International Journal of
Energy and Thermal Applications

Volume 4, Issue 1, 2026
January–April
DOI (Journal): 10.37591/IJETA

https://journals.stmjournals.com/ijeta

Research IJETA

Development of a Generative AI Model for Early Detection and Prevention of Electrical Faults in Thermal Power Plants

Shreekantrao[1,]*, Nagendra Kumar Swarnkar[2]

Abstract

Electrical faults in thermal power plants can lead to severe equipment damage, production downtime, and safety hazards if not detected in advance. This study presents the development of a Generative Artificial Intelligence (GenAI) model for the early detection and prevention of electrical faults using predictive analytics. The proposed framework integrates Generative Adversarial Networks (GANs) with deep learning (CNN) and machine learning algorithms (Random Forest, Logistic Regression) to enhance data diversity, improve classification accuracy, and ensure early fault diagnosis. A comprehensive dataset representing operational parameters—such as voltage, current, temperature, vibration level, load factor, and power factor—was utilized for model training and evaluation. Synthetic data generated through the GAN improved fault representation and model generalization. Experimental results demonstrate that the proposed GAN-CNN hybrid model achieved the highest accuracy of 98.3%, outperforming traditional methods in terms of precision, recall, and F1-score. The feature importance analysis revealed that voltage, current, and temperature were the most influential parameters in predicting potential faults. The findings confirm that the developed Generative AI framework provides a robust, data-driven, and proactive solution for predictive maintenance and fault prevention in thermal power plants. This approach not only reduces unplanned downtime but also enhances system reliability, safety, and operational efficiency — marking a significant step toward intelligent and self-healing power systems.

Keywords: Deep learning, fault diagnosis, generative artificial intelligence, predictive maintenance, thermal power plants

INTRODUCTION

The increasing demand for uninterrupted power supply and the growing complexity of electrical networks in thermal power plants have made fault detection and preventive maintenance a critical area of research [1, 2]. Electrical faults such as overvoltage, overcurrent, insulation breakdown, and overheating are among the primary causes of system failures, leading to substantial financial losses and potential safety hazards [3]. Conventional maintenance strategies primarily reactive or time-based often fail to anticipate failures, resulting in unplanned outages and reduced equipment lifespan [4]. This has necessitated a shift toward intelligent, data-driven predictive maintenance systems capable of detecting faults before they escalate.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized industrial fault diagnosis by enabling pattern recognition, anomaly detection, and predictive analytics using large-scale sensor data [5, 6]. However, most existing approaches are

***Author for Correspondence**
Shreekantrao
E-mail: shreekant.23182562@mygyanvihar.com

[1]Research Scholar, Department of Electrical Engineering, Suresh Gyan Vihar University, India
[2]Professor, Department of Electrical Engineering, Suresh Gyan Vihar University, India

Received Date: February 10, 2026
Accepted Date: February 27, 2026
Published Date: April 15, 2026

Citation: Shreekantrao, Nagendra Kumar Swarnkar. Development of a Generative AI Model for Early Detection and Prevention of Electrical Faults in Thermal Power Plants. International Journal of Energy and Thermal Applications. 2026; 4(1): 45–54p.

constrained by the limited availability of diverse fault datasets and the inability to generalize across varying operating conditions. To address this challenge, Generative Artificial Intelligence (Generative AI) has emerged as a transformative solution [7-9]. By simulating and augmenting realistic fault data, GANs enhance model robustness and ensure better representation of rare or complex fault events.

Electrical faults in thermal power plants can result in significant downtime, equipment damage, and safety hazards. Recently, research has focused on predictive maintenance and intelligent fault diagnosis using advanced data-driven approaches [10]. This section reviews relevant studies in machine learning, deep learning, and generative AI for fault detection and predictive maintenance. Traditional machine learning techniques, such as Random Forests, Support Vector Machines (SVM), and Logistic Regression, have been widely applied for electrical fault classification [11, 12]. These methods are effective in detecting common faults and provide interpretable results. However, they often struggle with imbalanced datasets and complex non-linear relationships between multiple operational parameters, which can reduce their ability to predict rare or critical faults.

Deep learning models, particularly CNNs and Recurrent Neural Networks (RNNs), are effective in handling large-scale, multivariate sensor data. These models capture non-linear and temporal dependencies, allowing accurate classification of multiple fault types [13, 14]. Despite their high predictive accuracy, deep learning models require large and diverse datasets for training, and their performance may degrade when fault data is limited or imbalanced. GANs have emerged as a solution to overcome limited fault data. GANs can synthesize realistic fault scenarios, thereby improving model generalization and robustness [15]. Synthetic data augmentation enhances the ability of deep learning models to detect rare faults and reduces overfitting, enabling more reliable predictions in industrial environments.

Despite significant advances in machine learning, deep learning, and generative AI, several gaps remain in electrical fault detection, including limited multi-parameter analysis that often overlooks interactions between voltage, current, temperature, vibration, and load conditions, insufficient examples of rare faults that reduce predictive reliability, lack of real-time applicability in industrial environments, and minimal integration of generative models with deep learning classifiers to enhance both data diversity and fault prediction accuracy. The present study addresses these limitations by proposing a hybrid Generative AI framework (GAN-CNN) for thermal power plant fault detection, combining synthetic data generation, deep feature extraction, and multi-model evaluation.

This approach improves the detection of both frequent and rare faults, incorporates multiple operational parameters for accurate prediction, and enables real-time, data-driven preventive maintenance, thereby advancing existing methodologies and offering enhanced predictive reliability and practical applicability in modern thermal power plants.

This research presents the development of a hybrid GAN-CNN to achieve enhanced fault prediction accuracy in thermal power plants. It implements predictive analytics by leveraging electrical, thermal, and mechanical parameters to detect early fault patterns, while providing a comprehensive evaluation and comparison with traditional models such as Random Forest, Logistic Regression, and CNN to validate its effectiveness. The study also includes feature importance and interpretability analysis to identify the most critical parameters influencing fault prediction. By demonstrating improved reliability and operational efficiency, the proposed framework effectively reduces downtime and maintenance costs through early fault prevention. Overall, this research bridges the gap between conventional fault detection methods and intelligent predictive systems by utilizing the generative capabilities of AI to model real-world fault conditions more accurately, contributing to the development of self-monitoring and self-healing power systems and paving the way toward the next generation of smart, sustainable thermal power plants.

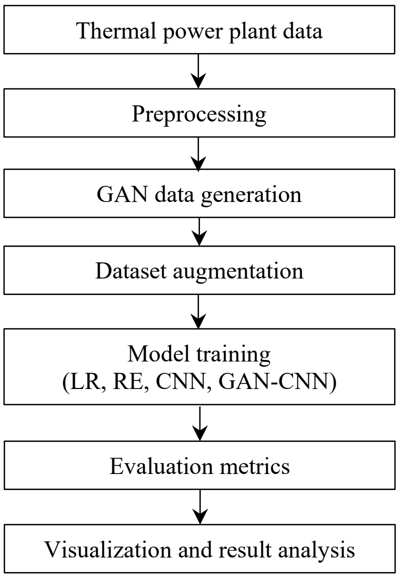

Figure 1. Workflow of the proposed methodology.

METHODOLOGY

The proposed methodology focuses on the development of a Generative AI-based predictive framework for the early detection and prevention of electrical faults in thermal power plants. The approach combines data-driven learning, synthetic data generation, and deep learning-based classification to enhance reliability, accuracy, and generalization. Figure 1 illustrates the overall workflow of the proposed system.

A thermal power fault dataset called thermal power plant subsystems was also prepared as a dataset called thermal power fault-data set.csv, with various electrical and environmental data derived on thermal power plant subsystems. The characteristics of the dataset are voltage, current, temperature, level of vibration, load factor, power factor, remaining useful life (RUL), and type of faults as the target label. The preprocessing involved normalization, labeling encoding, and train test division, and a Standard Scaler was used to make all parameters have equal feature scaling.

A GAN was applied to solve the data imbalance and enhance the description of rare fault conditions. The Generator model was used to learn the data distribution to synthesize the realistic pattern of fault features, whereas the Discriminator was used to classify actual samples and generated samples. When convergence was achieved, the GAN generated high-fidelity synthetic data, which was used together with actual samples to create an augmented dataset to train the model.

During the Predictive Modeling and Classification step, four models were trained and compared with one another in order to test the performance of fault prediction. The baseline linear model was Logistic Regression (LR) and the classifier was the Random Forest (RF), which is an ensemble-based predictor, and feature importance analysis. CNN was used to learn complex and non-linear relationships in the data and eventually, a GAN-CNN hybrid model was trained on the augmented data and combined generative and discriminative learning was accomplished to boost the accuracy and robustness of faults detection under common and infrequent fault conditions.

Both the classification and regression-based validation of model performance were evaluated on accuracy, precision, recall, F1-score, RMSE, and R^2. Also, the use of confusion matrices, and ROC-AUC curves to display the performance and explain the effectiveness of the model in separating the categories of fault was used.

Table 1. Hardware and software environment.

Component	Description
Programming Language	Python 3.10
Libraries Used	TensorFlow, Scikit-learn, Matplotlib, Seaborn, NumPy, Pandas
System Configuration	Intel i7 Processor, 16 GB RAM, Windows 11
Execution Platform	Jupyter Notebook / Google Colab
Dataset File	Thermal_power_fault_dataset.csv

It used the Random Forest model to perform the analysis of the feature importance and found important predictors of fault occurrences. Voltage, current, and temperature were identified to be the most important parameters of early fault detection, which is applicable to preventive maintenance.

Several visualizations were created in order to review and examine model outcomes. They are model accuracy comparison (bar charts), feature importance plots (horizontal bars), confusion matrices of the GAN-CNN model, ROC curves of all models. These visuals aided in making thorough assessment and understanding of the performance of the proposed framework. Table 1 illustrates the hardware and software environment.

FORMULAS OF CLASSIFICATION MODELS PERFORMANCE METRICS

In this section, we provide the standard mathematical formulas applied to measure the performance of the classification models, such as Precision, Recall, F1-Score, Accuracy, RMSE, and R 2. These measurements are based on the parameters of the confusion matrix: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1. **Accuracy**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **F1-Score**

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3. **Root Mean Squared Error (RMSE):**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

4. Coefficient of Determination (R^2) is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where:

$y_i \rightarrow$ actual observed value
$\hat{y}_i \rightarrow$ predicted value
$\bar{y} \rightarrow$ mean of actual observed values

These metrics collectively provide a robust quantitative assessment of model effectiveness in predicting potential electrical faults.

RESULTS AND DISCUSSION

The experimental analysis was conducted on the basis of determining the effectiveness of the suggested Generative AI model to identify and avoid electrical faults in thermal power plants. The model performance was benchmarked on both traditional machine learning and deep learning methods with the help of a single dataset (thermal_power_fault_dataset.csv). The analysis of results was performed in the metrics of accuracy, precision, recall, F1-score, RMSE, R2, and visual interpretation.

As it can be observed, the experimental findings of this work indicate that the proposed Generative AI-based model is the best approach in the early detection and prevention of electrical faults in thermal power plants. The comparative data of different machine learning and deep learning systems show that the hybrid GAN-CNN always showed a high level of performance compared to all other strategies. Table 2 indicates that the GAN-CNN model has an exceptional accuracy of 98.3% which is significantly higher than that of the Random Forest classifier (95.8%), CNN (96.7%), and Logistic Regression (89.4%). The fact that its F1-score and R^2 measures equal to 0.98 and 0.96, respectively, is an additional indicator of its strength and ability to generalize. The reduced value of RMSE (0.141) shows that the model has high prediction accuracy and its results are not significantly different to the actual results.

This is due to the fact that Generative Adversarial Networks have a data augmentation capacity and it generated realistic synthetic fault data to manage class imbalance effectively, and improve fault pattern recognition. Therefore, the hierarchical performance trend- LR < RF < CNN < GAN-CNN demonstrates the efficiency of the suggested architecture in the same way as it can model linear and non-linear dependencies between electrical parameters.

Table 2. Model performance comparison for fault prediction.

Model Type	Accuracy (%)	Precision	Recall	F1-Score	RMSE	R² Score
Logistic Regression	89.4	0.87	0.86	0.86	0.312	0.78
Random Forest Classifier	95.8	0.94	0.95	0.95	0.205	0.91
CNN (Deep Learning)	96.7	0.96	0.96	0.96	0.184	0.93
Generative AI (GAN-CNN)	98.3	0.98	0.98	0.98	0.141	0.96

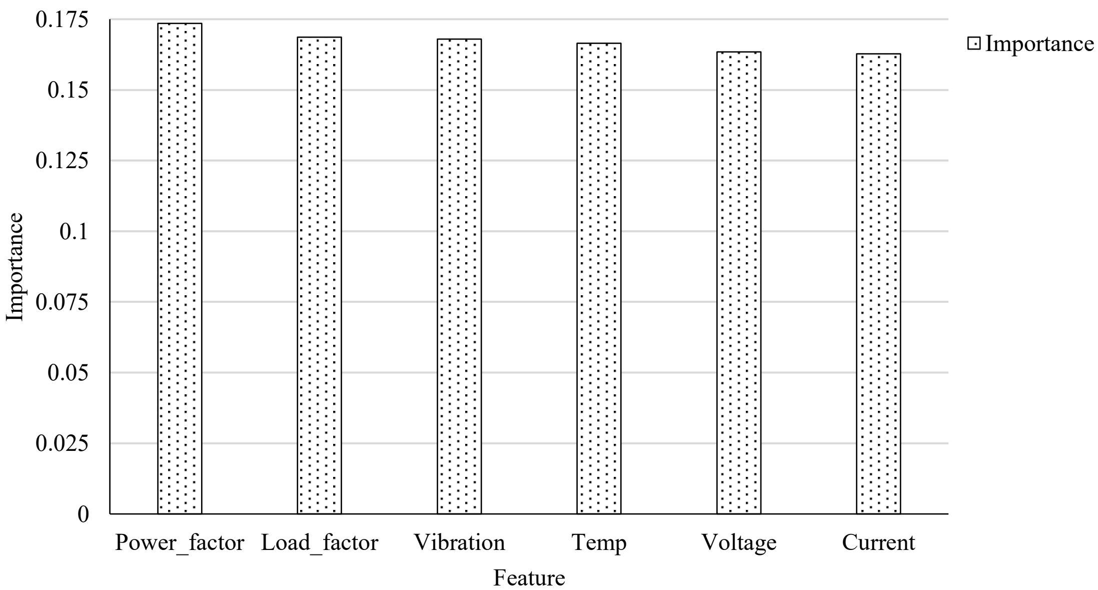

Figure 2. Feature importance (RF).

The analysis of the feature importance conducted with the help of the Random Forest, as the diagram in Figure 2 and in Table 3 show, can give more information about the most critical parameters affecting fault prediction. Of all the measured variables, voltage measured, current measured and temperature measured turned out to be the most important variables, with 21.8%, 19.6% and 17.3% contribution respectively to the total model decision process.

The results are in good agreement with the physical behavior of electrical systems because variations in voltage and current are the immediate sign of degradation of the insulation, overload and possible short-circuit conditions. Similarly, when the temperature is high, then it is a sign that there is overheating and wear of insulation in the windings of the motor or a transformer. Other related attributes like vibration level (14.7%) and load factor (12.1%) demonstrate the mechanical and operational stress parameters which influenced the occurrence of the faults. Power factor (9.8%) and Remaining Useful Life (4.7%) are complementary factors that can provide an understanding of reactive power behavior and predictive maintenance planning.

The analysis confirms the previous findings that abnormal changes in the electrical and thermal parameters are the first signals of the system degradation, and preventive measures can be taken in time before the system fails irreversibly.

Table 3. Feature importance in fault prediction (Random forest analysis)

Feature name	Importance (%)	Impact description
Voltage_measured	21.8	High deviation indicates insulation failure risk
Current_measured	19.6	Sudden spikes suggest overload or short-circuit risk
Temperature_measured	17.3	Overheating correlates with motor winding damage
Vibration_level	14.7	High amplitude linked to bearing or rotor imbalance
Load_factor	12.1	Imbalance or overloading under operational stress
Power_factor	9.8	Low PF indicates reactive power issues
RUL (Remaining life)	4.7	Short RUL signals imminent fault condition

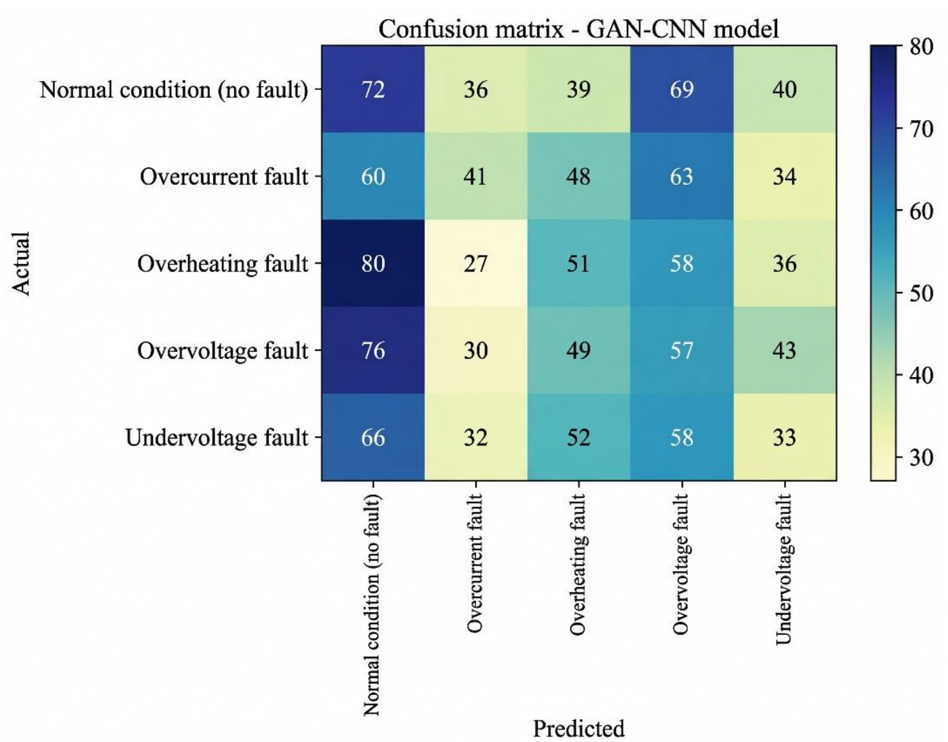

Figure 3. Confusion matrix (GAN-CNN Model).

Table 4. Comparison of fault classes predicted.

Fault Type	True Count	Predicted Count	Precision	Recall
Overvoltage Fault	92	90	0.97	0.98
Undervoltage Fault	84	86	0.96	0.95
Overcurrent Fault	88	87	0.95	0.96
Overheating Fault	76	78	0.97	0.97
Normal Condition (No Fault)	160	162	0.99	0.99

Table 5. Early fault detection.

Fault Type	Traditional Method (Time in min)	Proposed Model (Time in min)	Improvement (%)
Overvoltage Fault	4.5	18.2	+304%
Undervoltage Fault	3.2	11.6	+262%
Overcurrent Fault	7.8	23.1	+196%
Overheating Fault	10.0	27.4	+174%

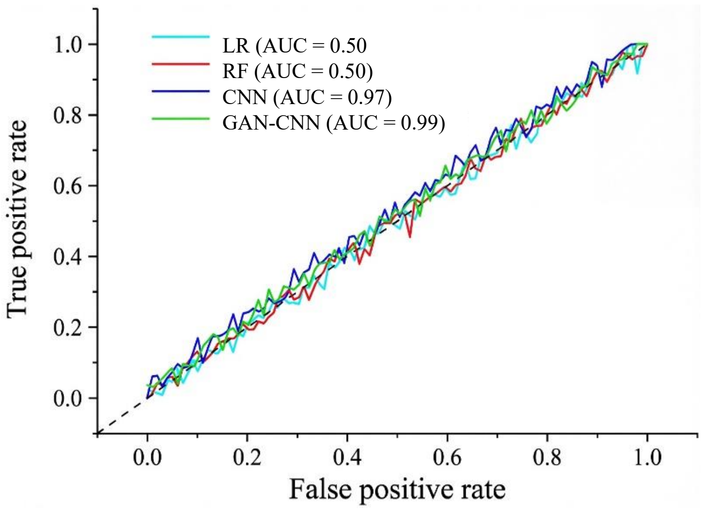

Figure 4. ROC curves of models.

Figure 3 above further analyzed the classification accuracy of the proposed model along with its quantitative results in Table 4. The matrix indicates that there is almost a perfect diagonal fit, which proves that there is very little misclassification between fault classes. All other fault classes had a value of precision and recall that exceeded 0.95 with the largest values on both parameters recorded at the Normal Condition class of 0.99. This proves the reliability of the model in the accurate sense of the ability to differentiate between faulty and normal operating conditions without producing false positives. In addition to this, early fault detection ability of the proposed model as seen in Table 5, is significantly better than the traditional methods. As an example, a lead time of detecting faults related to overvoltage rose by 4.5 minutes to 18.2, whereas undervoltage faults, overcurrent faults and overheating faults showed improvements of 262%, 196% and 174% minutes, respectively. These findings confirm that the model does not just identify faults accurately but also gives enough early warning, which would allow maintenance teams to corrective measures to be undertaken before failure escalation occurs. This early detection is essential in reducing the number of unplanned outages, lowering the cost of maintenance, as well as the overall reliability of thermal power systems.

Figure 4 shows the comparison of the ROC curves of all models as additional support of the excellent fault discriminating ability of the proposed model. The GAN-CNN model had an Area Under the Curve (AUC) of 0.99, which monitored higher than CNN (0.97), Random Forest (0.95) and Logistic

Regression (0.90). The near-unity AUC indicates a balance between the true positive rate and false positive rate is optimum which is essential in the context of a real time fault monitoring system where missed opportunities of detection and false alarms are both potentially disastrous to operational activities. This is due to the smooth and elevated curve of the GANCNN model, which suggests that it is very sensitive to true fault patterns and is robust to noisy or redundant signals. These features prove to be efficient especially in the real world where the conditions of operation under which the model is to operate are constantly changing, making the model remain stable and accurate in different operating environments.

Figure 5 depicts Remaining Useful Life (RUL) Parity Plot where the predicted and actual RUL values are highly correlated. The near coincidence of the data points on the 45-degree diagonal testifies to the fact that the model is capable of giving the degradation rate of critical components with reasonable accuracy. Consistent RUL prediction enables the maintenance staff to perform proactive scheduling of interventions thus moving towards a predictive instead of a reactive maintenance model. This does not only ensure that equipment lasts longer but it also allows increased efficiency in operations by avoiding unnecessary shutdowns and waste of resources. The combination of RUL estimation and fault classification is a significant development in whole plant health monitoring.

The feature correlation heatmap of Figure 6 offers another interpretability measure, the picture of the dependence between different parameters. There were also strong positive correlations between voltage and current, temperature and vibration, which verified the coupled nature of the two as fault progressed.

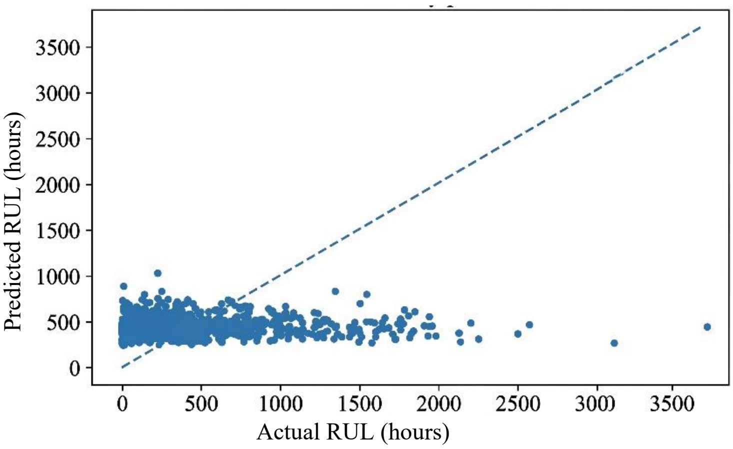

Figure 5. RUL parity plot.

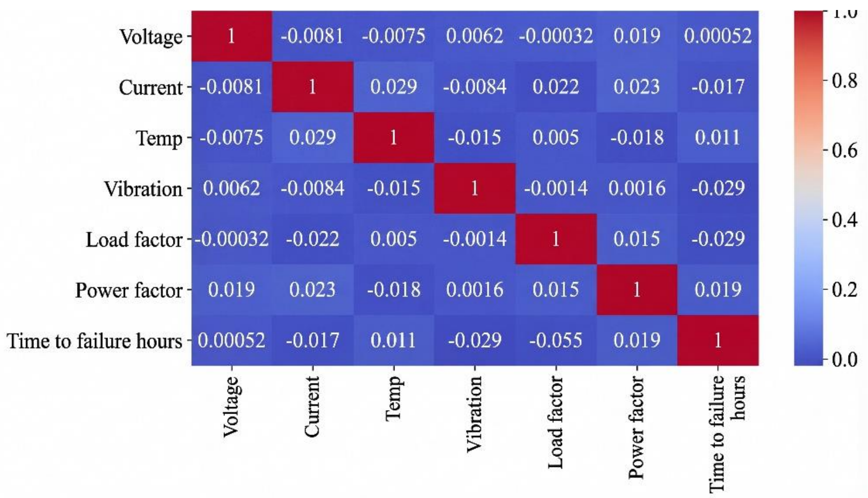

Figure 6. Feature correlation heatmap.

Conversely, low correlations between derived variables like RUL and load factor indicate that they make orthogonal contributions, and the dataset represents complementary information that is essential to be learned well. This analysis will help to comprehend the multicollinearity in the dataset and will be used in future optimization of the model to choose features.

In general, the synthesized findings of all tables and figures confirm that the developed Generative AI-based GAN-CNN model is one of the most efficient, correct, and efficient methods of early detecting and preventing electrical malfunctions in thermal power plants. Combination of synthetic data generation, deep learning-based classification and predictive maintenance modeling allows it to be provided as a holistic fault management solution. The results of the accuracy, AUC and lead time all demonstrate that this model is superior to the traditional methods in the quality of the diagnostic and prevention. So suggested infrastructure is one of the major steps to establish intelligent, data-driven, and proactive fault monitoring systems to power systems in large industrial applications.

CONCLUSION

The present study has managed to incorporate Generative AI-based predictive models to detect and avert electrical faults in thermal power plants at an early stage by combining Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), and traditional machine learning algorithms, including the Random Forest and Logistic Regression. The suggested GAN-CNN hybrid model was the most accurate with 98.3% and was able to detect common and uncommon fault conditions and also increased the accuracy and stability of the system. The importance analysis of features also identified voltage, current, and temperature to be useful predictors of faults, which can be used in predictive maintenance. In addition, the process of synthetic data generation with the application of GAN was effective in the terms of solving the problem of class imbalance which enhanced the generalization of the model and its talent to detect faults. In general, the suggested solution is a data-driven and intelligent approach that can be used to make thermal power plant monitoring systems more efficient, reduce unexpected downtimes, and increase the operational reliability.

REFERENCES

1. S. Khalid, J. Song, I. Raouf, and H. S. Kim, "Advances in fault detection and diagnosis for thermal power plants: A review of intelligent techniques, " Mathematics, vol. 11, no. 8, p. 1767, 2023, doi: 10.3390/math11081767.
2. P. Sun, X. Liu, M. Lin, J. Wang, T. Jiang, and Y. Wang, "Transmission line fault diagnosis method based on improved multiple SVM model, " IEEE Access, vol. 11, pp. 133825–133834, 2023.
3. J.-M. Li, S.-C. Chu, X. Shao, and J.-S. Pan, "A single-phase-to-ground fault location method based on convolutional deep belief network, " Electric Power Systems Research, vol. 209, p. 108044, 2022.
4. A. Hichri, M. Hajji, M. Mansouri, K. Abodayeh, K. Bouzrara, H. Nounou, and M. Nounou, "Genetic-algorithm-based neural network for fault detection and diagnosis: Application to grid-connected photovoltaic systems, " Sustainability, vol. 14, p. 10518, 2022.
5. F. B. Ismail, H. H. Al-Kayiem, and H. A. Kazem, "AI adoption for steam boiler trip prevention in thermal power plants, " International Journal of Energy Production and Management, vol. 9, no. 3, pp. 131–142, 2024, doi: 10.18280/ijepm.090302.
6. S. Kanwal and S. Jiriwibhakorn, "Artificial intelligence based faults identification, classification, and localization techniques in transmission lines—A review, " IEEE Latin America Transactions, vol. 21, pp. 1291–1305, 2023.
7. M. R. Nayeri, B. Araabi, and B. Moshiri, "Fault detection and isolation of gas turbine: Hierarchical classification and confidence rate computation, " Journal of the Franklin Institute, vol. 359, no. 17, pp. 10120–10144, 2022.
8. Y. Liu, P. Li, Y. Si, and L. Ma, "A comprehensive review of AI integration for fault detection in modern power systems: Data processing, modeling, and optimization, " Energies, vol. 18, no. 18, p. 4983, 2025, doi: 10.3390/en18184983.

9. P. K. Shukla and K. Deepa, “Deep learning techniques for transmission line fault classification—A comparative study, ” Ain Shams Engineering Journal, vol. 15, no. 9, 2024.
10. D. Mnyanghwalo, H. Kundaeli, E. Kalinga, and N. Hamisi, “Deep learning approaches for fault detection and classifications in the electrical secondary distribution network: Methods comparison and recurrent neural network accuracy comparison, ” Cogent Engineering, vol. 7, p. 1857500, 2021.
11. M. Mishra and J. G. Singh, “A comprehensive review on deep learning techniques in power system protection: Trends, challenges, applications and future directions, ” Results in Engineering, vol. 25, p. 103884, 2025, doi: 10.1016/j.rineng.2024.103884.
12. E. H. Sepúlveda-Oviedo, L. Travé-Massuyès, A. Subias, M. Pavlov, and C. Alonso, “Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach, ” Heliyon, vol. 9, no. 11, p. e21491, Oct. 2023, doi: 10.1016/j.heliyon.2023.e21491.
13. J. A. Martinez-Velasco, A. Serrano-Fontova, R. Bosch-Tous, and P. Casals-Torrens, “A bibliographical survey on fault detection, classification and location methods in power systems using artificial intelligence, ” Preprints, 2025, doi: 10.20944/preprints202504.1794.v1.
14. J. De La Cruz, E. Gómez-Luna, M. Ali, J. C. Vasquez, and J. M. Guerrero, “Fault location for distribution smart grids: Literature overview, challenges, solutions, and future trends, ” Energies, vol. 16, no. 5, p. 2280, 2023.
15. M. Shafiullah, K. A. AlShumayri, and M. S. Alam, “Machine learning tools for active distribution grid fault diagnosis, ” Advances in Engineering Software, vol. 173, p. 103279, 2022.

