Electrical Power and Energy Systems 175 (2026) 111607

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Hierarchical switch fault diagnosis based on transformer algorithm in four-leg inverters of stand-alone wind energy conversion systems

Jalal Heidari [a,b,*], Rasool Peykarporsan [c], Soroush Oshnoei [d], Tek Tjing Lie [c], Lieven Vandevelde [a,b], Guillaume Crevecoeur [a,b]

[a] Energy & Systems Lab, Department of Electromechanical, Systems and Metal Engineering, Ghent University, 9052 Ghent, Belgium
[b] FlandersMake@UGent – Corelab MIRO, Belgium
[c] School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand
[d] Department of Electrical and Computer Engineering, Aarhus University, 8200 Aarhus, Denmark

ARTICLE INFO

Keywords:
Fault detection and diagnosis
Four-leg inverter
Stand-alone wind energy system
Transformer algorithm

ABSTRACT

With the increasing development of renewable energy resources, stand-alone structures are gaining more attention. Among these, wind energy systems are particularly notable because of their advantages, including sustainability, low operational expenses, and minimal environmental impact. Due to the challenges of load balancing in such systems, four-leg inverters have emerged as a viable solution, offering improved performance under unbalanced load conditions. However, like all inverters, they remain susceptible to internal faults. Accordingly, this paper proposes a hierarchical two-level Transformer-based model to detect switch internal faults, including open-circuit and short-circuit in four-leg inverters. The OPAL-RT hardware-in-the-loop setup was used to generate data in various scenarios to validate the efficiency of the proposed framework. The results demonstrate that the developed technique can effectively classify fault types and identify faulty switches compared to state-of-the-art algorithms and single-level structures.

1. Introduction

In recent years, renewable energy resources have become increasingly popular as global energy needs have risen, necessitating a decrease in carbon emissions. Wind energy is one of the most viable renewable energy resources [1], due to its environmental benefits, economic competitiveness, and technological feasibility [2]. Wind energy conversion systems (WECSs) can use a number of different configurations. One of these configurations is wind turbines coupled with Permanent Magnet Synchronous Generators (PMSGs), which can be considered stand-alone structures for generating electricity [3]. In order to connect the PMSG to the loads, it is necessary to use inverters that convert the variable frequency and amplitude output voltage of the PMSG into a fixed frequency and amplitude voltage [4]. Moreover, since it is much more common to have unbalanced loads in these stand-alone structures, four-leg inverters can be installed after the PMSG, which show superior performance compared to three-leg inverters when dealing with unbalanced loads. The four-leg inverter has an extra neutral leg that can handle the unbalanced load currents and decrease harmonic distortions [5].

There have been several studies on four-leg inverters used in power systems. For example, for the four-leg inverter, a proportional resonant control technique is developed in *abc* reference frame [6]. In another study [7], a common mode voltage elimination technique is developed. Moreover, a direct predictive power control is another approach proposed in [8]. However, rather than only designing controllers, switch fault detection in such inverters is also of paramount importance, which was the focus of only a few studies. In general, four-leg inverters face two main types of internal faults: open-circuit (voltage loss due to switch malfunctions) and short-circuit (increased current caused by switch failure) [9].

Fault detection techniques in inverters can be divided into two categories: model-based and model-free techniques. In model-based approaches, fault diagnosis relies on an explicit model of the system, either through physics-based modeling or signal-model-based analysis [10,11]. However, model-free fault detection techniques are methods that do not require knowledge about the behavior or dynamics of a system. Model-free techniques (also known as data-driven techniques) can be more flexible and adaptable, making them more suitable to solve complex problems [12].

In the literature, model-based approaches have been investigated by researchers. In these studies, typically, an observer can detect faults in

* Corresponding author at: Energy & Systems Lab, Department of Electromechanical, Systems and Metal Engineering, Ghent University, 9052 Ghent, Belgium.
E-mail address: jalal.heidari@ugent.be (J. Heidari).

https://doi.org/10.1016/j.ijepes.2026.111607
Received 2 October 2025; Received in revised form 14 January 2026; Accepted 15 January 2026
Available online 27 January 2026

the system. For example, the authors in [13] developed a fault diagnosis approach using a voltage–current model flux observer, enabling timely and robust detection with low computational cost. In addition, a model-based fault detection approach for switches in inverters for PV applications is developed in [14], in which sliding observer banks incorporating PI control are employed to predict fault events. Moreover, in [15], a diagnosis strategy based on adaptive sliding mode observers is implemented to simultaneously detect IGBT open-circuit faults in inverters. Furthermore, in [16], an open-circuit fault diagnosis technique is presented for a three-level inverter using a predictive model based on the residual between the measured and estimated currents.

The development of fault detection systems based on model-free approaches, particularly those employing machine learning algorithms, has also gained significant attention in recent years [17]. These techniques typically rely on synthetic data and account for different fault scenarios within the system, enabling machine learning algorithms to achieve high detection accuracy [18]. For example, in [19], open-circuit faults in asymmetric multilevel inverters are detected using artificial neural networks (ANN). In addition, the authors in [20] developed a technique to detect faults in a three-phase inverter used in electric vehicles. Different machine learning algorithms, particularly the support vector machine (SVM) and decision tree (DT), have been investigated in [21] for a voltage source inverter to identify internal faults. Similarly, in [22], machine learning algorithms, specifically K-Nearest Neighbor (KNN) and Random Forest (RF), are applied to detect open-circuit faults in switches of a three-level Neutral Point Clamped inverter, which is connected to a Photovoltaic (PV) system. Moreover, an effective RF algorithm is also proposed in [23] based on a kernel principal component analysis technique to detect open-circuits and short-circuits within converter switches. Additionally, in [24,25], fault diagnosis methods using long short-term memory (LSTM) networks are proposed for three-level inverters.

Machine learning algorithms based on the transformer algorithm have also attracted increasing attention from researchers. For example, in [26], a transformer learning approach was developed to identify open-circuit faults of a three-level inverter connected to a permanent magnet synchronous motor (PMSM) drive for electric vehicles. In addition, another study [27] proposed a hybrid CNN and Transformer model with a multi-head attention mechanism for fault detection in high-voltage direct current (HVDC) networks under diverse operating and noise conditions, and [28] employed a Vision Transformer model for fault detection and classification using thermographic measurements in photovoltaic modules. Furthermore, authors in [29] combined a conditional generative adversarial network (CGAN) and a Vision Transformer by converting measured inverter signals into time-frequency images and using the generated samples to improve open-circuit fault diagnosis when only limited fault data are available.

As mentioned, research on fault detection in four-leg inverters remains limited. In [30], a fault diagnosis strategy is developed for a four-leg inverter operating under unbalanced load conditions; however, variations in the DC-link voltage are not considered, and the study focuses only on open-circuit faults, while short-circuit faults are not addressed. In our previous work [31], a KNN algorithm was employed for fault classification in a four-leg inverter with an ideal DC link and balanced load conditions. That study relied on a relatively limited set of operating points and was based solely on MATLAB/Simulink simulations, without hardware-in-the-loop experiments.

According to the literature review, the following points can be highlighted: Despite their growing adoption, limited research has addressed fault detection in the switches of four-leg inverters. In particular, to the best of the authors' knowledge, no existing studies consider the impact of voltage variations in the DC link. Furthermore, most existing studies are based on a limited number of operating points, whereas an effective fault diagnosis system must perform reliably across a broad range of conditions. Factors such as the timing of fault occurrence, instantaneous phase voltages, load conditions, and even wind speed can influence detection accuracy. To address this, the present study employs a comprehensive data extraction process using hardware-in-the-loop simulations, ensuring a diverse dataset. Furthermore, in this study, rather than more traditional classification machine learning algorithms such as KNN, the focus is on the transformer (TR) algorithm, which, with the advances of artificial intelligence, is showing superior performance in different applications. Additionally, rather than employing conventional machine learning methodologies, we implemented a two-level hierarchical fault detection structure to improve accuracy, rather than directly identifying the faulty switches. In summary, the main contributions of this study are as follows:

1. A WECS and a four-leg inverter, along with a standard control approach, are investigated to ensure the effect of the DC voltage link and neutral leg on fault detection performance.
2. The presence of both open-circuit and short-circuit switch faults has been studied for the implemented topology. A wide range of normal operating points (including balanced and unbalanced loads) have been considered.
3. A two-level hierarchical model is proposed based on the Transformer algorithm, which demonstrates improved performance compared to typical single-level structures and other state-of-the-art machine learning algorithms.
4. To verify the validity of the models, OPAL-RT has been used as the hardware-in-the-loop setup for testing the machine learning-based fault diagnosis system.

The rest of the paper is structured as follows: The model for the WECS and the four-leg inverter, as well as the control strategy, is presented in Section 2. The proposed fault detection strategy utilizing the transformer algorithm is introduced in Section 3. Results and discussions are presented in Section 4. Finally, the conclusion is outlined in Section 5.

2. System under study

The topology of the system under study is depicted in Fig. 1, which includes a PMSG, a battery, and a four-leg inverter that connects the DC bus to the load. The wind turbine is connected to the PMSG, and then it is integrated with an AC–DC converter. The battery is also attached to a DC–DC converter, and they share a common DC bus. The battery is charged when the available wind power exceeds the load demand, and it is discharged when the load demand is higher than the available wind power.

2.1. Permanent magnet synchronous generator

In this study, a WCES uses a wind turbine and a PMSG. In order to ensure optimal efficiency, the WCES continuously adjusts the generator speed to match changes in wind speed. The model of PMSG is similar to the model developed in [32].

2.2. Battery energy storage system

In all stand-alone structures, the storage of energy is an essential part of the overall design. In order to model the battery, a controlled voltage source connected in series to a constant resistance source can be used, which is discussed in more detail in [33].

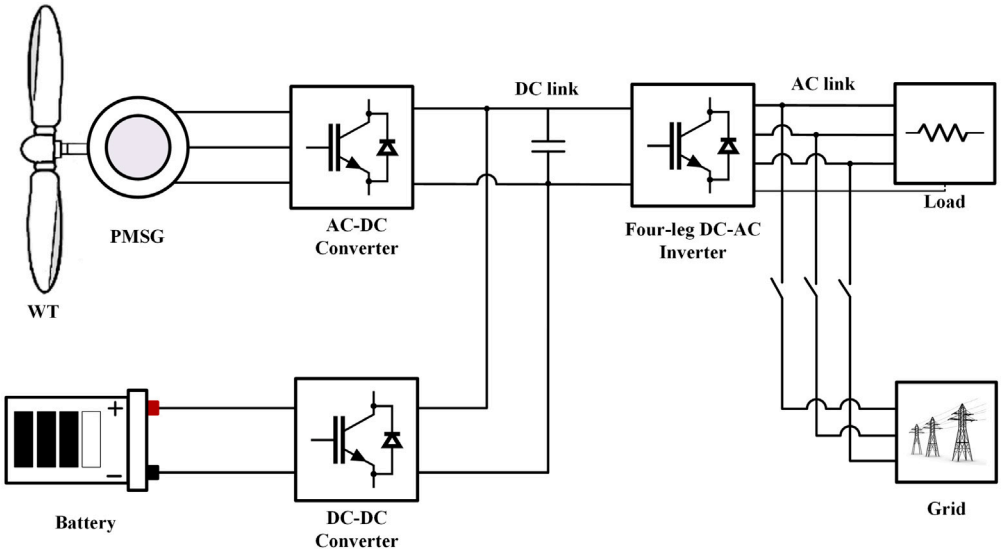

Fig. 1. Topology of the system under study.

2.3. Four-leg inverter

The four-leg inverter configuration is demonstrated in Fig. 2. In stand-alone applications, a low-pass filter follows the inverter to remove high-frequency switching voltages and currents. It is necessary to connect the fourth leg's midpoint to the load neutral point in order to achieve a zero component load current path [34]. The inverter input is powered by the DC bus, which is linked to the WECS and battery. When KVL is applied to phase legs, Eq. (1) can be obtained [35]:

$$\begin{bmatrix} V_{AN} \\ V_{BN} \\ V_{CN} \end{bmatrix} - \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = \begin{bmatrix} L_a & 0 & 0 \\ 0 & L_b & 0 \\ 0 & 0 & L_c \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_{La} \\ i_{Lb} \\ i_{Lc} \end{bmatrix} - L_n \frac{d}{dt} \begin{bmatrix} i_{Ln} \\ i_{Ln} \\ i_{Ln} \end{bmatrix} \tag{1}$$

In Eq. (1), V_{AN}, V_{BN}, and V_{CN} denote the inverter phase-to-neutral voltages, V_{an}, V_{bn}, and V_{cn} are the corresponding load voltages, L_a, L_b, and L_c are the phase filter inductances, L_n is the neutral-leg inductance, i_{La}, i_{Lb}, and i_{Lc} are the filter inductor currents of phases a, b, and c, and i_{Ln} is the neutral inductor current. The Eq. (2) can be determined by removing i_{Ln}:

$$\begin{bmatrix} V_{AN} \\ V_{BN} \\ V_{CN} \end{bmatrix} - \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = \begin{bmatrix} L_a + L_n & L_n & L_n \\ L_n & L_b + L_n & L_n \\ L_n & L_n & L_c + L_n \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_{La} \\ i_{Lb} \\ i_{Lc} \end{bmatrix} \tag{2}$$

For the four-leg inverter model to be complete, the KCL law must be satisfied at nodes a, b, and c in the abc frame. The following equation can be obtained, assuming the load voltage and the capacitor voltage are equal.

$$\begin{bmatrix} i_{La} \\ i_{Lb} \\ i_{Lc} \end{bmatrix} - \begin{bmatrix} i_{oa} \\ i_{ob} \\ i_{oc} \end{bmatrix} = \begin{bmatrix} C_a & 0 & 0 \\ 0 & C_b & 0 \\ 0 & 0 & C_c \end{bmatrix} \frac{d}{dt} \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} \tag{3}$$

Here, i_{oa}, i_{ob}, and i_{oc} denote the phase load currents, while C_a, C_b, and C_c are the filter capacitances for phases a, b, and c, respectively.

Eqs. (2) and (3) describe the inverter based on the abc reference frame incorporating the fourth leg inductor.

2.4. Converter controls

The DC link requires two control systems for the DC/DC bidirectional converter control of the battery and the AC/DC converter control of the PMSG. In this study, typical PI controllers have been used for the AC–DC converter of WECS and for the DC–DC bidirectional converter of the battery. The control systems are similar to those in [36]. Overall, the main goal of control systems in DC link is to maintain a constant DC voltage in the four-leg inverter. Furthermore, for the AC link, to ensure sinusoidal voltage output, a proportional resonant (PR) controller manages the four-leg inverter's output. Unlike PI controllers, PR uses a resonant term at a specific frequency (ω_0) to compensate for system resonances, leading to improved stability and dynamic response.

$$G_{PR}(s) = K_P + \frac{2K_R s}{s^2 + \omega_0^2} \tag{4}$$

In this transfer function, K_R is resonant gain, K_P is proportional gain, and ω_0 is the resonant frequency. It is noteworthy that the PR controller gain is infinite at the resonance frequency, and this affects the stability. Providing the damping term can effectively address this problem:

$$G_{PR}(s) = K_P + \frac{2K_R\omega_c s}{s^2 + 2\omega_c s + \omega_0^2} \tag{5}$$

where ω_c represents the cut-off frequency of the controller. In this equation, the proportional term K_P indicates system dynamics in terms of bandwidth, gain, and phase margins, whereas K_R specifies the resonant bandwidth near the frequency determined. Moreover, ω_c dampens the controller's sensitivity to changes in basic frequency during abnormal conditions. In practice, the controller implemented in this work corresponds to the damped (quasi-PR) structure in Eq. (5), which mitigates the potential instability of the ideal PR controller in Eq. (4) while introducing only a small steady-state error at the fundamental frequency that is acceptable for the considered operating conditions. In this study, a current and a voltage control loop for the PR approach are implemented, as shown in Fig. 3.

2.4.1. Current control loop

If the voltage drop in the inductance is controlled by a proportional controller, Eq. (2) could be written as follows:

$$\begin{bmatrix} V_{AN} \\ V_{BN} \\ V_{CN} \end{bmatrix} - \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = \begin{bmatrix} K_a & 0 & 0 \\ 0 & K_b & 0 \\ 0 & 0 & K_c \end{bmatrix} \begin{bmatrix} i^*_{La} - i_{La} \\ i^*_{Lb} - i_{Lb} \\ i^*_{Lc} - i_{Lc} \end{bmatrix} - K_n \begin{bmatrix} i^*_{Ln} - i_{Ln} \\ i^*_{Ln} - i_{Ln} \\ i^*_{Ln} - i_{Ln} \end{bmatrix} \tag{6}$$

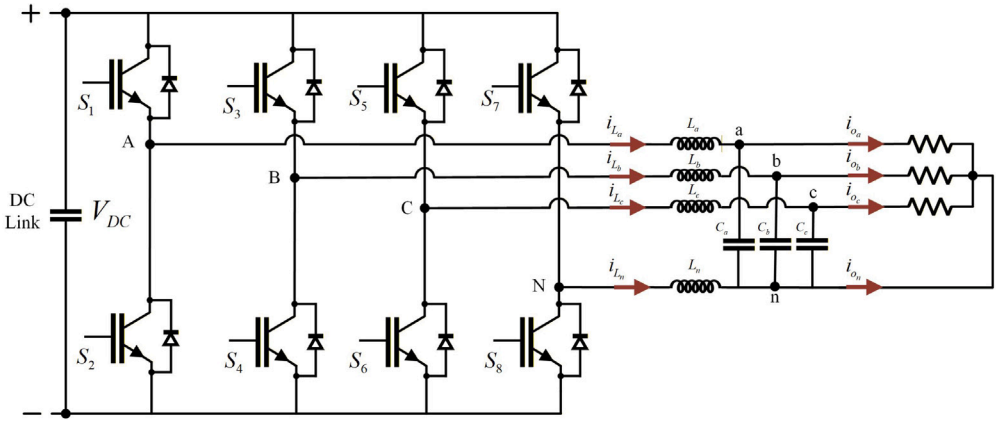

Fig. 2. Four-leg inverter structure.

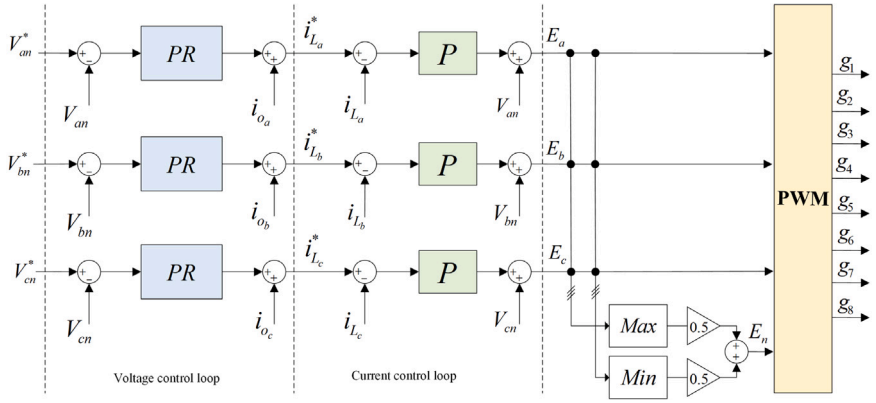

Fig. 3. Schematic diagram of PR control loops.

where K_a, K_b, K_c, and K_n are the proportional controller gains, and i^*_{La}, i^*_{Lb}, i^*_{Lc}, and i^*_{Ln} are the inverter output current references. Taking away i_{Ln} and i^*_{Ln} currents from this equation, the following equation can be derived:

$$\begin{bmatrix} V_{AN} \\ V_{BN} \\ V_{CN} \end{bmatrix} - \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = \begin{bmatrix} K_a + K_n & K_n & K_n \\ K_n & K_b + K_n & K_n \\ K_n & K_n & K_c + K_n \end{bmatrix} \begin{bmatrix} i^*_{La} - i_{La} \\ i^*_{Lb} - i_{Lb} \\ i^*_{Lc} - i_{Lc} \end{bmatrix} \tag{7}$$

2.4.2. Voltage control loop

In order to control the load voltage using the PR control, the following equation can be used:

$$\begin{bmatrix} C_a & 0 & 0 \\ 0 & C_b & 0 \\ 0 & 0 & C_c \end{bmatrix} \frac{d}{dt} \begin{bmatrix} V_{an} \\ V_{bn} \\ V_{cn} \end{bmatrix} = G_{PR}(s) \begin{bmatrix} V^*_{an} - V_{an} \\ V^*_{bn} - V_{bn} \\ V^*_{cn} - V_{cn} \end{bmatrix} \tag{8}$$

where V^*_{an}, V^*_{bn}, and V^*_{cn} are the inverter output voltage references. By substituting Eq. (8) in Eq. (3), the control loop is completed.

$$\begin{bmatrix} i^*_{La} \\ i^*_{Lb} \\ i^*_{Lc} \end{bmatrix} = G_{PR}(s) \begin{bmatrix} V^*_{an} - V_{an} \\ V^*_{bn} - V_{bn} \\ V^*_{cn} - V_{cn} \end{bmatrix} + \begin{bmatrix} i_{oa} \\ i_{ob} \\ i_{oc} \end{bmatrix} \tag{9}$$

Using *abc* reference, the proposed control system can independently regulate the voltage of each phase load [37].

3. Fault diagnosis system

This section discusses the three stages of implementing the fault diagnosis system as shown in Fig. 4: data extraction, data preprocessing, and training process.

3.1. Data extraction

The first step toward the implementation of a machine learning-based fault detection system is data extraction. This phase is extremely important, as it needs to cover a wide range of normal and fault operation points. Otherwise, it leads to a model with misleading accuracy, which, in practice, lacks generalization and cannot be used. In order to extract data, the four-leg inverter, WECS, and battery, along with all controllers, are developed in MATLAB/Simulink software. An experimental hardware-in-the-loop model is then built in order to extract data instead of acquiring the data directly from MATLAB/Simulink. The experimental setup includes the host PC, the OPAL-RT OP5700 real-time simulator (implemented using RT-LAB/Simulink code generation), and oscilloscopes. In the hardware-in-the-loop implementation, the model is partitioned into a Simulated Model (SM) subsystem that is compiled and executed on the OP5700 target in real time, and a Simulation Console (SC) subsystem that runs on the host PC for monitoring and data logging. The host target communication is handled through the RT-LAB interface. It should be noted that the analysis was conducted on a PC equipped with a 13th Gen Intel®Core™i7-1365U 1.80 GHz processor. The real-time simulation time step is set to $T_s = 0.0001$ s, corresponding to a sampling frequency of 10 kHz for the logged signals.

A wide range of operation points are integrated into the scenarios, considering different unbalanced loads and wind speeds. It is assumed

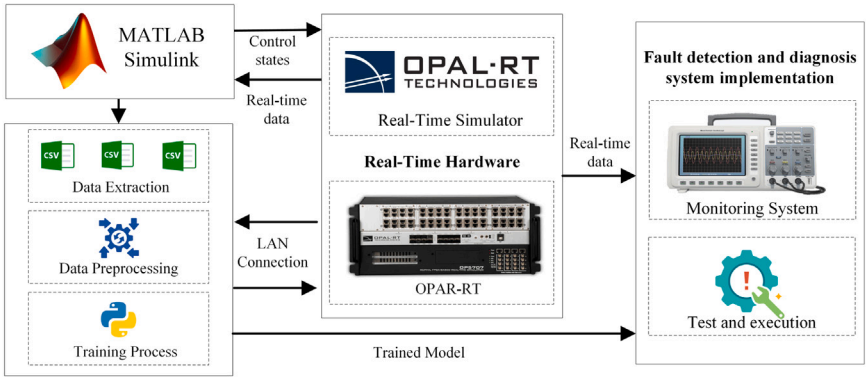

Fig. 4. Fault detection and diagnosis implementation.

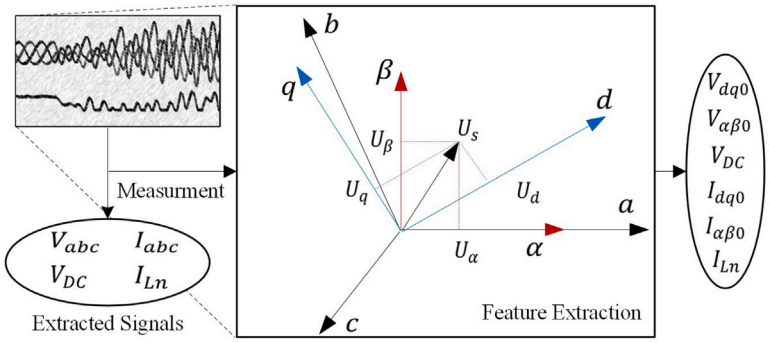

Fig. 5. Pseudo measurement.

that the unbalanced loads in all three phases, initially ranging from 10 Ω to 40 Ω, can vary between 20 Ω and 50 Ω, or become disconnected (no load). With a time step of 0.04 s, the load variations occur from $t = 0.08$ s to $t = 0.16$ s. The wind speed also varied from 10 m/s to 18 m/s between $t = 0.1$ s and $t = 0.18$ s, with the same time step.

For example, here is one normal operating condition: $R_a = 10\,\Omega$, $R_b = 20\,\Omega$, $R_c = 20\,\Omega$ to $R_a = 20\,\Omega$, $R_b = 35\,\Omega$, $R_c =$ no-load at $t = 0.12$ s, and wind speed from 12 m/s to 15 m/s at $t = 0.14$ s. It is assumed that various short-circuit and open-circuit fault types occur in all eight switches at different times (from 0.08 s to 0.16 s, with a time interval of 0.01 s), while the operating conditions are selected based on a normal random distribution of operating points. In total, approximately 4300 scenarios were generated by randomizing wind-speed profiles, load transitions, and fault type and occurrence time within the ranges described above. We assumed that, at a fault occurrence time, multiple simultaneous faults do not occur. At this stage, the inverter's current and voltage after the LC filter, the DC link's voltage, and the neutral current are stored. To augment the information content within the data and capture its full dynamic range and relevant features, we utilize pseudo-measurements, as shown in Fig. 5. These pseudo-measurements are constructed by applying the dq and the $\alpha\beta$ transformations to the original signals.

3.2. Data preprocessing

The preprocessing stage consists of two main steps: transforming the raw time series into sequence-to-label format and applying quantile normalization to standardize the features.

3.2.1. Sequence-to-label representation

Given the temporal nature of faults, we employ a sequence-to-label representation. For each case in our dataset, we create overlapping sequences. Formally, for a multivariate time series $\mathbf{X} = \{x_1, x_2, \ldots, x_T\}$ where each $x_t \in \mathbb{R}^d$ represents a d-dimensional measurement at time t, we construct sequences:

$$S_i = \{x_i, x_{i+1}, \ldots, x_{i+\tau-1}\} \quad \text{for} \quad i \in \{1, 2, \ldots, T-\tau+1\} \tag{10}$$

Each sequence S_i is paired with the corresponding label $y_{i+\tau-1}$, representing the system state (normal or specific fault type) at the end of the sequence. To address class imbalance, we implement a balanced sampling strategy where we define:

$$n_c = \min_{c \in C} \{|\{S_i \mid y_{i+\tau-1} = c\}|\} \tag{11}$$

where C is the set of all classes, and we sample exactly n_c sequences from each class to ensure balanced training. Fig. 6 shows how sequences and associated labels are handled. Fig. 7 illustrates the sinusoidal voltage measurements, the neutral current, and how sequences are labeled when a short-circuit fault occurs at Switch 7, as an example. Similar measurements are obtained for current signals and DC link voltage. Data samples are collected every 1 ms, creating multiple sequences. For example, sequences i and j (separated by 10 samples) both belong to the normal class before fault occurrence. When examining sequence k, which spans the moment of fault initiation, the entire sequence is classified as a short-circuit fault even though some earlier samples within that sequence were recorded during normal operation. By sequence n, all data samples within the sequence are from the post-fault condition, clearly exhibiting characteristic fault signatures.

In the implementation, we enforce a balanced sampling scheme at both levels of the hierarchy. At Level 1, we construct three equally sized

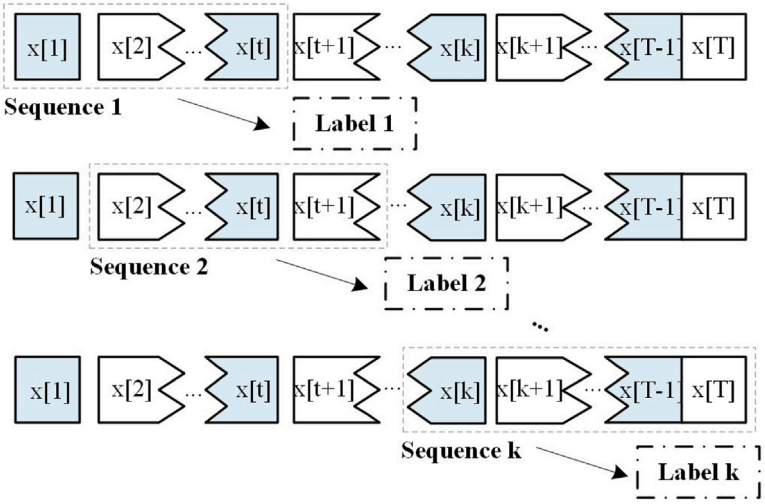

Fig. 6. Sequence-to-label representation.

classes (normal operation, short-circuit fault, and open-circuit fault) with $N_{\text{L1}} = 50{,}000$ sequences per class, resulting in $150,000$ sequences in total. At Level 2, we further balance the eight short-circuit fault classes and the eight open-circuit fault classes independently, assigning $N_{\text{fault}} = 6250$ sequences to each individual fault type. This yields $50,000$ sequences for the Level 2 short-circuit model and $50,000$ sequences for the Level 2 open-circuit model, each with a uniform distribution across the eight switches. Based on the duration and sampling of the original recordings across all scenarios, more than 10^6 overlapping sequences could be generated; however, we deliberately subsample this larger pool to the balanced set of $150,000$ sequences used in our experiments in order to control memory usage and training time while preserving representativeness of all fault classes.

3.2.2. Quantile normalization

Quantile normalization offers advantages for fault detection applications due to its robustness to outliers and effectiveness with non-Gaussian distributions often encountered in fault data. Based on the scikit-learn documentation, the quantile transformation process can be summarized as follows [38]:

Step 1: For each feature X, estimate its empirical cumulative distribution function (ECDF):

$$\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x) \tag{12}$$

Step 2: For each observed value, compute the estimated quantile:

$$q_i = \frac{\text{rank}(x_i) - 0.5}{n} \tag{13}$$

Step 3: Transform each quantile to the corresponding value in the target distribution:

$$y_i = \Phi^{-1}(q_i) \tag{14}$$

where Φ^{-1} is the inverse CDF of the standard normal distribution. When applying this to multivariate time series data, we transform each feature independently:

$$\tilde{X}_{j,t} = \Phi^{-1}(\hat{F}_{X_j}(X_{j,t})) \tag{15}$$

Remark. Applying quantile normalization compared to regular Min–Max normalization yields improvement across different models. Consequently, it is adopted as the normalization approach in our methodology.

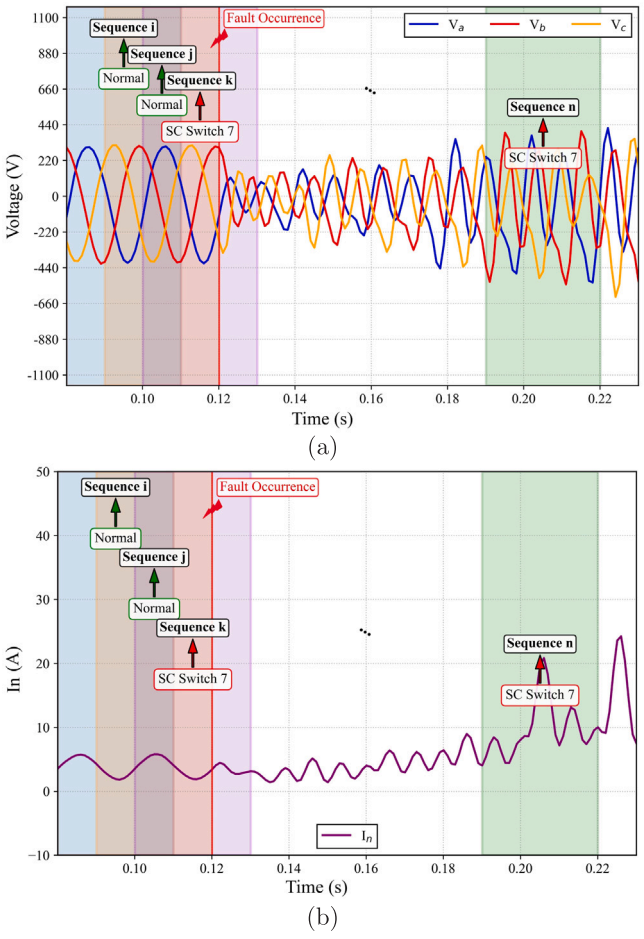

Fig. 7. Sequence-to-label data classification: (a) Three-phase output voltages (b) Neutral current.

3.3. Proposed 2-level hierarchical transformer-based model

To effectively classify faults in four-leg inverters, we propose a hierarchical approach consisting of two classification levels, as shown in Fig. 8. The first level determines whether the system is operating

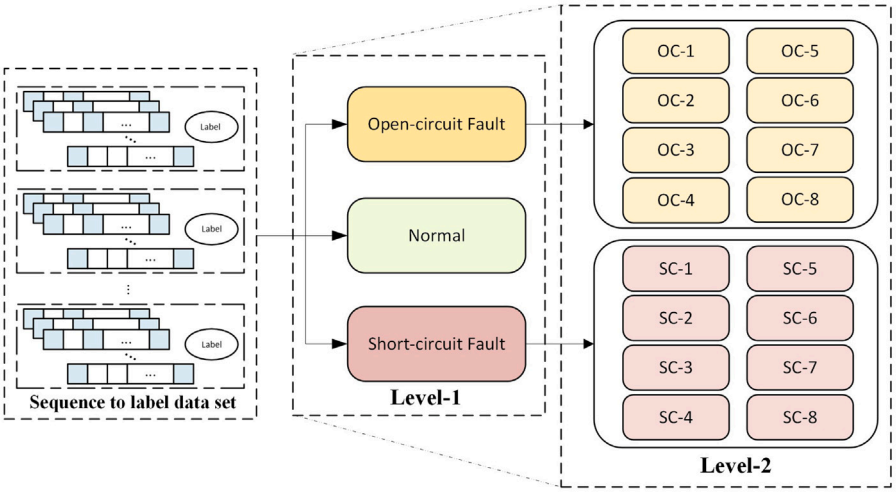

Fig. 8. Two-level hierarchical fault diagnosis.

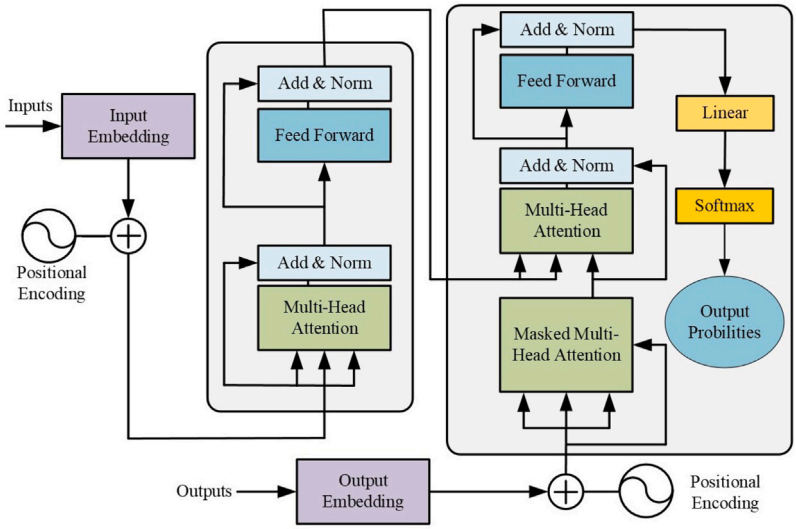

Fig. 9. Schematic diagram of Transformer model.

normally or experiencing a fault, and if a fault is detected, identifies its general category (short-circuit or open-circuit). The second level precisely identifies which of the eight switches has failed. This hierarchical structure simplifies the classification problem at each stage and improves overall accuracy. Each model in our hierarchical structure is implemented using a customized Transformer architecture optimized for time-series classification.

The Transformer algorithm is a technique for processing sequential data without the use of convolutional or recurrent layers. The model learns how elements of a sequence depend on each other using a mechanism called attention. The model has a sequence-to-sequence structure composed of an encoder and a decoder. An encoder replaces input tokens with context vectors based on their sequence of input tokens. A decoder receives a sequence of output tokens and creates a sequence of vectors. Using the output of the decoder, the final output is produced [39]. Fig. 9 shows the architecture of a baseline transformer model. The core block here is a transformer block, which contains a multi-head self-attention layer and then a feed-forward neural network.

To provide sequence order information, we applied sinusoidal positional encoding to the inputs before the attention mechanism, since Transformers do not inherently capture temporal context. The encoding is defined as

$$PE_{2i}(p) = \sin\left(\frac{p}{10000^{2i/d_{pos}}}\right), \tag{16}$$

$$PE_{2i+1}(p) = \cos\left(\frac{p}{10000^{2i/d_{pos}}}\right), \tag{17}$$

where p denotes the position in the sequence, d_{pos} is the embedding dimension, and $i \in \left\{0, 1, \ldots, \frac{d_{pos}}{2} - 1\right\}$. This encoding allows the model to distinguish between different time steps while maintaining a fixed dimensionality. After applying positional encoding, the attention mechanism is employed. Each transformer block can be formalized as:

$$\begin{aligned} \text{Attention}(Q, K, V) &= \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \\ \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O, \\ \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V). \end{aligned} \tag{18}$$

where Q, K, and V correspond to the queries, keys, and values matrices, respectively, while W_i^Q, W_i^K, W_i^V, and W^O are learnable parameter matrices. The division by $\sqrt{d_k}$ serves as a scaling factor to stabilize training, where d_k denotes the key vectors' dimension. Following each attention layer, we employ a two-layer feed-forward

Table 1
Transformer architecture and training hyperparameters used in the proposed fault diagnosis models.

Component	Configuration
Number of Transformer blocks	2
Embedding dimension	28 (equal to the input feature dimension)
Number of attention heads	2
Feed-forward dimension	128
Dropout rate	0.2 (applied to attention, feed-forward, and dense layers)
Pooling layer	Global average pooling over the temporal dimension
Feature dense layer	128 neurons, ReLU activation
Output layer (Level 1)	3 neurons, softmax (normal/short-circuit/open-circuit)
Output layer (Level 2, short-circuit)	8 neurons, softmax (switches S_1–S_8)
Output layer (Level 2, open-circuit)	8 neurons, softmax (switches S_1–S_8)
Optimizer	Adam, learning rate 10^{-3}
Loss function	Sparse categorical cross-entropy
Batch size	64
Number of epochs	25

network with a ReLU activation function:

$$\mathrm{FFN}(x) = \max(xW_1 + b_1, 0)W_2 + b_2 \tag{19}$$

Both the attention output and feed-forward network output are processed through layer normalization and residual connections to enhance training stability and convergence. Additionally, dropout with a rate of 0.2 is applied to prevent overfitting. To extract meaningful features from the sequence, a Global Average Pooling operation aggregates information across the temporal dimension, producing a fixed-size representation that is then passed through a dense layer with 128 neurons before the final classification layer. In this work, all three classifiers in the hierarchical structure (Level 1, Level 2 short-circuit, and Level 2 open-circuit) share the same Transformer backbone, and their main architectural and training hyperparameters are summarized in Table 1.

3.4. Performance evaluation metrics

To evaluate the proposed fault diagnosis system, we employ standard classification metrics as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{22}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{23}$$

where TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative predictions, respectively [40]. For the two-level hierarchical structure, the overall accuracy is defined as the proportion of samples correctly classified at both hierarchical levels. Furthermore, to visualize the high-dimensional feature spaces learned by our models, we employ t-distributed Stochastic Neighbor Embedding (t-SNE), a dimensionality reduction technique particularly effective for visualizing complex datasets. The t-SNE algorithm minimizes the Kullback–Leibler divergence between two distributions: a t-distribution representing pairwise similarities in the low-dimensional embedding space and a Gaussian distribution representing pairwise similarities in the high-dimensional space. In the high-dimensional space, the similarity between points x_i and x_j is determined by [41]:

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)} \tag{24}$$

where σ_i denotes the variance of the Gaussian centered at point x_i. The similarity between the corresponding points y_i and y_j in the low-dimensional space is modeled using a t-distribution:

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}} \tag{25}$$

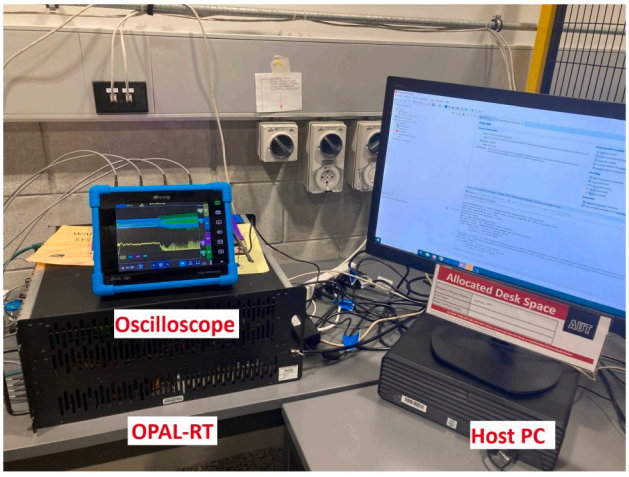

Fig. 10. Hardware-in-the-loop setup.

The optimal low-dimensional representation is obtained by minimizing the Kullback–Leibler divergence between these distributions:

$$\mathrm{KL}(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{26}$$

where p_{ij} is the symmetrized high-dimensional similarity ($p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$). In our implementation, we reduce the high-dimensional feature space to three dimensions (labeled as Dimension 1, Dimension 2, and Dimension 3) using t-SNE with a perplexity value of 30 in scikit-learn.

4. Results and discussion

The system under study is implemented in Simulink and OPAL-RT to extract data. The hardware-in-the-loop setup is shown in Fig. 10. Several operation points are considered, as explained in the Data Extraction section. Fig. 11 shows different current and voltage features when there is a short-circuit fault on switches S_1 and an open-circuit fault on switches S_2 as two examples.

4.1. Level-1 performance

The first level of the proposed hierarchical fault diagnosis system classifies the operating condition into three categories: normal operation, short-circuit fault, and open-circuit fault. Fig. 12 illustrates the effectiveness of the transformer algorithm in creating a separable embedding space, where normal operation (blue), short-circuit faults (orange), and open-circuit faults (green) form distinct clusters that can be easily distinguished by the classifier. The clear separation between

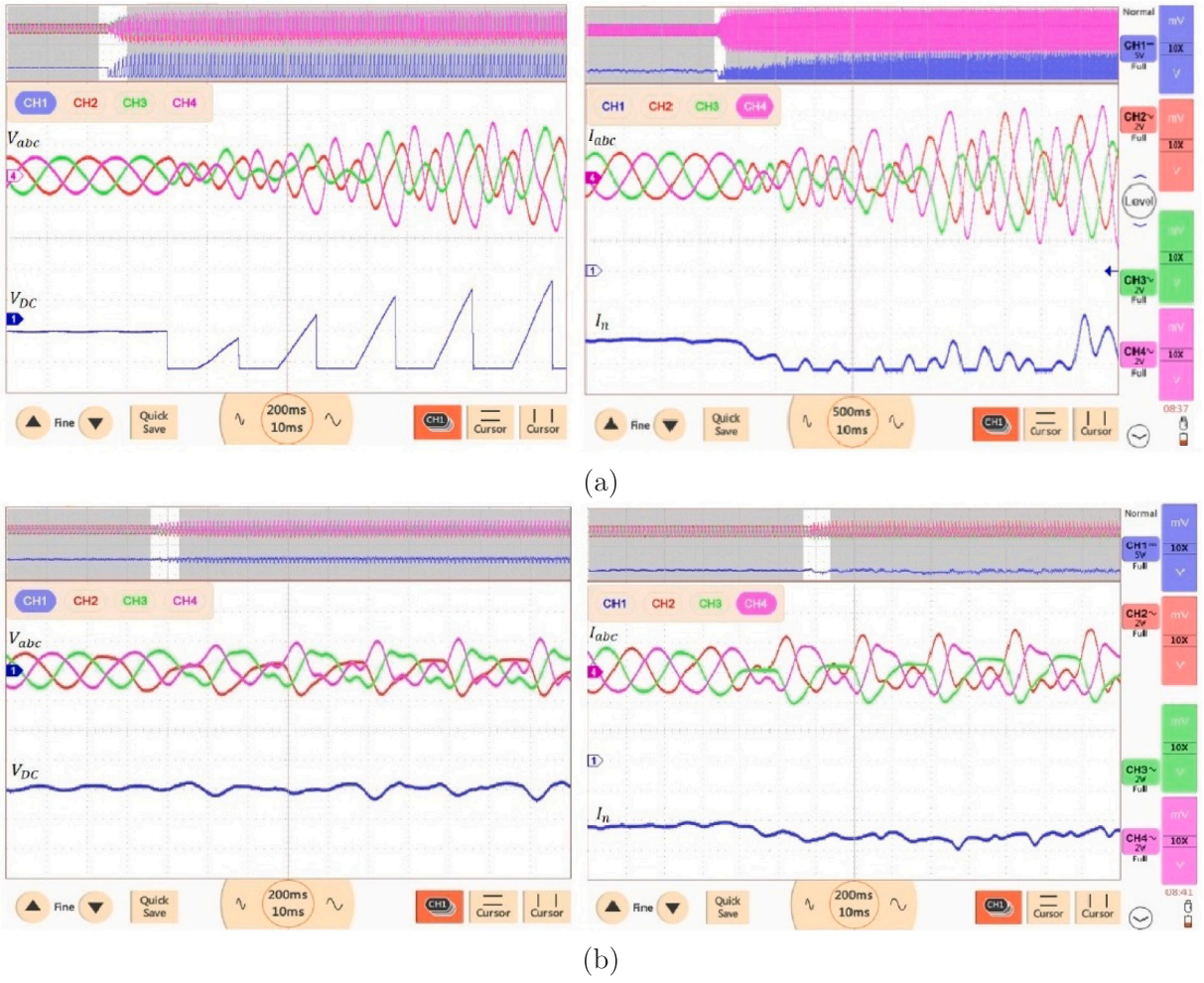

Fig. 11. Features under different fault conditions: (a) Short-circuit fault on Switch S_1, (b) Open-circuit fault on Switch S_2.

Table 2
Level-1 Transformer Model Performance.

Actual	Predicted			Recall (%)	F1 (%)
	N.O	S-C.F	O-C.F		
N.O	4997	0	3	99.94	99.68
S-C.F	11	4988	1	99.76	99.88
O-C.F	18	0	4982	99.64	99.78
Precision (%)	99.42	100.00	99.92	**Accuracy = 99.78%**	

N.O = Normal Operation; S-C.F = Short-Circuit Fault; O-C.F = Open-Circuit Fault.

these clusters demonstrates the model's ability to extract meaningful features from the data, enabling highly accurate fault classification. Table 2 presents the comprehensive performance metrics of the Level-1 Transformer model on the test dataset. The Level-1 model achieves a classification accuracy of 99.78% on the test dataset. The confusion matrix reveals that the model rarely misclassifies faults. Notably, the model achieves perfect precision (100%) in identifying short-circuit faults, indicating no false positives for this critical fault type. The high F1-scores across all three classes (99.68% for normal operation, 99.88% for short-circuit faults, and 99.78% for open-circuit faults) demonstrate the balanced performance of the model in terms of both precision and recall.

4.2. Level-2 short-circuit performance

After a fault is classified as a short-circuit, the Level-2 model identifies which specific switch has failed. Fig. 13 illustrates how the Transformer algorithm effectively separates the embedding space for different short-circuit fault locations. Table 3 presents the detailed performance metrics for the Level-2 short-circuit fault localization model. The model achieves an overall accuracy of 99.58% on the test dataset. The confusion matrix reveals reliable performance with very few misclassifications across all switches. The consistently high F1-scores across all switches (ranging from 99.44% to 99.68%) demonstrate the model's balanced performance in terms of both precision and recall.

4.3. Level-2 open-circuit performance

Following the identification of an open-circuit fault by the Level-1 model, the Level-2 open-circuit model precisely determines which of the eight switches has failed. Fig. 14 illustrates the Transformer algorithm's embedding space for open-circuit fault classification through t-SNE visualization. The visualization reveals well-separated clusters corresponding to faults in different switches. Table 4 presents the comprehensive performance metrics of the Level-2 open-circuit fault localization model. The model achieves an overall accuracy of 99.90%

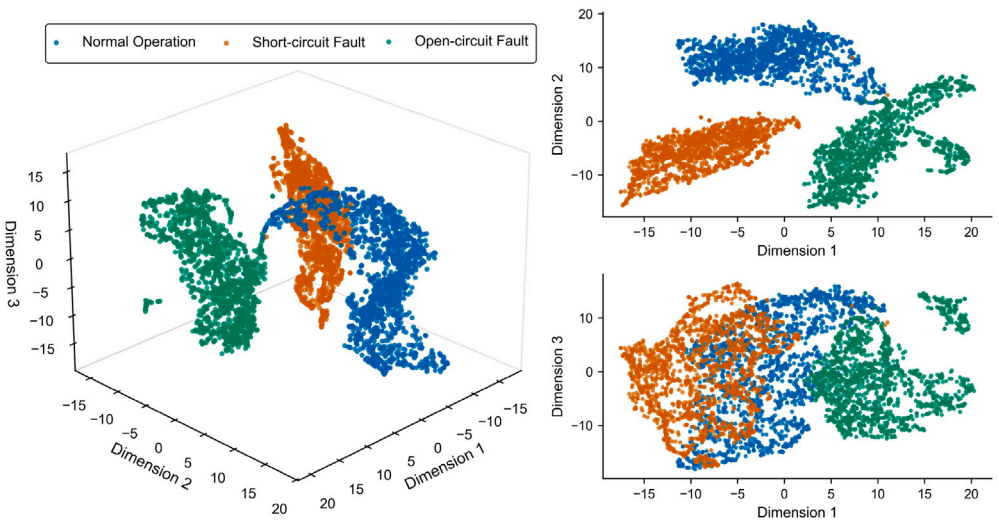

Fig. 12. Embedding space for Level-1.

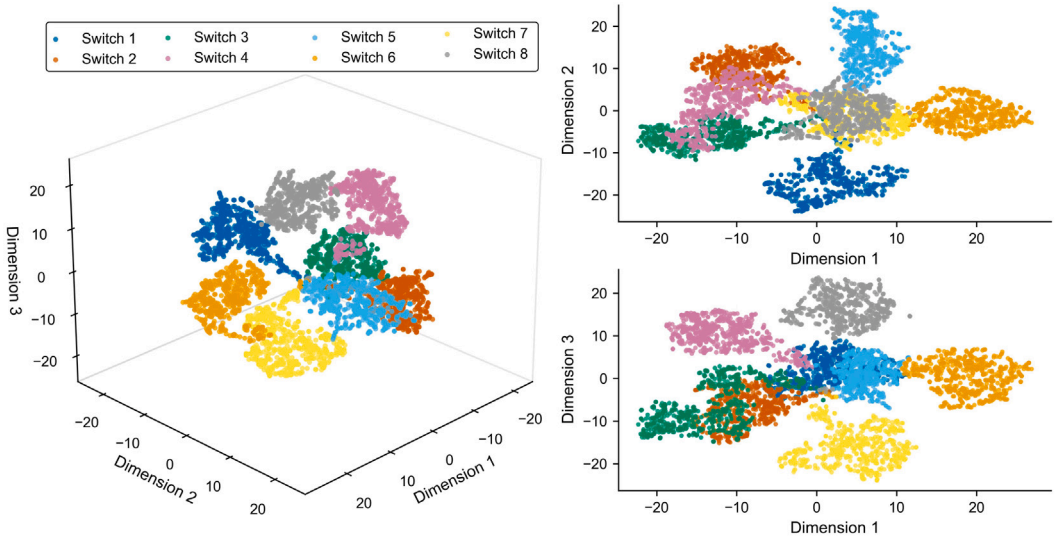

Fig. 13. Embedding space for level-2 short-circuit.

Table 3
Level-2 Transformer Model Performance for Short-circuit Fault Localization.

Actual Switch	Predicted Switch								Recall (%)	F1 (%)
	1	2	3	4	5	6	7	8		
1	620	0	0	0	0	3	0	2	99.20	99.52
2	0	620	0	5	0	0	0	0	99.20	99.44
3	1	1	621	0	0	0	2	0	99.36	99.68
4	0	0	0	625	0	0	0	0	100.00	99.52
5	0	0	0	1	622	2	0	0	99.52	99.60
6	0	0	0	0	0	625	0	0	100.00	99.52
7	0	1	0	0	0	1	623	0	99.68	99.68
8	0	0	0	0	2	0	0	623	99.68	99.68
Precision (%)	99.84	99.68	100.00	99.05	99.68	99.05	99.68	99.68	**Accuracy = 99.58%**	

on the test dataset, slightly outperforming even the short-circuit fault classification results. The confusion matrix reveals near-perfect classification across most switches, with several switches (1, 2, 4, 5, 7, and 8) achieving perfect recall rates of 100%, indicating that the model never misses faults in these components. The precision values are equally promising, with multiple switches (2, 3, 4, 5, and 6) achieving perfect precision scores (100%), indicating zero false positives for these components. The high F1-scores across all switches (ranging from 99.76% to 100.00%) demonstrate the model's balanced and consistent performance across all switches.

4.4. Performance comparison of different algorithms

In this section, a performance analysis of various machine learning algorithms for fault diagnosis is presented, including KNN (k-nearest neighbor algorithm), RF (random forest), MLP (multi-layer perceptron),

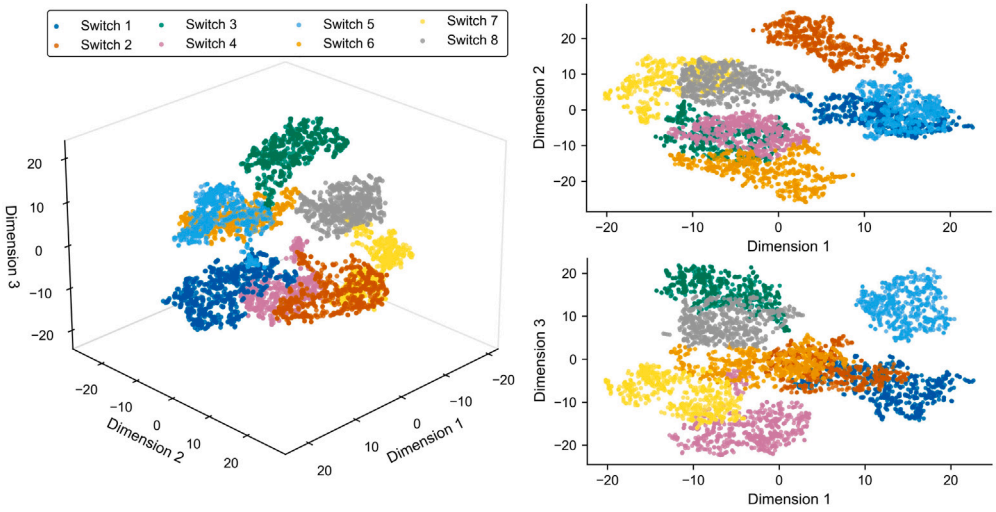

Fig. 14. Embedding space for level-2 open-circuit.

Table 4
Level-2 Transformer Model Performance for Open-circuit Fault Localization.

Actual Switch	Predicted Switch								Recall (%)	F1 (%)
	1	2	3	4	5	6	7	8		
1	625	0	0	0	0	0	0	0	100.00	99.84
2	0	625	0	0	0	0	0	0	100.00	100.00
3	0	0	623	0	0	0	2	0	99.68	99.84
4	0	0	0	625	0	0	0	0	100.00	100.00
5	0	0	0	0	625	0	0	0	100.00	100.00
6	2	0	0	0	0	622	0	1	99.52	99.76
7	0	0	0	0	0	0	625	0	100.00	99.84
8	0	0	0	0	0	0	0	625	100.00	99.92
Precision (%)	99.68	100.00	100.00	100.00	100.00	100.00	99.68	99.84	**Accuracy = 99.90%**	

CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), and TR (Transformer), with their associated sequence-to-label (SL) structure. As shown in Fig. 15 and Table 5, SL variants consistently outperform their conventional counterparts across all metrics, while requiring longer training times. It should be noted that the single-level KNN algorithm is conceptually similar to the approach developed in [31], but in this work it is evaluated under more demanding scenarios with unbalanced load conditions. To provide an indicative measure of model complexity, we also report the approximate model sizes used in this study, excluding KNN, which is non-parametric. The RF models are characterized by their tree structure and contain approximately 4×10^5 total leaf nodes at Level 1 and about 2×10^5 total leaf nodes in each Level 2 model. The MLP models have approximately 3×10^5 trainable parameters per level, the CNN models approximately 3.2×10^4 trainable parameters per level, the LSTM models approximately 1.6×10^5 trainable parameters per level, and the TR models approximately 3.2×10^4 trainable parameters per level. The reported parameter counts are obtained directly from the implemented TensorFlow/Keras models by summing all trainable weights and biases. More extensive architectures with a larger number of trainable parameters are considered for the MLP and LSTM models (to examine whether increased model capacity impacts performance), whereas the CNN and TR-based models are designed with approximately similar parameter counts.

The SL-TR model achieves the highest single-level accuracy (99.02%) and two-level accuracy (99.53%), followed closely by SL-LSTM (98.88% and 99.47%) and SL-CNN (98.74% and 98.63%). Comparing LSTM and TR models, the baseline LSTM demonstrates faster training and inference times than baseline TR; however, the strong accuracy of SL-LSTM is achieved with a larger parameter count than the TR models. Notably, increasing the MLP capacity does not yield a similar gain, indicating that explicit temporal modeling, rather than simply adding parameters, is the key driver of performance improvement. The CNN models provide a competitive accuracy level while maintaining relatively low inference times (0.08–0.11 ms across the three levels); compared to TR, CNN achieves lower accuracy but offers faster inference. The hierarchical approach affects algorithms differently: for models with limited sequential pattern recognition (KNN, MLP), the two-level approach yields reduced performance, whereas sequence-based deep models (CNN/LSTM/TR) maintain high accuracy, with LSTM and TR benefiting most and SL-CNN showing a small reduction in two-level accuracy relative to its single-level result. From a computational perspective, Table 5 indicates that the SL variants consistently improve accuracy at the cost of higher training time, while inference remains in the sub-millisecond range for all deep models and is particularly small for tree-based models (RF) due to their lightweight decision structure. In summary, SL-TR offers the best overall accuracy, SL-CNN provides a strong accuracy–latency trade-off, and SL-LSTM remains highly accurate with higher computational cost.

4.5. Discussion

A comparative analysis between the proposed approach and existing fault diagnosis methods is presented in Table 6. It should be noted that this comparison does not include all existing studies on fault detection systems in three-leg inverters, as the literature in this area is extensive. Rather, the focus is primarily on recent and relevant approaches, with particular emphasis on four-leg inverters. Most existing studies, including Diao et al. [13], Xu et al. [15], Singh et al. [22], Fezai et al. [23], Feng et al. [24], and Si et al. [25], considered three-leg inverters. Regarding DC link configurations, most research, including the works by Diao et al. [13], Xu et al. [15], Feng et al. [24], Si et al. [25], Xu et al. [30], and Peykarporsan et al. [31], employed ideal

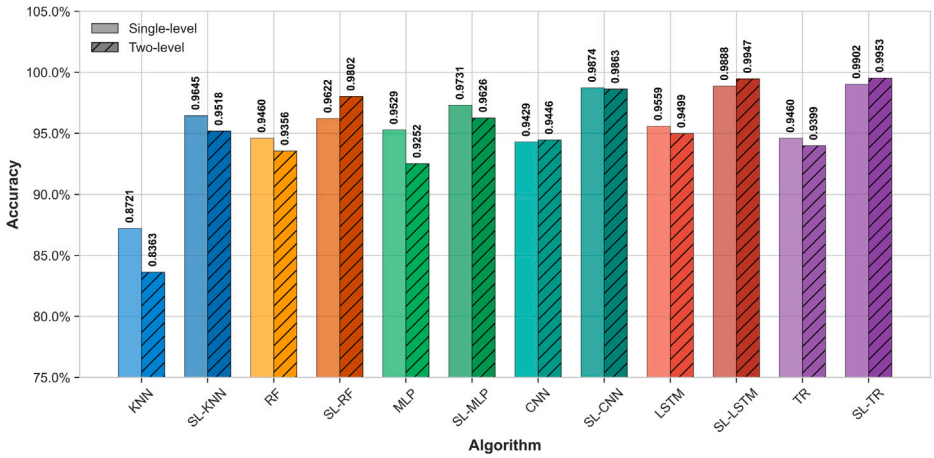

Fig. 15. Accuracy comparison of different algorithms.

Table 5
Performance comparison of different algorithms and their sequence-to-label variants.

Model	Single-level Acc.	Level-1			Level-2 S - C.F			Level-2 O - C.F			Two-level Acc.
		Acc.	F1	TT/Inf	Acc.	F1	TT/Inf	Acc.	F1	TT/Inf	
KNN	0.87213	0.9283	0.9277	0.18/0.12	0.8870	0.8873	0.11/0.06	0.9136	0.9135	0.11/0.08	0.8363
SL-KNN	0.96453	0.9716	0.9718	7.70/0.23	0.9812	0.9812	3.47/0.07	0.9886	0.9886	3.15/0.09	0.9518
RF	0.94600	0.9765	0.9765	23.38/0.0038	0.9514	0.9514	4.95/0.0085	0.9650	0.9651	4.97/0.0087	0.9356
SL-RF	0.96220	0.9822	0.9823	41.05/0.0090	0.9974	0.9974	7.91/0.0150	0.9986	0.9986	9.52/0.0178	0.9802
MLP	0.95287	0.9847	0.9847	137/0.04	0.9370	0.9372	43.2/0.06	0.9700	0.9700	44.7/0.07	0.9252
SL-MLP	0.97307	0.9855	0.9855	207/0.05	0.9780	0.9780	90.4/0.07	0.9814	0.9816	81.5/0.07	0.9626
CNN	0.9429	0.9873	0.9873	225/0.08	0.9406	0.9406	69/0.07	0.9730	0.9730	64/0.06	0.9446
SL-CNN	0.9874	0.9934	0.9934	430/0.08	0.9902	0.9901	143/0.10	0.9956	0.9956	159/0.11	0.9863
LSTM	0.95587	0.9837	0.9837	154/0.09	0.9462	0.9463	53.7/0.24	0.9732	0.9732	54.0/0.19	0.9499
SL-LSTM	0.98880	0.9976	0.9976	1545/0.31	0.9946	0.9946	442/0.42	0.9988	0.9988	411/0.41	0.9947
TR	0.94600	0.9860	0.9860	317/0.10	0.9420	0.9422	109/0.16	0.9674	0.9673	113/0.13	0.9399
SL-TR	0.99020	0.9978	0.9978	1023/0.28	0.9958	0.9958	374/0.32	0.9990	0.9990	386/0.33	0.9953

TT = Training time (s); Inf = Inference time (ms); Acc. = Accuracy; S-C.F = Short-Circuit Fault; O-C.F = Open-Circuit Fault.

DC supplies, which simplifies the fault diagnosis problem but reduces practical relevance. However, Singh et al. [22] and Fezai et al. [23] incorporated renewable energy sources in their DC link configurations. Regarding load conditions, Xu et al. [30] and the present work consider unbalanced loads, which represent a significant challenge in real-world stand-alone applications, while many other studies assume balanced loading. Concerning fault types, most studies primarily focus on open-circuit faults, while short-circuit faults—equally critical in practice—are addressed in works by Fezai et al. [23], Feng et al. [24], Peykarporsan et al. [31], and our proposed approach.

A closer comparison with the most relevant four-leg inverter studies, namely Xu et al. [30] and our previous work in Peykarporsan et al. [31], further clarifies the novelty of the present contribution. Xu et al. [30] propose a sliding-mode-observer-based diagnosis scheme for open-circuit faults in a four-leg inverter under unbalanced load conditions, but they employ an ideal DC source and do not consider DC-link voltage variations or short-circuit faults. In Ref. [31], a KNN-based classifier (similar to single-level KNN algorithm in Table 5) is used for four-leg inverter fault detection with an ideal DC supply and balanced load conditions, based solely on MATLAB/Simulink simulations and a limited set of operating points. In contrast, the present work considers a stand-alone WECS with a combined wind-turbine–battery DC link, explicitly exposing the fault diagnosis system to DC-link voltage variations, and covers both open-circuit and short-circuit faults across a wide range of balanced and unbalanced operating points. From a data extraction and evaluation perspective, the proposed method builds an extensive offline dataset from real-time hardware-in-the-loop fault scenarios and then trains and assesses the classifiers on this common dataset. By comparison, Xu et al. [30] focuses on online fault detection schemes, without systematically reporting performance across a similarly broad range of operating points. It should be noted that, to the best of the authors' knowledge, a hierarchical fault diagnosis system has not been investigated in the literature, despite our analysis showing that it can enhance overall accuracy.

From an algorithmic perspective, Table 5 shows that the sequence-to-label Transformer (SL-TR) achieves the highest single-level accuracy (99.02%), while its two-level hierarchical counterpart reaches an overall accuracy of 99.53%. The improvement is mainly due to decomposing the task into a high-level, three-class decision at Level 1 (normal/short-circuit/open-circuit), followed by specialized Level 2 models that focus solely on localizing the faulty switch within a given fault family. This decomposition reduces class confusion near decision boundaries and allows the Level 2 Transformers to learn more fine-grained patterns without having to simultaneously separate normal operation from all fault types. At the same time, the results indicate that sequence-to-label modeling is itself a dominant factor in achieving high accuracy. For temporal sequential models, the hierarchical structure primarily enhances the accuracy, while non-sequential classifiers such as KNN and MLP do not consistently benefit from hierarchy and, in some cases, exhibit slightly lower two-level accuracy than their single-level counterparts. Overall, these results suggest that the hierarchical design is most effective when combined with sliding-label, temporal feature learning, as realized in the SL-LSTM and SL-TR architectures.

The confusion matrices for Level 1 and Level 2 do not reveal any specific operating point (in terms of wind speed, load unbalance, or fault timing) where the proposed hierarchical Transformer-based

Table 6
Comparison with related works.

Ref.	Inverter	Technique	DC Link	Load	S-C.F	O-C.F	H.S
[13]	Three-leg	Model Flux Observer	DC Supply	Balanced	✗	✓	✗
[15]	Three-leg	Sliding Mode Observer	DC Supply	Balanced	✗	✓	✗
[22]	Three-leg	Random Forrest	PV	Balanced	✗	✓	✗
[23]	Three-leg	Random Forrest	WTG	Balanced	✓	✓	✗
[24]	Three-leg	Long Short-Term Memory	DC Supply	Balanced	✓	✓	✗
[25]	Three-leg	Long Short-Term Memory	DC Supply	Balanced	✗	✓	✗
[30]	Four-leg	Sliding Mode Observer	DC Supply	Balanced/Unbalanced	✗	✓	✗
[31]	Four-leg	K-nearest Neighbor	DC Supply	Balanced	✓	✓	✗
Proposed	Four-leg	Sequence-to-label Transformer	WTG+Battery	Balanced/Unbalanced	✓	✓	✓

S-C.F = Short-Circuit Fault; O-C.F = Open-Circuit Fault; H.S = Hierarchical Structure.

scheme systematically struggles; the few misclassifications are rare and not concentrated in a particular region of the operating space. Furthermore, regarding robustness to noise and sensor errors, the present setup emulates ideal measurement channels, so no explicit measurement noise or sensor bias has been injected into the data. As a consequence, the reported performance metrics primarily reflect the classification capability under numerically accurate signals rather than under realistic sensor distortions. Nonetheless, the combination of quantile-based normalization and sequence-to-label modeling provides some inherent robustness to moderate amplitude variations and outliers.

In terms of generalization to unseen operating profiles, the training, validation, and test sets are constructed from disjoint cases, each corresponding to different combinations of wind speeds, load configurations (including balanced and unbalanced conditions), and fault occurrence times. A large number of operating points are initially generated during the data extraction stage; however, to ensure balanced learning across classes, a balanced downsampling strategy is applied. This approach has two important implications. First, the retained dataset still spans a wide and diverse range of operating conditions, allowing the model to learn representative temporal patterns and achieve high generalization performance on the held-out test data. Second, because downsampling is performed across operating points, many additional operating scenarios present in the original dataset are intentionally excluded from the final training and testing sets. Despite this reduction, the consistently high test accuracies indicate that the proposed hierarchical Transformer-based scheme generalizes well within the considered operating conditions. Nevertheless, generalization to operating points that fall entirely outside these conditions, such as significantly different wind profiles, load dynamics, or converter parameter sets, has not been explicitly validated and would require further investigation.

5. Conclusion

In this study, a hierarchical, data-driven approach based on a transformer algorithm is proposed for diagnosing faults in a four-leg inverter connected to a stand-alone wind energy system. Initially, a diverse range of operating points is considered to ensure that the developed model is generalizable and capable of detecting both short-circuit and open-circuit faults under various scenarios. The datasets are obtained using a hardware-in-the-loop OPAL-RT simulator to closely mimic real-world conditions. The results demonstrate that the proposed transformer-based model outperforms other machine learning algorithms. Moreover, the two-level hierarchical structure is shown to be more effective than a single-level approach. For future work, potential extensions include the systematic study of measurement noise and sensor inaccuracies, the consideration of multiple simultaneous and cascading switch faults, the implementation of the proposed hierarchical Transformer-based diagnosis scheme on FPGA or similar real-time embedded platforms, and the investigation of additional fault types such as wear-out or degradation-related faults.

CRediT authorship contribution statement

Jalal Heidari: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Rasool Peykarporsan:** Visualization, Software, Methodology. **Soroush Oshnoei:** Writing – review & editing, Conceptualization. **Tek Tjing Lie:** Writing – review & editing, Supervision. **Lieven Vandevelde:** Writing – review & editing, Supervision, Project administration. **Guillaume Crevecoeur:** Writing – review & editing, Supervision, Project administration.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, authors used ChatGPT to check grammar and enhance readability. After that, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding sources

This work was conducted as part of the project "Smart, Aware, Integrated Wind Farm Control Interacting with Digital Twins (ICONIC)" and has received funding from the European Union's Horizon Europe Research and Innovation Actions under the grant agreement No. 101122329, the UKRI Horizon Europe Guarantee scheme under the grant No. 10095874, and the Flanders Make project MultiScale.

Declaration of competing interest

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.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

[1] Nilsson D, Sannino A. Efficiency analysis of low-and medium-voltage DC distribution systems. In: IEEE power engineering society general meeting, 2004.. IEEE; 2004, p. 2315–21.
[2] Roga S, Bardhan S, Kumar Y, Dubey SK. Recent technology and challenges of wind energy generation: A review. Sustain Energy Technol Assessments 2022;52:102239.
[3] Mayilsamy G, Palanimuthu K, Venkateswaran R, Antonysamy RP, Lee SR, Song D, Joo YH. A review of state estimation techniques for grid-connected PMSG-based wind turbine systems. Energies 2023;16(2):634.
[4] Safaeinejad A, Rahimi M. An enhanced modal controllability-observability driven power control scheme for active suppression of rotor in-plane loading in large-scale PMSG-based wind turbine systems. Int J Electr Power Energy Syst 2025;171:110978.
[5] Tan G, Wei J, Zhao W, Qi L, Sun X. Application of three-dimensional unbalanced coordinate transformation to stand-alone four-leg voltage-source inverter. IEEE Trans Power Electron 2022;37(10):11686–703.

[6] Demirkutlu E, Hava AM. A scalar resonant-filter-bank-based output-voltage control method and a scalar minimum-switching-loss discontinuous PWM method for the four-leg-inverter-based three-phase four-wire power supply. IEEE Trans Ind Appl 2009;45(3):982–91.
[7] Li C, Von Jouanne A, Oriti G, Julian AL, Agamloh E, Yokochi A. GaN four-leg inverter implementing novel common mode elimination using a hardware-in-the-loop system-level controller. IEEE Trans Ind Appl 2023.
[8] Bouzidi M, Barkat S, Krama A, Abu-Rub H. Generalized predictive direct power control with constant switching frequency for multilevel four-leg grid connected converter. IEEE Trans Power Electron 2022;37(6):6625–36.
[9] Chen T, Pan Y, Xiong Z. Fault diagnosis scheme for single and simultaneous open-circuit faults of voltage-source inverters on the basis of fault online simulation. J Power Electron 2021;21:384–95.
[10] Khan SS, Wen H. A comprehensive review of fault diagnosis and tolerant control in DC-DC converters for DC microgrids. IEEE Access 2021;9:80100–27.
[11] Zhi S, Wu H, Shen H, Cheng W, Xiao Y, Wang T. Weak time-varying fault indicator detection of the complex gearbox via the meshing order modulation-aided scaling reassigning chirplet transform. Appl Acoust 2025;232:110580.
[12] Malik A, Haque A, Kurukuru VB, Khan MA, Blaabjerg F. Overview of fault detection approaches for grid connected photovoltaic inverters. E-Prime-Advances Electr Eng Electron Energy 2022;2:100035.
[13] Diao N, Zhang Y, Sun X, Song C, Wang W, Zhang H. A real-time open-circuit fault diagnosis method based on hybrid model flux observer for voltage source inverter fed sensorless vector controlled drives. IEEE Trans Power Electron 2022;38(2):2539–51.
[14] Pecina Sánchez JA, Campos-Delgado DU, Espinoza-Trejo DR, Valdez-Fernández AA, De Angelo CH. Fault diagnosis in grid-connected PV NPC inverters by a model-based and data processing combined approach. IET Power Electron 2019;12(12):3254–64.
[15] Xu S, Chen X, Liu F, Wang H, Chai Y, Zheng WX, Chen H. A novel adaptive SMO-based simultaneous diagnosis method for IGBT open-circuit faults and current sensor incipient faults of inverters in PMSM drives for electric vehicles. IEEE Trans Instrum Meas 2023;72. Art. no. 3526915.
[16] Li G, Xu S, Sun Z, Yao C, Ren G, Ma G. Open-circuit fault diagnosis for three-level ANPC inverter based on predictive current vector residual. IEEE Trans Ind Appl 2023;59(6):6837–51.
[17] Wan A, Tong X, Al-Bukhaiti K, Zhou Z, Su Y, Cheng X, Ji X. Vibration prediction for abnormal elevator door system faults based on attention mechanism and neural networks with time-frequency domain features. Proc Inst Mech Eng Part C: J Mech Eng Sci 2025;239(18):7358–72.
[18] Wang M, Cheng F, Xie M, Qiu G, Zhang J. Intensive multiorder feature extraction for incipient fault detection of inverter system. IEEE Trans Power Electron 2025;40(2):3543–52.
[19] Raj N, Jagadanand G, George S. Fault detection and diagnosis in asymmetric multilevel inverter using artificial neural network. Int J Electron 2018;105(4):559–71.
[20] Alvarez-Gonzalez F, Sierra-Gonzalez A, Trancho E, Marcos MA. Online signal-based fault detection and diagnosis of EV inverter during WLTP driving cycle. In: 2021 IEEE vehicle power and propulsion conference. Gijon, Spain: IEEE; 2021.
[21] Mahafzah KA, Obeidat MA, Mansour AM, Al-Shetwi AQ, Ustun TS. Artificial-intelligence-based open-circuit fault diagnosis in VSI-fed PMSMs and a novel fault recovery method. Sustainability 2022;14(24):16504.
[22] Singh V, Yadav A, Gupta S, Abdelaziz AY. Switch fault identification scheme based on machine learning algorithms for PV-fed three-phase neutral point clamped inverter. E-Prime-Advances Electr Eng Electron Energy 2024;8:100582.
[23] Fezai R, Dhibi K, Mansouri M, Trabelsi M, Hajji M, Bouzrara K, Nounou H, Nounou M. Effective random forest-based fault detection and diagnosis for wind energy conversion systems. IEEE Sensors J 2020;21(5):6914–21.
[24] Feng L, Luo H, Xu S, Du K. Inverter fault diagnosis for a three-phase permanent-magnet synchronous motor drive system based on SDAE-GAN-LSTM. Electronics 2023;12(19):4172.
[25] Si Y, Wang R, Zhang S, Zhou W, Lin A, Wang Y. Fault diagnosis based on attention collaborative LSTM networks for NPC three-level inverters. IEEE Trans Instrum Meas 2022;71. Art. no. 3512416.
[26] Lang W, Hu Y, Zhang Z, Gan C, Si J, Wen H. Few-shot learning with residual current for EV inverter fault diagnosis of EV powertrain. IEEE Trans Transp Electrification 2024;10(4):9316–27.
[27] Mishra M, Govind Singh J. Fault detection and localization in an MT-VSC-HVDC network via a hybrid CNN-transformer model with a multi-head attention mechanism. IEEE Access 2025;13:26365–83.
[28] Ramadan E, Moawad NM, Abouzalm BA, Sakr AA, Abouzaid WF, El-Banby GM. An innovative transformer neural network for fault detection and classification for photovoltaic modules. Energy Convers Manage 2024;314:118718.
[29] Mahmoud MS, Salem A, Robbersmyr KG, et al. A few–shot open-circuit fault diagnosis of F-type inverters using CGAN-based vision transformer. IEEE J Emerg Sel Top Power Electron 2024.
[30] Xu S, Zhang Y, Hu Y, Chai Y, Wang H, Yang X, Ma M, Zheng WX. Multiple open-switch fault diagnosis for three-phase four-leg inverter under unbalanced loads via interval sliding mode observer. IEEE Trans Power Electron 2024;39(6):7607–19.
[31] Peykarporsan R, Heidary J, Oshnoei S, Lie TT. A machine learning approach for fault detection and diagnosis in four-legged inverters. In: 2023 IEEE kansas power and energy conference. IEEE; 2023.
[32] Heidary J, Gheisarnejad M, Khooban MH. Stability enhancement and energy management of AC-DC microgrid based on active disturbance rejection control. Electr Power Syst Res 2023;217:109105.
[33] Tremblay O, Dessaint L-A, Dekkiche A-I. A generic battery model for the dynamic simulation of hybrid electric vehicles. In: 2007 IEEE vehicle power and propulsion conference. IEEE; 2007, p. 284–9.
[34] Karimi MH, Taher SA, Guerrero JM. Independent predictive control with current limiting capability of three-phase four-leg inverter-interfaced isolated microgrids. Int J Electr Power Energy Syst 2022;134:107457.
[35] Yaramasu V, Rivera M, Narimani M, Wu B, Rodriguez J. Model predictive approach for a simple and effective load voltage control of four-leg inverter with an output *LC* filter. IEEE Trans Ind Electron 2014;61(10):5259–70.
[36] Gajewski P, Pieńkowski K. Control of the hybrid renewable energy system with wind turbine, photovoltaic panels and battery energy storage. Energies 2021;14(6):1595.
[37] Garganeev AG, Aboelsaud R, Ibrahim A. Voltage control of autonomous three-phase four-leg VSI based on scalar PR controllers. In: 2019 20th International conference of Young specialists on micro/nanotechnologies and electron devices. IEEE; 2019, p. 558–64.
[38] Scikit-learn. Preprocessing data. 2025, https://scikit-learn.org/stable/modules/preprocessing.html. [Accessed: 05 June 2025].
[39] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[40] Teta A, Medkour M, Chennana A, Chouchane A, Himeur Y, Gadhafi R, Belabbaci EO, Atalla S, Mansoor W. Enhanced fault diagnosis in grid-connected photovoltaic systems: Leveraging transfer learning and ensemble methods for superior accuracy. IEEE Access 2024.
[41] Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–605.