Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p30436-cd

Assessment of the Degradation of IGBTs used in Wind Turbine Power Converters by a Method based on Physics-informed Transformers

Chenyang Lai
Department of Energy, Polytechnic University of Milan, Italy. E-mail: chenyang.lai@polimi.it

Piero Baraldi
Department of Energy, Polytechnic University of Milan, Italy. E-mail: piero.baraldi@polimi.it

Oleksii Zinchenko
School of Engineering, University of Warwick, UK. E-mail: Oleksii.Zinchenko@warwick.ac.uk

Xu Zhang
School of Engineering, University of Warwick, UK. E-mail: Xu.Zhang.1@warwick.ac.uk

Li Ran
School of Engineering, University of Warwick, UK. E-mail: L.Ran@warwick.ac.uk

Enrico Zio
Department of Energy, Polytechnic University of Milan, Italy. E-mail: enrico.zio@polimi.it
MINES Paris-PSL, CRC, Sophia Antipolis, France. E-mail: enrico.zio@minesparis.psl.eu

Insulated Gate Bipolar Transistors (IGBTs) are key components in modern energy conversion systems such as power grids and wind turbines. Their degradation and failure can lead to significant downtime and maintenance costs. Among various degradation mechanisms, we consider thermal stress, which accelerates material fatigue and increases the IGBT thermal resistance (R_{th}), and, in turn, reduces heat dissipation efficiency, further increasing junction temperature and promoting failures. This work proposes a method based on physics-informed Transformers for assessing IGBT degradation. It combines features related to IGBT degradation, extracted by relying on physics-based knowledge with Transformer encoders. A case study about power converters of wind turbines demonstrates that the proposed method outperforms baseline approaches, including DNN, LSTM and CNN models.

Keywords: IGBT, degradation assessment, physics-informed features, Transformer, wind turbine, power converter, reliability.

1. Introduction

Wind energy has emerged as a cornerstone of global renewable power generation. Within Wind Energy Conversion Systems (WECSs), power converters are essential for transforming the variable-frequency Alternating Current (AC) produced by the generator into grid-synchronized power. Insulated Gate Bipolar Transistors (IGBTs) serve as the primary switching components within these converters. Therefore, their reliability is necessary to ensuring the safety and operational efficiency of the overall system (Liu et al. (2024)). The degradation of IGBTs is primarily driven by thermo-mechanical stressors, including thermal cycling, bond wire fatigue and solder joint delamination. These degradation mechanisms cause an increase of the on-state resistance and junction temperature, which may ultimately result in catastrophic failures (Yang et al. (2024)). Consequently, assessing the degradation state of IGBTs is essential for implementing robust condition-based maintenance strategies and minimizing unscheduled downtime (Lai et al. (2022)).

Degradation assessment of IGBTs primarily rely on the measurements of electrical and thermal parameters. Specifically, junction temperature (T_j) and case temperature (T_c) serve as critical indicators of thermal stress (Bahman et al. (2017)). As degradation evolves, the junction-to-case thermal resistance (R_{th}) typically increases due to mechanisms such as solder delamination and bond wire lift-off Eleffendi and Johnson (2017). As a result, the correlation between power dissipation and temperature gradient between T_j and T_c is altered. Despite the correlation of T_j and T_c with thermal stress, traditional physics-based methods for degradation assessment often fail to achieve satisfactory accuracy, due to the complex and nonlinear relationship between measurable signals and the evolution of degradation, further complicated by the continuous variation of the operating conditions (Sathik et al. (2020)).

With the rapid advancement of sensing technologies and the increased availability of operational data, data-driven methods based on deep learning architectures, such as Deep Neural Networks (DNNs), Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have been increasingly investigated for IGBT degradation assessment. Juybari et al. (2022) combined a stacked autoencoder and a DNN model to estimate the case temperature of IGBTs in induction cooktops, for monitoring the degradation of IGBTs. Bai et al. (2025) proposed a CNN-LSTM network to model the IGBT degradation process. Mamee et al. (2024) proposed a CNN model using the gate voltage waveform to predict the evolution of IGBT degradation due to the bond wire lift-off. These approaches circumvent the limitations of traditional physics-based models by extracting latent features from raw electrical and thermal signals. By leveraging the temporal dependency modeling capability of LSTM units and the spatial feature extraction strength of CNNs, these models have been shown to achieve satisfactory accuracy under diverse and dynamic operating conditions, such as those typically encountered by wind energy conversion systems. However, their performance often relies heavily on the availability of large-scale datasets, which may be difficult to obtain in practical applications. Also, their limited interpretability and lack of embedded physical knowledge hinder the trustworthiness of these methods for degradation assessment.

In contrast, physics-informed machine learning (PIML) approaches incorporate physical knowledge into data-driven models, thereby improving generalization capability and interpretability (Raissi et al. (2019)). Several strategies have been proposed to develop PIML models, including the integration of physics-informed loss functions (Lai et al. (2024)), physics-informed neural network architectures (Lai et al. (2024)), physics-guided data generation and feature engineering (Zhang et al. (2025)). In parallel, the Transformer architecture, which was originally developed for natural language processing, has recently attracted attention for time-series classification due to the capability of the self-attention mechanism of capturing long-range dependencies (Vaswani et al. (2017)). Specifically, differently to recurrent networks, Transformers process data sequences in parallel and can attend to arbitrary temporal positions, which can be advantageous for degradation assessment when long-range temporal dynamics are relevant.

This paper proposes a physics-informed Transformer method for degradation assessment of IGBTs in wind turbine power converters. The main contributions are: (i) the design of three physics-informed features derived from thermal and electrical measurements of IGBTs; (ii) the use of these features as input to Transformer encoders for degradation assessment.

A case study considering power converters used in wind turbines has been worked out to investigate the performance of the proposed method.

2. Problem Statement and Formulation

We consider IGBTs of wind turbine power converters, which are characterized by high operating frequency and power density. We formulate the task of assessing IGBT degradation as a supervised multi-class classification problem, where the objective is to develop a data-driven model $f(\cdot)$ that maps multi-dimensional $X \in \mathbb{R}^{T\times N}$ of N signals measured for T consecutive time steps, $T_p - T + 1, ..., T_p$, to the corresponding degradation level y of the IGBT at current time T_p. We assume to have available a dataset $\mathcal{D} = \{(X^r, y^r)\}_{r=1}^{R}$, containing R time series $X^r =$

$[\mathbf{x}_1^r, \mathbf{x}_2^r, \ldots, \mathbf{x}_N^r] \in \mathbb{R}^{T \times N}$, extracted from data collected in the past from different power converters of wind turbines. Each time window is formed by the sequence of T consecutive vectors of $[x_{n,1}^r, x_{n,2}^r, \ldots, x_{n,T}^r] \in \mathbb{R}^T$ and $y^r \in \{0, 1, ..., C\}$ denotes the degradation level of the IGBT at the the end of the rth time window, where 0 corresponds to the healthy state, and $1, ..., C$ correspond to degradation levels of increasing intensity. The degradation mechanism considered in this work is the increase in the thermal resistance between the junction and case (R_{th}) of the IGBT. Different degradation levels are defined according to the percentage increase of R_{th} with respect to its nominal value

Specifically, we consider the following $N = 7$ signals measurements:

- Origin wind speed (V_{ori}): original wind speed measured outside the turbine.
- Wind speed (V): actual wind speed applied to the turbine model, processed according to the turbine's rated wind speed limits.
- Peak current (I_{peak})
- Rotational speed (N_{rot})
- Ambient temperature (T_{amb})
- IGBT junction temperature (T_j)
- IGBT case temperature (T_c)

3. Method

Section 3.1 delineates the derivation of the Physics-Informed (PI) features extracted from the thermal and electrical measurements. Subsequently, Section 3.2 introduces the Transformer-based classification architecture developed to assess the IGBT degradation level.

3.1. *Physics-informed features*

According to consolidated knowledge, the IGBT thermal behavior provides information on degradation (Shi et al. (2022)). Degradation mechanisms such as bond wire fatigue and solder layer delamination cause an increase of the thermal resistance between junction and case (R_{th}), resulting in a larger junction temperature compared to that expected in the same power dissipation conditions (Qian et al. (2018)). To effectively capture this, we have designed the following three physics-informed features:

(1) Junction-to-case temperature difference:

$$\Delta T_{jc} = T_j - T_c \tag{1}$$

This feature quantifies the temperature differential across the thermal path of the IGBT, which is directly influenced by the evolution of the thermal resistance. Considering IGBTs in similar operating conditions, an increase of ΔT_{jc} can indicate an ongoing degradation.

(2) Junction-to-case temperature difference normalized by current:

$$\Delta T_{\text{norm}} = \frac{\Delta T_{jc}}{I_{\text{peak}} + \epsilon} \tag{2}$$

where I_{peak} is the peak current and ϵ is a small positive constant used to avoid division by zero. This feature provides a robust indication of abnormal thermal behavior by mitigating the fluctuations of the junction temperature caused by variations of the load current.

(3) Junction-to-case temperature difference normalized by power:

$$\Delta T_{\text{norm},P} = \frac{\Delta T_{jc}}{I_{\text{peak}}^2 + \epsilon} \tag{3}$$

The temperature difference is normalized by dividing it for the square of the current as the power dissipation in the IGBT is proportional to the square of the peak current.

By concatenating the above three physics-informed features with the original seven measured signals, we obtain the $N + 3$-dimensional time series, $X^{\text{+phy}} \in \mathbb{R}^{T \times (N+3)}$. This integration of physically meaningful information enhances both interpretability and predictive accuracy of the subsequent data-driven models.

3.2. *Transformer-based classifier*

The proposed architecture utilizes a Transformer-based classifier designed to map $X^{\text{+phy}} \in \mathbb{R}^{T \times (N+3)}$ to the IGBT degradation level y. The model architecture consists of an input projection layer, a positional encoding module, a stack of Transformer encoder layers, a dual-path temporal pooling layer, and a final fully connected (FC) classification head. The Transformer framework is employed to exploit the capability of its multi-head self-attention mechanism of extracting long-range temporal dependencies without the inductive biases and vanishing gradient limitations of recurrent neural networks (Vaswani et al. (2017)).

The model receives in input the $(N+3)$ dimensional time-series $X^{\text{+phy}} \in \mathbb{R}^{T\times(N+3)}$. An internal latent space is built using an input projection layer G_{proj}, which maps $X^{\text{+phy}}$ into a d-dimensional embedding space:

$$H^{(0)} = G_{\text{proj}}\left(X^{\text{+phy}}, \theta_{\text{proj}}\right) : \mathbb{R}^{T\times(N+3)} \to \mathbb{R}^{T\times d} \tag{4}$$

where θ_{proj} denotes the learnable parameters of G_{proj}. Since the self-attention mechanism is permutation-invariant, a learnable positional encoding matrix $E_{\text{pos}} \in \mathbb{R}^{T\times d}$ is element-wise added to $H^{(0)}$ to inject information regarding the temporal order of the sequence (Ke et al. (2020)):

$$H^{(0)} := H^{(0)} + E_{\text{pos}} \tag{5}$$

The embedded sequence $H^{(0)}$ is subsequently processed by K stacked Transformer encoder layers, denoted by G_{TE} with parameters θ_{TE}. Each layer facilitates feature extraction through multi-head self-attention and position-wise feed-forward networks:

$$H^{(K)} = G_{\text{TE}}\left(H^{(0)}, \theta_{\text{TE}}\right) : \mathbb{R}^{T\times d} \to \mathbb{R}^{T\times d} \tag{6}$$

To further extract the representative information, $H^{(K)}$ is passed through a dual-path pooling layer G_{pool}, which computes the global average and global maximum across the temporal dimension T:

$$\begin{aligned} H_{\text{pool}} &= G_{\text{pool}}(H^{(K)}) \\ &= \left[\frac{1}{T}\sum_{t=1}^{T} H_t^{(K)}; \max_{t=1,\ldots,T} H_t^{(K)}\right] \in \mathbb{R}^{2d} \end{aligned} \tag{7}$$

where $H_t^{(K)}$ denotes the tth row of $H^{(K)}$. The dual use of average and max pooling is motivated by their complementary roles: average pooling captures the typical temporal behavior across the sequence, whereas max pooling retains salient peaks that may indicate degradation events. Therefore, their concatenation provides a richer representation than either one alone. Finally, fully-connected layers G_{FC} with parameters θ_{FC} are used to map the pooled representation H_{pool} to the prediction of class logits:

$$\hat{y} = G_{\text{FC}}\left(H_{\text{pool}}, \theta_{\text{FC}}\right) : \mathbb{R}^{2d} \to \mathbb{R}^{C+1} \tag{8}$$

where the cth component $\hat{y}_c$ of $\hat{y}$ is a number proportional to the confidence that the time series $X^{\text{+phy}}$ is of class c. The overall workflow of Transformer-based classifier is shown in Figure 1. The combined mapping is:

$$\begin{aligned} \hat{y} = G_{\text{FC}}\Big(G_{\text{pool}}\Big(G_{\text{TE}}\Big(G_{\text{proj}}\big(X^{\text{+phy}}, \theta_{\text{proj}}\big) + E_{\text{pos}}, \theta_{\text{TE}}\Big)\Big), \theta_{\text{FC}}\Big) \\ : \mathbb{R}^{T\times(N+3)} \to \mathbb{R}^{C+1} \end{aligned} \tag{9}$$

The parameters $\theta = [\theta_{\text{proj}}, \theta_{\text{TE}}, \theta_{\text{FC}}]$ are obtained by minimizing the cross-entropy loss over the training set. Xavier (Glorot) initialization is applied to all trainable parameters to preserve the variance of activations and gradients across layers, facilitating stable training and mitigating the risk of vanishing or exploding gradients (Glorot and Bengio (2010)).

The training objective is to minimize the cross-entropy loss:

$$\begin{aligned} \mathcal{L}(\theta) &= -\sum_{r=1}^{R}\sum_{c=1}^{4} \mathbb{1}[y^r = c]\log p_c^{(r)}, \\ \text{with} \quad p_c^r &= \frac{\exp(\hat{y}_c^r)}{\sum_{c'=1}^{C}\exp(\hat{y}_{c'}^r)} \end{aligned} \tag{10}$$

where $\mathbb{1}[\cdot]$ is the indicator function, $\hat{y}_c^r$ denotes the cth logit for the rth sample, and y^r is the true degradation label.

4. Case Study

We consider a dataset of IGBTs of different levels of thermal degradation. The data are generated from simulations of wind turbine converters under various operating conditions taken from SCADA data of a real wind turbine. Further details on the simulation setup are provided in Zhong et al. (2023).

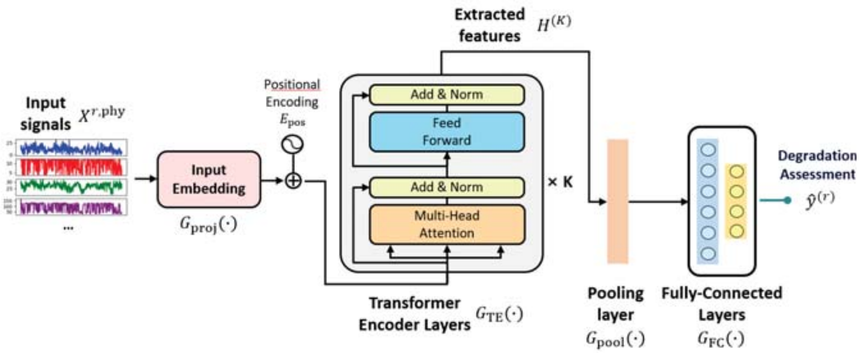

Fig. 1. Workflow of Transformer-based classifier.

4.1. *Data preprocessing*

The dataset is built by a simulation of the operation of an IGBT of a wind turbine power converter in healthy condition for 2000 minutes, and three simulations of the operation of IGBTs with light, medium and high degradation for 1000 minutes. All series are sampled at 10 points per minute. Degradation levels are defined according to the percentage increase of R_{th} with respect to its nominal value R_0. Specifically, light degradation corresponds to a 6.6% increase, $\frac{R_{th}-R_0}{R_0} = 6.6\%$, medium degradation to a 13.3% increase, $\frac{R_{th}-R_0}{R_0} = 13.3\%$, and high degradation to a 20% increase, $\frac{R_{th}-R_0}{R_0} = 20\%$.

The label $y \in \{0, 1, 2, 3\}$ denotes the degradation level of the IGBT, where 0 corresponds to the healthy state, 1 to light degradation, 2 to medium degradation and 3 to high degradation.

The data are split chronologically to avoid temporal leakage. Specifically, the first 70% of the time span is allocated to the training set, the next 10% to the validation set and the last 20% to the test set. Then, they have been subdivided into $R_{\text{train}} = 3474$, $R_{\text{val}} = 463$ and $R_{\text{test}} = 964$ time windows of length $T = 100$ with stride size of 10.

4.2. *Training strategy and hyperparameter setting*

Model architecture. $N^{+\text{phy}} = 10$ (seven original features and three physics-informed features) signals are projected by G_{proj} to a $d = 128$-dimensional embedding space. The Transformer encoder comprises $K = 4$ stacked layers, each with $d_{\text{model}} = 128$ hidden units, $n_{\text{head}} = 4$ attention heads, and a position-wise feed-forward sublayer of dimension $d_{\text{ff}} = 512$. Dropout with rate $p = 0.01$ is applied after the attention and feed-forward sublayers to improve generalization.

Training procedure. The model is trained for up to 500 epochs with batch size 64, using the Adam optimizer with initial learning rate 10^{-5} (Kingma and Ba (2014)). The learning rate is reduced by a factor of 0.8 at epochs 50, 100, 150, 200 and 250 via a multi-step scheduler. An early-stopping strategy is adopted to prevent overfitting with a patience of 30 epochs and a minimum improvement threshold of 10^{-4}. Xavier (Glorot) initialization is applied to all trainable parameters (Glorot and Bengio (2010)).

These hyperparameters were chosen through preliminary experiments and manual tuning to achieve satisfactory classification performance on the training set.

4.3. *Results*

The performance of the proposed method has been compared with that of baseline methods including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Deep Neural Networks (DNN). To assess the effectiveness of the physics-informed features, each method is evaluated under three input configurations: (1) the seven original signals only (ORI), (2) the three physics-informed features only (PHY) and (3) the combination of the original and physics-informed features (ORI+PHY). Hyperparameters of the baseline methods were selected by trials and errors using only the training set, whereas the validation set is used for early stopping.

Table 1 reports the obtained classification accuracy under each configuration. The results indicate that incorporating physics-informed features consistently improves performance across all backbone models. Notably, using only the physics-informed features (PHY) yields better accuracy than using only the seven original signals (ORI), which underscores the discriminative power of the physics-informed representation. Among the backbone architectures, the proposed Transformer-based classifier achieves the best performance in the ORI configuration, highlighting the effectiveness of its feature extraction capability. In the PHY and ORI+PHY configurations, the proposed method and the CNN-based method achieve comparable performance, outperforming LSTM and DNN. This is attributed to the fact that spatial relationships among patterns are more informative than temporal dynamics for capturing degradation patterns in this dataset. The confusion matrix obtained by the proposed method with ORI+PHY inputs is shown in Figure 2. Notice that all time series of the healthy class are correctly classified and all time series from degraded IGBTs are not assigned to the healthy class. Furthermore, all classification errors are conservatives, i.e. attributed to classes with larger degradation.

SHAP value analysis is performed to interpret the importance of input features for the proposed method. Because four degradation classes are considered, a separate SHAP analysis is conducted for each class. In all cases, ΔT_{jc} and ΔT_{norm} contribute most to the classification of the corresponding class. Figure 3 shows the SHAP results for the high-degradation class. The horizontal axis quantifies the contribution of each feature to the prediction: more positive values support the classification of high degradation, whereas more negative values support other classes. The color intensity indicates the feature value. The physics-informed features ΔT_{jc} and ΔT_{norm} are dominant; $\Delta T_{\mathrm{norm},P}$ is less important, possibly because the quadratic dependence on peak current

Table 1. Accuracy comparison (test set).

Config.	Proposed	CNN	LSTM	DNN
ORI	0.825	0.784	0.615	0.573
PHY	0.995	0.998	0.736	0.696
ORI+PHY	0.998	0.993	0.651	0.757

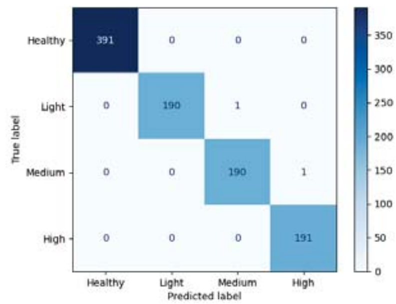

Fig. 2. Confusion matrix obtained by the proposed method with ORI+PHY inputs.

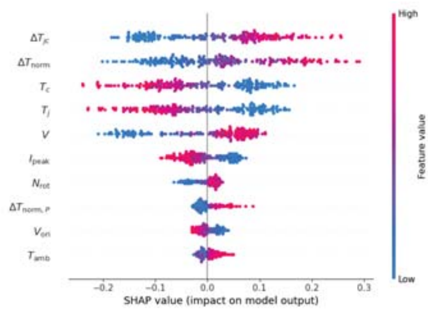

Fig. 3. SHAP value analysis results for the class of high degradation.

weakens its effect on the temperature differential, thereby reducing its sensitivity to degradation-induced changes.

5. Conclusions

This work has presented a physics-informed Transformer method for degradation assessment of IGBTs in wind turbine power converters. The approach integrates physics-informed features with a Transformer encoder. A case study based on simulated data from a wind turbine converter has demonstrated that the proposed method achieves satisfactory performance compared with other baseline approaches. The results indicate that incorporating physics-informed features consistently improves classification accuracy across all backbone architectures. Notably, using only the physics-informed features yields higher accuracy than using only the seven original signals, which underscores the discriminative power of the physically grounded representation. SHAP value analysis has confirmed that ΔT_{jc} and ΔT_{norm} are the most influential features for degradation classification, thereby providing interpretability and alignment with domain knowledge. Future work will be dedicated to the cross-validation of the obtained results considering different patterns of the data in training, validation and test sets. Also, we will address the problem that the junction temperature is difficult to measure in practical industrial scenarios.

Acknowledgement

This work has received funding from the European Union's Horizon Europe Research and Innovation Programme under grant agreement No. 101172940 (SAFE-POWER). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority. Neither the European Union nor the granting authority can be held responsible for them. The participation of Chenyang Lai is also supported by CINECA for the computing resources of Class C: RUL-PIDL - HP10CHF7KG.

References

Bahman, A. S., F. Iannuzzo, C. Uhrenfeldt, F. Blaabjerg, and S. Munk-Nielsen (2017). Modeling of short-circuit-related thermal stress in aged igbt modules. *IEEE Transactions on Industry Applications 53*(5), 4788–4795.

Bai, L., M. Huang, S. Pan, K. Li, and X. Zha (2025). Degradation prediction of igbt module based on cnn-lstm network. *Microelectronics Reliability 168*, 115639.

Eleffendi, M. A. and C. M. Johnson (2017). In-service diagnostics for wire-bond lift-off and solder fatigue of power semiconductor packages. *IEEE Transactions on Power Electronics 32*(9), 7187–7198.

Glorot, X. and Y. Bengio (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings.

Juybari, M. N., I. Ahmed, P. Baraldi, A. Del Cueto, J. Gil, C. Lai, S. Llorente, and E. Zio (2022). Estimation of the case temperature of insulated gate bipolar temperatures in induction cooktops by deep neural network. In *2022 6th International Conference on System Reliability and Safety (ICSRS)*, pp. 259–263. IEEE.

Ke, G., D. He, and T.-Y. Liu (2020). Rethinking positional encoding in language pre-training. *arXiv preprint arXiv:2006.15595*.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lai, C., P. Baraldi, I. Ahmed, E. Zio, J. Gil, S. Llorente, et al. (2022). Monitoring degradation of insulated gate bipolar transistors in induction cooktops by artificial neural networks. In *Proceedings of the 32nd European Safety and Reliability Conference-ESREL 2022*, pp. 1338–1345.

Lai, C., P. Baraldi, G. Quattrocchi, M. D. L. Dalla Vedova, L. Baldo, M. Bertone, and E. Zio (2024). Detection of abnormal conditions in electro-mechanical actuators by physics-informed long short-term memory networks. In *PHM Society European Conference*, Volume 8, pp. 1–8.

Lai, C., P. Baraldi, and E. Zio (2024). Physics-informed deep autoencoder for fault detection in new-design systems. *Mechanical Systems and Signal Processing 215*, 111420.

Liu, H., Y. Wang, T. Zeng, H. Wang, S.-C. Chan, and L. Ran (2024). Wind turbine generator failure analysis and fault diagnosis: A review. *IET Renewable Power Generation 18*(15), 3127–3148.

Mamee, T., Z. Lou, K. Hata, M. Takamiya, T. Sakurai, S.-I. Nishizawa, and W. Saito (2024). Estimating of igbt bond wire lift-off trend using convolutional neural network (cnn). *IEEE Access 12*, 96936–96945.

Qian, C., A. M. Gheitaghy, J. Fan, H. Tang, B. Sun, H. Ye, and G. Zhang (2018). Thermal management on igbt power electronic devices and modules. *Ieee Access 6*, 12868–12884.

Raissi, M., P. Perdikaris, and G. E. Karniadakis (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics 378*, 686–707.

Sathik, M. H. M., P. Sundararajan, F. Sasongko, J. Pou, and S. Natarajan (2020). Comparative analysis of igbt parameters variation under different accelerated aging tests. *IEEE Transactions on Electron Devices 67*(3), 1098–1105.

Shi, Y., J. Liu, Y. Ai, S. Chen, and Y. Bai (2022). Dynamic igbt three-dimensional thermal network model considering base solder degradation and thermal coupling between igbt chips. *IEEE Transactions on Transportation Electrification 9*(2), 2994–3011.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, Volume 30.

Yang, J., Y. Che, X. Guo, L. Ran, and Y. Chen (2024). In situ monitoring chip failure in multichip igbt modules using turn-on delay time extracted from auxiliary emitter voltage. *IEEE Journal of Emerging and Selected Topics in Power Electronics 12*(3), 2912–2922.

Zhang, S., M. Liu, R. Guo, J. Tian, Z. Man, and W. Shen (2025). Battery life prediction with scarce data using physics-informed data generation and adaptive autoencoder. *IEEE Transactions on Transportation Electrification*.

Zhong, Y., S. Lakshminarayan, L. Ran, P. Mawby, C. Jia, and C. Ng (2023). Detecting power module thermal resistance change in wind turbine converters with an attention-based lstm-autoencoder architecture. In *2023 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 314–320. IEEE.