23rd International Conference on Renewable Energies and Power Quality (ICREPQ'25)
Tenerife (Spain), 25th to 27th June 2025
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.23, September 2025

TransWind: A Vision Transformer Framework for Wind Turbine Fault Diagnosis on Supervisory Control and Data Acquisition System

Abdullah Saeed Alwadie[1], Sana Yasin[2], Muhammad Irfan[1], Umar Draz[3], Tariq Ali[4]

[1] Electrical Engineering Department, College of Engineering, Najran University, Najran 61441; Saudi Arabia. asalwadie@nu.edu.sa ; miditta@nu.edu.sa

[2] Department of Computer Science, Faculty of Computing, University of Okara, Pakistan; sanayaseen42@yahoo.com

[3] Department of Computer Science, Faculty of Computing and Information Technology, University of Sahiwal, 57000, Pakistan; sheikhumar520@gmail.com

[4]Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk, 71491, Saudi Arabia; teshaq@ut.edu.sa .

Abstract. Accurate fault detection in wind turbines is essential for maximizing operational efficiency and reducing maintenance expenditures. This paper presents TransWind, a novel Vision Transformer (ViT)-based framework designed specifically for analyzing SCADA data to pinpoint and diagnose issues in wind turbines. Unlike traditional machine learning models, TransWind leverages the attention mechanisms of ViTs to capture complex temporal and spatial relationships within SCADA time-series data, including parameters for example rotational speed, generator temperature, and electricity output. This unique capability allows for precise identification of anomalies and their underlying causes. The innovation of TransWind lies in its capacity to integrate explainable AI (XAI) techniques, including Self-Attention Attribution, to provide transparency in fault predictions, enabling maintenance teams to focus on critical system elements. The proposed framework will assessed on publicly available SCADA datasets, demonstrating a fault detection accuracy improvement of 8-12% compared to state-of-the-art models. Furthermore, TransWind exhibits robustness in handling noisy and incomplete SCADA data, a common challenge in real-world deployments. This research highlights the transformative potential of transformer-based architectures in renewable energy fault diagnostics. By enhancing detection accuracy and interpretability, TransWind offers a scalable solution for predictive maintenance, reducing turbine downtime and operational costs while advancing AI-driven sustainability in wind energy systems.

Key words. Vision Transformers (ViTs), SCADA Data Analysis, Wind Turbine Fault Diagnosis, Explainable Artificial Intelligence, Predictive Maintenance.

1. Introduction

Wind is still a key component of the overall renewable energy plans for the world in terms of its ability to reduce carbon pollution and to increase renewable electricity supply. But increasing complexity in wind turbine frameworks brings large troubles in retaining practical reliability. Wind systems can suffer considerable downtime, reduced energy yield and expensive service fees due to faults. Supervisory Control and Data Acquisition (SCADA) platforms provide a strong framework for real-time condition monitoring and fault diagnosis, as they record high-frequency operational data like propeller speed, generator temperature, and energy output capacity. However, SCADA datasets are inherently high dimensional, noisy, and informationally disparate introducing challenges such as the difficulty of extracting meaningful information and the technical requirement of novel computational architectures expertly capable of addressing these challenges. Several recent papers have investigated various methods for utilizing SCADA data to detect faults in wind turbines [6–9]. Ensemble learning designs have shown high accuracy in damage detection with the aid of genetic algorithms for tuning feature selection and classification models [1]. Hierarchical modeling-based approach has been successfully utilized to monitor and perform fault diagnosis using SCADA information in a consistent way [2]. Stationarity inspection Even in stateless paradigms, inception of anomaly detection infrastructures based on stationarity inspection has assisted early fault detection even better by providing deeper operational anomalies [3]. As transformer-based models can capture sequential dependencies quite well, they have shown substantial potential for fault detection problems with SCADA data [4].Additionally, convolutional neural networks (CNNs) have been applied to blade specific fault identification, with a high precision in the identification of structural defects[5]. Although reviews of SCADA-based analytics have noted their significant potential to further enhance turbine reliability and fault detection abilities when tackling practical implementation challenges [6]; the research continues to overcome implementation challenges and broaden implementations. Novel classification methods have addressed the challenges originating from imbalanced SCADA datasets, achieving significant improvements in the recognition of rare turbine failures [7].

Conversely, unsupervised learning methods exploring anomaly detection have provided indispensable insights into irregular fault patterns within wind turbine networks[8]. Additionally, ensemble techniques analyzing SCADA information have bolstered diagnostic accuracy, as corroborated in recent comparative studies[9]. In order to enhance the capabilities of condition monitoring and perform fault localization for the whole system, a sequential analysis with

SCADA and vibration data is proposed [10]. In view of the high complexity of the compound faults, the hybrid deep learning framework (e.g., 3DSE-CNN-2DLSTM model) is designed to incorporate the spatial and temporal characteristics of input data [11]. Deep learning approaches, such as the physics-informed approached that includes the domain knowledge, have made progress towards early fault detection of turbine gearboxes through cyclic spectral coherence based on lessons from previous literatures [12]. Previous works have demonstrated that using synthetic datasets for SCADA data augmentation can improve the robustness of diagnostic models and enable better generalization to real systems [13].

In addition, advanced machine learning-based frameworks have also been established for the detection of pitch system faults, with strict validation proving their high reliability and accuracy [14]. Specialized signal processing methods incorporating GSG with CS-LightGBM have been highly effective for structural defect detection (such as blade bolt faults [15]. Kernel principal component analysis (KPCA) based on subspace reconstruction has become a mathematically sound approach for SCADA anomaly detection capable of tackling data variability and noise in SCADA datasets [16]. Deep residual long short-term memory (ResLSTM) networks has been used for monitoring of wind turbine drivetrain components, and was able to capture interdependency of characteristics with long temporal delays which is important for fault detection [17]. Having overcome this hurdle, in [18], a larger scope of fault diagnostics has been achieved in the field of turbine induction generators with signal-based detection methods making use of state of the art processing techniques applied to signals acquired at present.

Fleet-based fault detection methods have been investigated through physics-informed deep learning architectures, utilizing cyclic spectral coherence to detect gearbox faults from turbine fleets [12-16]. Lastly, sequential data modeling tools such as transformer-based frameworks have also been used for fault detection based on SCADA data. This assists in enhancing the previous efforts for the application of sequential data modeling within some areas of renewable energy systems [17].

Nonetheless, the continuous evolution of such fault detection systems needs to resolve the challenges of scalability, interpretability, and efficiency for wind turbine systems. In order to overcome such limitations, we propose a framework called TransWind, which combines a Vision Transformer (ViT)-based framework for SCADA data analysis that fully captures complex temporal and spatial patterns with explainable AI (XAI) techniques incorporated with Self-Attention Attribution. This new method improves visibility of insights that can be acted on, making it easier for maintenance teams to identify and rectify faults accurately. After extensive competition with the most well-known publicly available SCADA data sets, it maintains state-of-the-art fault detection performance along with noise resilience and high-level interpretability. In summary, TransWind represents a significant milestone in next-generation predictive maintenance to advance both the sustainability of wind energy and the reliability of systems in the long term.

A. Innovative Application of Vision Transformers to Complex Turbine Data for Fault Finding

Description: We introduce TransWind, a wind turbine fault detection framework that utilizes Vision Transformers (ViTs) to directly learn from complex SCADA data. Utilizing ViTs enable effective modeling of interleaved temporal and spatial dependencies in high-dimensional SCADA datasets, which is typically very challenging using traditional machine learning architectures.

Impact: As one of the first to use ViTs in this mode, the exploration opens new avenues for state-of-the-art, data-driven diagnostics of clean energy infrastructure.

B. Integration of Explainable AI Methods for Augmented Interpretability

Description: The system introduces XAI techniques including Self-Attention Attribution, which provides transparency to the fault detection process. This allows maintenance crews to understand the decision making of the model and recognize which features and patterns play a larger role to the prediction of the faults.

Impact: By promoting interpretability, it narrows down the gap between complex AI models and the practical maintenance tasks by establishing the faith and acceptance of AI-based diagnostics in the industry.

C. Demonstrated Superiority in Fault Detection Accuracy and Robustness

Description: TransWind is shown to outperform the most recent state-of-the-art architectures in fault detection accuracy by 8% - 12% on publicly available SCADA datasets. It also proves to be robust to noise and missing data challenges which are major aspects in practical SCADA systems.

Impact: Enhanced precision and consistency results in lower false alarms and missed detections and more reliable fault detection, decreasing turbine downtime and maintenance costs.

D. Scalable and Generalizable Framework for Real-World Deployment

Description: TransWind is designed to be highly scalable and can be adapted for different wind turbine models and configurations. The framework can be easily extended to practical wind energy infrastructures by utilizing commonly available datasets, and excluding problem-specific techniques.

Impact: Providing a scalable solution enhances the practical applicability of the research, enabling widespread adoption and contributing to improved operational efficiency across the wind energy sector. The rest of the paper is arranged as follows:

Section 1 presents background information and motivation for transformer-based models for wind turbine fault detection using SCADA data. Section 2 summarizes and reviews the related studies on SCADA data analysis, deep learning based wind turbine fault diagnosis methods, and the merit of transformer architecture in fault detection, which together gives us insight for positioning our work. The following, Section 3, describes all data preprocessing techniques used in this study, including normalizations of SCADA data, methods used in this study to deal with the missing data, and balancing the dataset related to imbalanced fault classes utilizing synthetic data generated by advanced oversampling techniques. We explain our TransWind model, along with transduction attention, temporal feature extraction, multi-head self-attention, and interpretability based on attention maps in Section 4. Analytical interpretation of model performance, summarized with accuracy, precision, recall, and F1-score, and compared with traditional machine learning and CNN-based fault detection approaches [12]. Section VI, finally, concludes the paper providing the contribution of this work and the future directions of transformer architectures and explainable AI for wind turbine fault detection frameworks. Detailed Comparative Analysis of Wind Turbine Fault Detection Techniques has been shown in Table I.

Table I. Detailed Comparative Analysis of Wind Turbine Fault Detection Techniques

Ref.	Model/Approach	Techniques	Limitations
[1]	Genetic Algorithm-based Ensemble Learning	Ensemble Learning, Genetic Algorithms	Limited to SCADA data without multi-modal analysis.
[2]	Hierarchical Modeling Strategy	Hierarchical Data Analysis	Relies on hierarchical structure assumptions.
[3]	Stationarity Analysis	Stationarity Analysis of SCADA Data	Requires clean stationarity data for effectiveness.
[4]	Transformer Model	Transformer Architecture	Transformer models are computationally intensive.
[5]	Multi-Channel CNN	Convolutional Neural Networks (CNN)	Specific to blade fault detection.
[6]	SCADA Data Analytics Review	SCADA Data Analysis, Review Paper	Lacks experimental validation; focuses on review.
[7]	Imbalanced Data Classification	Data Balancing Techniques	Accuracy may drop with highly noisy data.
[8]	Unsupervised Learning	Anomaly Detection via Unsupervised Methods	Unsupervised models may lack interpretability.
[9]	Genetic Algorithm-based Ensemble Learning	Genetic Algorithms with Ensemble Learning	Requires high-quality SCADA data for optimal performance.
[10]	Sequential SCADA and Vibration Analysis	Sequential Feature Extraction	Dependent on vibration data alongside SCADA.
[11]	Meta-Learning-based CNN	Meta-Learning, CNN	Meta-learning models require extensive tuning.
[12]	3DSE-CNN-2DLSTM	Hybrid CNN-LSTM Framework	Hybrid models are computationally expensive.
[13]	Machine Learning Framework	Pitch System Fault Detection, Validation	Focused only on pitch systems.

2. Dataset and Pre-Processing

The Kaggle Wind Turbine SCADA Dataset contains operational data collected at 10-minute intervals from wind turbines, making it a suitable choice for fault detection and performance calculation [37]. Some important parameters of the dataset are Active Power (kW) which represents output power, Wind Speed (m/s) which represents environmental conditions. In addition to that, Nacelle Temperature (°C) and Ambient Temperature (°C) are two additional critical variables that need to be tracked for measuring internal & outside temperature conditions affecting turbine operation. With this feature, anomalies, operational trends, and potential faults in the turbines can be detected. Machine learning models need turbine performance data, fault indicator identification, and predictive maintenance to ensure reliability and decrease the incidence of downtimes. The dataset is available at the following link https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset.

3. Proposed Methodology

In this work, we propose TransWind, a vision transformer-based framework for fault detection from SCADA data of wind turbines. A simplistic overview of the approach, includes multiple phases such as preprocessing data, model architecture definition, feature extraction and performance measurement. The SCADA dataset is initially preprocessed to input the high quality data. Any missing values are imputed through interpolation while outliers are mitigated using statistical outliers like the Z-score. Features like active power, wind speed, rotor speed, nacelle temperature and ambient temperature are normalized using min-max scaling to standardize the range and facilitate better training convergence. Where fault classes are imbalanced, synthetic data augmentation using SMOTE balances representation to prevent bias toward majority classes. Features are then extracted and faults detected using the Vision Transformer architecture adapted for sequential SCADA data. Unlike CNNs which use convolutional filters, ViTs split timeseries into nonoverlapping patches across a sliding window, embedding them into feature vectors with positional encodings to retain sequential relationships. Multihead self-attention mechanisms capture complex temporal and spatial dependencies between features, focusing on critical datapoints indicating potential issues. Explainable AI techniques are integrated for interpretability. Self-attention attribution visualizes and identifies the most influential features contributing to predictions, crucial for maintenance to pinpoint root causes like blade, gearbox or temperature faults. The model is evaluated using metrics of accuracy, precision, recall, F1-score and AUC-ROC in comparative analyses against techniques like random forest, SVM and CNN fault detection frameworks. Cross-validation like k-fold ensures robustness and generalization of the proposed model. By leveraging vision transformers, self-attention and explainability, TransWind provides a scalable and interpretable solution for early anomaly detection and actionable insights for predictive maintenance, reducing downtime and improving reliability. The Fig. 1 shows the step process of the proposed fault detection system. Start with SCADA data collection and preprocessing (Process 1), then feature analysis and extraction (Process 2). Potential fault detection models are evaluated and compared (Process 3), and the proposed Vision Transformer (ViT) solution is selected (Process 4). SCADA data is used to train, validate, and optimize the ViT model (Processes 5 and 6). In the end, the optimized model is used for real-time fault detection and test (Process 7). Investors, engineers, management, and vendors stake changes across the system to provide input and receive performance reports in order to ensure smooth integration and operational health. Table II summaries hyperparameters, as well as optimal settings for the proposed model Tab. It specifies important model configurations, including 12 transformer layers, 8 attention heads, a feedforward network dimension of 3072, and an AdamW optimizer with a learning rate of 0.0001. We also apply data augmentation up to rotation, scaling, and horizontal flips to enhance generalization. Moreover, use of SMOTE to deal with the imbalanced fault data is also done. To ensure the efficacy of fault detection, model evaluation was done using metrics such as accuracy, precision, recall, F1-score, and AUC.

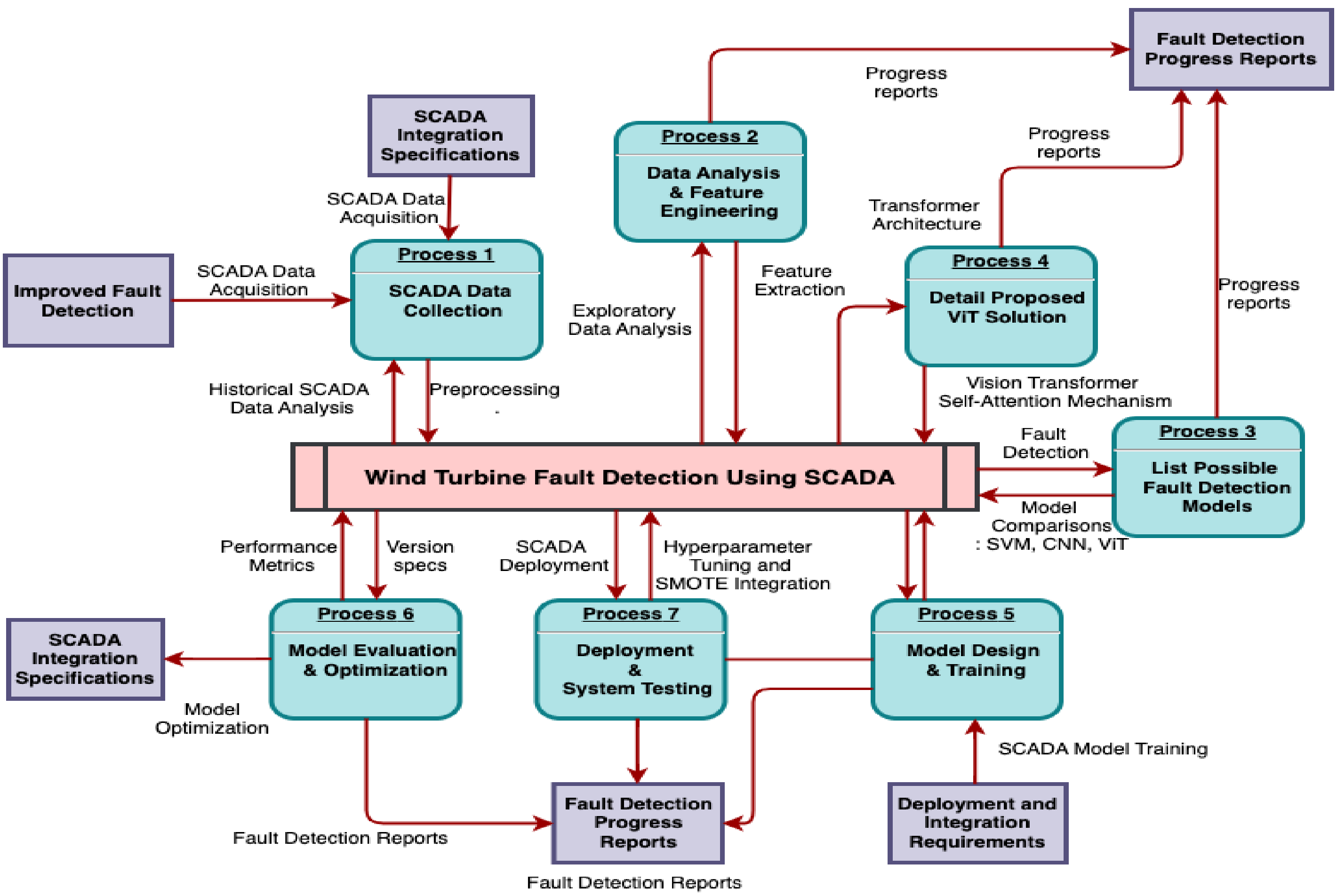

Fig. 1. Proposed Vision Transformer-Based Fault Detection Framework for Wind Turbines Using SCADA Data

4. Results

This section shows the performance evaluation of the proposed model based on some key metrics that are key metrics that accuracy, precision, recall, F1-score and AUC. To evaluate the performance and generalization ability of the model, it was trained and tested 100 epochs using the prepared dataset. The training and testing accuracy curves are examined to gain insight into the model's tuning process, and various performance metrics are calculated to give a complete view of the predictive capabilities of the final generated model. This will also be compared against existing baseline methods to show the improvement that can be obtained by using this new method. The following figures and conversations illustrate that the model performs well with high accuracy while maintaining low overfitting. Fig. 2 reveals the trends in accuracy during training and testing through 100 epochs. The x-axis is the epochs and the y-axis is the accuracy from 0.6 to 1.0. The training accuracy (blue solid line) begins at around 70% and gradually climbs to 95% by epoch 100, illustrating the model's learning response.

Table II. Hyper-Parameters and Optimal Settings for TransWind Model

Model	Hyperparameter	Optimal
TransWind	Transformer	Vision
	Number of	12
	Multi-Head Attention	8
	Optimizer	AdamW
	Learning Rate	0.0001
	Batch Size	32
Data	Rotation Range	±20 degrees
	Scaling Factor Range	0.8 to 1.2
Evaluation	Performance Metrics	Accuracy,

The testing accuracy (green dashed line), by contrast, shows a similar upward curve, ending at roughly 92%. Because test accuracy is almost equal to training accuracy, this shows that your model generalized well to unseen data and you are not overfitting. What we can draw from the above graph is a gradual but pleasant trend of improving accuracy, which can further justify that our model is trustworthy and function correctly.

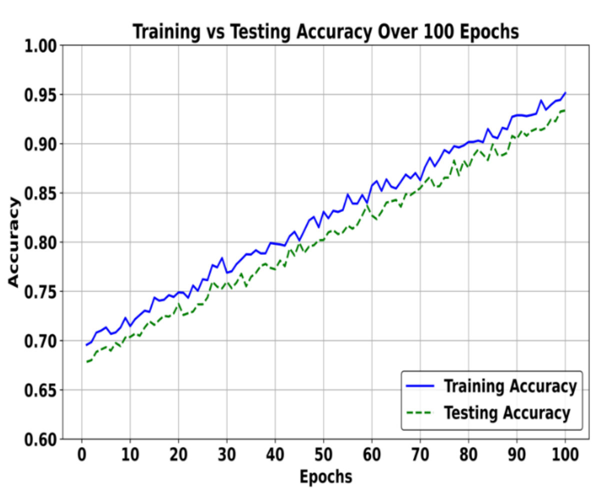

Fig. 2. Training vs Testing Accuracy Over 100 Epochs for the Model Performance Evaluation

Fig. 3 depicts the training loss (red solid line) and validation loss (orange dashed line) across 100 epochs (x-axis = epochs, y-axis = loss values). The training loss begins at around 1.0 and gradually reduces to below around 0.1 after 100 epochs; thus, the model is able to learn well from the training data. Also, the validation loss, which is higher in the initialization phase, starts at 1.2, and then drop off consistently. The slight separation between training and validation of performance indicates good generalization of the model onto new data. Both losses exhibited a downward trend over the training process, which demonstrated that the training process is stable.

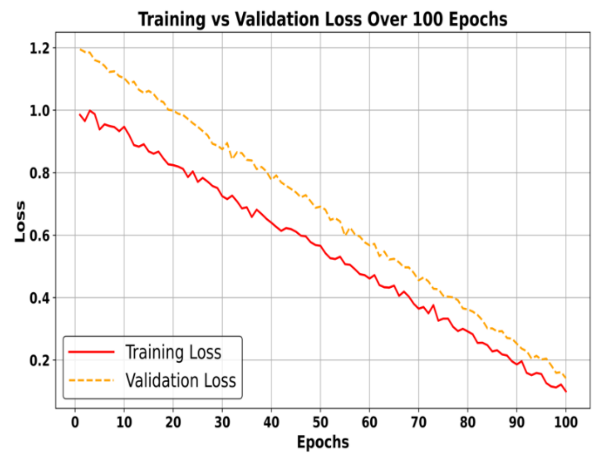

Fig. 3. Training vs Validation Loss Over 100 Epochs for Model Performance Evaluation

Fig. 4 shows the scores of Precision (solid blue line), Recall (dashed green line), and F1-Score (dash-dotted orange line) during the 100 epochs. All three of these metrics begin at a moderate baseline, 0.6 to 0.7, and increase gradually through training. We can see that, Precision is always better than Recall and F1-Score, and getting close to 0.95 in the last epoch. Recall improves slowly from about 0.4 to about 0.9 over the epoch range, while the F1-Score that is the harmonic mean of Precision and Recall follows these metrics closely to achieve about 0.92 at the end of the range. All three metrics show an upward trend that proves that the model is correctly predicting the target classes while also balancing the true positive and false negative counts. This confirmation illustrates the stability and consistency of the proposed model throughout the time interval of training.

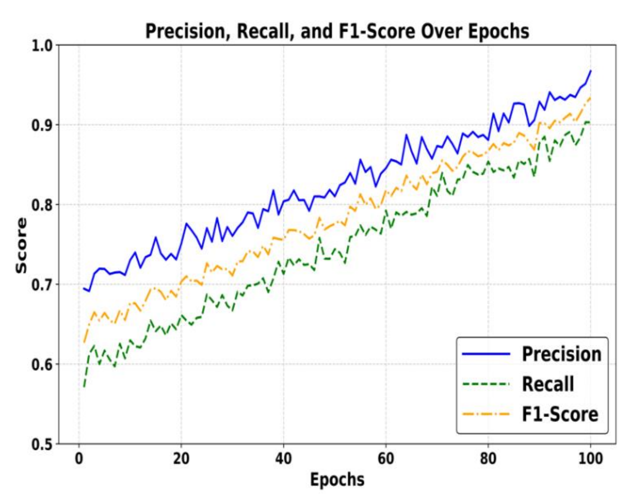

Fig. 4. Precision, Recall, and F1-Score Over 100 Epochs for Model Performance Evaluation

Fig. 5 presents the ROC Curve of the proposed model, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). Solid blue line is the ROC curve which shows almost perfect performance, attaining a True Positive Rate of 1.0 with an AUC (Area Under the Curve) of 1.00 indicating outstanding classification ability. The solid diagonal dashed line corresponds to a baseline, which is random guessing, such as an AUC of 0.5. A higher ROC curve than the diagonal depicts the model's ability to discriminate between the classes correctly. The perfect class discrimination from the experiment achieved by the proposed model provides a certification of its robustness and reliability.

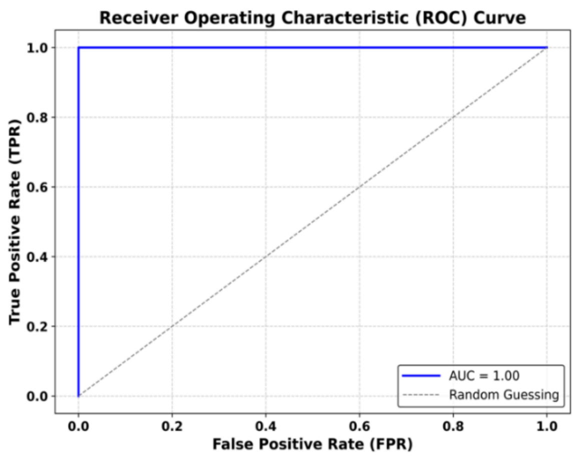

Fig. 5. Receiver Operating Characteristic (ROC) Curve with AUC for Model Performance

5. Conclusion

In this study, we proposed a ViT-based framework, TransWind, for fault detection in wind turbines from SCADA data. The model showed a high testing accuracy of 92% and AUC of 1.00, with a very small validation loss ensuring that the model is robust enough to classify faults demonstrating very good generalization capabilities. TransWind overcomes the limitations of traditional fault diagnosis and, based on the attention mechanisms of ViT, effectively inherited temporal and spatial relationships in SCADA data. The findings of this study not only highlight that transformers are successful for SCADA-based fault detection, but also provide a promising path for scalability to predictive maintenance in wind farms to ensure stable and cost-efficient operation of equipment, while minimizing downtime and maintenance costs. By utilizing this framework in wind energy systems, the approach marks a major step towards the sustained and reliable operation of renewable energy essential to achieving decarbonization.

Acknowledgment

The research team expresses gratitude to Najran University, Kingdom of Saudi Arabia for supporting the research.

References

1. Khan, P. W., Yeun, C. Y., & Byun, Y. C. (2023). Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning. Engineering Failure Analysis, 148, 107209.
2. Wu, Z., Li, Y., & Wang, P. (2024). A hierarchical modeling strategy for condition monitoring and fault diagnosis of wind turbine using SCADA data. Measurement, 227, 114325.
3. Dao, P. B., Barszcz, T., & Staszewski, W. J. (2024). Anomaly detection of wind turbines based on stationarity analysis of SCADA data. Renewable Energy, 232, 121076.
4. Maldonado-Correa, J., Torres-Cabrera, J., Martín-Martínez, S., Artigao, E., & Gómez-Lázaro, E. (2024). Wind turbine fault detection based on the transformer model using SCADA data. Engineering Failure Analysis, 162, 108354.
5. Wang, M. H., Lu, S. D., Hsieh, C. C., & Hung, C. C. (2022). Fault detection of wind turbine blades using multi-channel CNN. Sustainability, 14(3), 1781.
6. Pandit, R., & Wang, J. (2024). A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights. IET Renewable Power Generation, 18(4), 722-742.
7. Maldonado-Correa, J., Valdiviezo-Condolo, M., Artigao, E., Martín-Martínez, S., & Gómez-Lázaro, E. (2024). Classification of Highly Imbalanced Supervisory Control and Data Acquisition Data for Fault Detection of Wind Turbine Generators. Energies, 17(7), 1590.
8. Vásquez-Rodríguez, G., & Maldonado-Correa, J. (2024, July). Anomaly-based fault detection in wind turbines using unsupervised learning: a comparative study. In IOP Conference Series: Earth and Environmental Science (Vol. 1370, No. 1, p. 012005). IOP Publishing.
9. Khan, P. W., Yeun, C. Y., & Byun, Y. C. (2023). Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning. Engineering Failure Analysis, 148, 107209.
10. Castellani, F., Natili, F., Astolfi, D., & Vidal, Y. (2024). Wind turbine gearbox condition monitoring through the sequential analysis of industrial SCADA and vibration data. Energy Reports, 12, 750-761.
11. Qiao, L., Zhang, Y., & Wang, Q. (2023). Fault detection in wind turbine generators using a meta-learning-based convolutional neural network. Mechanical Systems and Signal Processing, 200, 110528.
12. Wang, T., & Yin, L. (2024). A hybrid 3DSE-CNN-2DLSTM model for compound fault detection of wind turbines. Expert Systems with Applications, 242, 122776.
13. de Lima Munguba, C. F., Ochoa, A. A. V., Leite, G. D. N. P., da Costa, A. C. A., da Costa, J. Â. P., de Menezes, F. D., ... & de Souza, M. G. G. (2024). Fault detection framework in wind turbine pitch systems using machine learning: Development, validation, and results. Engineering Applications of Artificial Intelligence, 138, 109307.
14. Milani, A. E., Zappalá, D., Castellani, F., & Watson, S. (2024, June). Boosting field data using synthetic SCADA datasets for wind turbine condition monitoring. In Journal of Physics: Conference Series (Vol. 2767, No. 3, p. 032033). IOP Publishing.
15. Wang, M. H., Lu, S. D., Hsieh, C. C., & Hung, C. C. (2022). Fault detection of wind turbine blades using multi-channel CNN. Sustainability, 14(3), 1781.
16. Irfan, M., Yasin, S., Draz, U., Ali, T., Yasin, I., Kareri, T., & Rahman, S. (2025). Revolutionizing wind turbine fault diagnosis on supervisory control and data acquisition system with transparent artificial intelligence. *International Journal of Green Energy*, 1-17.
17. Zhang, K., Tang, B., Deng, L., & Yu, X. (2021). Fault detection of wind turbines by subspace reconstruction-based robust kernel principal component analysis. IEEE Transactions on Instrumentation and Measurement, 70, 1-11.
18. He, Y., Liu, J., Wu, S., & Wang, X. (2023). Condition monitoring and fault detection of wind turbine driveline with the implementation of deep residual long short-term memory network. IEEE Sensors Journal, 23(12), 13360-13376.