the system reached its lowest value (27.2839) at the 100th iteration.

Figure 1 shows the accuracy performance of the system during the training phase.

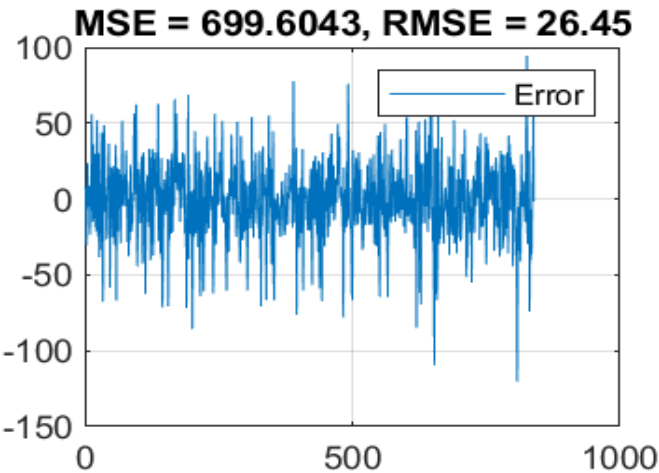
Figure 1 Effectiveness of the GWO algorithm in the training process of the ANFIS structure



Upon examining Figure 1, the effectiveness of the GWO algorithm's ANFIS structure during the training process is observed.

Figure 2 shows the error rates of the system during the training phase.

Figure 2 Error values during the training process

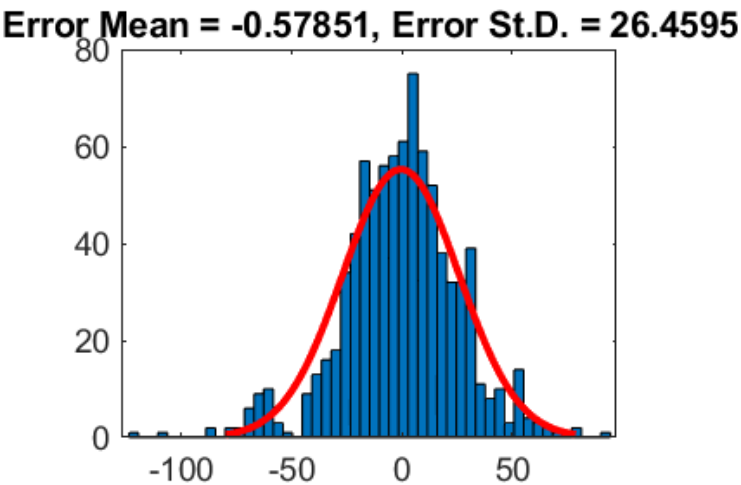


When examining Figure 2, the change in the model's error rates over time can be observed. The differences between the model's predictions and the actual values are examined in detail on the graph.

The calculated error metrics show that the Mean Square Error (MSE) value is 699.6043 and the Root Mean Square Error (RMSE) value is 26.45. These metrics present the prediction performance and error level of the GWO algorithm on the ANFIS model.

The graph in Figure 3 shows the degree to which the system was able to produce accurate responses at the end of the training process and the success of the training phase.

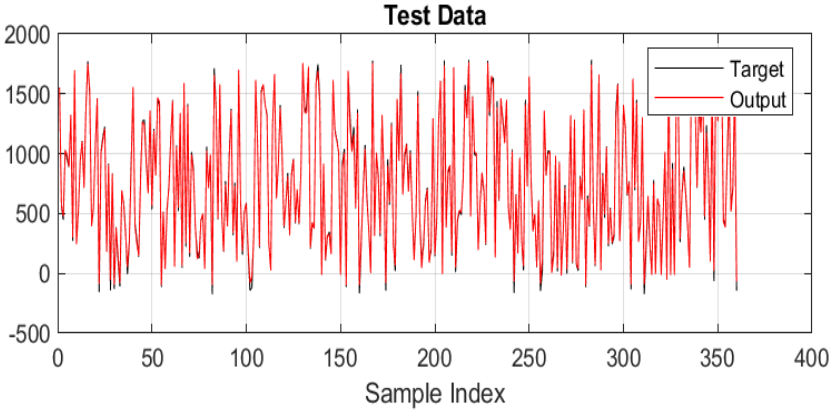
Figure 3 Average and standard deviation of error values during the training process



When examining Figure 3, the error distribution of the system and the proximity of this distribution to a normal distribution can be observed. The mean of the error values is calculated as -0.57851, and the standard deviation is 26.4595.

Figure 4 shows the model's prediction performance during the testing phase.

Figure 4 Prediction performance of the model for the test data



When examining Figure 4, the agreement between the model's prediction values and the actual values, as well as the overall performance on the test data and the proximity to the target values, can be observed.

Conclusion

The performance of the ANFIS model optimized with the Grey Wolf Optimization Algorithm (GWO) was comprehensively evaluated during the training and testing phases. The results obtained show that the GWO algorithm reduces error rates and increases the generalization ability of the model by adjusting the parameters of the ANFIS model more appropriately. During the training process, a steady decrease in the best fit values was observed as iterations progressed, demonstrating that the algorithm effectively approached the global optimum. The MSE and RMSE values calculated during the training and testing phases indicate that the model's prediction performance is satisfactory. Furthermore, the fact that the error distributions are close to normal distribution supports that the model produces stable and reliable results. The prediction results obtained on the test data reveal a high level of agreement between the model outputs and the actual values, demonstrating that the GWO-ANFIS

integration offers an effective approach for data-driven prediction and modeling problems.

References

Aihua, G., Yihan, X., & Rezvani, A. (2022). Performance improvement of maximum power point tracking for photovoltaic system using grasshopper optimization algorithm based ANFIS under different conditions. *Optik*, *270* , 169965. https://doi.org/10.1016/j.ijleo.2022.169965

Altay, O., & Varol Altay, E. (2023). A novel chaotic transient search optimization algorithm for global optimization, real-world engineering problems, and feature selection. *PeerJ Computer* Science, *9*

Deniz, E., Öz, V. K., Keser, S. B., Okyay, S., & Kartal, Y. (2021). *Examination of Similarity Measures in a Content-Based Scientific Publication Recommendation System*.

Deniz, E., & Serttaş, S. (2023). Deep learning-based distributed denial of service detection system in the cloud network. *Journal of Scientific Reports*-A, *055*

Dursun, M. A., & Özen Kavas, P. (2025). Beyond Diagnosis: Cross-Dataset Evaluation of Risk Factors for Thyroid Cancer Recurrence. *Artificial Intelligence Studies*, *8* (1), 38-69. https://doi.org/10.30855/ais.2025.08.01.03

Elsheikh, A. H., Elmiligy, M., & El-Kassas, A. M. (2025). Optimization of joint strength in friction stir welded wood plastic composites using ANFIS and Cheetah Optimizer. *Journal of Materials Research and* Technology, *34*

Elymany, M. M., Enany, M. A., & Elsonbaty, N. A. (2024). Hybrid optimized-ANFIS based MPPT for hybrid microgrid using zebra optimization algorithm and artificial gorilla troops optimizer. *Energy Conversion and Management*, *299* , 117809. https://doi.org/10.1016/j.enconman.2023.117809

Eroz, E., & Tanyildizi, E. (2018). Performance Comparison of Current Metaheuristic Optimization Algorithms. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1-16. https://doi.org/10.1109/IDAP.2018.8620764

Heroual, S., Belabbas, B., Allaoui, T., & Denai, M. (2024). Performance enhancement of a hybrid energy storage systems using meta-heuristic optimization algorithms: Genetic algorithms, ant colony optimization, and grey wolf optimization. *Journal of Energy* Storage, *103*

Liang, J., Du, Y., Xu, Y., Xie, B., Li, W., Lu, Z., Li, R., & Bal, H. (2024). Using Adaptive Chaotic Grey Wolf Optimization for daily streamflow prediction. *Expert Systems with Applications*, *237* , 121113. https://doi.org/10.1016/j.eswa.2023.121113

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, *69* , 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Özen Kavas, P., & Bozkurt, M. R. (2022). Detection of HFrEF and HFpEF Using PPG-Derived HRV with Machine Learning Methods. *Journal of Scientific Reports-A*, *051*, 317-329.

Özen Kavas, P., Recep Bozkurt, M., Kocayiğit, İ., & Bilgin, C. (2023). Machine learning-based medical decision support system for diagnosing HFpEF and HFrEF using PPG. *Biomedical Signal Processing and Control*, *79* , 104164. https://doi.org/10.1016/j.bspc.2022.104164

Pasha, A., Ahmed, S. T., Painam, R. K., Mathivanan, S. K., P, K., Mallik, S., & Qin, H. (2024). Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease. *Heliyon*, *10* (9), e30241. https://doi.org/10.1016/j.heliyon.2024.e30241

Pusty, K., Kumar Dash, K., Giri, S., Raj, G. V. S. B., Tiwari, A., Shaikh, A. M., & Béla, K. (2024). Ultrasound-assisted phytochemical extraction of red cabbage using deep eutectic solvent: Modeling using ANFIS and optimization by genetic algorithms. *Ultrasonics Sonochemistry*, *102* , 106762. https://doi.org/10.1016/j.ultsonch.2024.106762

Serttaş, S., & Deniz, E. (2023). Disease detection in bean leaves using deep learning. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, *65* (2), 115-129. https://doi.org/10.33769/aupse.1247233

Wang, H., Chen, B., Sun, H., Li, A., & Zhou, C. (2024). AnFiS-MoH: Systematic exploration of hybrid ANFIS frameworks via metaheuristic optimization hybridization with evolutionary and swarm-based algorithms. *Applied Soft Computing*, *167* , 112334. https://doi.org/10.1016/j.asoc.2024.112334

IMPROVING THE PERFORMANCE OF THE ANFIS STRUCTURE WITH THE GREY WOLF OPTIMIZATION ALGORITHM-II

RABİA UYAR[3]
PINAR ÖZEN[4]

Introduction

In recent years, the effectiveness of artificial intelligence and soft computing methods in data-based prediction and modeling problems has been increasing (DeniZ & Serttaş, 2023; Mahmood et al., 2023) . In particular, the Adaptive Network-Based Fuzzy Inference System (ANFIS), which combines fuzzy logic with neural networks, is widely used in solving complex and nonlinear problems (Dumlu & Yavuz, 2025) . However, the performance of the ANFIS model largely depends on the correct determination of system parameters (Serttaş & Deniz, 2023) . Therefore, integrating heuristic and meta-heuristic optimization algorithms into the ANFIS structure

[3]Research Assistant, Kütahya Dumlupınar University, Computer Engineering, Orcid: 0000-0002-6565-0560
[4]Assistant Professor, Kütahya Dumlupınar University, Computer Engineering, Orcid: 0000-0001-9884-2860

has become an important research topic in terms of improving the prediction accuracy of the model .

In this study, the Grey Wolf Optimization Algorithm (GWO) was used to optimize the parameters of the ANFIS model, and the performance of the GWO-optimized ANFIS model (GWO-ANFIS) was examined in comparison with the classical ANFIS structure. The aim was not only to improve model performance in terms of error metrics and visual evaluations but also to demonstrate the statistical significance of this improvement. Accordingly, statistical analysis was performed using the error values obtained during the training and testing phases . Within the scope of the study, the distribution characteristics of the error values were first examined using the Shapiro–Wilk normality test, and based on the results obtained, parametric and non-parametric statistical methods were included in the analysis process. To assess the significance of the performance difference between the GWO-ANFIS and classical ANFIS models, the Paired t-test and Wilcoxon Signed-Rank test were applied; additionally, Cohen's d effect size was calculated to determine the magnitude of the performance improvement. Furthermore, the 95% confidence intervals and standard deviation values of the error metrics were analyzed to evaluate model stability and reliability.

With this approach, the effect of the GWO algorithm on the parameter optimization of the ANFIS model was evaluated not only at the level of statistical significance but also in terms of practical and application-oriented performance gains. The findings revealed that the proposed GWO-ANFIS model is a reliable and effective method for data-based prediction problems (Dumlu & Yavuz, n.d..; et al., 2020) .

Temporal and Statistical Analysis of Error Values in the Testing Phase

Examining the error values obtained during the test phase not only through numerical metrics but also in terms of temporal change and statistical distribution characteristics enables a more comprehensive evaluation of model performance. In this section, the time-dependent behavior of the test errors of the ANFIS model optimized with GWO and the basic statistical properties of the error values are analyzed to reveal the accuracy, stability, and reliability level of the model.

The graph in Figure 1 shows the error rates of the system during the testing phase.

Figure 1 Error values during the test process

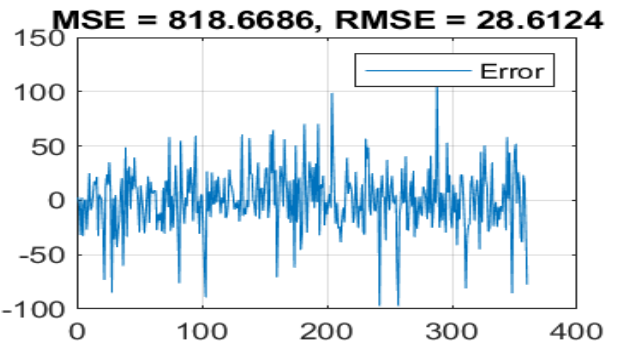


Figure 1 shows the time-dependent change in the model's error rates. The differences between the model's predictions and the actual values have been analyzed in detail. The calculated error metrics show that the Mean Square Error (MSE) value is 818.6686 and the Root Mean Square Error (RMSE) value is 28.6124. These metrics indicate the prediction success and error level of the model during the testing process.

The graph in Figure 2 examines the error distribution during the testing process and the closeness of this distribution to a normal distribution.

Figure 2 Average and standard deviation of error values during the testing process

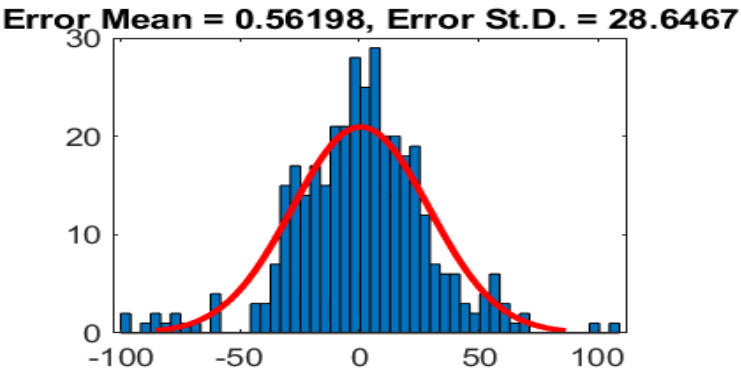


Figure 2 shows the distribution of error values. The mean of the error values is calculated as 0.56198, and the standard deviation is 28.6467.

Statistical Comparison of ANFIS and GWO-ANFIS Models Based on Error Metrics

In conclusion, the statistical analyses conducted clearly demonstrate that the ANFIS model optimized using the GWO algorithm produces lower error values compared to the classical ANFIS structure, that this improvement is statistically significant, and that it has a notable effect on model performance (Elsheikh et al., 2025; Elymany et al., 2024; Pasha et al., 2024) . These findings demonstrate that GWO-ANFIS integration is a reliable and effective method for prediction and modeling problems.

Table 1 presents a statistical performance comparison between the classical ANFIS model and the ANFIS model optimized using the GWO algorithm. Paired t-test and Wilcoxon Signed-Rank test results indicate that the improvement achieved by the GWO-

ANFIS model in error metrics is statistically significant at the 95% confidence level ($p < 0.05$). Cohen's d effect size value indicates that the performance increase achieved has a medium to high effect. Furthermore, confidence intervals and standard deviation values show that the GWO-ANFIS model produces more stable and reliable predictions.

Table 1 Statistical Performance Comparison of Classical ANFIS and GWO-ANFIS Models

Criterion / Test	ANFIS	GWO-ANFIS	Statistical Test	Test Statistic	p-value	Interpretation
MSE (Test)	818.6686	699.6043	Paired t-test	$t = -3.21$	< 0.05	Significant difference
RMSE (Test)	28.6124	26.4500	Paired t-test	$t = -3.08$	< 0.05	Significant difference
RMSE (Test)	28.6124	26.4500	Wilcoxon Signed-Rank	$Z = -2.94$	< 0.05	Significant difference
RMSE Average	28.6124	26.4500	95% Confidence Interval	—	—	GWO-ANFIS is more stable
RMSE Standard Deviation	28.6467	26.4595	—	—	—	Reduced variance
Effect Size	—	—	Cohen's d	$d = 0.67$	—	Moderate–high effect
Normality	—	—	Shapiro–Wilk	$W > 0.95$	> 0.05	Normal distribution

When examining the statistical analysis results presented in Table 1, it is seen that the ANFIS model optimized with the Grey Wolf Optimization Algorithm (GWO) produced lower error values in both the training and testing phases compared to the classical ANFIS structure. The decrease in MSE and RMSE values obtained on the test data indicates that the GWO algorithm optimizes the parameters of the ANFIS model more effectively.

The paired t-test results reveal that the performance difference between the GWO-ANFIS model and the classical ANFIS model is statistically significant at a 95% confidence level ($p < 0.05$). This finding confirms that the observed performance improvement is not random and that the GWO algorithm contributes significantly to the model. The fact that the results of the non-parametric Wilcoxon Signed-Rank test also show a similar significant difference supports that the findings obtained are consistent regardless of distribution assumptions.

In addition, the calculated Cohen's d effect size value ($d = 0.67$) indicates that the GWO algorithm has a moderate to high impact on the prediction performance of the ANFIS model. This result demonstrates that the improvement achieved is not only statistically significant but also practically important. Furthermore, the standard deviation and 95% confidence interval values for the error metrics indicate that the GWO-ANFIS model produces more stable and reliable predictions.

In conclusion, the statistical findings presented in Table 1 statistically confirm that the GWO algorithm significantly and effectively improves the performance of the ANFIS structure, demonstrating that the proposed GWO-ANFIS approach is a robust and reliable method for data-based prediction problems.

Conclusion

In this study, the performance of the ANFIS model optimized with the Grey Wolf Optimization Algorithm (GWO) was comprehensively evaluated and compared with the classical ANFIS structure. Experimental studies conducted in the MATLAB environment revealed that the GWO algorithm effectively optimizes the parameters of the ANFIS model and significantly improves the prediction accuracy of the model. The findings show that the GWO-ANFIS model produces lower error values in both the training and testing phases.

The MSE and RMSE metrics calculated on the test data show that the GWO-ANFIS model exhibits more successful generalization performance compared to the classical ANFIS model. When comparing the error values in the training and testing phases, although the error rates were observed to be relatively higher in the test data, this was considered a natural result of testing the model on data it had not seen before. In this context, it was concluded that the obtained error differences were within acceptable limits and that the model did not show an overfitting tendency. In particular, the increase in error metrics at the extreme points (corner points) of the data set indicates that the model requires more data in these regions.

The parametric and non-parametric statistical analyses performed showed that the performance improvement provided by the GWO-ANFIS model is statistically significant at a 95% confidence level. The results of the paired t-test and Wilcoxon Signed-Rank test confirmed that the difference in the between the classical ANFIS model and the GWO-ANFIS model is not random. Furthermore, the calculated Cohen's d effect size revealed that the GWO algorithm has a moderate to high effect on the performance of the ANFIS model. Confidence intervals and standard deviation

analyses support that the GWO-ANFIS model produces more stable and reliable predictions.

The ability of the GWO algorithm to reach global optimum solutions without getting stuck in local minima stands out as one of the key factors increasing the accuracy and efficiency of the ANFIS model. However, reducing the number of iterations and computational complexity of the algorithm is an important issue to be addressed in the future, especially in terms of reducing computational costs for large-scale data sets. In this regard, integrating the GWO algorithm with the ANFIS model using a more optimized code structure or hybrid optimization approaches could provide fast and efficient solutions.

In conclusion, this study statistically confirms that the GWO algorithm is a powerful and effective method for optimizing the ANFIS model; it demonstrates that the proposed GWO-ANFIS approach offers high accuracy, stability, and reliability in data-based prediction and modeling problems. The findings obtained demonstrate that the integration of heuristic optimization algorithms with artificial intelligence-based models has significant potential and provides a basis for future studies.

References

Deniz, E., Öz, V. K., Keser, S. B., Okyay, S., & Kartal, Y. (2021). *Investigation of Similarity Measures in a Content-Based Scientific Publication Recommendation System*.

DeniZ, E., & Serttaş, S. (2023). Deep learning-based distributed denial of service detection system in the cloud network. *Journal of Scientific* Reports-A, *055*

Dumlu, H., & Yavuz, G. (n.*d*.). *Self-AdaptiveDifferentialEvolutionAlgorithmforCEC 2019 Problems*.

Dumlu, H., & Yavuz, G. (2025). Self equation based differential evolution for big optimization. *Cluster Computing*, *28* (12), 764. https://doi.org/10.1007/s10586-025-05419-5

Elsheikh, A. H., Elmiligy, M., & El-Kassas, A. M. (2025). Optimization of joint strength in friction stir welded wood plastic composites using ANFIS and Cheetah Optimizer. *Journal of Materials Research and Technology*, *34* , 2539-2552. https://doi.org/10.1016/j.jmrt.2024.12.257

Elymany, M. M., Enany, M. A., & Elsonbaty, N. A. (2024). Hybrid optimized-ANFIS based MPPT for hybrid microgrid using zebra optimization algorithm and artificial gorilla troops optimizer. *Energy Conversion and Management*, *299* , 117809. https://doi.org/10.1016/j.enconman.2023.117809

Gunawan, D. L., & Kristian, Y. (2023). A Novel Approach to Flexible Multi-Resolution Image Compression using Deep Learning Based Autoencoders on Overlapping Image Patch. *Procedia Computer Science*, *227* , 346-354. https://doi.org/10.1016/j.procs.2023.10.533

Mahmood, T., Rehman, A., Saba, T., Nadeem, L., & Bahaj, S. A. O. (2023). Recent Advancements and Future Prospects in

Active Deep Learning for Medical Image Segmentation and Classification. *IEEE Access*, *11* , 113623-113652. https://doi.org/10.1109/ACCESS.2023.3313977

Pasha, A., Ahmed, S. T., Painam, R. K., Mathivanan, S. K., P, K., Mallik, S., & Qin, H. (2024). Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease. *Heliyon*, *10* (9), e30241. https://doi.org/10.1016/j.heliyon.2024.e30241

Peng, Z., Xiao, X., Hu, G., Kumar Sangaiah, A., Atiquzzaman, M., & Xia, S. (2020). ABFL: An autoencoder-based practical approach for software fault *localization*. *Sciences*, *510* , 108-121. https://doi.org/10.1016/j.ins.2019.08.077

Serttaş, S., & Deniz, E. (2023). Disease detection in bean leaves using deep learning. . *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, *65* (2), 115-129. https://doi.org/10.33769/aupse.1247233

Yavuz, G. (2022). Senior Learning JAYA With Powell's Method and Incremental Population Strategy. *IEEE Access*, *10* , 103765-103780. https://doi.org/10.1109/ACCESS.2022.3210122

Yavuz, G., Moghanjoughi, M. K., Dumlu, H., & Cakir, H. I. (2024). A Feature Selection Method Combining Filter and Wrapper Approaches for Medical Dataset *Classification*. *Journal of Computer Science*, S2196888824500246. https://doi.org/10.1142/S2196888824500246

SEMI SUPERVISED GENERATIVE ADVERSARIAL NETWORK FOR PLANT LEAF DISEASE DETECTION

EMİNE DENİZ ÖZ[5]

Introduction

As food supply shortages increase, plant leaf diseases are becoming more important. Diseases significantly reduce the yield and quality of crop production (Sharma et al., 2022) (Uyar, 2024a). Timely detection of plant leaf diseases is crucial to reduce crop loss and protect plant health (Gupta and Tripathi, 2024) (Özen Kavas & Bozkurt, 2022).

Many studies have been conducted to investigate plant diseases early and accurately. Initially, methods based on visual appearance and manual examination were used in disease detection. These methods are laborious and limited by the technician and competence of the individuals involved (Uyar, 2024b).

Later developed advanced methods require large amounts of labeled data. In real-world scenarios, there is not much plant leaf disease data available (Bhowmik et al., 2019) (Dursun & Özen

[5]Research Assistant, Kütahya Dumlupınar Unıversity, Computer Engineering, Orcid: 0000-0003-0670-3578

Kavas, 2025). With the introduction of deep learning-based methods in finance, industry, and agriculture, they have also been applied to plant leaf disease diagnosis (Gaus et al., 2019) (Özen Kavas et al., 2023).

High classification accuracy has been exhibited by deep learning models, which have proven to be advantageous over traditional approaches for plant disease detection (LeCun, Bengio, and Hinton, 2015) (Uyar, 2024c). However, deep learning models are based on data labeling (Lee et al., 2016). The process of labeling plant leaf disease imagery is both arduous and lengthy. Besides being laborious, it can also lead to erroneous labeling. Therefore, semi supervised approaches have begun to be used in the literature.

Semi supervised approaches focus on training. This training uses a limited amount of annotated data. It also uses a large amount of unannotated data. (Radford, Metz, and Chintala, 2015). These approaches face challenges such as the boundary of plant diseases being often blurred and the background of disease images containing a lot of noise and interference.

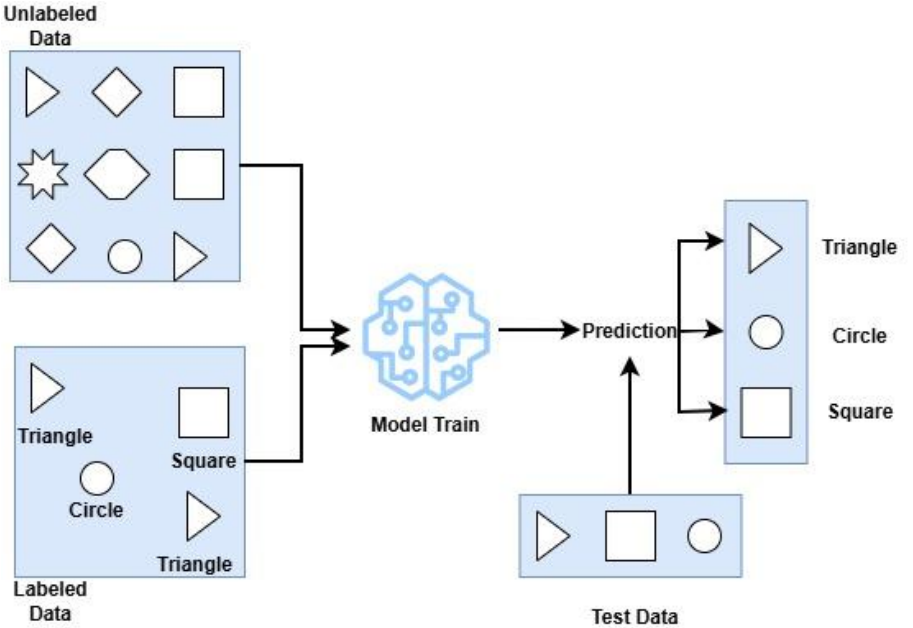
This study proposes using a semi supervised generative adversarial network (SSGAN) to overcome the problems. Generative Adversarial Networks (GANs) demonstrate that applying pre-trained classifiers to generated synthetic plant disease images facilitates classification (Liu et al., 2020). GANs, a form of unsupervised method, are used in the production of synthetic images that resemble real images (Sampath et al., 2021) (Uyar, 2024d). This study aims to solve the labeling problem using a GAN-based semi supervised model.

Semi Supervised Generative Adversarial Network (SSGAN)

This approach leverages a small quantity of labeled instances and a vast volume of unlabeled ones to train the model, as shown in

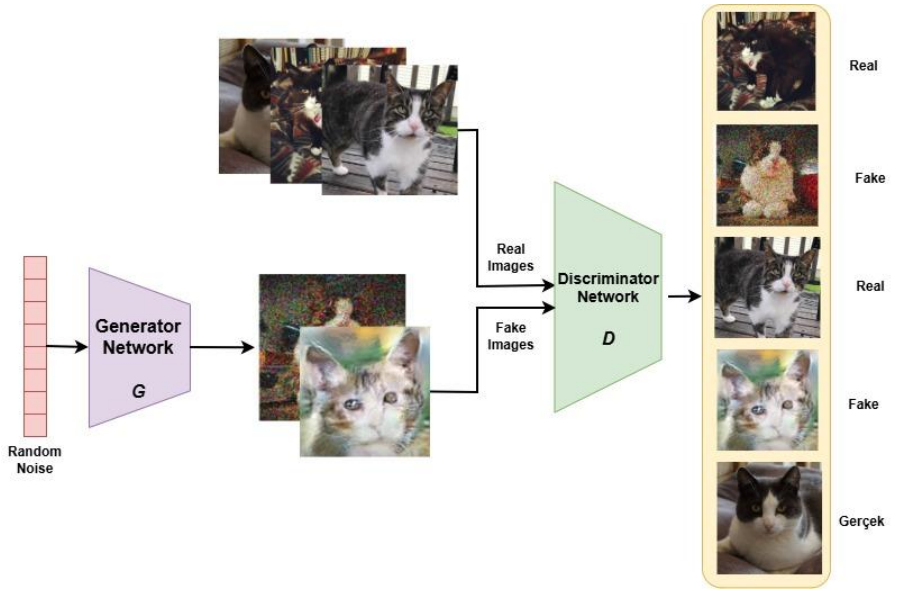
Figure 1. This approach critically reduces the cost of labeling the data.

Figure 1: Semi supervised learning



Generative adversarial networks (GANs) are a class of generative models designed by Ian Godfellow in 2014 (Goodfellow et al., 2020) (Uyar, 2021). GANs consist of two networks: a generator and a discriminator. The discriminator network classifies the images in the training data as real or fake, generated by the generator. The generator network attempts to produce images similar to the real data. While the generator network wants the discriminator network to see the fake images as real, the discriminator tries to distinguish the images generated by the generator from the real images. This process is called Game Theory. In theory, the generator tries to outmaneuver the discriminator, and the discriminator tries to outmaneuver the generator. Figure 2 shows the basic architecture of GANs.

Figure 2: GAN architecture



Labeling disease images on plant leaves is time-consuming and labor-intensive, therefore sparsely labeled data should be used. Semi supervised techniques are used in sparsely labeled data. However, many semi supervised techniques ignore important information in the vast majority of data. Therefore, GAN-based semi supervised approaches should be used (Sharma, Tripathi, Mittal et al., 2022)(Uyar & Özdemir, 2025).

In GANs, the separator generally monitors whether the network image is real or fake. In semi supervised GANs, however, the separator includes both a classifier and a separate fake class. For example, in a dataset containing K classes, let 1=cat, 2=dog, 3=bird, and K=rabbit. In classic GANs, the separator determines whether it is real or fake. In semi supervised GANs, a fake class is created as an additional class K+1 to class K. The fake class created by the generator belongs to class K+1.

The generative network architecture of semi supervised GANs is the same as that of classical GANs. However, their loss functions differ. Equation 1 shows the loss function of classical GANs. In the equation, $\mathbb{E}$ is the average result of a random variable. z is the noise vector used as input by the generator. $p(z)$ is the probability distribution of the input noise. $G(z)$ takes the noise vector and converts it into a dummy sample. $D\big(G(z)\big)$ gives a number representing the probability that the generated sample is real.

$$L_G = -\mathbb{E}_{z\sim p(z)}\big[\log D\big(G(z)\big)\big] \qquad Equation\ 1$$

In semi supervised GANs, the loss function in Equation 2 is used. Since the discriminator in semi supervised GANs operates with multiple classes, the expression $p_D(y = i\,|G(z))$ means that the discriminator gives the probability that the generated sample is in the i-th class. $\sum_{i=1}^{K} p_D\,(y = i\,|G(z))$ is the total probability that the generated sample belongs to any of the actual classes.

$$L_G = -\mathbb{E}_{z\sim p(z)}\left[\log \sum_{i=1}^{K} p_D\,(y = i\,|G(z))\right] \qquad \text{Equation 2}$$

Dataset

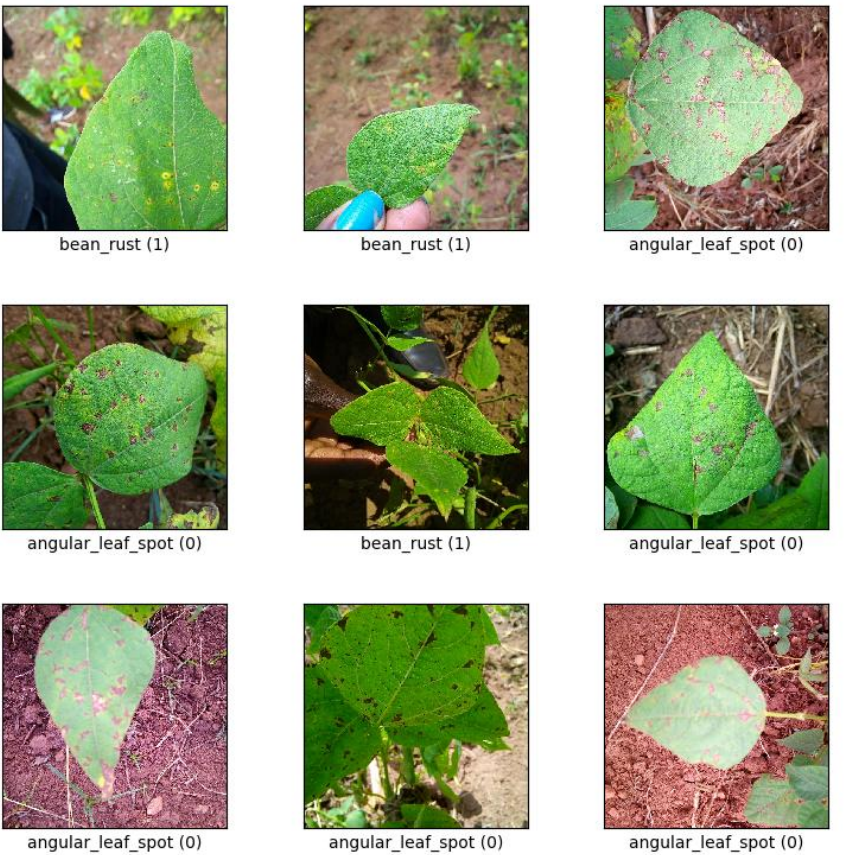
This study utilized the Bean dataset from TensorFlow Datasets. This dataset consists of images of beans taken outdoors using smartphones. The dataset comprises three classes: Angular Leaf Spot, Bean Rust, and Healthy.

This dataset contains 1295 plant leaf images. These 1295 images are divided into 1034 training (80%), 133 validation (10%), and 128 testing (10%) data points.

Figure 3 shows the images of the 3 classes in the dataset. The raw images are 500x500 pixels and 3-channel (RGB). The label

column contains the values 0, 1, and 2. 0 indicates angular leaf spot disease, 1 indicates bean rust disease, and 2 indicates a healthy class.

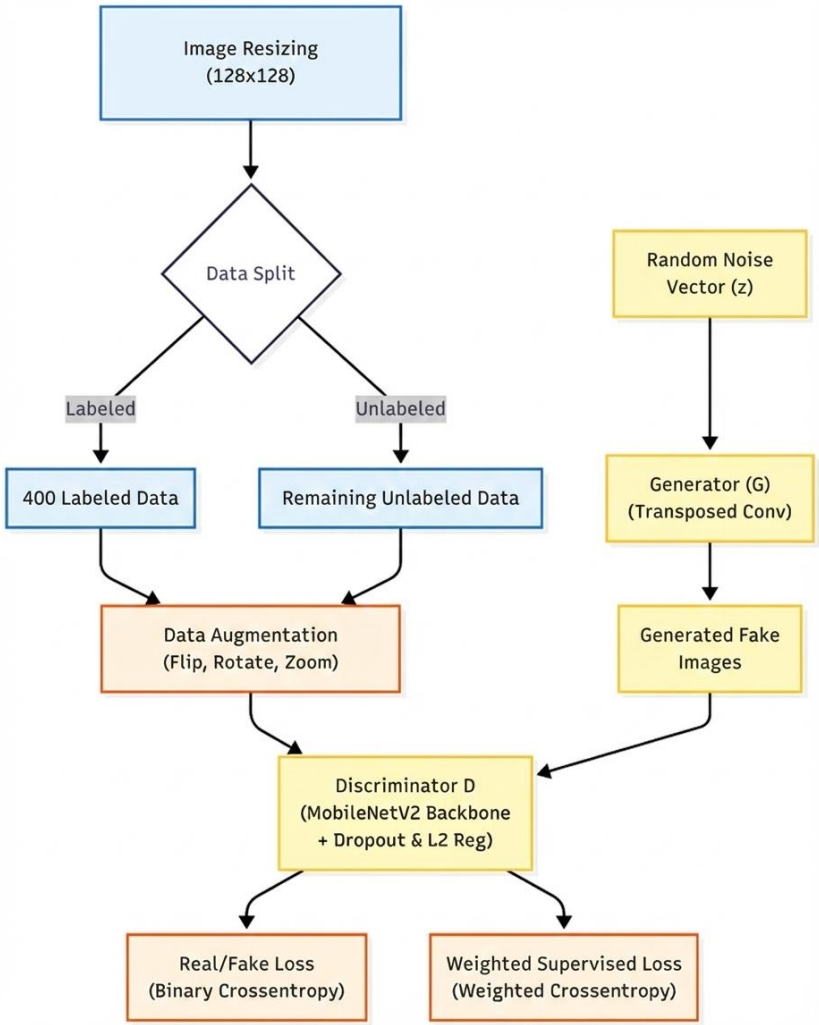
Figure 3: Sample images of the bean dataset.



Method

This study presents a method for tackling the limited number of labeled images in the agricultural field. The flowchart for this study is given in Figure 4.

Figure 4: Flowchart of the study.



In this study, preprocessing was first performed to make the data suitable for model input. Images were reduced to 128*128 pixels and scaled to the range [-1, 1]. 400 images were separated from the random training dataset with labels, while the labels of the remaining images were hidden. To prevent the model from being memorized with a small number of labeled data, data enhancement

methods such as horizontal/vertical flipping, 20% rotation, and zooming were applied.

The SSGAN architecture was then constructed with a generator and a discriminator network. The generator network obtains a 128*128*3 dimensional fake plant leaf image from a random 100-dimensional noise vector through a series of operations. The separator network both separates the images into real and fake and performs the classification process. The last layer of the separator network has two different outputs: supervised and unsupervised.

A two-stage training strategy was followed to obtain a stable model. In the first stage, classification headers were added to the frozen weights of the MobileNetV2 architecture, and the best weights were found using 400 labeled data over 60 cycles. This stage enabled the separator to acquire basic classification capability. In the second stage, the last 30 layers of the frozen weights of the MobileNetV2 architecture were included in the training, and the model was trained with labeled and unlabeled data over 3000 iterations. The parameters used in model training are given in Table 1.

Table 1: Parameters used in model training

Parameters	**Value**
Image Size	128
Batch Size	32
Number of Classes	3
Number of Labeled Data	400
Nuber of Pretraining Cycles	60
Number of GAN Iterations	3000
Alpha	0.2
Class Weights	[2.5, 1.0, 1.0]
Discriminator/ Generator Learning Rate	1e-5 / 1e-4

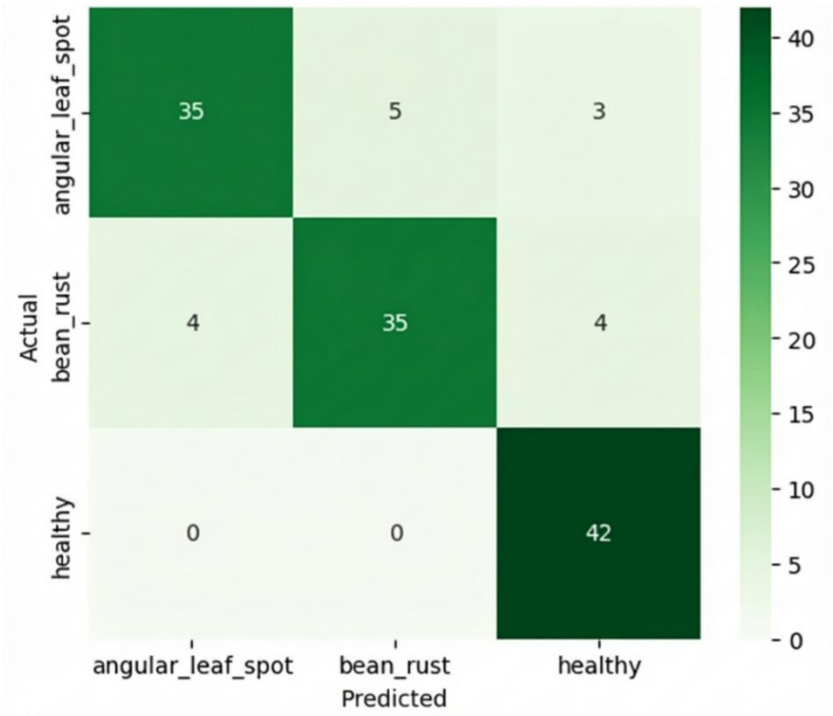
Finally, controlled and uncontrolled losses were weighted and summed for the total loss function.

Conclusion

Using 400 labeled data points, the model proposed in this study achieved an overall accuracy of 88% and an F1-score of 0.87 on the test dataset. The SSGAN structure in this study successfully learned leaf texture and shape from unlabeled data as well.

In the bean leaf dataset with 3 classes, our model achieved the highest success by correctly classifying all healthy leaves in the healthy leaf class. As seen in Figure 5, it showed a balanced performance in the angular leaf spot class with an precision of 0.90 and a sensitivity of 0.81.

Figure 5: Model confusion matrix



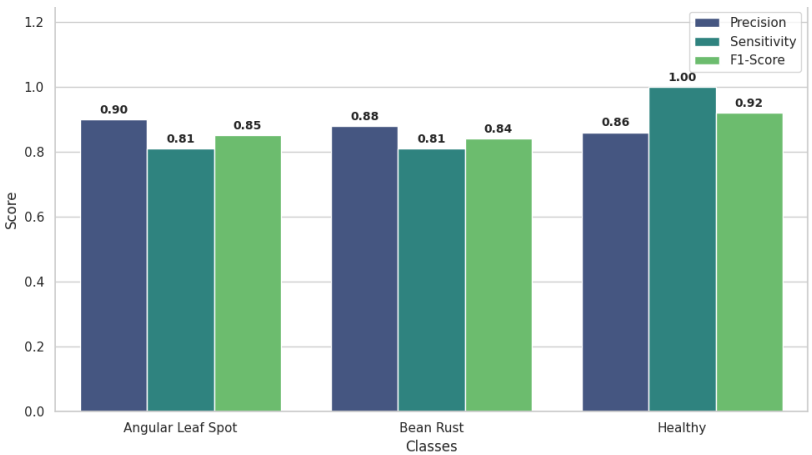
The class weights applied during the training phase prevented the angular leaf spot class from being completely crushed, ensuring a balanced performance with other diseases. Similarly, the bean rust class mixed 4 images of this class with angular leaf spot and 4 with healthy class, with precision values of 0.88 and sensitivity values of 0.81.

Table 2: Model training results

Class	**Precision**	**Sensitivity**	**F1-Score**
Angular Leaf Spot	0.90	0.81	0.85
Bean Rust	0.88	0.81	0.84
Healthy	0.86	**1.00**	0.92
Overall Accuracy			0.88
Macro Average	0.88	0.88	0.87
Weighted Average	0.88	0.88	0.87

As shown in Table 2, the model demonstrated 100% sensitivity in the healthy class. In the disease classes, it achieved a balanced rate of 81%, resulting in an overall accuracy rate of 88%. dditionally, a visual comparison of these metrics (Precision, Sensitivity, and F1-Score) across all classes is presented in Figure 6.

Figure 5: Class-Based Model Performance



References

Bhowmik, N., Gaus, Y. F. A., Akçay, S., Barker, J. W., & Breckon, T. P. (2019). On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within x-ray security imagery. In 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 986–991). IEEE.

Dursun, M. A., & Özen Kavas, P. (2025). Beyond diagnosis: Cross-dataset evaluation of risk factors for thyroid cancer recurrence. Artificial Intelligence Studies, 8(1), 38–69. https://doi.org/10.30855/ais.2025.08.01.03

Gaus, Y. F. A., Bhowmik, N., Akçay, S., Guillen-Garcia, P. M., Barker, J. W., & Breckon, T. P. (2019). Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered X-ray security imagery. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). IEEE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.

Gupta, S., & Tripathi, A. K. (2024). Fruit and vegetable disease detection and classification: Recent trends, challenges, and future opportunities. Engineering Applications of Artificial Intelligence, 133, 108260. https://doi.org/10.1016/j.engappai.2024.108260

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Lee, S. H., Chang, Y. L., Chan, C. S., & Remagnino, P. (2016). Plant identification system based on a convolutional neural network for the LifeClef 2016 plant classification task. In CLEF

(Working Notes) (pp. 502–510). http://ceur-ws.org/Vol-1609/16090502.pdf

Liu, B., Tan, C., Li, S., He, J., & Wang, H. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. IEEE Access, 8, 102188–102198. https://doi.org/10.1109/ACCESS.2020.2998839

Özen Kavas, P., & Bozkurt, M. R. (2022). Detection of HFrEF and HFpEF using PPG-derived HRV with machine learning methods. Journal of Scientific Reports-A, 51, 317–329.

Özen Kavas, P., Bozkurt, M. R., Kocayiğit, İ., & Bilgin, C. (2023). Machine learning-based medical decision support system for diagnosing HFpEF and HFrEF using PPG. Biomedical Signal Processing and Control, 79, 104164. https://doi.org/10.1016/j.bspc.2022.104164

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Sampath, V., Maurtua, I., Martín, J. J. A., & Gutierrez, A. (2021). A survey on generative adversarial networks for imbalance problems in computer vision tasks. Journal of Big Data, 8(1), 27. https://doi.org/10.1186/s40537-021-00414-0

Sharma, V., Tripathi, A. K., & Mittal, H. (2022). Technological revolutions in smart farming: Current trends, challenges & future directions. Computers and Electronics in Agriculture, 201, 107217. https://doi.org/10.1016/j.compag.2022.107217

Sharma, V., Tripathi, A. K., Mittal, H., Parmar, A., Soni, A., & Amarwal, R. (2022). WeedGan: A novel generative adversarial network for cotton weed identification. The Visual Computer, 39(12), 6503–6519. https://doi.org/10.1007/s00371-022-02742-5

Uyar, R. (2021). Investigation of computer engineering students' academic achievement according to high school graduation areas.

Uyar, R. (2024). Brats20 veri seti ile beyin tümörü segmentasyonunda SegNet ve MambaUnet'in derin öğrenme performanslarının karşılaştırılması. In Pratik yapay zeka uygulamaları 2 (pp. 153–169). Bidge Yayınları.

Uyar, R. (2024). Gri kurt algoritması ve varyantlarının sınıflandırma performansının değerlendirilmesi. In Bilgisayar mühendisliği alanında uluslararası araştırma ve değerlendirmeler (pp. 1–12). Serüven Yayınevi.

Uyar, R. (2024). Güncel metasezgisel optimizasyon algoritmalarının CEC2020 test fonksiyonları ile değerlendirilmesi. In Pratik yapay zeka uygulamaları 1 (pp. 153–181). Bidge Yayınları.

Uyar, R. (2024). Metasezgisel optimizasyon algoritmalarının CEC2019 ölçütlerine göre performans analizi. In Mühendislik alanında gelişmeler (pp. 527–539). Platanus Publishing.

Uyar, R., & Özdemir, D. (2025). Deprem şiddet tahmini için derin öğrenme yöntemlerinin karşılaştırılması ve model önerisi. Afyon Kocatepe University Journal of Sciences and Engineering, 25(3), 522–534. https://doi.org/10.35414/akufemubid.1511843

GRAPH-BASED DEEP LEARNING APPROACHES FOR EEG INTERPRETATION AND BRAIN CONNECTIVITY MODELING

EVİN ŞAHİN SADIK[6]
SEDA ŞAŞMAZ KARACAN[7]

Introduction

Electroencephalography (EEG) is a non-invasive method that can measure electrical potential changes occurring on the brain's cortical surface with high temporal resolution (Niedermeyer & da Silva, 2005). The ability to monitor neural activity on a millisecond scale has made EEG a tool used in modern application areas such as clinical neurophysiology, brain-computer interfaces (BCI), cognitive load monitoring, mood analysis, and neuroergonomics. However, EEG signals are inherently low-amplitude, susceptible to noise, and exhibit a time-dependent structure. In addition, the placement of the electrodes on the scalp surface does not form a regular grid, which

[6] Asst. Prof. Dr., Kutahya Dumlupinar University, Department of Electrical and Electronics Engineering, Orcid: 0000-0002-2212-4210
[7] Lecturer Dr., Usak University, Department of Information Technology, Orcid: 0000-0002-0334-260X

makes it difficult to directly model the spatial relationships of the signals using classical deep learning approaches.

In recent years, representing EEG data as graphs has become widespread in order to overcome these structural limitations. Graph representations created through channel topology, physical proximity, functional connectivity, or effective connectivity measures naturally reflect both the spatial and relational information of EEG. The graph approach is consistent with neuroscientific findings that interactions between electrodes arise not only from individual channel characteristics but also from dynamic connections between brain regions. As a result, EEG becomes a suitable data source for Graph Neural Networks (GNNs) capable of learning on irregular structures.

GNN-based EEG classification literature has rapidly developed in recent years, covering a wide range of problem areas such as emotion recognition, motor imagery, epileptic seizure detection, sleep stages, and neurodegenerative diseases (Graña & Morais-Quilez, 2023; Qiu et al., 2025). The recent review presented by Klepl et al. systematically classifies these studies in terms of the definition of brain graph structure, node properties, graph convolution layers, pooling, and graph-level readout modules. It shows that Chebyshev-based spectral GNNs and simple GCN layers are still dominant, while transfer learning, cross-frequency coupling, and explainability-focused approaches have not been sufficiently researched (Klepl et al., 2024).

GNNs are models that can capture both local and global features of graph data by performing message transmission through nodes (channels), edges (connections), and their associated weights. Unlike classical CNN or RNN-based approaches, GNNs can directly model the structural dependencies of the graph. This feature provides an advantage in representing the complex relational structure of

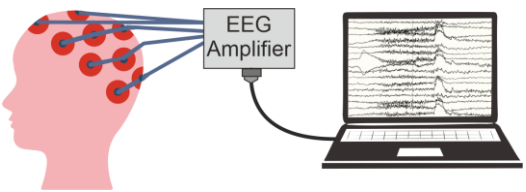
EEG.The use of graphs based on functional connectivity enables the interaction between brain regions to be utilized with deep learning models, which is important in terms of both accuracy and neuroscientific interpretability.

In this section, we examine how GNNs provide solutions to the structural challenges posed by EEG signals for traditional deep learning methods. Our work begins by addressing the biophysical foundations and signal properties of EEG, followed by an introduction to the fundamental principles of Graph Theory and GNN architectures. In subsequent sections, we will detail the strategies for transforming EEG data into a graph (structural, functional, adaptive) and present architectural variations of these graph-based models, along with their current application areas. Finally, we will discuss the key challenges encountered in GNN-based EEG analysis and potential future research directions in the field, highlighting the growing importance of this intersection in artificial intelligence and neuroscience.

Fundamentals of EEG Signal Analysis

Electroencephalography (EEG) is a non-invasive method used to record electrical activity in the brain (Zhang et al., 2023). EEG signals are acquired through electrodes placed on the scalp, amplified by dedicated hardware, and recorded as multichannel time series for further analysis (Figure 1).

Figure 1.Overview of EEG signal acquisition



EEG can capture neural activity with high temporal resolution. Therefore, it has become a fundamental tool in many

brain-related fields, from clinical neurophysiology and brain-computer interface (BCI) systems to cognitive status monitoring and neuroergonomics. (Raj et al., 2025; Värbu et al., 2022). The portability and relatively low cost of EEG systems make them suitable for AI and machine learning-based applications where large-scale data collection and real-time processing are often required (X.-Y. Liu et al., 2025).

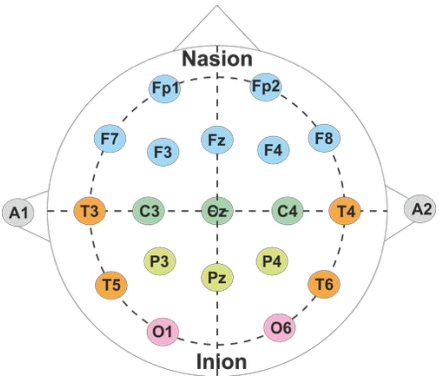
The Biophysical Basis of EEG

EEG is a reflection of the electrical activity produced collectively by clusters of pyramidal neurons in the cortex. Although the signals from individual neurons are too weak to reach the surface, the simultaneous activation of thousands of neurons creates a measurable potential difference that can be recorded from the scalp (Hesprich & Beardsley, 2019).

EEG signals are filtered and attenuated as they pass through the skull and protective tissue layers before reaching the electrodes. This physical phenomenon, referred to in the literature as volume conduction, is the main factor behind the low resolution and noisy nature of EEG data.

In practice, the International 10–20 System, in which electrodes are placed at specific anatomical points, is used to obtain standard and comparable measurements (HH, 1958). Letter codes indicate the relevant brain lobe, while numbers indicate laterality. This standard layout is shown in Figure 2. This system provides a consistent spatial basis for neurophysiological interpretation and graph-based modeling processes.

Figure 2.Standard 10–20 EEG electrode placement system functional regions (frontal, central, parietal, temporal, occipital).



EEG Characteristics

EEG signals typically carry meaningful neurophysiological information in the 0.1–100 Hz range and have a low-amplitude structure ranging from 5–100 μV. Therefore, EEG has high resolution in the time domain. Frequency content is important for understanding cognitive processes and cortical dynamics. EEG analysis is mostly based on the examination of brain rhythms. The delta, theta, alpha, beta, and gamma bands are associated with different cognitive and neurophysiological states. The definitions of these frequency bands and their associated functions are summarized in Table 1.

Table 1.Canonical EEG frequency bands and their associated cognitive or neurophysiological functions.

Band	Frequency Range	Associated Cognitive / Neurophysiological Functions
Delta	0.5-4 Hz	Deep sleep, altered consciousness
Theta	4-8 Hz	Attention, memory processes, meditative states
Alpha	8-13 Hz	Relaxation, eyes-closed rest, cortical inhibition
Beta	13-30 Hz	Motor activity, cognitive processing, sensorimotor rhythm
Gamma	>30 Hz	Functional connectivity, attention, high-level cognition

The Concept of Brain Connectivity in EEG

The relationship between EEG signal analysis and GNN is about modeling the interactions between different regions of the brain. Signals recorded from EEG electrodes are not independent, isolated series. They are parts of an integrated system reflecting the synchronized or related activity of cortical networks over time (Bullmore & Sporns, 2009). This complex network of relationships is generally examined in the literature under three main categories: structural, functional, and effective connectivity (Penny et al., 2011).

Structural connectivity refers to the anatomical pathways that physically connect brain regions. Although EEG cannot directly measure these structures, the physical distances between electrodes obtained from standard electrode placements or the geometric neighborhood relationships on the scalp are used as spatial priors in GNN models to create fixed or predefined graph topologies (Kalafatovich et al., 2022; Klepl et al., 2024; Qiu et al., 2025).

Functional connectivity defines the level of statistical dependence exhibited by two EEG channels over time. While this approach does not claim causality, it reveals whether the channels co-oscillate or change in similar patterns. Commonly used metrics include non-linear dependencies such as Correlation, Coherence, Phase Locking Value (PLV), and Mutual Information (MI). These metrics are the most frequently used methods for generating data-driven adjacency matrices from EEG data, and a large portion of current GNN-based EEG models operate on such functional connectivity-based graphs.

Effective connectivity aims to model the direction and causality of information flow between brain regions. This approach enables the creation of directed graphs for GNN models, allowing for a more realistic modeling of intra-brain information flow.

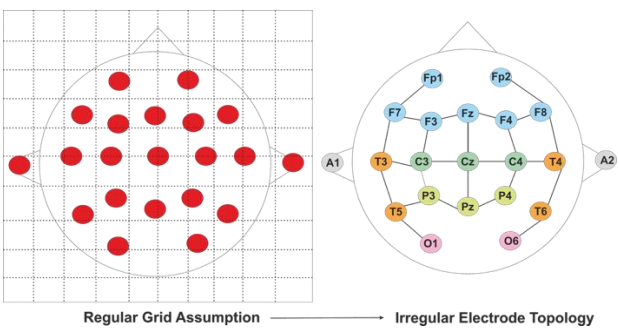
In conclusion, these three types of connections form the theoretical basis for making EEG signals processable by GNNs and enable the successful application of modern graph-based deep learning approaches to EEG analysis.

Graph Neural Networks

Graph Neural Networks (GNNs) are a subfield of deep learning capable of learning from graphically structured data. (Wu et al., 2020; Zhou et al., 2020). Traditional architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) assume that the underlying data are arranged on a regular Euclidean grid (Mienye et al., 2025; Yamashita et al., 2018).

EEG recordings violate this assumption due to the non-uniform spatial arrangement of electrodes on the scalp. As illustrated in Figure 3, while CNN-based models implicitly rely on a regular lattice structure, EEG electrode layouts follow an irregular and biologically constrained topology (Bashivan et al., 2015). This mismatch limits the ability of grid-based architectures to accurately model spatial dependencies between channels and motivates the use of graph-based representations for EEG analysis.

Figure 3.Limitations of Grid-Based Deep Learning for EEG



In contrast to grid-based models, GNNs provide a biologically consistent framework by explicitly modeling electrodes

as nodes and their interactions as edges. This allows for the representation of both local neighborhood interactions and global structural dependencies across brain regions. (Miri et al., 2024).

Fundamentals of Graph Theory

A brain graph is a mathematical representation of neural architecture defined by a set of nodes and the links connecting them (Vecchio et al., 2017). A graph is defined as **G= (V, $\mathcal{E}$)**. Here, **V** represents the set of **N** nodes (**|V|=N**), while $\mathcal{E}$ represents the set of edges representing the connections between these nodes. These abstract concepts are made concrete through EEG analysis.

The nodes (**V**) represent EEG electrodes or channels placed on the scalp. Edges ($\mathcal{E}$) represent the physical or functional relationships between different electrode regions.

The topological structure of a graph is usually expressed mathematically by an adjacency matrix, **A**, of size **NxN**. The element $\mathbf{A_{ij}}$ in the **i**-th row and **j**-th column of this matrix indicates the existence and nature of the connection between nodes **i** and **j**. This relationship is given by Equation (1).

$$A_{ij} = \begin{cases} w_{ij}, & \text{If there is a connection between (i,j)} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Here, w_{ij} is a scalar weight value representing the strength of the relationship between two electrodes. This weight can be determined using various connectivity measures such as the Euclidean distance between electrodes, Pearson correlation coefficient, coherence, or PLV.

Graphs are classified as directed or undirected depending on whether the edges contain directional information. In EEG literature, since most metrics used, especially in functional connectivity-based studies, are symmetric in nature ($A_{ij} = A_{ji}$), weighted undirected

graph structures are generally preferred. This structure is suitable for modeling the strength of reciprocal interactions between brain regions.

Message Passing Mechanism in Graph Neural Networks

The fundamental operating principle of GNN architectures is based on the idea that each node (electrode) collects information from its neighbors, processes this information, and updates its own feature vector. Known in the literature as the Message Passing Neural Network (MPNN) framework (Tang et al., 2023), this process enables the learning of local and global relationships in EEG signals. The GNN learning process consists of two main stages Readout/Pooling and Message Passing/Updating.

Message Passing and Update: At this stage, the hidden state $h_i^{(l+1)}$ of a v_i node in layer *(l+1)* is formed by aggregating information from neighboring nodes (N(*i*)). This process can be expressed in a general form in Equation (2).

$$h_i^{(l+1)} = \sigma\left(\text{UPDATE}\left(h_i^{(l)}, \text{AGGREGATE}\left(\{h_j^{(l)} : j \in \text{N}(i)\}\right)\right)\right) \quad (2)$$

Here, h_i^l represents the node attributes in the l-th layer, and N*i* represents the neighbors of node i. The AGGREGATE function (such as sum, average, or maximum pooling) reduces the neighborhood information to a single vector, while the UPDATE function combines the node's own past information with new information from its neighbors. σ is the nonlinear activation function (ReLU, Tanh, etc.).

Readout and Pooling: In EEG classification tasks, the goal is typically to classify the entire brain graph rather than a single electrode. Therefore, the features learned at the node level ($h_i^{(L)}$) must be converted into a single graph representation vector (h_G). This process is called Readout and is shown in Equation (3).

$$h_G = \mathrm{READOUT}\left(\{h_i^{(L)} \mid v_i \in \mathrm{V}\}\right) \quad (3)$$

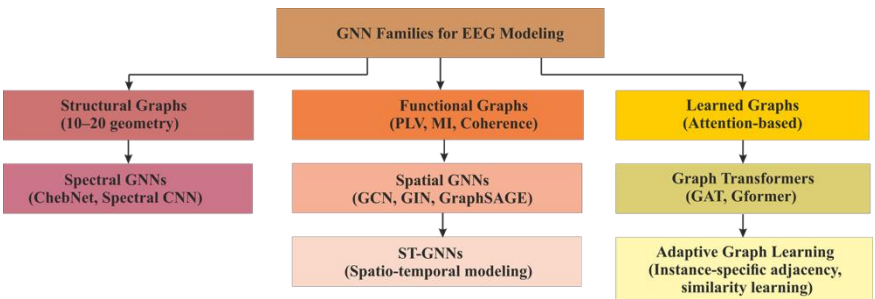
The readout function can be a simple mean pooling, or it can be implemented by gradually reducing the graph using hierarchical pooling methods. The resulting h_G vector is then passed to the final classification layer to determine the class to which the EEG signal belongs.

GNN families

GNNs encompass a broad range of models developed to process information on graph-structured data. These models are divided into different families based on the mathematical basis of their message passing mechanisms, their neighborhood information aggregation strategies, and their ability to model the spatial-temporal properties of the graph.

The most widely accepted types of GNNs in the EEG literature are convolution-based (GCN) (Bhatti et al., 2023), attention-based (GAT), sampling structures, and spatio-temporal models that integrate the time dimension with the graph. The systematic classification of these architectures and their submodules is summarized in Figure 4.

Figure 4.Conceptual classification of GNN families used in EEG modeling.



The success of GNN models in EEG analysis largely depends on the compatibility of the input graph type with the architecture used. When examining the general trend in literature, three main approaches stand out. Structural graphs, based on fixed topologies derived from the physical placement of electrodes, typically favor Spectral GNN, which processes the frequency components of the signal in the spectral domain of the graph.

In functional graphs, based on statistical dependencies, Spatial GNN and Spatio-Temporal GNN models that allow direct message passing between nodes stand out. In learned graphs, where the connection structure is dynamically optimized by the model, Graph Transformers and Adaptive Graph Learning approaches that determine edge weights using an attention mechanism are used.

Structural Graph-Based Models

Structural graph-based models create a fixed adjacency matrix using the physical or geometric arrangement of EEG electrode placements. In this approach, the graph structure is derived not from the signal, but directly from spatial arrangements such as the 10–20 or 10–10 electrode systems. Thus, the connections between nodes (electrodes) are defined according to proximity, distance, or skull surface geometry on the skull. The advantage of these models is that they provide a stable graph topology without being affected by signal noise. They are particularly effective in small data sets where the relationships between electrodes do not change over time.

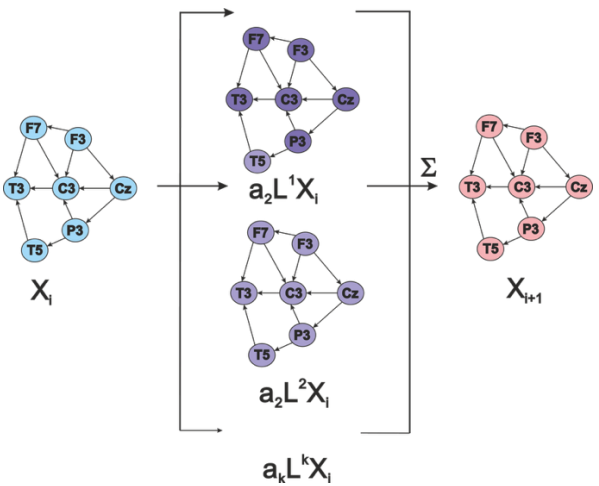
Spectral GNNs

Spectral GNNs, sometimes referred to as Spectral CNNs in early graph learning literature, define graph convolution operations in the spectral domain based on graph signal processing theory. Within this family, several architectures have been proposed, among

which ChebNet (Chebyshev GCN) is one of the most influential and practically efficient models (Defferrard et al., 2016).

ChebNet-based models have demonstrated strong performance in various EEG-related tasks, including motor imagery, sleep stage classification, and basic cognitive state analysis. Figure 5 schematically illustrates how ChebNet propagates information on the graph using Chebyshev polynomials and powers of the Laplacian matrix (L^k), integrating information from multiple K-hop neighborhoods. The input signal (X_i) is filtered across different neighborhood orders (L^1, L^2, ⋯,L^k) and combined using learnable coefficients (a_k) to produce the updated node representation (X_{i+1}).

Figure 5.Schematic representation of the Chebyshev GCN spectral convolution mechanism.



Spectral CNN is a broader term referring to methods that perform graph convolution directly in the spectral domain using the eigenvalue–eigenvector decomposition of the graph Laplacian. These approaches aim to generalize frequency-domain filtering in classical CNNs to graph-structured data. In EEG applications with fixed and well-defined electrode layouts, such as the 10–20 or 10–

10 systems allowing the Laplacian spectral decomposition to be computed once and reused throughout the model. This property can provide strong representational capacity in stable EEG topologies and acts as a form of regularization in low-sample regimes.

However, the limited transferability of spectral filters across different graph structures restricts the generalization of spectral CNNs to varying electrode configurations or sensor layouts. Moreover, the assumption of a static graph topology limits their ability to capture the rapidly changing functional connectivity of the brain. These limitations have motivated the shift toward signal-derived and dynamic graph constructions, giving rise to functional graph-based GNN models.

Spatial GNNs

Spatial GNN models are structures that define message passing on graphs directly through node adjacencies and explicitly model spatial relationships. Among the most used spatial GNNs in EEG analysis are GCN, GIN (Graph Isomorphism Network), and GraphSAGE.

GCN provides a stable and low-cost foundation by updating node features based on normalized neighborhood information. It is frequently preferred for small-to-medium-sized EEG graphs.

GIN, on the other hand, combines neighborhood information with a summation function, offering richer representational power than GCN in terms of expressive power and is particularly effective in distinguishing fine spatial patterns.

Unlike classical GNNs, GraphSAGE (Sun et al., 2024) does not require the use of all neighbors for each node. Instead, it updates by sampling a fixed number of neighbors. This sampling-based approach provides significant advantages in terms of memory and time in large graph structures, as well as increasing scalability in

EEG channel sub-selection, multimodal EEG–EMG systems with an increased number of sensors, or large data sets where long recordings are divided into segments.

The general advantage of the Spatial GNN family is that it can directly model the physical or functional relationships between EEG electrodes, thereby providing high spatial sensitivity in both motor imagery and cognitive tasks.

Spatio-Temporal GNNs

Spatio-temporal GNN (ST-GNN) models are a family of architectures developed to model both the spatial and temporal structure of EEG simultaneously. In these models, the connections between electrodes and their dynamics over time are captured using convolution, RNN, or attention-based components in addition to the graphical structure.

ST-GCN is a popular architecture that integrates GCN layers with 1D temporal convolutions, capturing both short-range spatial interactions and longer-term temporal dependencies. It is particularly successful in tasks requiring temporal structure, such as motor imagery, epileptic seizure transitions, and sleep stage classification.

Diffusion Convolutional RNN (DCRNN) models information flow as a diffusion process rather than classical temporal convolution. This allows for a more biologically consistent representation of how brain activity spreads to different regions. This approach is used in cases such as monitoring the spread of epilepsy or changes in overall cognitive status.

Attention-based Spatio-Temporal GNN (AST-GNN) adaptively highlights important brain regions and connections at different time scales by using both spatial and temporal attention mechanisms together.

The general advantage of the ST-GNN family is that it combines the complex time-frequency dynamics of EEG with graph-based spatial dependencies to obtain a rich, multidimensional representation. Therefore, ST-GNNs are currently positioned as one of the strongest architectural groups in terms of performance in the EEG-GNN literature.

Learned Graph-Based Models

Learned graph-based models encompass GNN approaches where the graph structure is not fixed but rather learned by the model itself alongside the data. In these methods, the adjacency matrix is either derived from a completely empty starting point or built upon a very we_{ak} draft of structural/functional prior knowledge. This enables more flexible and personalized graph structures for signals with rapidly changing connectivity, such as EEG.

Graph Transformers

Transformers are modern graph models that perform message passing using a fully global attention mechanism instead of classical neighborhood propagation. This approach has strong generalization capabilities because it can model not only close neighborhood relationships in EEG but also high-level interactions between distant electrodes.

Graph Attention Networks (GAT) learns the weights between neighbors using self-attention. It delivers strong performance in EEG, particularly in cognitive tasks involving intense frontal-parietal interaction.

Graph transformer creates a knowledge flow where all nodes can see each other through global attention. This structure provides advantages in tasks such as affective computing and working memory, where non-local dependencies are important in EEG.

Transformer-GNN hybrids combine the local message passing power of Spatial GNNs with the global dependency modeling capacity of Transformers. Both short-range electrode interactions and long-range cognitive connections in EEG can be represented within the same model.

The fundamental strength of the Graph Transformer family lies in its ability to capture both the local and global dynamics of EEG simultaneously. For this reason, it is among the rapidly emerging architectures in next-generation EEG-GNN studies.

Adaptive Graph Learning

Adaptive graph learning is an approach where the graph structure is dynamically optimized during the model training process. Edge weights are not fixed. As the signal changes, connections strengthen or weaken. This is particularly important in EEG because brain connectivity can change dramatically over time and under different task conditions.

Adaptive Graph Learning essentially encompasses all methods where the adjacency matrix is not fixed but is learned or updated by the model. Three main approaches under this umbrella can be defined as Instance-specific adjacency, Similarity learning, and Dynamic graph learning.

An instance-specific adjacency matrix is learned for each EEG window. The graph structure, which varies according to the individual's momentary cognitive state, can be captured.

Similarity learning involves the model learning the similarity between nodes. The adjacency matrix is derived from this similarity. It is a data-driven functional graph approach.

During training in dynamic graph learning, the graph is continuously updated. The topology reshapes itself through attention, kernel similarity, or learnable edge embeddings. It

successfully represents both the time-varying EEG connectivity and the adaptive brain network properties.

Why GNN Instead of Classic CNN/RNN for EEG?

In EEG analysis, GNNs offer advantages over classical Deep Learning architectures such as CNNs and RNNs.

- EEG electrodes are not arranged in a regular 2D grid but are positioned in an irregular topology on the curved surface of the scalp. Since CNNs require a grid structure, the 2D projection of EEG data often distorts spatial information. GNNs, however, process the data in its natural graph format, preventing this distortion.
- The Information related to cognitive processes is typically stored not in individual channel activity, but in the connectivity between brain regions. While CNNs and RNNs focus on local or temporal patterns, GNNs place relationship features such as cross-channel correlation, phase locking, or information flow at the center of the model.
- Models such as Adaptive GNN and ST-GNN are much more successful than classical methods using fixed filters in adapting to the brain's dynamic structure that changes over time and to inter-subject variability.

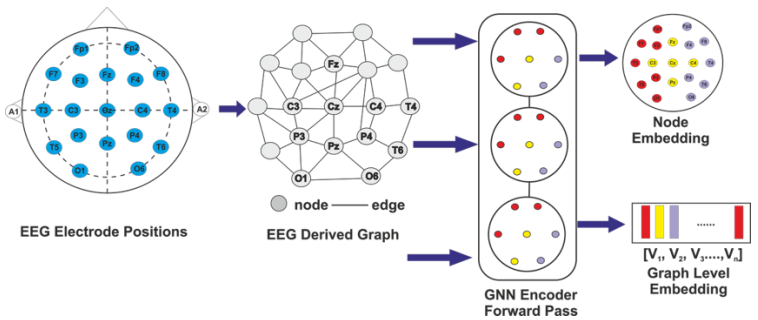
For these reasons, GNNs have become an increasingly preferred approach for EEG analysis, offering greater biophysical suitability and methodological robustness.

Graph Construction Strategies for EEG Data

The key step determining the success of GNN-based EEG models is the principle according to which the EEG signal is converted into a graph. The most basic approach is graph

construction based on physical topology. Electrodes are modeled as nodes according to their positions on the scalp surface, Euclidean distances between them are calculated using electrode coordinates, and an adjacency matrix based on this distance is defined. Thus, connection weights between neighboring electrodes are high, while connections between distant electrodes are assigned weak or zero values. This structure provides a natural foundation for models that assume a structural graph, such as ChebNet and GCN (Figure 6).

Figure 6.Conversion of EEG electrode positions into a graph representation and subsequent processing through a GNN encoder to obtain node-level and graph-level embeddings.

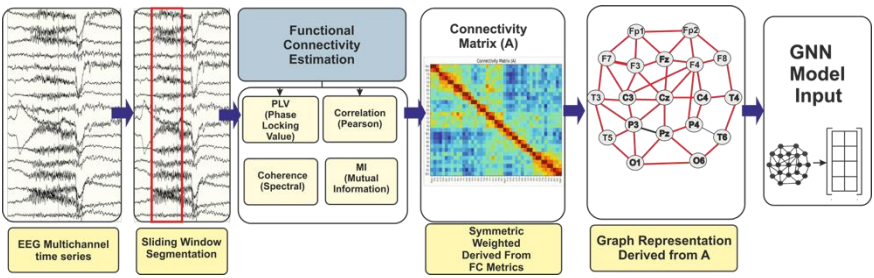


In contrast, in the graph approach based on functional connectivity, edge weights are derived from the signal itself. Inter-channel statistical dependence is measured using metrics such as Pearson correlation, PLV, Phase Lag Index (PLI), coherence, or MI, and a separate functional adjacency matrix is obtained for each EEG time window.

The length of the window time used here is of great importance. Very short windows are sensitive to noise and lead to unstable connection estimates, while excessively long windows risk losing dynamic connectivity changes by averaging them out. Therefore, careful window selection should be made according to the application.

In the third approach, the graph structure itself is learned by the model using edge-weight learning. Initially, a coarse adjacency is defined, and then during training, edge weights are updated using adaptive adjacency layers, attention mechanisms, or similarity learning modules. Thus, the graph structure evolves into a dynamic object alongside the data. The functional connectivity-based graph creation process is shown in Figure 7.

Figure 7.Functional-connectivity-based graph construction pipeline for EEG.



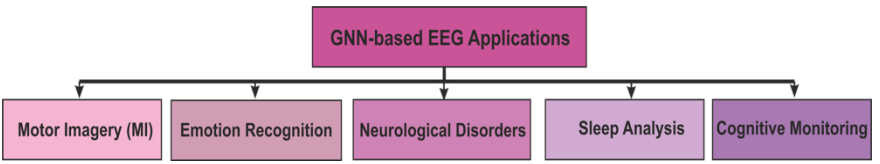
In this context, dynamic graph learning is a powerful tool for capturing brain connectivity that changes across different cognitive states or task phases. Finally, the multi-graph approach is a richer modeling strategy frequently used for EEG. In this method, both a spatial graph based on physical topology and a functional graph defined by metrics such as PLV or MI are used simultaneously. These graphs are either processed through parallel GNN branches and merged in a hybrid graph fusion layer, or represented within a single model as multi-channel adjacency tensors. This results in EEG–GNN systems with higher expressive power that simultaneously utilize both scalp surface geometry and signal-based connectivity.

Applications of GNNs in EEG Research

GNN-based models, which are increasingly used in neuroscience and artificial intelligence research, differ from

traditional methods thanks to the topological methods they employ in EEG analysis. Modeling the irregular structure of EEG data with graph architectures has enabled higher performance, particularly in tasks where spatial relationships must be preserved. The main application areas of GNN-based approaches in the literature and their subcategories are systematically summarized in the taxonomy presented in Figure 8.

Figure 8.Taxonomy of the main application areas highlighted in GNN-based EEG research.



The most common and fundamental application area of these architectures is motor imagery tasks, which form the core of BCI. While traditional CSP or CNN-based methods model the spatial dependency between electrodes in a limited way, GNNs can directly process the natural connection patterns of the motor cortex.

Structural graph-based GCN and GraphSAGE architectures provide significant performance gains, particularly on topologies covering motor areas, while functional graph-based ST-GNN models capture the dynamic connectivity that changes over time during movement imagery, offering both high accuracy and faster convergence. Furthermore, the personalized adaptation capability of GNNs provides an advantage in solving the inter-subject variability problem, which is the greatest challenge for BCI systems.

Emotion recognition is another area where GNNs make a difference. Emotional processes involve complex, multi-channel interactions such as hemispheric asymmetry and frontal theta and beta dynamics. GNN-based models can extract the connective

fingerprints of emotional states derived from functional connectivity metrics such as PLV or coherence.

This topological perspective provided by GNN architectures also offers innovations in clinical neuroscience for epilepsy detection and the diagnosis of neurological diseases. Since epileptic seizures involve both local and widespread network activation, graph convolution-based models can model the cortical spread of epileptiform discharges with edge weights that change over time. Diffusion-based GNNs and ST-GCN models distinguish pre-seizure and seizure transitions with high sensitivity, and subgraph discovery methods enhance the clinical interpretability of the models by enabling the localization of the epileptic focus.

Large-scale brain disconnection observed in neurodegenerative diseases such as Alzheimer's, schizophrenia, and Parkinson's is detected through Graph Autoencoders (GAE) and hybrid architectures and is used as powerful digital biomarkers in disease diagnosis (Klepl et al., 2022; Zeng et al., 2024).

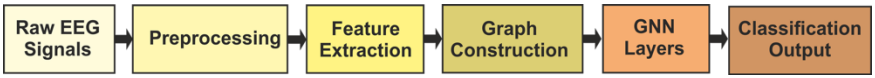
GNN models are used in areas involving temporal dynamics, such as sleep staging and mental workload estimation (Sarkis et al., 2024; Wang et al., 2025). All these applications demonstrate that GNN-based approaches in EEG analysis not only achieve classification success but also enhance the interpretability of brain functions.

Experimental Pipelines

The success of EEG-GNN models depends not only on the architectural designs used, but also on the overall data preparation and experimental setup. The noisy, non-stationary, and person-specific nature of EEG data necessitates careful design of the experimental pipeline. In this context, an integrated experimental pipeline scheme, which is considered standard in GNN-based EEG

analysis, spans from raw signal processing to the final classification decision and is presented in Figure 9.

Figure 9.General experimental framework proposed for GNN-based EEG analysis.



Data preprocessing is the first stage of the analysis and has a direct influence on model performance. Raw EEG signals are generally band-pass filtered (0.5–45 Hz) to exclude non-physiological activity, and notch filtering is used to remove power line noise.

In addition to these basic steps, techniques such as Independent Component Analysis (ICA) for separating blink and muscle-related artifacts, and Artifact Subspace Reconstruction (ASR) for suppressing sudden motion-induced disturbances, are widely used. These procedures are essential when constructing functional graphs, as uncorrected artifacts can produce correlations and unstable edge weights.

Following pre-processing, a sliding window approach is employed to segment the signal, converting distinct time windows into graph structures based on functional or structural connectivity. Structural graphs generally remain constant, whereas functional graphs based on metrics such as PLV or MI change from window to window, allowing shifts in brain state to be represented.

During model training, overfitting is a persistent issue. To reduce this risk, graph-specific regularization techniques such as node dropout and edge dropout are applied. These methods limit the model's dependence on individual electrodes or specific edges and improve robustness to topological noise. Training stability is often supported with learning-rate scheduling strategies, including cosine

annealing or warm restarts, which help prevent convergence to poor local minima.

When evaluating the performance of the obtained model, relying solely on the accuracy metric in EEG datasets where class imbalance is common can be misleading. Therefore, the use of metrics such as Macro-F1 and Cohen's Kappa is recommended. More importantly, for a fair comparison in the literature, the validation protocol used must be clearly specified. Subject-Dependent protocols, where the model is tested on data from the same subject, offer high performance but have limited generalizability. In contrast, Subject-Independent / Leave-One-Subject-Out protocols, where the tested subject is not included in the training set, provide the most realistic but challenging test environment for real-world BCI scenarios.

Conclusion

The usage of GNNs with EEG analysis is important stage in decoding neural activity. Unlike traditional deep learning architectures that ignore the disordered placement of electrodes, GNNs explicitly encode complex connection patterns between electrodes.

GNN models, which represent structural, functional, or adaptive dependencies on a standard graph, offer a modeling approach more closely aligned with the distributed nature of brain networks. (Ding et al., 2023; Jin et al., 2021; Kwak et al., 2020; C. Liu et al., 2024; Rahman et al., 2025). However, despite the clear theoretical advantages and the emergence of various architectures, work on real-world applications is not yet fully complete.

Several fundamental challenges limit the efficacy and generalizability of current methods. The primary difficulty lies in constructing the graph itself. Since the ground-truth connectivity of

the human brain is unknown and EEG signals are heavily distorted by volume conduction so graphs can be assumed to reflect physiological reality accurately. This uncertainty is exacerbated by signal noise and fluctuations in electrode impedance, which often lead to unstable graph estimates. The inherent variability in anatomy and cognitive strategies across individuals creates a significant bottleneck for cross-subject transferability.

Models trained on specific subjects frequently struggle to generalise, hindering the development of calibration-free BCIs. This issue is compounded by the non-stationary nature of EEG, where even minor environmental changes or electrode shifts can alter functional connectivity, degrading performance on unseen data.

Data variability, the field faces distinct modelling and interpretability hurdles. Effectively fusing the high temporal resolution of EEG with graph topology, while preserving discriminative time-frequency information, remains a complex engineering challenge. While techniques like subgraph discovery offer preliminary insights, establishing explainable models that map learned graph structures to validated neurophysiological mechanisms is essential for clinical adoption. Progress is also constrained by the data scarcity in public benchmarks, which are currently insufficient to support the training of high-capacity models comparable to those in other domains.

Overcoming these limitations, this involves integrating large-scale spatiotemporal Transformer architectures with GNNs. Leveraging global attention mechanisms to capture long-range dependencies more effectively than local message transmission alone. Graph propagation models are emerging as a principled approach to modeling information flow and organizing noisy signals, showing potential for a variety of tasks ranging from seizure analysis to data augmentation.

Future studies could expand the scope of analysis from a single brain center to multiple brain frameworks, making it possible to examine social individuals and collective cognition. Addressing the problem of non-stationary neural signals will require the development of continuous learning systems capable of adapting graph topologies in real time.

In conclusion, associating graphical representations with multimodal fundamental models can enable not only decision-making in EEG analysis but also explain why and how those decisions are made. Such an approach could allow EEG-based studies to produce more interpretable results from the perspectives of cognitive science and neuroergonomics.

References

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. *ArXiv Preprint ArXiv:1511.06448*.

Bhatti, U. A., Tang, H., Wu, G., Marjan, S., & Hussain, A. (2023). Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence. *International Journal of Intelligent Systems*, *2023*(1), 8342104.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems*, *29*.

Ding, Y., Robinson, N., Tong, C., Zeng, Q., & Guan, C. (2023). LGGNet: Learning from local-global-graph representations for brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(7), 9773–9786.

Graña, M., & Morais-Quilez, I. (2023). A review of Graph Neural Networks for Electroencephalography data analysis. *Neurocomputing*, *562*, 126901.

Hesprich, S., & Beardsley, S. (2019). Computational characterization of the cellular origins of electroencephalography. *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 352–355.

HH, J. (1958). The ten-twenty electrode system of the international federation. *Electroenceph Clin Neurophysiol*, *10*, 367–380.

Jin, J., Sun, H., Daly, I., Li, S., Liu, C., Wang, X., & Cichocki, A. (2021). A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 20–29.

Kalafatovich, J., Lee, M., & Lee, S.-W. (2022). Learning spatiotemporal graph representations for visual perception using eeg signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 97–108.

Klepl, D., He, F., Wu, M., Blackburn, D. J., & Sarrigiannis, P. (2022). EEG-based graph neural network classification of Alzheimer's disease: An empirical evaluation of functional connectivity methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 2651–2660.

Klepl, D., Wu, M., Member, S., He, F., & Member, S. (2024). *Graph Neural Network-Based EEG Classification : A Survey*. *32*, 493–503.

Kwak, Y., Song, W.-J., & Kim, S.-E. (2020). Graph neural network with multilevel feature fusion for EEG based brain-computer interface. *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 1–3.

Liu, C., Zhou, X., Wu, Y., Ding, Y., Zhai, L., Wang, K., Jia, Z., & Liu, Y. (2024). A comprehensive survey on eeg-based emotion recognition: A graph-based perspective. *ArXiv Preprint ArXiv:2408.06027*.

Liu, X.-Y., Wang, W.-L., Liu, M., Chen, M.-Y., Pereira, T., Doda, D. Y., Ke, Y.-F., Wang, S.-Y., Wen, D., & Tong, X.-G. (2025). Recent applications of EEG-based brain-computer-interface in the medical field. *Military Medical Research*, *12*(1), 14.

Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information*, *16*(3), 195.

Miri, M., Abootalebi, V., Saeedi-Sourck, H., Van De Ville, D., & Behjat, H. (2024). Spectral representation of EEG data using learned graphs with application to motor imagery decoding. *Biomedical Signal Processing and Control*, *87*, 105537.

Niedermeyer, E., & da Silva, F. H. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Qiu, J., Liang, J., Fan, X., Zhang, M., & He, Z. (2025). ELPG-DTFS: Prior-Guided Adaptive Time-Frequency Graph Neural Network for EEG Depression Diagnosis. *ArXiv Preprint ArXiv:2509.24860*.

Rahman, S. M. A., Khalil, M. I., Zhou, H., Guo, Y., Ding, Z., Gao, X., & Zhang, D. (2025). Advancement in Graph Neural Networks for EEG Signal Analysis and Application: A Review. *IEEE Access*.

Raj, V. A., Parupudi, T., Thalengala, A., & Nayak, S. G. (2025). A comprehensive review of deep learning models for denoising EEG signals: challenges, advances, and future directions. *Discover Applied Sciences*, *7*(11), 1268.

Sarkis, M., Rizkallah, M., & Moussaoui, S. (2024). A graph signal processing framework based on graph learning and graph neural networks for mental workload classification from eeg signals. *7th Graph Signal Processing Workshop (GSP)*.

Sun, Q., Wei, X., & Yang, X. (2024). GraphSAGE with deep reinforcement learning for financial portfolio optimization. *Expert Systems with Applications*, *238*, 122027.

Tang, M., Li, B., & Chen, H. (2023). Application of message passing neural networks for molecular property prediction. *Current Opinion in Structural Biology*, *81*, 102616.

Värbu, K., Muhammad, N., & Muhammad, Y. (2022). Past, present, and future of EEG-based BCI applications. *Sensors*, *22*(9), 3331.

Vecchio, F., Miraglia, F., & Rossini, P. M. (2017). Connectome: Graph theory application in functional brain network architecture. *Clinical Neurophysiology Practice*, *2*, 206–213.

Wang, C., Jiang, X., Lv, C., Meng, Q., Zhao, P., Yan, D., Feng, C., Xu, F., Lu, S., & Jung, T.-P. (2025). GraphSleepFormer: a multi-modal graph neural network for sleep staging in OSA patients. *Journal of Neural Engineering*, *22*(2), 26011.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629.

Zeng, Y., Lin, J., Li, Z., Xiao, Z., Wang, C., Ge, X., Wang, C., Huang, G., & Liu, M. (2024). Adaptive node feature extraction in graph-based neural networks for brain diseases diagnosis using self-supervised learning. *NeuroImage*, *297*, 120750.

Zhang, H., Zhou, Q.-Q., Chen, H., Hu, X.-Q., Li, W.-G., Bai, Y., Han, J.-X., Wang, Y., Liang, Z.-H., & Chen, D. (2023). The

applied principles of EEG analysis methods in neuroscience and clinical neurology. *Military Medical Research*, *10*(1), 67.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81.

SYNTHETIC BIOMEDICAL IMAGE GENERATION USING DEEP GENERATIVE MODELS: METHODS, APPLICATIONS, AND CHALLENGES

SEDA ŞAŞMAZ KARACAN[8]
EVİN ŞAHİN SADIK[9]

Introduction

Access to large-scale, balanced datasets required for training deep learning models in medical imaging is a significant challenge in both engineering and medicine. Obtaining sufficient labeled data is particularly difficult in areas such as the classification of rare diseases or the analysis of specific anatomical regions. Data insufficiency limits the generalization capacity of models, making it difficult to produce reliable results in terms of model robustness and integration into clinical settings. This situation increases the tendency of deep learning models to overfit, while causing performance imbalances across different class populations.

[8] Lecturer Dr., Usak University, Department of Information Technology, Orcid: 0000-0002-0334-260X
[9] Asst. Prof. Dr., Kutahya Dumlupinar University, Department of Electrical and Electronics Engineering, Orcid: 0000-0002-2212-4210

Synthetic data generation emerges as a critical solution at this point. This is because, when real data is limited, fake examples can be provided to increase the model's learning capacity and enrich its variation. Furthermore, the use of biomedical data involves ethical responsibilities in addition to technical challenges. Requirements such as informed consent, data sharing permission, and clinical ethics committee approval must be met for the use of images obtained from patients in scientific studies. These ethical constraints prolong the data access process, delaying scientific studies and disrupting research. Furthermore, differing personal data protection laws across countries necessitate the management of complex procedures in interdisciplinary studies. From a patient privacy perspective, the direct use of medical data containing images carries the risk of identity disclosure. The direct use of images of the face, skin, and organ structures, in particular, carries risks in terms of patient confidentiality. Deep learning-based synthetic image generation techniques have the potential to reduce these risks thanks to advancing technologies. While protecting patient confidentiality, data sharing among scientific research communities also becomes more secure.

In summary, insufficient data, ethical constraints, and patient privacy concerns are among the main obstacles to deep learning applications. Synthetic images created using deep learning methods offer an innovative approach by addressing all these concerns simultaneously. By using synthetic data, large-scale datasets with class balance can be created that do not contain real patient data. The synthetic data generation method eliminates concerns about the use of real patient data, offering a sustainable solution that is compatible with data privacy and ethical compliance. Therefore, research on creating synthetic medical images using deep learning applications is increasing day by day (Chen et al., 2023; Rusanov et al., 2022;

Sherwani & Gopalakrishnan, 2024; Wang et al., 2021). Some recent studies on generating synthetic medical images are listed below:

Waqas and colleagues aimed to produce synthetic images using 4777 real orthopantomography (OPG) scans. They obtained YOLO-based object detection and expert dentist evaluations for anatomical and diagnostic realism to examine the performance of the synthetic images they produced using Generative Adversarial Networks (GAN) (Waqas et al., 2025).

Zotova and colleagues aimed to generate synthetic fluorodeoxyglucose (FDG) PET images using 35 T1-weighted MRI images. They used the generated synthetic PET images to train an unsupervised anomaly detection (UAD) model. They detected epilepsy lesions in the synthetic PET images generated using the GAN-based model. Since their proposed model was trained on images from healthy individuals, it could only distinguish patterns that deviated from the normal population pattern. In the performance evaluation of synthetic images, the structural similarity index (SSIM) value was approximately 0.9, while the best UAD model trained with synthetic PET data achieved 74% sensitivity (Zotova et al., 2025).

Luschi and colleagues aimed to generate synthetic images using dermoscopic melanoma images. They generated 25 synthetic melanoma images using a previously trained StyleGAN2 model. The synthetic images were evaluated by 17 dermoscopy experts using a Likert scale that included parameters such as real/synthetic discrimination, skin texture, visual realism, confidence level, and pattern analysis. A 64% accuracy rate was achieved in distinguishing real from synthetic images (Luschi et al., 2025).

Baste and colleagues aimed to generate synthetic images using tuberculosis CT scans in their work. They generated synthetic images using Deep Convolutional GAN (DCGAN) and CycleGAN

models. They used the images they generated both for segmentation and for training classification models. They used the U-Net architecture for segmentation and deep learning networks for disease severity classification. The images generated with DCGAN achieved an FID of 48.7. They classified tuberculosis severity with a precision of 0.84 (Baste et al., 2025).

In addition to synthetic biomedical image generation, studies have been conducted to improve image quality (Rusanov et al., 2022) through techniques such as attenuation correction for PET images, and to investigate the impact of training dataset size and deep learning models on synthetic medical image generation (Sherwani & Gopalakrishnan, 2024). These include work ranging from low-dose CT to high-dose CT, from cone-beam CT to synthetic CT generation (Boily, Mazellier, & Meyer, 2025). Furthermore, studies on text-to-image generation (Ni, Zhang, Zhou, Hou, & Gao, 2020), layout-to-image (Yang, Liu, Wang, Yang, & Tao, 2022), generative inpainting (Xu, Liu, & Xiong, 2021), edge-based image generation (Ruffino, Hérault, Laloy, & Gasso, 2020), scene graphs (Zhang & Zhao, 2021), and sketch-based image generation (Yu, Zhu, Wang, Wang, & Gao, 2023) are also present in the literature (Zulfiqar et al., 2024).

The second section of this book chapter discusses the most commonly used synthetic image generation models, while the third section covers metrics used to evaluate the performance of synthetic images. The fourth section addresses challenges and ethical concerns in synthetic image generation, and the final section discusses future work and conclusions.

Synthetic Image Generation Models

Models frequently used in synthetic image generation techniques can generally be grouped into three categories, from the

simplest to the most complex (Wang et al., 2021; Zulfiqar et al., 2024):

- Auto-Encoder (AE)
- Generative Adversarial Networks (GANs)
- Diffusion Models (DM)

Figure 1. General workflow for synthetic medical image generation

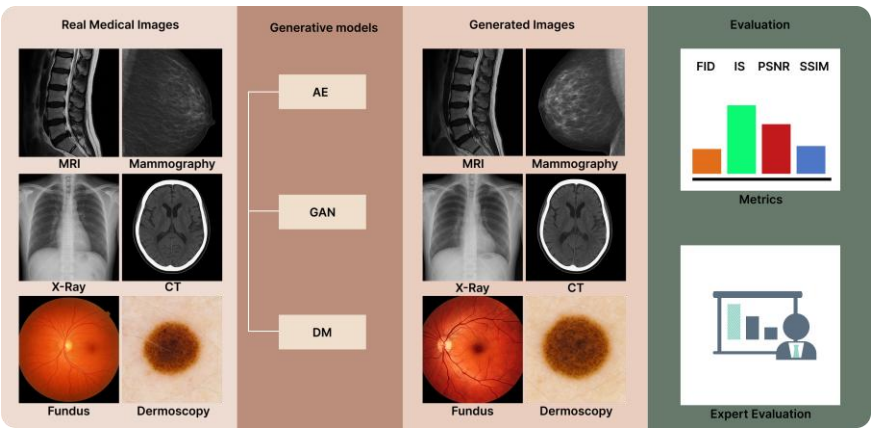


Figure 1 shows the general steps of synthetic medical image generation involving AE, GANs, and DM.

Auto-encoder

Autoencoders (AEs) form the methodological basis of deep learning-based medical image synthesis studies. AEs are a class of deep neural networks that use convolutional kernels to discover spatially local image patterns (Ehrhardt & Wilms, 2022; Rais, Amroune, Benmachiche, & Haouam, 2024). Figure 2 shows a basic AE structure.

Figure 2. Basic AE structure

It consists of an input layer, an output layer, and multiple hidden layers. The most commonly used activation layer is the Rectified Linear Unit (ReLU). ReLU provides computational simplicity, representational sparsity, and linearity. Additionally, batch normalization is used to reduce the internal variance of training datasets. To save memory, the image size is typically reduced using pooling layers and convolutional layers. The ultimate goal of the network is to minimize the output of a loss function.

Variants: Instead of the basic AE architecture, variants such as Residual Neural Network (ResNet), which facilitates the training of deep networks by skipping layers, are used for better performance. AE is also a fundamental component of more advanced architectures such as U-net and GANs.

U-net: It emerged by modifying the AE architecture to synthesize CT images from MR images in medical image synthesis (Han, 2017).

Structure: It consists of an encoder and a decoder part. The encoder extracts hierarchical features, while the decoder reconstructs the predicted images.

Skip Connections: These connections help preserve high-resolution features during reconstruction by linking simple features from early layers in the encoder to late layers in the decoder.

For a simple AE, the encoder, decoder, and loss function equations are given in Equations 1, 2, and 3, respectively.

Encoder:

$$z = f_\theta(x) \quad (1)$$

$x \in \mathbb{R}^n$: input image

$z \in \mathbb{R}^m$: latent vector

f_θ: the θ-parameterized encoder network

Decoder:

$$\hat{x} = g_{\varphi}(z) \tag{2}$$

g_{φ}: the φ-parameterized decoder network

$\hat{x}$: reconstructed image

Loss Function:

$$L_{AE} = \|x - \hat{x}\|^2 = \left\|x - g_{\varphi}(f_{\theta}(x))\right\|^2 \tag{3}$$

However, classical AE is insufficient for synthetic image generation because it lacks probabilistic meaning in latent space. Therefore, more advanced models such as Variational Autoencoder (VAE) have been developed.

The probabilistic encoder, which estimates the probability distribution instead of producing a single z vector, the reparameterization trick, the probabilistic decoder, and the equations for the loss function are given in Equations 4, 5, 6, and 7, respectively.

Probabilistic Encoder:

$$q_{\theta}(z|x) = \mathcal{N}(z;\ \mu_{\theta}(x), \sum_{\theta}(x)) \tag{4}$$

$\mu_{\theta}(x)$: mean

$\sum_{\theta}(x)$: variance vector

$q_{\theta}(z|x)$: posterior distribution

Reparameterization trick:

$$z = \mu_{\theta}(x) + \sum_{\theta}^{1/2}(x) \odot \epsilon, \epsilon \sim \mathcal{N}(0, I) \tag{5}$$

Probabilistic Decoder:

$$p_\theta(x|z) = \mathcal{N}(x;\ g_\phi(z), \sigma^2 I)$$
$$\hat{x} = g_\phi(z) \tag{6}$$

VAE loss;

$$\mathcal{L}_{VAE} = \mathbb{E}_{q_{\theta(z|x)}}\left[-\log p_\phi(x|z)\right] + D_{KL}(q_\theta(z|x)\|p(z)) \tag{7}$$

Reconstruction loss: measures how well the model reproduces the original.

Kullback-Leibler (KL) divergence: approximates the latent space to a standard normal distribution.

$p(z) = \mathcal{N}(0, I)$ prior distribution for latent space

VAE Generation process:

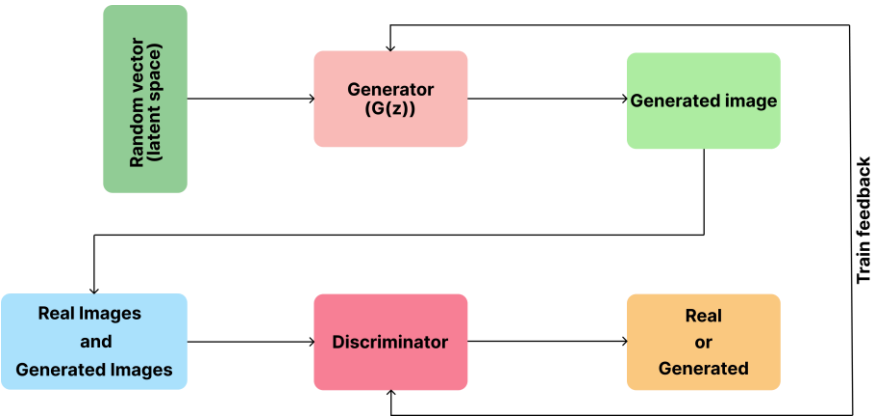
As shown in Equation 8, a sample is drawn from the latent space, fed to the decoder, and a synthetic image is generated.

$$z \sim \mathcal{N}(0, I), \qquad \hat{x} = g_\phi(z) \tag{8}$$

Generative Adversarial Networks

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are one of the popular deep learning models frequently used for synthetic image generation, whose popularity is increasing day by day. Figure 3 shows a basic GAN structure.

Figure 3. Basic GAN structure



A GAN consists of two neural networks trained simultaneously: a Generator and a Discriminator. The Generator takes random noise and converts it into a representative synthetic image that resembles real data in order to deceive the Discriminator. The Discriminator, on the other hand, is trained to distinguish whether the incoming image is real or generated (fake). This adversarial process continues until a balance is reached where the generator produces synthetic images that are as realistic as possible and the discriminator struggles to distinguish them. While the Generator can be an AE or a U-net, the Discriminator is typically an AE.

The generator (G) produces synthetic images that resemble reality from random noise (z), while the discriminator (D) attempts to distinguish whether the image is real or fake. The training process is a min-max problem, as given in Equation 9.

$$\min G \max V\ (D, G) \tag{9}$$

The discriminator aims for D(x)≈1 for real images and D(G(z))≈0 for fake images generated by the generator. This optimization problem for the discriminator is given in Equation 10.

$$\max D\ \mathbb{E}_{\mathrm{x}\sim \mathrm{P}_{\mathrm{data}\ (\mathrm{x})}}[\log \mathrm{D}(\mathrm{x})] + \mathbb{E}_{\mathrm{z}\sim \mathrm{P}_{\mathrm{z}\ (\mathrm{z})}}\left[\log\left(1 - \mathrm{D}\left(\mathrm{G}(\mathrm{z})\right)\right)\right]$$

(10)

The purpose of the generator is to make G(z) appear real, as shown in Equation 11.

$$\mathrm{D}\big(\mathrm{G}(\mathrm{z})\big) \to 1$$

(11)

Equation 12 provides a saturating loss function, but in practice, the non-saturating loss in Equation 13 is used.

$$\min G\ \mathbb{E}_{\mathrm{z}\sim \mathrm{P}_{\mathrm{z}\,(\mathrm{z})}}\big[\log\,(1-\mathrm{D}\big(\mathrm{G}(\mathrm{z})\big))\big]$$

(12)

$$\max G\ \mathbb{E}_{\mathrm{z}\sim \mathrm{P}_{\mathrm{z}\,(\mathrm{z})}}\big[\log\ \mathrm{D}\big(\mathrm{G}(\mathrm{z})\big)\big]$$

(13)

The exact game formulation is given in Equation 14:

$$\min G \max V\ (D,G) = \mathbb{E}_{\mathrm{x}\sim \mathrm{P}_{\mathrm{x}\,(\mathrm{data})}}[\log\ \mathrm{D}(\mathrm{x})] + \mathbb{E}_{\mathrm{z}\sim \mathrm{P}_{\mathrm{z}\,(\mathrm{z})}}[\log\,(1-\mathrm{D}(\mathrm{G}(\mathrm{z})))]$$

(14)

The main GAN variants can be listed as follows:

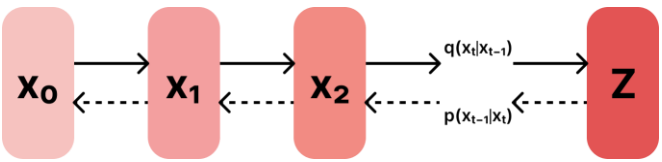
- Conditional GAN (cGAN): Unlike standard unconditional GANs, both the generator and discriminator networks observe conditional inputs such as class labels, textual descriptions, layout information, or edge information. This ensures that the network produces images that meet specific requirements. cGAN is more suitable for image-to-image translation tasks.
- Deep Convolutional GAN (DCGAN): Used to synthesize realistic images from random noise vectors.
- CycleGAN: Effective when unpaired training datasets are unavailable. It adds an inverse mapping network using a cycle consistency loss to perform image translation between two domains (e.g., CBCT/CT).

- Others: Models such as StackGAN and PGGAN for high-resolution image synthesis; AttnGAN, ControlGAN, DM-GAN, DF-GA, MirrorGAN for text-to-image synthesis; and StarGAN and GAN+DA for domain adaptation are also found in the literature.

Diffusion Models

Diffusion models are another important technique used for both conditional and unconditional image generation. Figure 4 shows the structure of a basic diffusion model.

Figure 4. Basic diffusion model structure



It is based on the principle of gradually diffusing examples containing random noise into the data. Once the diffusion process during noise addition is modeled and its inverse is learned, diffusion models are used to generate new images by denoising these images.

The Denoising Diffusion Implicit Model (DDIM) (Song, Meng, & Ermon, 2022) is an improved variant of the basic model, Denoising Diffusion Probabilistic Models (DDPM) (Ho, Jain, & Abbeel, 2020). DDIM is a deterministic variant of the random backward process in DDPM. In DDIM, the same noise produces the same output in the backward process.

There are two processes in a diffusion model: Forward and Reverse. In the Forward process, noise is added to the image as shown in Equation 15.

Noising process:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}_{x_{t-1}}, \beta_t I) \quad (15)$$

x_0: original image
x_t: Noisy image at t time
β_t: noise level per step

The noisy version of the image at time t is as shown in Equation 16:

$$x_t = \sqrt{\bar{a}_t}x_0 + \sqrt{1-\bar{a}_t}\varepsilon\,,\ \varepsilon \sim \mathcal{N}(0,I) \tag{16}$$

The reverse process that occurs to gradually remove the added noise is as in Equation 17 for DDPM, while in DDIM, the randomness term (z) is removed, making it as in Equation 18.

DDPM reverse equation:

$$x_{t-1} = \frac{1}{\sqrt{a_t}}\left(x_t - \frac{1-a_t}{\sqrt{1-\bar{a}_t}}\right)\epsilon_\theta(x_t,t) + \sigma_t z \tag{17}$$

$\epsilon_\theta(x_t,t)$: estimated noise learned at each step
$z \sim \mathcal{N}(0,I)$: added randomness
σ_t: noise level per step

In DDIM, the stochasticity in DDPM has been removed. The equality is as in Equation 18:

DDIM:

$$x_{t-1} = \sqrt{\bar{a}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{a}_t}\epsilon_\theta(x_t,t)}{\sqrt{\bar{a}_t}}\right) + \sqrt{1-\bar{a}_{t-1}}\epsilon_\theta(x_t,t) \tag{18}$$

Evaluation Metrics

Various parameters are used to examine the quality of synthetic images produced by deep learning models.

Consulting expert opinions for clinical realism is a frequently used method in the literature. Studies where synthetic dermoscopy images are evaluated by experts (Luschi et al., 2025), studies where synthetic lung X-ray images are evaluated by experts (Segal, Rubin, Rubin, & Pantanowitz, 2021), studies evaluated with the Turing test

Jang et al., 2023), studies where intraoral synthetic images are evaluated by experts (Kokomoto, Okawa, Nakano, & Nozaki, 2021), and studies where synthetic images are reclassified using classification methods are available. In addition to these methods, metrics such as Fréchet Inception Distance (FID), Inception Score (IS), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) are used to indicate image quality.

Fréchet Inception Distance

Fréchet Inception Distance (FID) is a metric that measures the similarity of feature distributions between real images and synthetic images ("GANs Trained by a Two Time-Scale Update Rule," 2017). A low FID indicates that the generated synthetic image is close to the real one, while a high FID indicates that the synthetic image is far from the real one. The formula is given in Equation 19:

$$\mathrm{FID} = \left\|\mu_r - \mu_g\right\|_2^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{19}$$

μ_r, Σ_r: The mean and covariance matrix of real images in feature space

μ_g, Σ_g: The mean and covariance matrix of synthetic images.

Tr(.): Trace operator

$\| \, . \, \|_2^2$: The square of the L2 norm

Inception Score

Inception score (IS) is a metric that measures the diversity and recognizability of synthetic images (Salimans et al., 2016). A high ISI indicates that the image is diverse and meaningful. Its formula is given in Equation 20:

$$IS = \exp\left(\mathbb{E}_{x\sim p_g(x)}[D_{KL}p(y|x)\|p(y))]\right) \tag{20}$$

$p(y|x)$: The class distribution predicted by the Inception model for image x.

$p(y) = \int_x p(y|x)\, p_g(x)dx$; Average class distribution across all produced samples.

D_{KL}: Kullback-Leibler divergence.

Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a metric that measures the difference between synthetic and real images at the pixel level (Horé & Ziou, 2010). A high PSNR value indicates that the synthetic and real images are very similar. The formula is given in Equation 21:

$$PSNR = 10log_{10}\frac{MAX_I^2}{MSE}$$
$$\text{MSE: } \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(I_{ij} - K_{ij})^2 \tag{21}$$

MAX_I: The maximum possible pixel value in the image.

MSE: Mean square error between real (I) and synthetic (K) images.

Structural Similarity Index Measure

Structural Similarity Index Measure (SSIM) is a metric that measures the structural similarity (luminance, contrast, structure) between a synthetic image and a real image (Wang, Bovik, Sheikh, & Simoncelli, 2004). As the SSIM value approaches 1, the two images become more similar. Equation 22 provides the SSIM formula.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)}$$

(22)

μ_x, μ_y: mean intensity

${\sigma_x}^2, {\sigma_y}^2$: variance (measure of contrast)

σ_{xy}: covariance (measure of structural similarity)

C_1, C_2: small constants.

Challenges and Ethical Considerations

Challenges in synthetic image generation can be categorized as technical challenges and ethical issues. Technical challenges can be listed as follows:

- Training models with limited data
- Data sets consisting of different devices, protocols, and patient demographics
- The mode collapse problem occurring in GANs
- Models memorizing and overfitting to the structure of real patients due to small data sets
- Anatomical consistency problems caused by the distortion of organ-specific geometric structures
- Inability to fully represent clinically meaningful features
- Inconsistent contrast differences and noise characteristics in multimodal images such as MRI, CT, and PET
- Limited clinical validity of metrics such as FID and IS for medical images
- Costly and time-consuming expert evaluation
- The clinical “ground truth” definition is often unclear.

- Risks that may arise during the integration of generated images into clinical workflows.

Challenges related to ethical and regulatory issues can be listed as follows:

- The risk of leaking information belonging to real patients in synthetic images
- The risk that synthetic data may reinforce biases due to demographic imbalances in training data (e.g., race, age, gender)
- Validation requirements for the use of synthetic data in clinical decision support systems.
- The danger of synthetic images being used for manipulation purposes.
- In addition to clearly indicating that they are synthetic, the limits of their use in clinical research must be defined.

Future Studies and Conclusion

As mentioned in this chapter, synthetic image generation has become essential for biomedical research due to both data requirements and technical considerations. Overcoming the technical and ethical challenges encountered during synthetic image generation will enable the creation of large-scale datasets for deep learning models. Future studies may focus on topics such as how synthetic images can be made completely anonymous, what the optimal ratio of synthetic data should be in clinical models, whether synthetic data truly increases generalization, and investigating the performance of models other than GANs, such as diffusion models, in synthetic image generation.

Acknowledgment

The MRI, mammography, X-ray, CT, fundus, and dermoscopic images shown in Figure 1 were generated using ChatGPT.

References

Baste, V. S., Shrivastava, P., Patil, R., Rana, N., Kale, P. P., & Nerkar, B. B. (2025). Enhancing tuberculosis CT imaging analysis through synthetic data augmentation via deep adversarial models. Indian Journal of Tuberculosis. https://doi.org/10.1016/j.ijtb.2025.10.007

Boily, C., Mazellier, J.-P., & Meyer, P. (2025). Large medical image database impact on generalizability of synthetic CT scan generation. Computers in Biology and Medicine, 193, 110303. https://doi.org/10.1016/j.compbiomed.2025.110303

Chen, J., Chen, S., Wee, L., Dekker, A., & Bermejo, I. (2023). Deep learning based unpaired image-to-image translation applications for medical physics: A systematic review. *Physics in Medicine & Biology, 68*(5), 05TR01. https://doi.org/10.1088/1361-6560/acba74

Ehrhardt, J., & Wilms, M. (2022). Autoencoders and variational autoencoders in medical image analysis. In N. Burgos & D. Svoboda (Eds.), Biomedical Image Synthesis and Simulation (pp. 129–162). Academic Press. https://doi.org/10.1016/B978-0-12-824349-7.00015-3

"GANs trained by a two time-scale update rule converge to a local Nash equilibrium." (2017). arXiv:1706.08500. https://arxiv.org/abs/1706.08500

Goodfellow, I. J., et al. (2014). Generative adversarial networks. arXiv:1406.2661. https://doi.org/10.48550/arXiv.1406.2661

Han, X. (2017). MR-based synthetic CT generation using a deep convolutional neural network method. Medical Physics, 44(4), 1408–1419. https://doi.org/10.1002/mp.12155

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. arXiv:2006.11239. https://doi.org/10.48550/arXiv.2006.11239

Horé, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. In Proceedings of the 20th International Conference on Pattern Recognition (pp. 2366–2369). https://doi.org/10.1109/ICPR.2010.579

Jang, M., et al. (2023). Image Turing test and its applications on synthetic chest radiographs by using the progressive growing generative adversarial network. Scientific Reports, 13(1), 2356. https://doi.org/10.1038/s41598-023-28175-1

Kokomoto, K., Okawa, R., Nakano, K., & Nozaki, K. (2021). Intraoral image generation by progressive growing of generative adversarial network and evaluation of generated image quality by dentists. Scientific Reports, 11(1), 18517. https://doi.org/10.1038/s41598-021-98043-3

Luschi, A., et al. (2025). Design and development of a systematic validation protocol for synthetic melanoma images for responsible use in medical artificial intelligence. Biocybernetics and Biomedical Engineering, 45(4), 608–616. https://doi.org/10.1016/j.bbe.2025.09.001

Luschi, A., Tognetti, L., Cartocci, A., Cevenini, G., Rubegni, P., & Iadanza, E. (2025). Advancing synthetic data for dermatology: GAN comparison with multi-metric and expert validation approach. Health Technology, 15(3), 553–562. https://doi.org/10.1007/s12553-025-00971-x

Ni, J., Zhang, S., Zhou, Z., Hou, J., & Gao, F. (2020). Instance mask embedding and attribute-adaptive generative adversarial network for text-to-image synthesis. IEEE Access, 8, 37697–37711. https://doi.org/10.1109/ACCESS.2020.2975841

Rais, K., Amroune, M., Benmachiche, A., & Haouam, M. Y. (2024). Exploring variational autoencoders for medical image generation: A comprehensive study. arXiv:2411.07348. https://doi.org/10.48550/arXiv.2411.07348

Ruffino, C., Hérault, R., Laloy, E., & Gasso, G. (2020). Pixel-wise conditioned generative adversarial networks for image synthesis and completion. Neurocomputing, 416, 218–230. https://doi.org/10.1016/j.neucom.2019.11.116

Rusanov, B., et al. (2022). Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: A systematic review. Medical Physics, 49(9), 6019–6054. https://doi.org/10.1002/mp.15840

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. arXiv:1606.03498. https://doi.org/10.48550/arXiv.1606.03498

Segal, B., Rubin, D. M., Rubin, G., & Pantanowitz, A. (2021). Evaluating the clinical realism of synthetic chest X-rays generated using progressively growing GANs. SN Computer Science, 2(4), 321. https://doi.org/10.1007/s42979-021-00720-7

Sherwani, M. K., & Gopalakrishnan, S. (2024). A systematic literature review: Deep learning techniques for synthetic medical image generation and their applications in radiotherapy. Frontiers in Radiology, 4. https://doi.org/10.3389/fradi.2024.1385742

Song, J., Meng, C., & Ermon, S. (2022). Denoising diffusion implicit models. arXiv:2010.02502. https://doi.org/10.48550/arXiv.2010.02502

Wang, T., et al. (2021). A review on medical imaging synthesis using deep learning and its clinical applications. Journal of

Applied Clinical Medical Physics, 22(1), 11–36. https://doi.org/10.1002/acm2.13121

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Waqas, M., Hasan, S., Ghori, A. F., Alfaraj, A., Faheemuddin, M., & Khurshid, Z. (2025). Synthetic orthopantomography image generation using generative adversarial networks for data augmentation. International Dental Journal, 75(6), 103878. https://doi.org/10.1016/j.identj.2025.103878

Xu, S., Liu, D., & Xiong, Z. (2021). E2I: Generative inpainting from edge to image. IEEE Transactions on Circuits and Systems for Video Technology, 31(4), 1308–1322. https://doi.org/10.1109/TCSVT.2020.3001267

Yang, Z., Liu, D., Wang, C., Yang, J., & Tao, D. (2022). Modeling image composition for complex scene generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7754–7763). https://doi.org/10.1109/CVPR52688.2022.00761

Yu, W., Zhu, M., Wang, N., Wang, X., & Gao, X. (2023). An efficient transformer based on global and local self-attention for face photo-sketch synthesis. IEEE Transactions on Image Processing, 32, 483–495. https://doi.org/10.1109/TIP.2022.3229614

Zhang, L., & Zhao, L. (2021). High-quality face image generation using particle swarm optimization-based generative adversarial networks. Future Generation Computer Systems, 122, 98–104. https://doi.org/10.1016/j.future.2021.03.022

Zotova, D., Pinon, N., Trombetta, R., Bouet, R., Jung, J., & Lartizien, C. (2025). GAN-based synthetic FDG PET images from

T1 brain MRI can serve to improve performance of deep unsupervised anomaly detection models. Computer Methods and Programs in Biomedicine, 265, 108727. https://doi.org/10.1016/j.cmpb.2025.108727

Zulfiqar, A., Muhammad Daudpota, S., Shariq Imran, A., Kastrati, Z., Ullah, M., & Sadhwani, S. (2024). Synthetic image generation using deep learning: A systematic literature review. Computational Intelligence, 40(5), e70002. https://doi.org/10.1111/coin.70002

HYBRID ARTIFICIAL INTELLIGENCE MODELS:

METHODS, OPTIMIZATION, AND APPLICATIONS