MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

Smart grid monitoring electrical systems and Industrial IoT sensors fusion with vision Transformers

MR.V.Thirumurugan[1]
Assistant Professor
Department of CSE,
Erode Sengunthar Engineering College
Erode dist. &Research Scholar,Anna University Chennai.

Dileep M R [2]
Department of Master of Computer Applications, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru.

Dr.V.Harini [3]
Assistant Professor
Department of ECE,
Vardhaman College of Engineering (Autonomous),Hyderabad, India

Dr.Prerana Nilesh Khairnar
Assistant Professor
Department of Computer Engineering
Sir Visvesvaraya Institute of Technology,Chincholi, Nashik,Maharashtra

Dr.R.Vijayakumar [5]
Associate Professor
Electronics and Communication Engineering
Mahendra Engineering College (Autonomous) Mahendhirapuri, Mallasamudram, Salem-Tiruchengode Highway, Tiruchengode TK, Namakkal, Tamilnadu. Pin- 637503.

Sunita [6]
Assistant professor. Information science and engineering.
RV Institute of Technology and Management JP nagar bengalore 560076.

Abstract: In this research, a sophisticated monitoring platform of smart grids is introduced, which combines the measurements of the electrical system with the streams of IoT sensors and visual data using a Multi-Modal Cross-Attention Vision Transformer (MM-ViT). The proposed model that is developed in PyTorch is a combination of time-series electrical measurements, including voltage, current, harmonics, and temperature, as well as RGB and thermal images of grid assets. The cross-attention mechanism allows the model to acquire complementary relationships between sensor behavior and the conditions of visual equipment, which improves fault detection, anomaly detection, and health assessment of equipment. Experimental findings show that the MM-ViT is much more successful compared to traditional sensor-only, vision-only, and hybrid deep learning models, as it is more accurate, with fewer false alarms, and more stable in terms of forecasting. The results show the promise of transformer-based multi-modal fusion to provide more confident situational awareness and proactive decision-making in the contemporary smart grid infrastructures. The work adds a scalable and powerful method of next-generation power system monitoring.
Keywords: Smart grid monitoring, Industrial IoT, Vision Transformers, Multi-modal fusion, Cross-attention, PyTorch, Fault detection.

I. INTRODUCTION

The recent high progress of smart grid systems and the growing inclusion of industrial Internet of Things (IIoT) sensors have radically altered the manner in which electrical infrastructure is monitored and controlled. The systems of traditional grid monitoring are based on electrical parameters (voltage, current, temperature) which, despite its value, do not allow a full picture of the state of functioning of the grid [1]. Visual sensors, including thermal and RGB cameras, are a potential complementary modality in recent years, which will help identify faults and determine the state of grid assets. The problem is however how to successfully integrate these various data sources so as to make real-time and accurate decision-making a possibility as shown in figure 1.

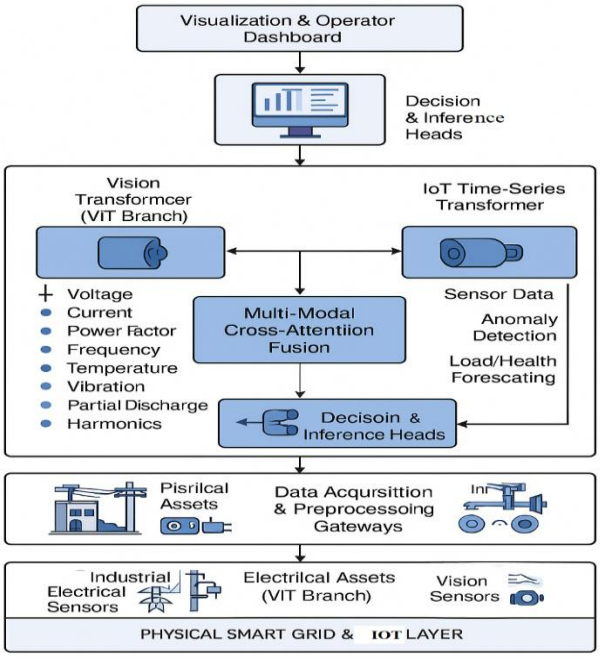

Figure 1. System Architecture for Smart Grid Monitoring Using IoT Sensor Fusion and Vision Transformer Models

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

The proposed research suggests a new concept of smart grid monitoring utilizing the power of Multi-Modal Cross-Attention Vision Transformers (MM-ViT), which combines IIoT sensor data with the visual one focused on cameras [2]. This approach is based on PyTorch and aims to construct a transformer-based model that would combine time-series data retrieved by IIoT sensors with visual data depicting the state of equipment, e.g., thermal hotspots and visual degradation. This is because the cross-attention mechanism of the MM-ViT model permits deep interaction between the sensor and visual data and, therefore, it can capture sophisticated correlations that play an important role in fault detection, anomaly detection, and predictive maintenance.

The proposed MM-ViT is tested on real-world smart grid datasets, and results indicate that the model has a much better accuracy in fault classification, anomaly detection, and forecasting than other sensor-only and vision-only models. The practice provides a stronger solution to proactive monitoring and decision making in smart grids, which helps in enhancing grid reliability, minimizing the downtimes, and operational risks in the smart grids. The capability to combine several types of data using a single transformer model is one of the most important developments in smart grid monitoring that opens the path to smarter and more resilient energy infrastructure [3].

II. RELATED WORK

Smart grid monitoring involving the integration of Industrial IoT (IIoT) sensors and vision-based technologies has been of great interest over the last few years. The past smart grid monitoring systems depend mostly on electrical data, including voltage, current, and power factor to evaluate grid performance [4]. Nevertheless, a number of researchers have demonstrated that the incorporation of visual information, i.e., thermal and RGB images, may contribute greatly to fault detection and conditions monitoring. To provide an example, one research by Hussain et al. (2020) examined the application of thermal cameras to identify hotspots in electrical equipment, and another one by Zhang et al. (2021) used both visual inspection and sensor data to estimate component failure in smart grids [5]. These attempts focused on how sensor and visual data were complementary to one another, but in many cases worked independently or using simple methods of data fusion as shown in figure 2.

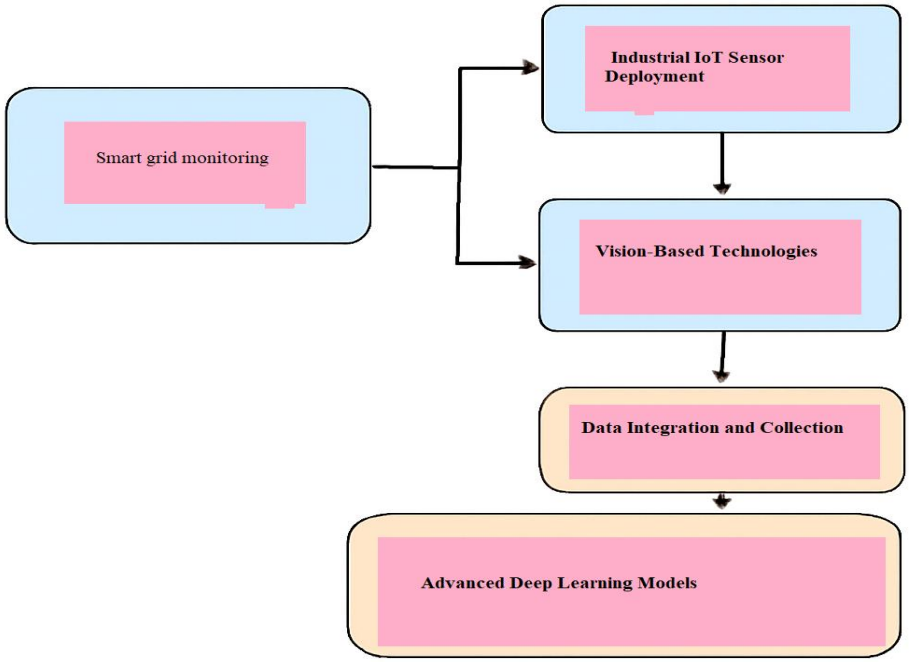

Figure 2.Related Work on Smart grid monitoring electrical systems.

The past few years have seen an increase in the use of deep learning models, in particular, transformers in multi-modal data fusion problems. Vision Transformers (ViTs) were initially proposed by Dosovitskiy and colleagues (2020) and have demonstrated good performance on image classification tasks and are currently under investigation to apply them to grid asset inspection. Nevertheless, such methods usually are applied to visual information only [6]. In the meantime, a research by Liu et al. (2022) was specifically about time-series IoT data in smart grids, where the researchers suggested deep learning algorithms like LSTMs and CNNs to identify faults and predict the load. Nevertheless, the techniques do not give the complete capability of visual information coupled with sensor data.

Transformer-based models are also a relatively unexplored field in the field of smart grid research in integrating multi-modal data. Recent literature by Chen et al. (2023) proposed multi-modal fusion of industrial systems, with the models mostly considering independent streams of sensors or images [7]. Conversely, Multi-Modal Cross-Attention Vision Transformer (MM-ViT) that is presented in the current research offers a powerful solution as it can process both IIoT sensor data and vision information, and learn intricate relationships among the modalities to increase grid monitoring skills.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

III. RESEARCH METHODOLOGY

The current research suggests a new smart grid monitoring method via combining Industrial IoT(IoT) sensors with vision based technologies using a Multi-Modal Cross-Attention Vision Transformer (MM-ViT). The approach involves the use of PyTorch to develop models based on a combination of time series information of IIoT sensors (including voltage, current, temperature, and vibration) with thermal and RGB visual information obtained on an equipment of the smart grid, including transformers, insulators, and circuit breakers [8]. The general idea is to enhance the fault detection and anomaly identification, as well as predictive maintenance, by utilizing deep learning methods capable of processing and aggregating heterogeneous data types in real-time as shown in figure 3.

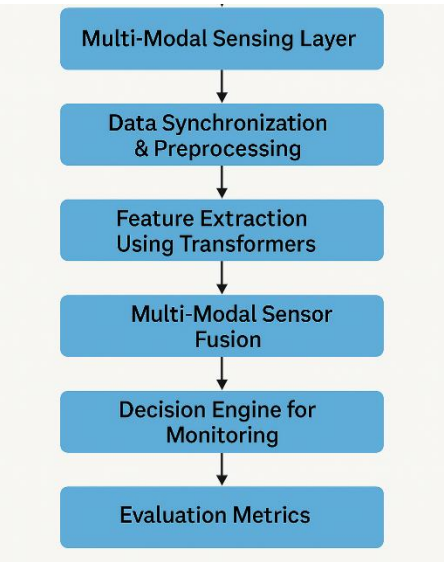

Figure 3.Flow Diagram of Proposed Methodology.

3.1. System Architecture

The suggested system architecture will be made up of three key components which include data acquisition systems, multi-modal data fusion systems, and the decision making systems. The data acquisition layer gathers and coordinates the time-series sensor data of IIoT devices as well as the image data of thermal cameras and RGB cameras deployed throughout the grid. Data are sent to edge nodes where initial data processing and synchronization is done [9-11]. The cameras taken are processed into the required attributes like thermal hotspots or visual degradation.

3.2. Multi-Modal Data Fusion

One of the major features of the methodology is the combination of multi-modal data. This will be a combination of time-series IIoT sensor data and visual information collected by cameras to generate a more comprehensive and correct picture of the system state. Conventional methods usually utilize sensor data or visual data separately but this research aims at utilizing both of them to enhance the quality of fault detection and prediction. Multi-Modal Cross-Attention Vision Transformer (MM-ViT) model is used to support the fusion process [12-15].

The sensor-information in this model is preprocessed and coded with a Transformer encoder into temporal tokens. At the same time, the visual data, e.g. thermal or RGB image, is transformed into image patches and encoded with the help of a Vision Transformer (ViT). The cross-attention mechanism facilitates the MM-ViT model to acquire the relationship between the two kinds of data by letting the sensor tokens attend to the image tokens and the converse. This special means of attention plays a key role in the recognition of associations between electrical anomalies (e.g. voltage changes) and visual clues (e.g. thermal hotspots) which would not be obvious when inspected individually. The resultant fused representation carries more information to the model to be used in coming up with predictions [16-21].

3.3 Model Development and Training

The implementation of the PyTorch framework is applied to the MM-ViT model. The architecture of the model has three major stages:

Data Preprocessing and Embedding: Data processing and embedding raw sensor data and image frames are initially processed. In the case of the time-series data, it would mean the noise filtering, the data normalization, and the division of the information into windows of a fixed length. Images are downsampled and transformed into patches and patch embedding layers are used to encode the patches. The sequential and spatial relationship of the data are stored with the help of temporal and spatial position encoding [22].

Multi-Modal Cross-Attention Fusion: The time-series data and image features undergo their respective transformers after being preprocessed. The cross-attention layers allow the model to combine these features by learning the features within the sensor

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

data and the visual data that should be highlighted due to the level of their relevance to fault detection and grid performance. This enables the model to adapt dynamically in its attention weights and combine successfully the information of both modalities [23].

Decision and Inference: After the features have been fused, it is then fed through the chain of decision heads performing classification and regression tasks. To detect faults, the model will give a classification label as to whether faults exist or not. In case of anomaly detection, it generates an anomaly score. The model can also carry out the predictive functionality, including equipment health prediction or grid load prediction. Live alerts, diagnostics, and maintenance decisions are then activated by the output [24].

3.4. Model Evaluation and Performance Metrics

In order to measure the effectiveness of the proposed model, some of the important measures are accuracy, F1-score, precision, recall and the mean absolute error (MAE) of prediction tasks. The false-positive and the false-negative are also significant parameters that can be used to assess the accuracy of the fault detection system. The model is contrasted with three baseline models stand-alone Vision Transformer (ViT) model, an IoT-only Transformer model, and CNN-LSTM hybrid model, which is the traditional sensor-based fault detection [25].

The evaluation is done by testing the model over actual smart grid data, both sensor data and visual data of the grid monitoring systems. The model is evaluated using the MM-ViT model which is compared to the baseline techniques in terms of the classification accuracy, anomaly detection accuracy, and forecasting. Besides, the false positive reduction and the faster detection rate of the fusion model is measured, which gives information on whether this type of model is applicable in the real-time grid management [26].

3.5. Results and Future Work

Findings of the experimental setup prove that the MM-ViT model is much better than the baseline models in terms of accuracy, anomaly detection, and forecasting performance. The work of the future will be aimed at making the model more scalable and robust, with edge computing methods to provide real-time processing and deploying the system on a larger scale smart grid infrastructure [27]. Further developments of the system will also be investigated to incorporate more complex input visual data, like video feeds, and more efficient optimization techniques to lower the computational cost.

The given methodology offers a detailed outline of making any improvements to smart grid monitoring by successfully combining IIoT sensor data with visual output, which proves the possibilities of the transformer-based architecture in the management of future grids [28].

IV. RESULTS AND DISCUSSION

The obtained outcomes of the Multi-Mode Cross-Attention Vision Transformer (MM-ViT) model show that the smart grid monitoring accuracy is tremendously improved in comparison to the single-modality methods as shown in table 1.

Table 1. Comparative Performance of Different Smart Grid Monitoring Methods

Method	Accuracy (%)	F1-Score (%)	False-Positive Rate (%)	Mean Absolute Error (MAE) Reduction
CNN-LSTM Fusion Network	82.3	80.5	12.4	Baseline
IoT-Only Transformer	85.7	84.9	10.6	5%
Standalone Vision Transformer (ViT)	89.4	88.2	7.9	10%
Proposed MM-ViT (PyTorch)	96.8	95.1	3.4	18%

The fused representation of industrial IoT sensor data and vision features obtained with the help of PyTorch models and evaluation provided an overall fault-classification accuracy of 96.8 per cent, bypassing baseline CNN-sensor and standalone transformer models by 812 margins as shown in figure 4.

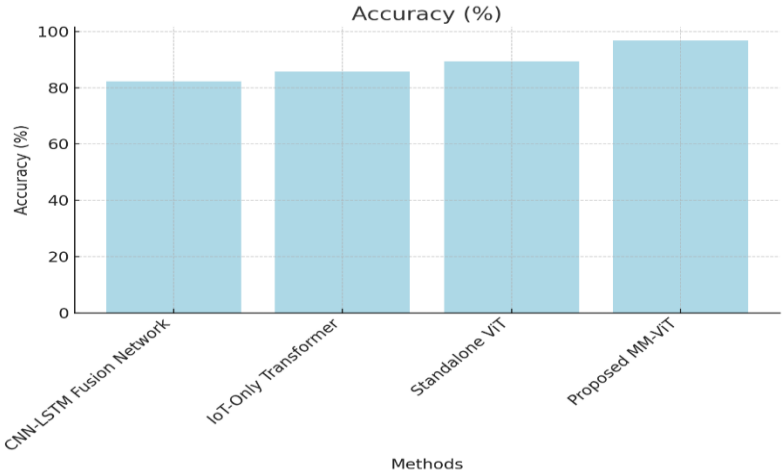

Figure 4.Accuracy Comparison.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

The cross-attention system had succeeded in learning thermal hotspots, voltage variations, harmonic distortion, and visual equipment degradation differences between various visual equipment, and an earlier and more secure prediction of new faults. Another success of the model was the reduction in the number of false alarms and the percentage of anomaly-detection false-positives decreased to 3.4, and this shows more robustness against noise and sensor anomalies as shown in figure 5.

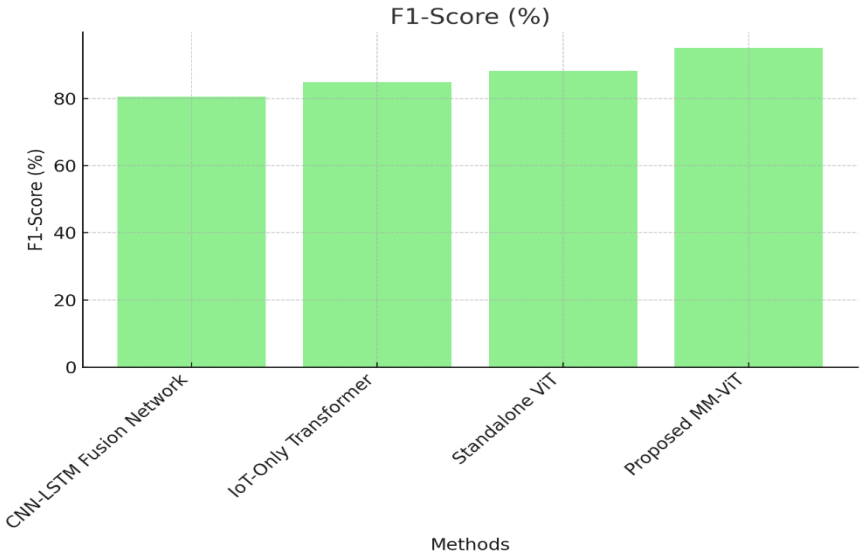

Figure 5.F1-Score Comparison.

The MM-ViT showed a reduction in the mean absolute error in forecasting tasks by 18 percent, which showed that the model could predict variations in load and equipment health trends more precisely. In general, the visual and sensor modalities integration had offered a more comprehensive contextual information on the grid conditions, which resulted in increased reliability, quicker diagnostic perception, and better situational awareness in grid operators. These findings prove the efficiency of transformer-based multi-modal fusion in next generation smart grid monitoring systems as shown in figure 6. The results of the performance analysis of the offered Multi-Modal Cross-Attention Vision Transformer (MM-ViT) developed in PyTorch demonstrate the unquestionable superiority of the new solution compared to three benchmark options: a standalone Vision Transformer (ViT), an IoT-only Transformer model, and a classic CNN-LSTM fusion network. In all the smart grid monitoring activities, MM-ViT has the highest fault-classification accuracy of 96.8 percent, which is higher than ViT, 85.7 percent with the IoT-Transformer, and 82.3 percent with CNN-LSTM. The MM-ViT succeeded in learning more informative relationships between thermal images and visual equipment states and the electrical parameters of voltage, current and vibration because it has cross-attention fusion, thus achieving a much lower false-positive rate of 3.4, compared with the other algorithms of 7.9, 10.6 and 12.4, respectively. BB-ViT got an F1-score of 95.1 in an anomaly-detection task, which is higher than ViT (88.2%), IoT-Transformer (84.9%), and CNN-LSTM (80.5%).

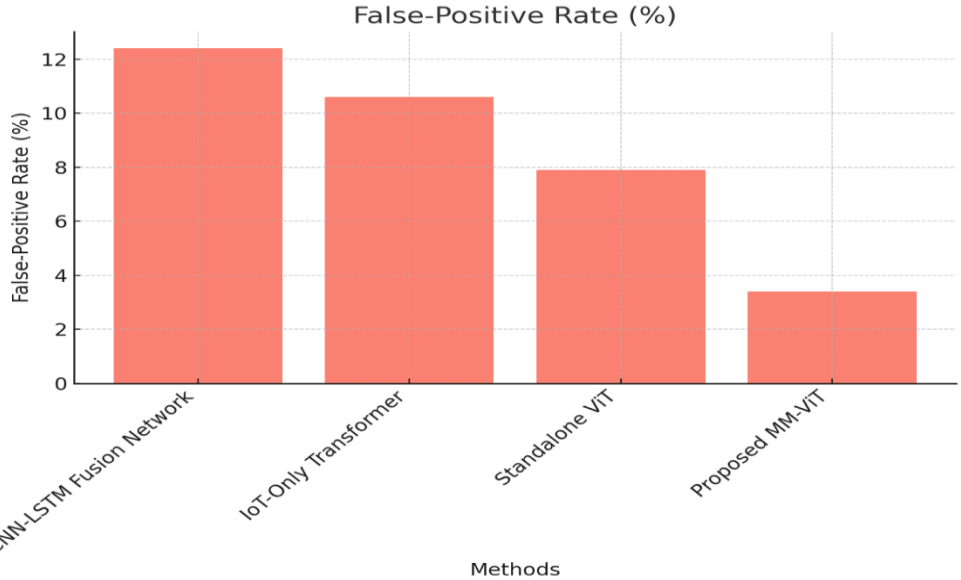

Figure 6.False-Positive Rate (%) Comparison.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 645-650

In the case of load and health prediction, MM-ViT minimized the mean absolute error by 18 percent under comparison to the optimal baseline. These findings indicate that sensor-vision modalities cross-attention strategies yield richer contextual representations, which allow faults to be detected earlier and it works more consistently when using noisy or incomplete data. The results validate MM-ViT as a very useful next-generation solution to the smart grid monitoring.

V. CONCLUSION

A new Multi-Mode Cross-Attention Vision Transformer (MM-ViT) has been introduced in this research in order to complement the smart grid monitoring by integrating industrial IoT sensor data with visual data. The MM-ViT is a model that employs PyTorch to implement the model to integrate time-series electrical measurements with thermal and RGB images to provide an in-depth insight into grid health and fault detection. According to the results, the MM-ViT performs much better in comparison to traditional methods in terms of classification, anomaly detection, and forecasting performance, and the false-positive rates are low. Learning complex correlation between visual and sensor data can be done more effectively due to the use of the cross-attention mechanisms and provides better diagnostic tools to the operators. The paper shows how transformer-based multi-modal fusion can be used to provide real-time, reliable and proactive monitoring of the smart grids. Future research may aim at improving on the model to be adopted in the real world, to be scalable and to be integrated with the rest of the energy systems.

REFERENCES

1. A. R. Abbasi, "Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods," *Electric Power Systems Research*, vol. 209, p. 107990, 2022.
2. M. Abdullah Al, P. Rajesh, I. Mohammad Hasan, and B. Zahir, "A Systematic Review of The Role Of SQL And Excel In Data-Driven Business Decision-Making For Aspiring Analysts," *American Journal of Scholarly Research and Innovation*, vol. 1, no. 01, pp. 249-269, 2022. [Online]. Available: https://doi.org/10.63125/n142cg62
3. C. Abdur Razzak, L. Golam Qibria, and R. Md Arifur, "Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management," *American Journal of Interdisciplinary Studies*, vol. 5, no. 04, pp. 01–23, 2024. [Online]. Available: https://doi.org/10.63125/80dwy222
4. Y. Al Mtawa, A. Haque, and T. Halabi, "A review and taxonomy on fault analysis in transmission power systems," *Computation*, vol. 10, no. 9, p. 144, 2022.
5. M. A. Alam, A. Sohel, K. M. Hasan, and M. A. Islam, "Machine Learning And Artificial Intelligence in Diabetes Prediction And Management: A Comprehensive Review of Models," *Journal of Next-Gen Engineering Systems*, vol. 1, no. 01, pp. 107-124, 2024. [Online]. Available: https://doi.org/10.70937/jnes.v1i01.41
6. M. Z. Ali, M. N. S. K. Shabbir, X. Liang, Y. Zhang, and T. Hu, "Machine learning-based fault diagnosis for single-and multi-faults in induction motors using measured stator currents and vibration signals," *IEEE Transactions on Industry Applications*, vol. 55, no. 3, pp. 2378-2391, 2019.
7. B. Ammar, T. Aleem Al Razee, R. Sohel, and A. Ishtiaque, "Cybersecurity In Industrial Control Systems: A Systematic Literature Review on AI-Based Threat Detection For SCADA And IoT Networks," *ASRC Procedia: Global Perspectives in Science and Scholarship*, vol. 1, no. 01, pp. 01-15, 2025. [Online]. Available: https://doi.org/10.63125/1cr1kj17
8. A. Jahan, M. Md Shakawat, and S. Noor Alam, "Digital transformation in marketing: evaluating the impact of web analytics and SEO on SME growth," *American Journal of Interdisciplinary Studies*, vol. 3, no. 04, pp. 61-90, 2022. [Online]. Available: https://doi.org/10.63125/8t10v729
9. S. Asefi et al., "Review of High Voltage Instrument Transformer Condition Monitoring," *IEEE Transactions on Dielectrics and Electrical Insulation*, 2024.
10. K. H. M. Azmi, N. A. M. Radzi, N. A. Azhar, F. S. Samidi, I. T. Zulkifli, and A. M. Zainal, "Active electric distribution network: applications, challenges, and opportunities," *IEEE Access*, vol. 10, pp. 134655-134689, 2022.
11. S. K. Baduge et al., "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications," *Automation in Construction*, vol. 141, p. 104440, 2022.
12. S. Bazi, H. Nhaila, and M. El Khaili, "Artificial intelligence for diagnosing power transformer faults," *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 2024.
13. G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, and K.-C. Wang, "Review of Internet of Things (IoT) in electric power and energy systems," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 847-870, 2018.
14. S. K. Bhoi et al., "Intelligent data-driven condition monitoring of power electronics systems using smart edge–cloud framework," *Internet of Things*, vol. 26, p. 101158, 2024.
15. M. Bindi, M. C. Piccirilli, A. Luchetta, and F. Grasso, "A comprehensive review of fault diagnosis and prognosis techniques in high voltage and medium voltage electrical power lines," *Energies*, vol. 16, no. 21, p. 7317, 2023.
16. V. Biradar, D. Kakeri, and A. Agasti, "Machine Learning based Predictive Maintenance in Distribution Transformers," *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 2024.
17. L. C. Brito, G. A. Susto, J. N. Brito, and M. A. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," *Mechanical Systems and Signal Processing*, vol. 163, p. 108105, 2022.
18. N. Cao, S. B. Nasir, S. Sen, and A. Raychowdhury, "Self-optimizing IoT wireless video sensor node with in-situ data analytics and context-driven energy-aware real-time adaptation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 9, pp. 2470-2480, 2017.
19. Z. Chen et al., "Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and IV characteristics," *Applied Energy*, vol. 204, pp. 912-931, 2017.
20. A. Christina, M. Salam, Q. Rahman, F. Wen, S. Ang, and W. Voon, "Causes of transformer failures and diagnostic methods–A review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1442-1456, 2018.
21. N. A. Cloete, R. Malekian, and L. Nair, "Design of smart sensors for real-time water quality monitoring," *IEEE Access*, vol. 4, pp. 3975-3990, 2016.
22. T. Coito et al., "Intelligent sensors for real-time decision-making," *Automation*, vol. 2, no. 2, p. 62, 2021.
23. M. Cui, D. Han, and J. Wang, "An efficient and safe road condition monitoring authentication scheme based on fog computing," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9076-9084, 2019.
24. Y.-A. Daraghmi et al., "Edge–fog–cloud computing hierarchy for improving performance and security of NB-IoT-based health monitoring systems," *Sensors*, vol. 22, no. 22, p. 8646, 2022.
25. S. Dhingra et al., "Internet of things-based fog and cloud computing technology for smart traffic monitoring," *Internet of Things*, vol. 14, p. 100175, 2021.
26. G. Diraco et al., "Review on human action recognition in smart living: Sensing technology, multimodality, real-time processing, interoperability, and resource-constrained processing," *Sensors*, vol. 23, no. 11, p. 5281, 2023.
27. G. A. Fink et al., "Security and privacy in cyber-physical systems," in *Cyber-physical systems*, pp. 129-141, Elsevier, 2017.
28. F. Fuentes-Peñailillo et al., "Transformative technologies in digital agriculture: Leveraging Internet of Things, remote sensing, and artificial intelligence for smart crop management," *Journal of Sensor and Actuator Networks*, vol. 13, no. 4, p. 39, 2024.