I.J. Wireless and Microwave Technologies, **2026, 5, 316-329**
Published Online on October 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2026.05.19

Deep Dual Masked Transformer-Based Self-Supervised Anomaly Detection for Wireless Sensor Network Node Fault Diagnosis

Sayeekumar Madheswaran
Department of Computer Science and Engineering – AIML, Global Campus, Jain Deemed-to-be-University, Bengaluru – 562112, Karnataka, India
E-mail: sayee.academic@gmail.com
ORCID iD: https://orcid.org/0000-0003-2190-3009

Karthik Govindan Manoharan*
School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Katpadi – Thiruvalam Road, Vellore – 632014, Tamil Nadu, India
E-mail: karthik.gm@vit.ac.in
ORCID iD: https://orcid.org/0000-0002-8425-7828
*Corresponding Author

Syed Shameem
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur District – 522302, Andhra Pradesh, India
E-mail: shameemsyed@kluniversity.in
ORCID iD: https://orcid.org/0000-0001-8718-0734

M. Arumugam
Department of Computer Technology–PG, Kongu Engineering College (Autonomous), Perundurai, Erode – 638060, Tamil Nadu, India
E-mail: maructpg@gmail.com; arumugam.ctpg@kongu.edu
ORCID iD: https://orcid.org/0000-0002-1999-6798

Kasukurthi Rambabu
Department of Electrical and Electronics Engineering, Aditya University, Surampalem – 533437, Andhra Pradesh, India
E-mail: rambabu.kasukurthi@gmail.com
ORCID iD: https://orcid.org/0000-0001-6542-1322

V. Gokula Krishnan
Department of Computer Science and Engineering, Easwari Engineering College, Ramapuram, Chennai – 600089, Tamil Nadu, India
E-mail: gokul_kris143@yahoo.com
ORCID iD: https://orcid.org/0009-0005-6819-6729

Received: July 9, 2026; Revised: August 7, 2026; Accepted: September 15, 2026; Published: October 8, 2026

Abstract: Wireless sensor networks enable real-time monitoring in critical applications such as industrial automation, environmental sensing, healthcare, and infrastructure management. However, sensor node faults caused by hardware degradation, communication failures, and harsh operating conditions can significantly reduce data reliability and system performance. Existing anomaly detection approaches often depend on large amounts of labelled fault data and suffer from class imbalance, noise sensitivity, computational complexity, and limited generalization in dynamic environments, highlighting the need for adaptive and scalable fault diagnosis frameworks. To address these limitations, this study proposes a self-supervised deep anomaly detection framework for wireless sensor network node fault diagnosis. The proposed method integrates preprocessing and sliding-window segmentation with a Deep Dual Masked Transformer (DDMT) architecture to capture temporal dependencies and learn discriminative features without extensive labelled data. A self-supervised reconstruction objective with a 15% masking strategy is employed to optimize latent representations,

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

while an adaptive anomaly-scoring mechanism generates reliable fault predictions. Experimental results on the Zidi wireless sensor network dataset demonstrate that the proposed method achieves an accuracy of 99.0%, a precision of 98.5%, a recall of 98.3%, an F1-score of 98.4%, and an ROC-AUC of 99.5%, outperforming existing baseline methods. By combining temporal modelling and representation learning within a unified framework, the proposed approach improves anomaly detection accuracy, robustness, scalability, and generalization capability while maintaining low false alarm rates in challenging wireless sensor network environments.

Index Terms: Wireless Sensor Networks, Anomaly Detection, Self-Supervised Learning, Fault Diagnosis, Deep Representation Learning, Temporal Data Analysis

1. Introduction

Wireless Sensor Networks (WSNs) have become essential for environmental monitoring, industrial automation, smart agriculture, healthcare, and infrastructure management. These networks consist of spatially distributed sensor nodes that continuously collect and transmit data such as temperature, humidity, vibration, and pressure. Due to harsh environments, limited energy resources, hardware degradation, and communication interference, sensor nodes are highly prone to faults [1]. Undetected failures can propagate erroneous data, resulting in inaccurate decisions, reduced reliability, and operational risks. Traditional fault diagnosis methods rely on labelled datasets and manually defined thresholds, which are often impractical because labelled fault data are scarce and costly. Consequently, anomaly detection has emerged as a practical alternative, enabling identification of abnormal behaviour without exhaustive supervision. Recent advances in deep learning have enabled automatic feature extraction from raw telemetry signals [2]. Self-supervised learning leverages intrinsic data structures to learn meaningful representations without requiring extensive labelled data, improving scalability and adaptability in dynamic WSN environments. Recent advances in deep learning have significantly improved fault diagnosis in wireless sensor networks by enabling automatic feature extraction and temporal modelling from sensor data [3]. Self-supervised learning techniques have further attracted attention because they can learn meaningful representations without requiring extensive labelled samples. However, many existing approaches still rely heavily on annotated fault data, which are expensive and difficult to obtain in practical WSN deployments. Furthermore, current methods often face challenges related to class imbalance [4], noise sensitivity, computational complexity, and limited generalization to unseen fault patterns and changing environmental conditions. In resource-constrained WSN scenarios, these limitations can reduce detection reliability and hinder the practical deployment of intelligent fault diagnosis systems. Therefore, there is a need for adaptive and scalable anomaly detection frameworks that can effectively operate under dynamic network conditions while minimizing dependence on labelled data [5]. Although existing methods have achieved promising performance, they still suffer from several limitations, including dependence on labelled datasets, static threshold selection, computational complexity, and limited adaptability to dynamic WSN environments. Threshold determination remains challenging, as static thresholds may not generalize across operational conditions. Deep models can also be computationally intensive, limiting deployment in resource-constrained environments [6]. Although edge-cloud strategies mitigate this issue, latency and communication overhead may affect real-time detection. Additionally, self-supervised models may capture environmental variations rather than true fault characteristics, increasing false positives. Interpretability remains limited due to the opaque nature of deep embeddings. Furthermore, WSN datasets are often imbalanced, with few rare fault instances, complicating validation. This research proposes a Self-Supervised Deep Anomaly Detection framework for WSN node fault diagnosis, integrating segmentation-based temporal modelling with anomaly scoring and classification. The method begins with structured preprocessing and sliding window segmentation to preserve temporal continuity while stabilizing signals. A deep encoder network is trained using a self-supervised objective that leverages intrinsic data consistency rather than labelled fault instances, enabling robust representation learning in label-scarce environments. The model optimizes latent space representations to improve separability between normal and faulty patterns [7]. An anomaly score is computed from learned representations, followed by adaptive thresholding to generate binary fault labels. Unlike conventional supervised classifiers, the proposed framework unifies representation learning and decision modelling within a single pipeline, reducing dependence on annotated datasets. By combining segmentation-driven temporal context with self-supervised optimization, the approach enhances sensitivity to subtle degradations while maintaining low false alarm rates.

The main contributions of this work are summarized as follows:

- A self-supervised deep anomaly detection framework based on the Deep Dual Masked Transformer (DDMT) is proposed for wireless sensor network node fault diagnosis.
- A segmentation-based preprocessing strategy is introduced to capture temporal dependencies and improve feature representation.
- A self-supervised reconstruction mechanism and adaptive anomaly-scoring strategy are employed to reduce dependence on labelled fault data.

- Extensive experiments on the Zidi dataset demonstrate the effectiveness, robustness, and generalization capability of the proposed method.

This paper is structured as follows. Section 1 introduces the background and objectives. Section 2 reviews related work on WSN fault diagnosis and self-supervised anomaly detection. Section 3 presents the proposed method, including preprocessing, segmentation, representation learning, and classification. Section 4 discusses experimental results and evaluation. Section 5 concludes the paper and outlines future directions. Section 6 provides references supporting this study.

2. Literature Survey

The field of self-supervised deep anomaly detection for Wireless Sensor Network (WSN) node fault diagnosis has gained significant attention due to the increasing reliance on sensor-driven intelligent systems across industrial and environmental applications. Ensuring reliable fault detection in resource-constrained and dynamic WSN environments remains a critical challenge. In recent years, research efforts have expanded rapidly, encompassing diverse methodologies such as deep representation learning, graph-based modelling, hybrid machine learning frameworks, and adaptive statistical techniques. This literature survey systematically examines existing approaches to identify methodological progress, prevailing challenges, and research gaps that motivate further advancements in this domain. This category includes research that leverages graph structures and deep spatio-temporal feature learning for anomaly or intrusion detection in WSNs. Ye et al. [8] propose GPamba-AD, a graph-prompt fine-tuning framework integrating an improved Mamba model with multi-scale dilated convolution and a variational graph convolution network to capture complex spatial–temporal correlations. Their method enhances generalization while reducing training cost, though correlation anomaly detection remains challenging. Mohan et al. [9] introduce OOA-GNN, combining the Osprey Optimization Algorithm with Graph Neural Networks for optimized feature selection and classification. The graph-based representation enables modelling of inter-node dependencies, but computational overhead increases due to two-stage optimization. Wang et al. [10] develop ST-DMLAD, which integrates Graph Attention Networks and Transformers with triplet-loss metric learning to improve sensitivity under small-sample conditions. While effective in learning discriminative embeddings, triplet loss increases computational complexity. These graph-centered approaches share a focus on modelling node relationships and temporal dependencies; however, they commonly face challenges related to scalability, computational burden, and optimization efficiency in large or resource-constrained WSN deployments. This category comprises hybrid architectures that combine deep learning models with ensemble or machine learning classifiers. Haque et al. [11], present a Transformer-based Autoencoder integrated with Isolation Forest and XGBoost to distinguish between malfunctions and intrusions. Recent self-supervised anomaly detection approaches for time-series data have adopted contrastive learning and masked modelling strategies to reduce dependence on labelled samples. Contrastive learning methods focus on maximizing agreement between positive pairs, whereas masked modelling techniques reconstruct missing segments to capture temporal dependencies. Unlike these approaches, the proposed Deep Dual Masked Transformer integrates segmentation-based preprocessing, self-supervised representation learning, and adaptive anomaly scoring within a unified framework for wireless sensor network fault diagnosis. The layered validation strategy reduces false alarms, yet real-time adaptability remains limited under high noise conditions. Salmi et al. [12], evaluate multiple deep learning modelsDNN, CNN, RNN, and CNN+RNNfor DoS detection, concluding that CNN achieves superior accuracy, though recurrent architectures incur significantly higher training time. Talukder et al. [13], propose a hybrid pipeline using KMeans-SMOTE for data balancing, PCA for dimensionality reduction, and Random Forest for classification, achieving strong precision but lacking adaptability to dynamic network changes. Nguyen et al. [14], introduce GSWO-Cat Boost, integrating metaheuristic optimization with gradient boosting to achieve high accuracy and rapid inference; however, performance depends heavily on optimization parameters. Collectively, these hybrid models emphasize feature engineering, imbalance handling, and classifier optimization, but often encounter challenges in computational efficiency, adaptability, and robustness under evolving WSN conditions. Federated and distributed learning strategies form another methodological category focused on privacy preservation and scalable collaboration. Anwar et al. [15], propose a Federated Learning framework integrated with LSTM networks for intrusion detection across multiple IoT-based WSN datasets. By employing the Federated Averaging algorithm, the system trains models collaboratively without exchanging raw data, enhancing privacy and reducing transmission energy. Their approach achieves improved accuracy and lower false positive rates compared to centralized systems, demonstrating suitability for distributed sensor environments. However, the reliance on LSTM networks introduces significant computational and training overhead, particularly under extreme resource constraints. Additionally, federated settings remain vulnerable to dynamic network shifts and heterogeneous device capabilities. While this category emphasizes decentralized intelligence and energy efficiency, challenges persist in communication overhead, synchronization delays, and maintaining model consistency across non-identically distributed data sources. Such approaches are promising for large-scale WSN deployments but require further optimization to balance privacy, efficiency, and robustness. This category includes approaches prioritizing computational efficiency and adaptive optimization. Alauthman et al. [16], propose a Light Gradient Boosting Machine model combined with the Flying Fox Optimization Algorithm for intelligent fault detection and self-healing. The framework enhances

resilience and minimizes downtime, yet optimization costs increase with network size and environmental variability. Ahmad et al. [17], introduce Online Adaptive Kalman Filtering (OAKF), a lightweight statistical method that dynamically adjusts noise covariances and detection thresholds in real time. Recent studies have increasingly explored self-supervised representation learning and transformer-based architectures for time-series anomaly detection. Sun et al. [18] proposed a self-supervised tri-domain framework that improves anomaly discrimination by jointly modeling temporal, frequency, and representation spaces. Yang et al. [19] introduced masked autoencoders with temporal-frequency masking strategies to learn robust features from unlabeled time-series data. Marino et al. [20] developed an interpretable transformer-based anomaly detection framework that combines self-supervised learning with network-aware representations. Similarly, Dai et al. [21] presented a dual self-supervised representation learning strategy to capture anomalies from multiple perspectives. Although these methods reduce dependence on labelled data and improve generalization, their applicability to wireless sensor network fault diagnosis remains limited. Motivated by these advances, the proposed DDMT framework integrates self-supervised learning, temporal segmentation, and adaptive anomaly scoring for robust WSN node fault diagnosis. The approach is particularly suitable for resource-constrained nodes due to minimal processing time, though parameter tuning remains dataset-specific. Compared to deep learning frameworks, these methods emphasize simplicity, rapid processing, and practical deployment feasibility. However, they may struggle to capture complex nonlinear spatio-temporal dependencies present in large-scale WSNs. While computationally attractive, their generalization capacity and robustness under diverse attack or fault scenarios require further investigation.

3. Proposed Method

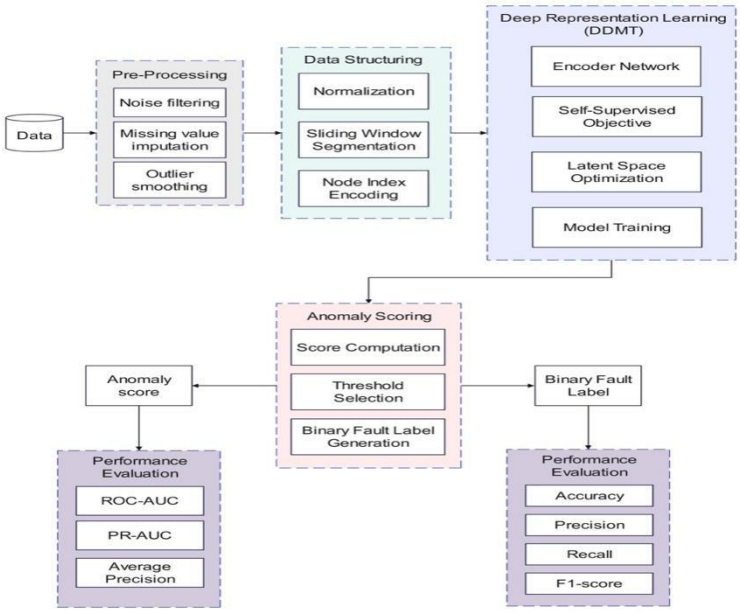


Fig. 1. Proposed self-supervised deep anomaly detection framework

Figure 1 illustrates the proposed self-supervised deep anomaly detection framework for Wireless Sensor Network (WSN) node fault diagnosis. The architecture is designed to systematically transform raw sensor telemetry into reliable anomaly scores and binary fault labels through a structured, modular pipeline. The system consists of five major stages: Pre-Processing, Data Structuring, Deep Representation Learning, Anomaly Scoring, and Performance Evaluation. The process begins with raw sensor data collected from distributed WSN nodes. Due to environmental noise, packet loss, and hardware instability, raw data often contain inconsistencies. The pre-processing module enhances signal quality through noise filtering, missing value imputation, and outlier smoothing. This ensures stability and reliability before feature extraction. Next, the Data Structuring module prepares the cleaned data for deep learning. Normalization scales sensor readings to a consistent range, while sliding window segmentation converts continuous streams into fixed-length temporal sequences. Node index encoding preserves spatial identity information, enabling the model to distinguish node-specific behaviour patterns.

The Deep Representation Learning block implements a Deep Data Modelling and Training strategy. An encoder network extracts compact latent embeddings from structured inputs. A self-supervised objective function guides training without requiring labelled fault data, while latent space optimization enhances separability between normal and abnormal behaviour. This design reduces dependency on annotated datasets and improves generalization. The learned representations are forwarded to the Anomaly Scoring module, where anomaly scores are computed based on reconstruction error or embedding deviation. A threshold selection mechanism converts continuous scores into binary

fault labels. Notably, the architecture supports parallel evaluation: anomaly score-based metrics (ROC-AUC, PR-AUC, Average Precision) and classification metrics (Accuracy, Precision, Recall, F1-score) are computed independently. The flow of information is sequential from data acquisition to evaluation; however, threshold tuning can form an implicit feedback loop during validation to optimize decision boundaries. The modular design enhances scalability, interpretability, and adaptability, making the framework suitable for resource-constrained and dynamic WSN environments.

The pre-processing module ensures data quality before representation learning. Its primary input is raw multivariate sensor data $X = \{x_{i,t}\}$, where i denotes the node index and t represents time. The module performs three key operations: noise filtering, missing value imputation, and outlier smoothing. Noise filtering reduces high-frequency disturbances using smoothing filters such as moving average or low-pass filters. Missing values, caused by communication failures or power interruptions, are imputed using interpolation or statistical estimation methods. Outlier smoothing mitigates extreme spikes that may distort training, typically through z-score clipping or median filtering. The output is a refined dataset X' that maintains structural integrity while reducing variance unrelated to faults. This module directly feeds into the Data Structuring block. By stabilizing data distribution, it prevents representation models from learning noise patterns instead of meaningful operational characteristics. The design rationale emphasizes robustness and generalization, ensuring that subsequent deep models focus on genuine node behavior patterns rather than artifacts.

The Data Structuring module transforms cleaned data into structured representations suitable for deep learning. The input X' is first normalized to ensure consistent scaling across sensor types. Sliding window segmentation partitions time-series data into fixed-length sequences S_k, preserving temporal dependencies. Node index encoding embeds spatial identity information, enabling the model to differentiate node-specific anomalies. This is particularly important in multi-node WSNs where fault patterns may vary spatially. The transformation produces a tensor $X_s \in \mathbb{R}^{N\times T\times F}$, where N is the number of windows, T is the window size, and F represents the features. This structured tensor becomes the input to the deep encoder. The rationale behind segmentation is to capture localized temporal patterns while maintaining computational efficiency. By structuring both temporal and spatial information, this module enhances the model's ability to detect gradual degradations and intermittent faults.

The Deep Dual Masked Transformer (DDMT) module performs self-supervised feature learning. It receives structured sequences X_s and passes them through an encoder network f_θ, parameterized by θ. The encoder extracts latent embedding H, representing compressed spatio-temporal features. A self-supervised objective such as reconstruction loss or contrastive loss is applied to optimize θ. Latent space optimization encourages separation between normal operational patterns and anomalous deviations. The output is a low-dimensional embedding $Z \in \mathbb{R}^d$, where d is the embedding dimension. These embeddings are forwarded to the anomaly scoring module. The advantage of this design lies in reduced reliance on labelled data. By learning intrinsic data structures, the model generalizes better to unseen fault types and dynamic network conditions.

This module converts latent representations into actionable fault decisions. The anomaly score A_i is computed for each window using reconstruction error or distance-based metrics in latent space. A threshold τ is selected either statistically or via validation to generate binary fault labels. If $A_i \geq \tau$, the node window is classified as faulty. This module enables dual evaluation: continuous anomaly scores are used for ROC-AUC and PR-AUC analysis, while binary labels are used for accuracy and F1-score computation. The separation of scoring and thresholding increases flexibility and supports adaptive tuning.

The raw multivariate sensor observations collected from all wireless sensor nodes are first represented as a temporal feature matrix, as expressed in Eq. (1). This representation serves as the input to the proposed self-supervised anomaly detection framework.

$$X = \{x_{i,t}\} \in \mathbb{R}^{N\times T\times F} \tag{1}$$

where Ndenotes the total number of sensor nodes, Trepresents the number of temporal observations, and Findicates the number of sensing features recorded at each time instant. To eliminate scale variations among heterogeneous sensor measurements, each feature is normalized using min-max normalization, as defined in Eq. (2).

$$\hat{x}_{i,t} = \frac{x_{i,t} - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where $x_{\min}$and $x_{\max}$denote the minimum and maximum values of the corresponding sensor feature, respectively. The normalized sensor stream is partitioned into fixed-length temporal segments using a sliding window operation, as formulated in Eq. (3).

$$S_k = [\hat{x}_k, \hat{x}_{k+1}, \dots, \hat{x}_{k+L-1}] \tag{3}$$

where Ldenotes the window length and $k$represents the starting index of each temporal segment. To preserve spatial information among distributed sensor nodes, each node is transformed into a learnable embedding vector, as described in Eq. (4).

$$e_i = E(i) \tag{4}$$

where $E(\cdot)$is the node embedding function and $e_i$represents the embedding corresponding to the i^{th}sensor node. The temporal features and node embeddings are fused to generate a unified input representation before deep feature learning, as expressed in Eq. (5).

$$Z_0 = S_k \oplus e_i \tag{5}$$

where $\oplus$ denotes feature concatenation and $Z_0$is the fused representation supplied to the encoder. A random masking strategy is applied to the fused representation to enable self-supervised learning through masked reconstruction, as formulated in Eq. (6).

$$Z_m = M \odot Z_0 \tag{6}$$

where $M$is the binary masking matrix, $\odot$ denotes element-wise multiplication, and masked elements are reconstructed during training. The masked sequence is encoded by the Transformer encoder to obtain compact latent representations, as defined in Eq. (7).

$$H = f_\theta(Z_m) \tag{7}$$

where $f_\theta(\cdot)$denotes the Transformer encoder parameterized by θ, and Hrepresents the learned latent feature space. The decoder reconstructs the original sequence from the latent representation to preserve meaningful temporal dependencies, as shown in Eq. (8).

$$\hat{Z} = g_\phi(H) \tag{8}$$

where $g_\phi(\cdot)$ represents the decoder network and $\hat{Z}$ denotes the reconstructed feature sequence. The reconstruction objective is optimized by minimizing the Mean Squared Error (MSE) between the original and reconstructed sequences, as presented in Eq. (9).

$$\mathcal{L}_{rec} = \frac{1}{n}\sum_{i=1}^{n} \left\| Z_0^{(i)} - \hat{Z}^{(i)} \right\|^2 \tag{9}$$

where $n$denotes the number of training samples and $\mathcal{L}_{rec}$measures the reconstruction quality of the self-supervised model. To improve latent feature discrimination and avoid overfitting, L2 regularization is incorporated into the learning objective, as given in Eq. (10).

$$\mathcal{L}_{reg} = \lambda \parallel \theta \parallel_2^2 \tag{10}$$

where $\lambda$is the regularization coefficient and θdenotes the trainable parameters of the encoder network. The overall optimization objective jointly minimizes the reconstruction loss and regularization term, as expressed in Eq. (11).

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{reg} \tag{11}$$

where $\mathcal{L}_{total}$is minimized using the Adam optimizer to learn generalized feature representations. The anomaly score for each temporal segment is computed from the reconstruction error obtained after model inference, as formulated in Eq. (12).

$$A_i = \left\| Z_0^{(i)} - \hat{Z}^{(i)} \right\|_2 \tag{12}$$

where $A_i$represents the anomaly score of the i^{th}segment, with larger values indicating a higher probability of sensor node faults. Finally, binary fault labels are generated using an adaptive threshold applied to the anomaly scores, as defined in Eq. (13).

$$y_i = \begin{cases} 1, & A_i \geq \tau \\ 0, & A_i < \tau \end{cases} \tag{13}$$

where τdenotes the adaptive decision threshold, $y_i = 1$indicates a faulty sensor node, and $y_i = 0$corresponds to normal operating conditions.

Algorithm 1: DDMT-Based Self-Supervised Fault Diagnosis

Input:
Raw sensor observations X
Output:
Binary fault labels y

Step 1: Normalize the raw sensor data using min-max normalization.
Step 2: Partition the normalized data into fixed-length temporal windows.
Step 3: Generate node embeddings for each sensor node.
Step 4: Concatenate temporal features and node embeddings.
Step 5: Apply random masking to the fused representation.
Step 6: Feed the masked sequence into the DDMT encoder.
Step 7: Reconstruct the original sequence using the decoder.
Step 8: Compute the reconstruction loss and regularization loss.
Step 9: Update the model parameters using the Adam optimizer.
Step 10: Compute anomaly scores from the reconstruction error.
Step 11: Apply adaptive thresholding to obtain binary fault labels.

Return y.

4. Result and Discussion

The proposed classification and grading framework was implemented in Python using the Google Colab environment, leveraging an NVIDIA A100 GPU for accelerated computation. The setup provided an efficient platform for training deep neural networks, enabling efficient processing of wireless sensor data and ensuring optimal utilization of computational resources for model optimization.

4.1. Model Specification

The proposed framework adopts a Deep Dual Masked Transformer (DDMT) architecture for self-supervised anomaly detection in wireless sensor network (WSN) nodes. The model is designed to learn robust temporal representations from multivariate sensor time-series data without requiring labelled fault information. Raw sensor readings are first segmented using a sliding window mechanism and projected into a 128-dimensional embedding space. The core architecture consists of three stacked Transformer encoder layers with four multi-head self-attention heads and a feedforward dimension of 256. A masking strategy is applied during training, where 15% of the input tokens are randomly masked to enforce contextual reconstruction learning. The model is optimized using the Adam optimizer with a learning rate of 0.001 and a dropout rate of 0.1 to prevent overfitting. Anomaly scores are computed using reconstruction error, where higher deviations indicate potential node faults. Binary fault labels are generated through adaptive thresholding of the anomaly scores. Table 1 summarizes the architectural and training hyperparameters used in the DDMT model, including encoder configuration, self-supervised settings, and optimization details to ensure stable learning and effective anomaly detection.

Table 1. Parameter Specification of the Proposed DDMT Framework

Parameter	Value
Input Window Size	20 time steps
Embedding Dimension	128
Transformer Encoder Layers	3
Number Of Attention Heads	4
Feedforward Network Dimension	256
Dropout Rate	0.1
Masking Ratio (Self-Supervised)	15%
Reconstruction Loss Function	Mean Squared Error (MSE)
Optimizer	Adam
Learning Rate	0.001
Batch Size	64

All experiments were conducted using a fixed random seed to ensure reproducibility. The model was trained with a batch size of 64 using the Adam optimizer and a learning rate of 0.001. The hyperparameters of the proposed DDMT

framework were selected empirically through preliminary experiments and validation performance. The window length was chosen to balance temporal dependency modelling and computational efficiency. A masking ratio of 15% was adopted to encourage contextual reconstruction while preserving sufficient information in the input sequence. The regularization coefficient λ was tuned experimentally to achieve an appropriate trade-off between reconstruction accuracy and model generalization. The final parameter settings shown in Table 1 provided the best overall performance on the validation dataset.

4.2. Datasets

For experimental evaluation, we employ the publicly available WSN fault detection dataset introduced by Zidi *et al[22].* and derived from multi-hop wireless sensor network measurements. The original data were collected using TelosB motes that recorded temperature and humidity readings every 5 s over a 6-hour period in an outdoor multi-hop network. The prepared version contains 4 688 observation vectors, each comprising measurements at three successive time instances, with two temperature and two humidity values per instance (total of 12 features per vector). To facilitate fault detection research, faults of four types (Offset, Gain, Stuck-at, and Out-of-bounds) were injected at varying rates (10 %, 20 %, 30 %, 40 %, 50 %) to create 60 labelled datasets. Each dataset consists of paired observation and label files, where the label y = 1 denotes normal behaviour and y = −1 denotes a faulty sample. Experimental fairness was ensured by applying identical preprocessing, normalization, and sliding-window segmentation procedures to all compared methods. The same training and validation splits were used across baseline models to provide an unbiased comparison. All experiments were conducted under identical hardware and software settings.

4.3. Classification matrix

The performance of the proposed classification model is quantitatively assessed using standard evaluation metrics derived from the confusion matrix. These metrics measure the model's capability to correctly distinguish between positive and negative classes.

Overall classification accuracy measures the proportion of correctly classified samples among all testing instances and is calculated using Eqn. (14).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{14}$$

where TP denotes correctly classified damaged samples, TN represents correctly classified non-damaged samples, FPindicates falsely predicted damaged samples, and FNcorresponds to damaged samples incorrectly classified as non-damaged. Precision evaluates the reliability of positive predictions by measuring how many predicted damaged samples are actually damaged, as defined in Eqn. (15).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{15}$$

where TPrepresents true positive predictions and FPdenotes false positive predictions. Higher precision indicates fewer false alarms during disaster damage assessment. Recall, also known as sensitivity, measures the capability of the proposed framework to correctly identify damaged samples and is expressed in Eqn. (16).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

where TPdenotes correctly detected damaged samples and FNrepresents damaged samples that were missed during prediction. A higher recall indicates improved detection of actual disaster damage. The F1-score provides a balanced evaluation by combining precision and recall into a single metric through their harmonic mean, as presented in Eqn. (17).

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where Precision measures prediction reliability and Recall evaluates detection capability. The F1-score provides a balanced assessment when class distributions are imbalanced. Specificity evaluates the capability of the proposed model to correctly recognize non-damaged samples and is calculated using Eqn. (18).

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{18}$$

where TN denotes correctly identified non-damaged samples and FP represents non-damaged samples incorrectly classified as damaged. Higher specificity indicates reduced false-positive predictions. The False Positive Rate (FPR) measures the proportion of non-damaged samples incorrectly predicted as damaged and is formulated in Eqn. (16).

Table 2 shows consistent performance between cross-validation and overall evaluation, indicating stable learning, strong generalization capability, and reliable fault detection accuracy across data splits.

Table 2. Performance of classification under training and validation evaluations.

Training / Testing	F1 Score	Spe	Sen	Acc	Dsc	Precision
5-FOLD CV (MEAN ± STD)	0.984	0.992	0.983	0.989	0.984	0.985
Overall Model	0.984	0.992	0.983	0.990	0.984	0.985

Both losses decrease steadily, indicating stable convergence and effective learning, while the small gap between curves suggests good generalization without significant overfitting.



Fig. 2. Training and validation reconstruction loss over training epochs.

The model achieves high and balanced performance for both fault and normal classes, demonstrating reliable discrimination capability and consistent accuracy in identifying abnormal and healthy sensor behavior.

Table 3. Performance of classification based on each class.

Classes	F1 Score	Spe	Sen	Acc	Dsc	Precision
Fault	0.984	0.992	0.983	0.990	0.984	0.985
Normal	0.991	0.983	0.992	0.990	0.991	0.991

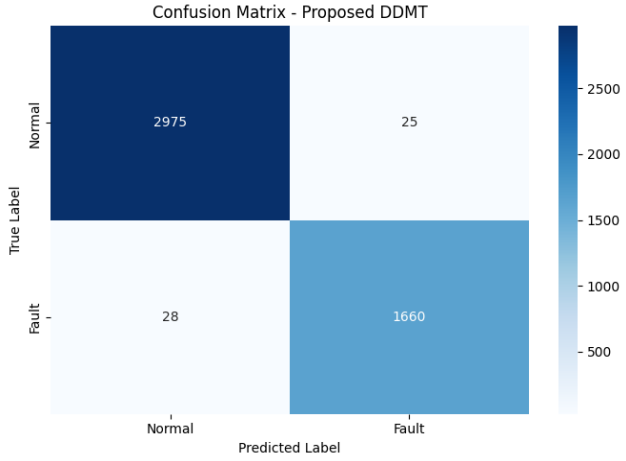


Fig. 3. Confusion matrix of the proposed DDMT model for fault classification.

The model correctly identifies the majority of normal and fault instances, with minimal misclassifications, demonstrating high detection accuracy and reliable fault discrimination. The proposed DDMT framework outperforms baseline approaches across all evaluation metrics, demonstrating superior accuracy, balanced precision–recall performance, and improved robustness for reliable fault detection.

Table 4. Comparison of the suggested method's performance with current techniques.

METHOD	ACCURACY	PRECISION	RECALL	F1	ROC-AUC
OOA-GNN [9]	0.987	0.946	0.941	0.944	0.981
ST-DMLAD [10]	0.979	0.950	0.890	0.918	0.975
DEEP-ANN [18]	0.964	0.963	0.958	0.961	0.972
RNN-LSTM [19]	0.980	0.956	0.952	0.954	0.983
PROPOSED DDMT	0.990	0.985	0.983	0.984	0.995

The proposed DDMT framework was compared with representative baseline methods, including OOA-GNN [9], ST-DMLAD [10], Deep-ANN [18], and RNN-LSTM [19]. The implementations followed the configurations reported in the corresponding studies whenever available. All methods were evaluated using the same dataset partitions and performance metrics to ensure a fair comparison.

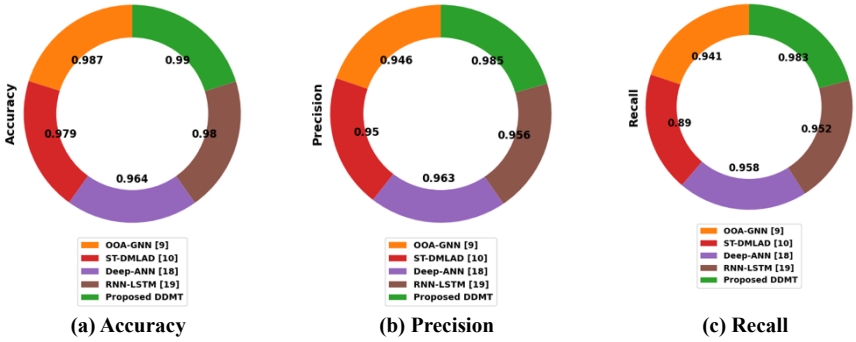


Fig. 4. Comparative performance of classification models.

The donut charts compare (a) Accuracy, (b) Precision, and (c) Recall across baseline methods and the proposed DDMT, highlighting its superior and consistent performance. The model maintains high accuracy as fault intensity increases, demonstrating strong robustness and stability even under progressively challenging and noisy operating conditions.

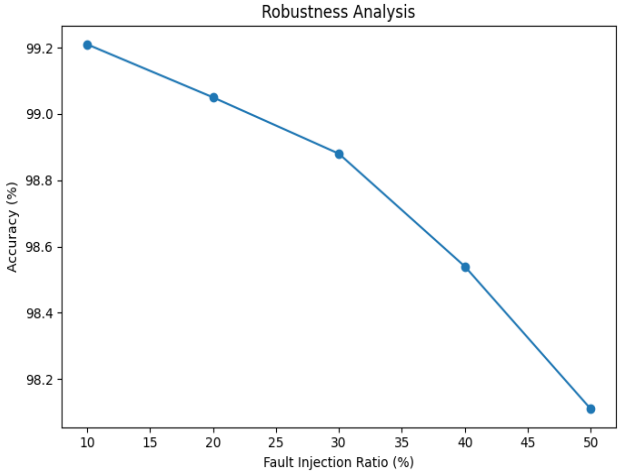


Fig. 5. Robustness analysis under varying fault injection ratios.

5. Conclusion

In conclusion, this study presents a structured self-supervised deep anomaly detection framework for Wireless Sensor Network (WSN) node fault diagnosis, integrating preprocessing, temporal segmentation, deep representation learning, and adaptive anomaly scoring within a unified pipeline. The proposed method transforms raw sensor data into

latent representations without relying heavily on labelled fault samples. By combining sliding window segmentation with a self-supervised encoder and latent space optimization, the framework enhances sensitivity to subtle and gradual node degradations while maintaining robustness against noise and missing data. The framework aims to improve sensitivity to subtle and gradual node degradations while maintaining robustness against noise and missing data. The anomaly scoring and adaptive thresholding mechanism further ensure reliable binary fault identification. The proposed approach addresses key limitations of prior methods, such as dependency on extensive labelled datasets, poor generalization to unseen faults, and sensitivity to environmental variations. Its modular design also improves scalability and adaptability for large-scale WSN deployments. Despite the promising results, several limitations remain. The proposed framework may experience performance degradation under extreme noise conditions, heterogeneous sensor configurations, or previously unseen fault patterns that differ significantly from the training data distribution. In addition, the Transformer-based architecture introduces computational overhead, which may limit real-time deployment on highly resource-constrained sensor nodes. The current model has been evaluated on a single public dataset and therefore requires further validation across cross-domain environments and large-scale wireless sensor networks. Consequently, the proposed approach is most suitable for scenarios in which sensor characteristics and operational conditions are reasonably consistent with the training setting. Future work will focus on four directions. First, adaptive window-sizing strategies and reinforcement-learning-based threshold optimization will be investigated to improve robustness under varying operating conditions. Second, lightweight and edge-deployable model variants will be developed to reduce computational complexity. Third, explainable learning mechanisms will be incorporated to improve the interpretability of latent representations and anomaly decisions. Finally, graph-based modelling and federated self-supervised learning will be explored and evaluated on heterogeneous and cross-domain WSN datasets. Additionally, integrating graph-based spatial correlation modelling and federated self-supervised learning could further improve robustness and privacy preservation in distributed WSN environments.

All the Declarations and Statements

Author Contributions Statement

Sayeekumar M.: Conceptualization, Methodology, Supervision, Project Administration, and Validation. Conceived the research idea, designed the overall framework, supervised the study, and reviewed the final manuscript.
Karthik G.M.: Data Curation, Software, and Investigation. Performed data collection, preprocessing, implementation, and experimental setup.
Syed Shameem: Formal Analysis, Model Training, Validation, and Performance Evaluation. Conducted model training, comparative analysis, and interpreted the experimental results.
M. Arumugam: Visualization, Statistical Analysis, and Investigation. Prepared figures, tables, performance visualizations, and assisted in evaluating the results.
K. Rambabu: Writing—Original Draft, Literature Survey, and Documentation. Prepared the initial manuscript, conducted the literature review, and documented the proposed methodology.
Dr. V. Gokula Krishnan: Writing—Review & Editing, Resources, and Project Coordination. Critically reviewed and edited the manuscript and coordinated the overall research activities.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research received no external funding.

Data Availability Statement

The datasets used in this study are publicly available wireless sensor network datasets. The processed data and implementation details used during the study are available from the corresponding author upon reasonable request.

Ethical Declarations

Ethical Approval: N/A
Consent to Participate: N/A
Consent for Publication: N/A
This study does not involve human participants, animals, or identifiable personal data.

Acknowledgments

The authors sincerely thank the Editor and the anonymous reviewers for their valuable comments and constructive

suggestions, which significantly improved the quality, clarity, and presentation of this manuscript. The authors also acknowledge the researchers and developers who made the public wireless sensor network datasets available to the research community.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used OpenAI's ChatGPT to assist with English language improvement, grammar correction, formatting, and manuscript organization. All scientific content, methodology, experimental design, analysis, interpretation of results, and conclusions were independently developed, verified, and approved by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the final published work.

Abbreviations

None.

Appendix A/B/C…, with appendix title

This appendix may include additional implementation details, hyperparameter settings, ablation studies, and supplementary experimental results that support the findings presented in the main manuscript.

Reference

[1] K.-X. Shi, S.-M. Li, G.-W. Sun, Z.-C. Feng, and W. He, “Fault diagnosis of wireless sensor network nodes based on belief rule base with adaptive attribute weights,” *Scientific Reports,*2024, vol. 14, Art. no. 4038.

[2] M. M. Feghhi, R. M. Alsharfa, and M. H. Majeed, “Efficient Fault Detection in WSN Based on PCA-Optimized Deep Neural Network Slicing Trained with GOA,” *arXiv preprint arXiv*: 2505.07030, 2025.

[3] J. Xia, D. Zhan, and X. Wang, “Machine learning diagnosis of node failures based on wireless sensor networks,” *Applied Mathematics and Nonlinear Sciences*, 2024. vol. 9, no. 1.

[4] P. Iswarya and K. Manikandan, "Spatio-temporal multi-fault and severity aware diagnosis in WSN via GNN transformers," Scientific Reports, vol. 16, Art. no. 23945, 2026, doi: 10.1038/s41598-026-52619-z.

[5] G. Jiang, K. Shen, X. Liu, X. Cheng, and P. Xie, "Hierarchical spatiotemporal graph network for fault diagnosis of industrial processes," IEEE Internet of Things Journal, vol. 12, no. 3, pp. 3043-3054, 2025, doi: 10.1109/JIOT.2024.3476287.

[6] Y. Fan, T. Fu, N. I. Listopad, P. Liu, S. Garg, and M. M. Hassan, "Utilizing correlation in space and time: Anomaly detection for Industrial Internet of Things (IIoT) via spatiotemporal gated graph attention network," Alexandria Engineering Journal, vol. 106, pp. 560-570, 2024, doi: 10.1016/j.aej.2024.08.048.

[7] M. Ye, Q. Zhang, X. Xue, Y. Wang, Q. Jiang, and H. Qiu, "A novel self-supervised learning-based anomalous node detection method based on an autoencoder for wireless sensor networks," IEEE Systems Journal, vol. 18, no. 1, pp. 256-267, 2024, doi: 10.1109/JSYST.2023.3347435.

[8] M. Ye, J. Cui, Y. Huang, Q. He, Y. Wang, and J. Zhang, “A graph prompt fine-tuning method for WSN spatio-temporal correlation anomaly detection,” *arXiv preprint arXiv*: 2601. 2026, 12745.

[9] D. B. Mohan, P. Arumugam, and A. R. Anand, “Osprey optimization algorithm integrated with graph neural networks for intrusion detection in wireless sensor networks,” Scientific Reports, 2025, vol. 15, pp. 44849, doi: 10.1038/s41598-025-28359-x.

[10] Z. Wang, M. Ye, J. Cheng, C. Zhu, and Y. Wang, “An anomaly node detection method for wireless sensor networks based on deep metric learning with fusion of spatial–temporal features,” *Sensors,* 2025, vol. 25, no. 10, pp. 3033, doi: 10.3390/s25103033.

[11] A. Haque and H. Soliman, “A transformer-based autoencoder with isolation forest and XGBoost for malfunction and intrusion detection in wireless sensor networks for forest fire prediction,” *Future Internet,* 2025, vol. 17, no. 4, pp. 164, doi: 10.3390/fi17040164.

[12] S. Salmi and L. Oughdir, “Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network,” *Journal of Big Data,* 2023, vol. 10, pp. 17, doi: 10.1186/s40537-023-00692-w.

[13] M. A. Talukder, M. Khalid, and N. Sultana, “A hybrid machine learning model for intrusion detection in wireless sensor networks leveraging data balancing and dimensionality reduction,” *Scientific Reports,* 2025, vol. 15, pp. 4617, doi: 10.1038/s41598-025-87028-1.

[14] T. M. Nguyen, H. H.-P. Vo, and M. Yoo, “Enhancing intrusion detection in wireless sensor networks using a GSWO-CatBoost approach,” *Sensors,* 2024, vol. 24, no. 11, pp. 3339, 2024, doi: 10.3390/s24113339.

[15] R. W. Anwar, M. Abrar, A. Salam, and F. Ullah, “Federated learning with LSTM for intrusion detection in IoT-based wireless sensor networks: A multi-dataset analysis,” *PeerJ Computer Science*, 2025, vol. 11, pp. e2751, doi: 10.7717/peerj-cs.2751.

[16] A. Alauthman and A. Al-Hyari, “Intelligent fault detection and self-healing mechanisms in wireless sensor networks using machine learning and flying fox optimization,” *Computers*, 2025, vol. 14, no. 6, pp. 233, doi: 10.3390/computers14060233.

[17] R. Ahmad and E. H. Alkhammash, “Online adaptive Kalman filtering for real-time anomaly detection in wireless sensor networks,” *Sensors,* 2024, vol. 24, no. 15, pp. 5046, doi: 10.3390/s24155046.

[18] Y. Sun, G. Pang, G. Ye, T. Chen, X. Hu, and H. Yin, “Unraveling the ‘Anomaly’ in Time Series Anomaly Detection: A Self-Supervised Tri-Domain Solution,” in Proc. IEEE 40th Int. Conf. Data Engineering (ICDE), Utrecht, Netherlands, May 2024, pp. 981–994, doi: 10.1109/ICDE60146.2024.00080.

[19] Y. Fang, J. Xie, Y. Zhao, L. Chen, Y. Gao, and K. Zheng, "Temporal-Frequency Masked Autoencoders for Time Series Anomaly Detection," in Proc. IEEE 40th Int. Conf. Data Engineering (ICDE), Utrecht, Netherlands, May 2024, pp. 1228–1241, doi: 10.1109/ICDE60146.2024.00099.
[20] D. L. Marino, C. S. Wickramasinghe, C. Rieger, and M. Manic, "Self-Supervised and Interpretable Anomaly Detection Using Network Transformers," IEEE Transactions on Industrial Informatics, vol. 21, no. 5, pp. 4252–4261, 2025, doi: 10.1109/TII.2025.3534443.
[21] Y. Dai, I. Spence, K. Rafferty, B. Quinn, J. Huang, and H. Wang, "TDSRL: Time Series Dual Self-Supervised Representation Learning for Anomaly Detection from Different Perspectives," IEEE Internet of Things Journal, vol. 12, no. 17, pp. 35078–35096, 2025, doi: 10.1109/JIOT.2025.3577931.
[22] M. M. Zidi, T. Moulahi, and B. Alaya, "Fault detection in wireless sensor networks through SVM classifier," IEEE Sensors Journal, Jan.2018, vol. 18, no. 1, pp. 340–347. doi: 10.1109/JSEN.2017.2771226.

Authors' Profiles



Dr. M. Sayeekumar is an Associate Professor in the Department of Computer Science and Engineering (AI & ML) at Jain Deemed-to-be University, Bangalore, India, with over 24 years of experience in teaching, research, and academic administration. He earned his Ph.D. from Anna University, Chennai, specializing in secure and scalable communication architectures. His research interests include Network Security, Machine Learning, Internet of Things (IoT), Cloud Computing, and Software-Defined Networks. He has published numerous research articles in SCI and Scopus-indexed journals, along with book chapters in reputed international publications such as Springer, Wiley, and IGI Global. He is also a patent holder in IoT-based intelligent systems. Dr. Sayeekumar has received research funding from AICTE and MSME for innovative projects. He actively mentors research scholars and contributes as a reviewer for international journals. He is a member of professional bodies including IEEE, ISTE, IAENG, and IACSIT.



Dr. G. M. Karthik Born in Madurai, Tamil Nadu state in India, in 1981, received the B.E. in Computer Science and Engineering from SACS MAVMM Engineering College, Madurai, M.E. in Computer Science and Engineering from PSNA College of Engineering and Technology, Dindugal, in 2003 and 2005 respectively. He has completed Ph.D in Information and Communication Engineering from Anna University, Chennai. He is active life member of ISTE and IEE. He is having 17 years of teaching experience in more than five engineering colleges in India. His primary research interests lie in the areas of Data Mining, Big data Analytics, Time series data, Web Mining, Modelling and Complexity Analysis. Currently, he is working as Associate Professor in School of Computer Science Engineering (SCOPE), Vellore Institute of Technology, Vellore, India.



Dr. Syed Shameem is currently serving as an associate professor in the Department of Electronics and Communication Engineering at Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, A.P.-522302. He received B.Tech. from JNTU, Hyderabad in 2001. He received M.Tech. from ANU, Guntur in 2007. He received his Ph.D. in the field of Bio Sensors from K L University in 2021. He has published 50 papers in international, national journals,and conferences. He is a Fellow in IETE. He is a life member of the Indian Science Congress and IAENG. His research interests include image processing, machine learning, and Bio Sensors.



Dr. M. Arumugam received his M.Sc. in Computer Science from the University of Madras, Chennai**,** in 1999. He obtained his M.Phil. in Computer Science **f**rom Manonmaniam Sundaranar University, Tirunelveli**,** in 2003, and was awarded his Ph.D. in Computer Science **by** Bharathidasan University, Tiruchirappalli**,** in 2026.He is currently serving as an Assistant Professor (Senior Grade) in the CT-PG Department **at** Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India**.** He has published 20 research papers in reputed SCI-, Scopus-, and UGC-indexed international journals**.** His research interests include Big Data Analytics, Machine Learning, and Deep Learning**.**



Kasukurthi Rambabu is an Assistant Professor in the Department of Electrical Engineering at Aditya University. He holds an M.Tech degree in power systems from Acharya Nagarjuna University, Guntur. He is pursuing Ph.D in Andhra University. His research interests include power systems, smart grids, and the development of novel control and converter topologies. He has published several research papers on topics such as modular multilevel converters, demand-side management using machine learning, deep neural techniques for breast tumor detection and IoT-enabled smart power systems. Currently, he is actively engaged in teaching, research, and academic development at Aditya University.



Dr. V. Gokula Krishnan is currently working as Professor in the Department of Computer Science and Engineering in Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India. He has completed his Under-Graduation (BE) in Anna University, Post-Graduation (M.Tech) in Dr. MGR University and Ph.D in Sathyabama Institute of Science and Technology, Chennai. He has more than 18 years of teaching experience in various colleges in Chennai and Hyderabad. He has published several papers in SCI/Scopus/WoS indexed journals and also he has presented various papers in National/International Conferences. His area of interest includes Computer Networks, Computer Architecture, Data Structures and Software Engineering etc. He serves as the guest editor, editorial member and also as reviewer in many reputed international journals. He is a member of Professional Bodies like ISTE, IAENG, CSI and IEEE etc.

How to cite this paper

Sayeekumar Madheswaran, Karthik Govindan Manoharan, Syed Shameem, M. Arumugam, Kasukurthi Rambabu, V. Gokula Krishnan, "Deep Dual Masked Transformer-Based Self-Supervised Anomaly Detection for Wireless Sensor Network Node Fault Diagnosis", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.16, No.5, pp. 316-329, 2026. DOI:10.5815/ijwmt.2026.05.19