Physics-informed Neural Network for Transformer Thermal Modelling

Abhinav Bhattarai, Anuradha Abeysekara, Chandima Ekanayake, Hui Ma, Tapan Kumar Saha

School of Electrical Engineering and Computer Science
The University of Queensland, Brisbane, Australia
{a.bhattarai, abeysekara.anuradha}@uq.edu.au, {chandima, huima, saha}@eecs.uq.edu.au

***Abstract*—Accurate prediction of transformer hot-spot temperature (HST) over a future time horizon is essential for assessing thermal stress and gaining valuable insights into transformer ageing trends. This, in turn, supports proactive and effective asset management strategies. The reliability of HST prediction is influenced by the precision of transformer top-oil temperature (TOT) prediction. Hence, this study proposes a physics-informed neural network (PINN) framework for TOT prediction, which integrates data-driven learning with domain-specific physical knowledge. A thermal model is developed using historical operational data, including load, ambient temperature, solar irradiance, and water activity in transformer oil, which are collected from a transformer operating at a grid-connected solar farm. Physical constraints are embedded into the learning process by incorporating Susa's thermal model. The values of TOT obtained from the proposed PINN are compared with measured data, as well as calculated TOT from a conventional artificial neural network (ANN) and the IEEE thermal model. The proposed PINN framework significantly outperforms both traditional ANN and IEEE models in prediction accuracy, demonstrating its potential for enhancing transformer thermal monitoring and supporting the development of transformer digital twins.**

***Index Terms*—Hot-spot temperature (HST), Physics-informed neural network (PINN), Top-oil temperature (TOT), Transformer thermal modelling, Digital Twin**

I. Introduction

Transformers are among the most critical components in power systems, and their lifespan is significantly dependent on the condition of the insulation, which is primarily affected by thermal stresses. The hot-spot temperature (HST), i.e. the highest temperature within a winding of a transformer, is a key indicator of insulation ageing and a major determinant of transformer life expectancy. Although its precise location is not directly measurable, it is widely accepted that the hot-spot typically forms in the upper region of the winding. However, HST can be accurately measured using fibre-optic temperature sensors [1] but, their adoption is limited due to high costs, maintenance requirements, and the difficulty of installing them in existing units. As a result, utilities often estimate hot-spot temperature indirectly using a formula, which requires the values of TOT, the temperature rise from oil to the hot spot, and the transformer's load.

To assess transformer insulation ageing over a future time horizon, it is necessary to forecast its TOT and, subsequently, hot-spot temperature within that time frame. Therefore, developing an accurate and reliable TOT prediction model is essential for improving future HST estimation and enabling proactive transformer thermal management.

There are several approaches for predicting the TOT of transformers. The IEEE and IEC standards provide first-order thermal models that estimate TOT as a function of ambient temperature and transformer loading [2]. These models need specific thermal parameters for each transformer, such as oil time constants and thermal exponents. While the standards offer recommended default values for transformers lacking detailed thermal data [3], using these generic values often leads to discrepancies between predicted and actual TOT [4].

To improve model accuracy, several enhancements have been proposed in the literature. Susa et al. [5] incorporated the effect of oil viscosity variation with temperature, significantly improving the thermal model. Additionally, the impact of water activity in transformer oil-paper insulation system on transformer temperature has been investigated in [6], and it was shown that water activity affects the thermal conductivity of insulation materials. These findings suggest that incorporating the water activity into thermal models can further enhance TOT prediction accuracy because water activity can indirectly influence observed thermal behaviour through change in heat transfer.

Recently, machine learning techniques have been adopted to transformer thermal modelling. He et al. [7] demonstrated that transformer temperature prediction can be improved using artificial neural networks (ANNs). Other studies have explored support vector machines and non-linear auto-regressive models to study transformer thermal behaviour, showing promising results [4]. Wu et al. [8] employed an online extreme learning machine to predict transformer operating temperature, incorporating weather parameters such as solar irradiance, wind speed, and ambient temperature. However, water activity, which can significantly influence internal temperature through its effect on insulation thermal conductivity, was not considered in that study. Moreover, these machine learning models are typically black-box in nature, and their internal decision-making processes are not easily interpretable, which limits their applicability in transformer condition monitoring.

In the context of evaluating transformer insulation ageing condition over a certain future time horizon, this study aims to enhance transformer TOT prediction by developing a physics-informed neural network (PINN) framework. By

979-8-3315-8138-1

embedding the governing differential equations of TOT into the learning process, the PINN is trained to learn from data while adhering to known physical principles. Unlike traditional models, the PINN does not rely on recommended parameter values, unlike conventional machine learning models, it is not a black-box system. This results in predictions that are not only more accurate but also physically interpretable. The proposed model takes input operational variables, including transformer load, ambient temperature, solar irradiance, and water activity measurement in transformer insulation system. The predicted TOT values are validated against measured data and predictions from a conventional ANN and the IEEE thermal model. Additionally, water activity is included as an input to the data-driven component of the PINN, further enhancing the thermal model's accuracy and relevance for real-world transformer predictive maintenance.

The following section presents the implementation of PINN for transformer thermal modelling. Section II describes the design of PINN incorporating ANN and known physics. Section III briefly explains the dataset used in this study. In Section IV, the results of the proposed models and conventional models are analysed. Finally, the last section concludes this study.

II. Physics-informed Neural Network (PINN) Design for Transformer Thermal Modelling

This section provides a brief review of the IEEE thermal model, Susa's model, and the ANN-based model, followed by a detailed description of the proposed PINN framework for transformer thermal modelling.

A. IEEE Model

As the transformer's winding current increases, both the winding and oil temperatures rise. The TOT rise above ambient can be calculated as [2]:

$$\tau_{\text{oil},R}\frac{d\Delta\theta_{\text{oil}}}{dt} = \Delta\theta_{\text{oil},U} - \Delta\theta_{\text{oil},I} \tag{1}$$

$$\Delta\theta_{\text{oil},U} = \Delta\theta_{\text{oil},R}\cdot\left(\frac{1+K^2\times R}{R+1}\right) \tag{2}$$

where,

$\tau_{oil,R}$	Oil time constant
$\Delta\theta_{oil}$	TOT rise over ambient temperature
$\Delta\theta_{oil,U}$	Ultimate TOT rise over ambient temperature
$\Delta\theta_{oil,I}$	Initial TOT rise over ambient temperature
$\Delta\theta_{oil,R}$	Rated TOT rise over ambient temperature
R	Ratio of copper loss to iron loss
K	Ratio of load current to rated winding current

B. Susa's Model

Susa's TOT model accounts for the nonlinear thermal resistance of transformer oil, which arises from temperature-dependent changes in oil viscosity. In the model, TOT can be calculated as [5].

$$\frac{1+K^2R}{1+R}\cdot\Delta\theta_{\text{oil},R} = \tau_{\text{oil},R}\frac{d\theta_{\text{oil}}}{dt} + \frac{(\theta_{\text{oil}}-\theta_{\text{amb}})^{1+n}}{\Delta\theta^n_{\text{oil},R}\cdot\mu^n_{\text{pu}}} \tag{3}$$

where,

μ^n_{pu}	Ratio of oil viscosity at any temperature to the ratio of oil viscosity at rated TOT
n	Constant
θ_{amb}	Ambient temperature
μ_{pu}	Per-unit oil viscosity

The values and ranges of the constants and quantities used in 1 - 3 are available in [9].

The oil viscosity at a particular temperature is given by 4 [10].

$$log(log(\mu_\theta+0.7)) = A - (B\times log(\theta)) \tag{4}$$

where, μ_θ is the viscosity of oil at a given oil temperature θ. A and B are constants, and for FR3 oil, these values are 7.102 and 2.772, respectively [10].

C. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) Fig. 2 are machine learning models composed of interconnected layers of neurons (nodes). A standard ANN has an input layer, one or more hidden layers, and an output layer. Raw data is fed to the input layer and processed through weighted connections and biases in the hidden layers. Each neuron applies an activation function such as $tanh$, $relu$, $silu$, $sigmoid$ to produce an output.

$$n^{k+1}(i) = \begin{cases} \sum_{j=1}^{N^k} b_i^{k+1} + w_{ij}^{k+1}x_j, & \text{if } k=0 \\ \sum_{j=1}^{N^k} b_i^{k+1} + w_{ij}^{k+1}a_j^k, & \text{if } k>0 \end{cases} \tag{5}$$

where,

i	Index of neuron
j	Index of input parameter
N	Total number of input parameters provided to each layer
k	Index of the layer

The weights and biases are optimised to minimise the sum of squared errors between the measured and predicted TOT. Optimisation can be performed using algorithms such as gradient descent, Levenberg–Marquardt, Adam, AdamW, or RMSprop [11].

D. Physics-Informed Neural Network (PINN)

PINNs integrate traditional neural network architectures with physical laws governed by differential equations. During training, a residual term derived from the differential equations is integrated into the loss function as a soft regularisation term [12].

The PINN architecture employed in this study, illustrated in Fig. 1, is built using a fully connected feed-forward neural network that maps input features, including load, ambient temperature, solar irradiance, and water activity to the TOT, as described in Section II-C. The differential equation-based residual enforces the heating of the transformer due to load. No

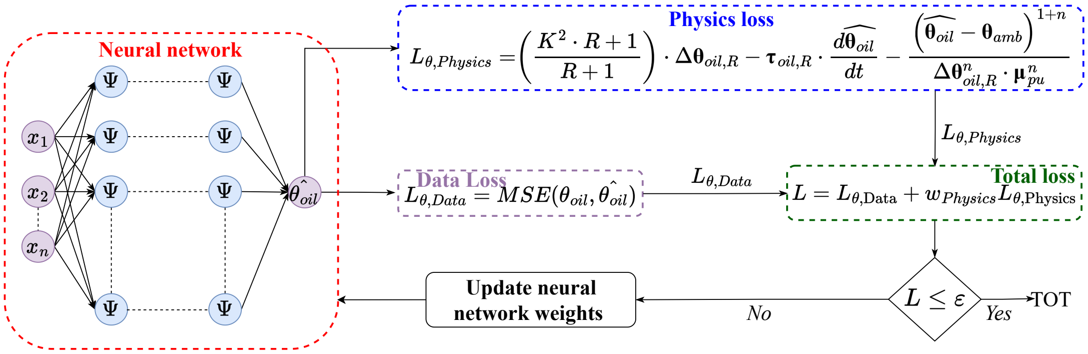


Fig. 1: Proposed PINN framework for transformer TOT prediction over future time horizon.

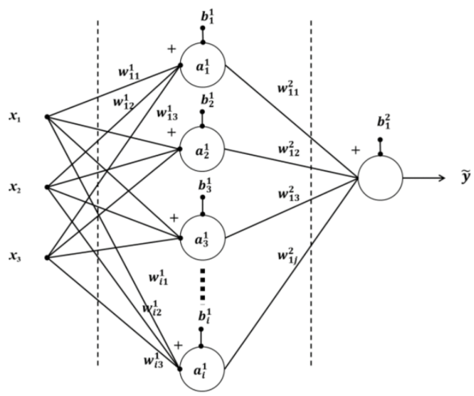


Fig. 2: A Typical ANN architecture [4]. x, w, b, a and $\tilde{y}$ denote the inputs, weights, bias, output generated from the activation function, and model output, respectively.

separate radiative heating term is integrated into the equation. Therefore, solar irradiance is treated as the input to the data-driven/neural network part of the proposed architecture, which accounts for direct solar heating effects not captured by the governing equation. The standard backpropagation algorithm is used to update the hidden layer weights during training. It leverages both data-driven learning and embedded physical knowledge to accurately capture transformer thermal behaviour.

The proposed PINN architecture minimises the loss function L, which integrates both the data loss $(L_{\theta,Data})$ and physics-based loss $(L_{\theta,Physics})$, as defined in Equation 6. When the physics weighting term $w_{Physics} = 0$, this architecture reduces to an ANN model, relying solely on data-driven learning.

$$L = L_{\theta,Data} + w_{Physics} \times L_{\theta,Physics} \tag{6}$$

$L_{\theta,Physics}$ is calculated using 7, which is the residual of 3.

$$\begin{aligned} L_{\theta,\text{Physics}} = {} & \frac{K^2 R + 1}{R+1} \Delta\theta_{\text{oil},R} - \tau_{\text{oil},R} \frac{d\hat{\theta}_{\text{oil}}}{dt} \\ & - \frac{\left(\hat{\theta}_{\text{oil}} - \theta_{\text{amb}}\right)^{1+n}}{\Delta\theta^n_{\text{oil},R}\, \mu^n_{\text{pu}}} \end{aligned} \tag{7}$$

where, $L_{\theta,Data}$ represents the mean squared error (MSE) between measured and predicted TOT, and $L_{\theta,Physics}$ denotes the residual of the differential equation 7 governing the TOT dynamics of the transformer, and $\hat{\theta_{oil}}$ is the TOT produced by ANN block.

The optimal hyperparameter configurations were determined using the Bayesian hyperparameter tuning [13] and the results are shown in Table I. The tuned network consists of two hidden layers with 16 and 10 neurons, respectively. The activation function is the $relu$. The output layer consists of a single neuron that predicts the value of TOT. A dropout rate of 0.078 is applied between layers to avoid overfitting of the model.

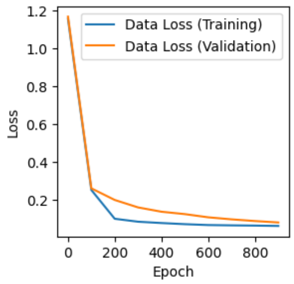
The physics loss weighting factor $(w_{Physics} = 5.90)$ balances the contribution of the physics-based residual relative to the data loss in the total loss function. This value of $w_{physics}$ was obtained through Bayesian hyperparameter tuning using Optuna over a search range of 0.5–50, and was jointly optimised with the network architecture. A sensitivity analysis around 5.90 showed that both smaller and larger values than this increased the error. Raising the weight to 6.00 increased the RMSE by 2.5%, while reducing it to 5.80 increased the RMSE by 2.7%, indicating that 5.90 provides the most stable performance. For the baseline ANN model, the same architecture and hyperparameter settings were used, but with $w_{Physics} = 0$.

Network parameters were optimised using the AdamW optimiser with a learning rate of 6.17×10^{-4} and weight decay of 1.89×10^{-5} as identified by the Bayesian tuning process. AdamW optimiser was selected because it has enhanced stability and predictive performance on time-series data. Training was performed with an early stopping patience of 15 epochs for 4000 epochs.

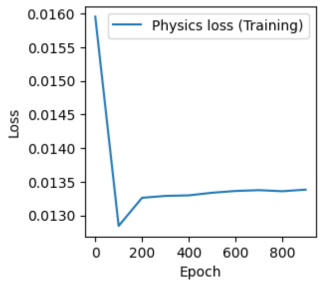
The study was performed on a workstation equipped with an Intel Core i7-14700 CPU and 32 GB RAM. Training of the PINN required 146 minutes. Average inference latency measured on the same CPU was 2 ms per sample, indicating that the model is suitable for real-time transformer monitoring.

The evolution of averaged data loss during training and validation Fig. 3a, and averaged physics loss during validation, Fig. 3b are also examined for the verification of convergence and possible overfitting. Further studies using extended and multiple time periods are also planned to evaluate their impact on the error metrics reported in Table II and also to assess

979-8-3315-8138-1



(a)



(b)

Fig. 3: Losses during training and validation of the model. (a) Data losses during training and validation. (b) Physics loss during training.

TABLE I: Optimised hyperparameters of the proposed PINN obtained using Bayesian hyperparameter tuning

Parameter	Symbol / Notation	Optimized Value
Hidden layer configuration	$[n_1, n_2]$	[16, 10]
Activation function	-	$relu$
Learning rate	η	6.17×10^{-4}
Weight decay	w_d	1.89×10^{-5}
Dropout rate	p	0.078
Total epochs	E	4000
Physics loss weight	w_{Physics}	5.90
Warm-up fraction	-	0.241
Early stopping patience	-	15
Scheduler patience	-	10
Optimizer	-	AdamW
Training algorithm	-	Backpropagation

potential gains in accuracy and generalisation.

III. Transformer Condition Monitoring Dataset

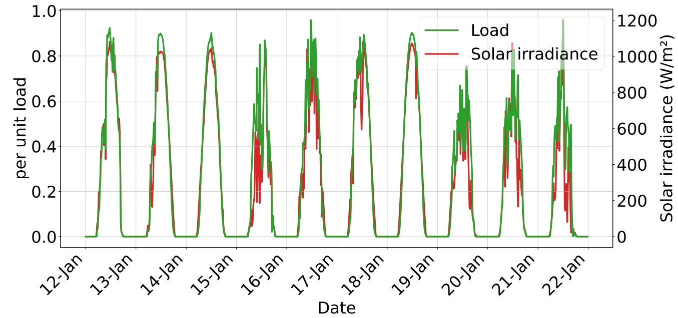
This study focuses on a 1.4 MVA inverter transformer T_1 operating at a 3.275 MW solar photo-voltaic plant [14]. Comprehensive operational and technical datasets for this transformer are obtained [9].

The dataset comprises time-series measurements of transformer load, solar irradiance, measured TOT, ambient temperature, and water activity in the transformer oil. Water activity was measured using commercially available capacitive moisture sensors [15]. The operating load data exhibits frequent and rapid fluctuations due to generation variability. Although the unseen event dataset was not separately designed, the training and evaluation data already included highly fluctuating load profiles. It spans from 1 May 2016 to 1 January 2018, with a sampling interval of 15 minutes. Prior to analysis, missing values and outliers were systematically filtered to ensure data consistency and reliability.

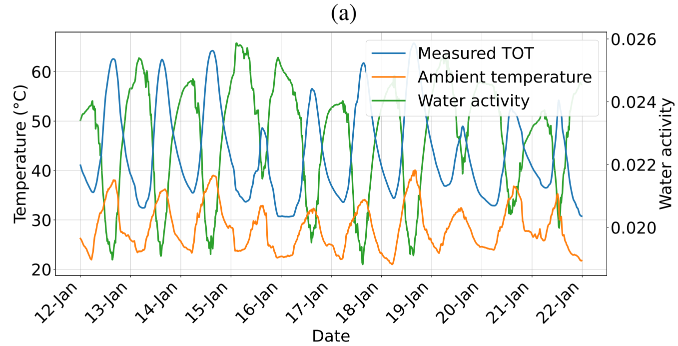
The first 80% of the dataset is used to train the ANN and PINN models, while the remaining 20% is reserved for validation. Fig. 4 presents a representative sample of data during training.

IV. Results and Discussions

Fig. 5 presents the TOT predictions from the ANN and PINN models, alongside the IEEE model calculations and measured TOT values, over a representative 10-day period.



(a)



(b)

Fig. 4: Snapshots of the data used in training (from 2017) (a) Load and Solar irradiance. (b) Measured TOT, ambient temperature and water activity.

All models effectively capture the dynamic trends of TOT during both rising and falling temperature phases. However, deviations from the measured values are observed at peak and valley points, with all models tending to underestimate the maximum temperatures. Among the three, the proposed PINN model demonstrates superior accuracy in tracking the measured TOT.

A statistical comparison of model performance was conducted. As shown in Table II, the PINN model achieved the lowest values for MAE, MSE, RMSE, and MAPE, indicating the highest prediction accuracy among the evaluated models. Additionally, the PINN yielded the highest R^2 value. Overall, the combined results indicate that the PINN offers the most consistent performance relative to the ANN and IEEE models.

Fig. 6 illustrates the error duration curves for the IEEE,

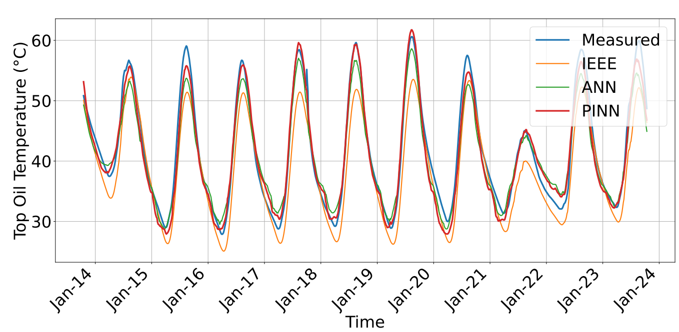


Fig. 5: Measured and model predicated TOT over Period from 2018-Jan-13 to 2018-Jan-23.

979-8-3315-8138-1

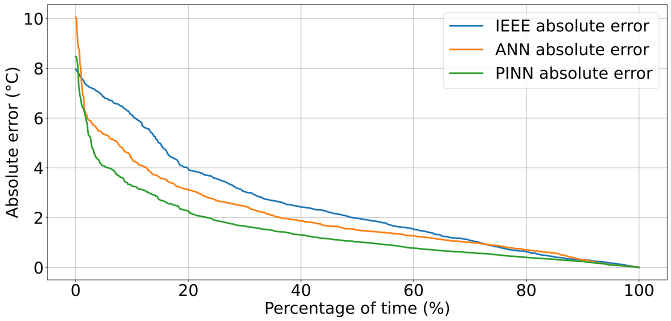


Fig. 6: Error duration curve for the proposed PINN, ANN and IEEE models.

TABLE II: Performance of different models

	MAE	MSE	RMSE	MAPE (%)	R^2
IEEE	2.49	10.44	3.23	5.76	0.82
NN	1.99	6.74	2.60	4.67	0.88
PINN	1.45	4.01	2.00	3.39	0.93

ANN, and PINN models based on absolute temperature error. On the left side of the graph (low percentage of time), all models exhibit their maximum errors, with ANN showing the largest peak error. The curves drop sharply, indicating that high errors occur only for a small fraction of time. At the 50% mark, the absolute errors reduce to approximately 1.5°C, 1.3°C, and 0.8°C for the IEEE, ANN, and PINN models, respectively. These results demonstrate that the PINN model has a superior ability to capture the nonlinear dynamics of transformer thermal behaviour. Furthermore, PINN maintains the lowest absolute error across the entire range, confirming its better alignment with actual thermal performance compared to IEEE and ANN.

Integrating physics-based differential equations into the PINN learning framework enables superior predictive performance compared to both purely data-driven approaches and traditional empirical thermal models.

V. Conclusions

This study proposed a novel PINN-based thermal model for predicting the TOT of transformers. Although both PINN and ANN were trained using identical hyperparameter settings, the integration of physics into the PINN learning framework significantly enhanced prediction accuracy. The proposed model incorporates key parameters influencing transformer thermal behaviour, including load, ambient temperature, solar irradiance, and water activity. Statistical analysis confirms that PINN outperforms ANN and the IEEE method across all performance metrics, demonstrating its potential for reliable temperature prediction.

This study is based on data from a single solar farm transformer that is mostly operated below its nameplate rating. Thus, the available measurements represent normal operating conditions and do not include sustained overload operation or unforeseen events like faults. Therefore, while the proposed approach demonstrates strong performance within the normal operating range represented in the dataset, its extrapolation capability under extreme loading conditions cannot be quantitatively evaluated from the available data. In future, this work can be extended to evaluate the proposed framework on a variety of transformers with overload events and fault conditions (subjected to availability of data) to: (i) assess the robustness of the proposed approach under extreme operating conditions, (ii) quantify predictive uncertainty, and (iii) improve generality across transformer types, ratings, and operating environments.

Acknowledgement

The authors acknowledge the use of Microsoft Copilot for assistance in language refinement during manuscript preparation. The intellectual content, analysis, and conclusions presented in this paper remain solely the responsibility of the authors.

References

[1] M. A. Ansari, D. Martin, and T. K. Saha, "Advanced online moisture measurements in transformer insulation using optical sensors," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 6, pp. 1803–1810, 2020.

[2] "IEEE guide for loading mineral-oil-immersed transformers and step-voltage regulators," *IEEE Std C57.91-2011 (Revision of IEEE Std C57.91-1995)*, pp. 1–123, 2012.

[3] *Power Transformers – Part 7: Loading guide for oil-immersed power transformers*, International Electrotechnical Commission (IEC) Std. IEC 60 076-7, 2018, standard.

[4] A. Doolgindachbaporn, G. Callender, P. L. Lewin, E. Simonson, and G. Wilson, "Data driven transformer thermal model for condition monitoring," *IEEE Transactions on Power Delivery*, vol. 37, no. 4, pp. 3133–3141, Aug 2022.

[5] D. Susa, M. Lehtonen, and H. Nordman, "Dynamic thermal modelling of power transformers," *IEEE transactions on Power Delivery*, vol. 20, no. 1, pp. 197–204, 2005.

[6] Y. Cui, H. Ma, T. Saha, C. Ekanayake, and D. Martin, "Moisture-dependent thermal modelling of power transformer," *IEEE Transactions on Power Delivery*, vol. 31, no. 5, pp. 2140–2150, Oct 2016.

[7] Q. He, J. Si, and D. J. Tylavsky, "Prediction of top-oil temperature for transformers using neural networks," *IEEE Transactions on Power Delivery*, vol. 15, no. 4, pp. 1205–1211, 2002.

[8] T. Wu, F. Yang, U. Farooq, X. Li, and J. Jiang, "An online learning method for constructing self-update digital twin model of power transformer temperature prediction," *Applied Thermal Engineering*, vol. 237, p. 121728, 2024.

[9] X. Zhong, C. Ekanayake, H. Ma, and T. Saha, "Operational data and detailed information of solar farm inverter transformers," *IEEE Dataport*, 2020. [Online]. Available: https://dx.doi.org/10.21227/77f3-g257

[10] D. Martin, J. Wijaya, N. Lelekakis, D. Susa, and N. Heyward, "Thermal analysis of two transformers filled with different oils," *IEEE Electrical Insulation Magazine*, vol. 30, no. 1, pp. 39–45, 2014.

[11] M. S. Osigbemeh, C. Osuji, M. O. Onyesolu, and U. P. Onochie, "Comparison of the artificial neural network's approximations for the levenberg-marquardt algorithm and the gradient descent optimization on datasets," *Artificial Intelligence*, vol. 5, no. 1, p. 25, 2024.

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

[13] A. H. Victoria and G. Maragatham, "Automatic tuning of hyperparameters using bayesian optimization," *Evolving Systems*, vol. 12, no. 1, pp. 217–223, 2021.

[14] A. Abeysekara, C. Ekanayake, and H. Ma, "Harmonic emission and aggregation behaviour at an inverter-transformer in a grid-connected pv plant," in *2025 IEEE PES 35th Australasian Universities Power Engineering Conference (AUPEC)*, 2025, pp. 1–5.

[15] D. Martin, O. Krause, and T. Saha, "Measuring the pressboard water content of transformers using cellulose isotherms and the frequency components of water migration," *IEEE Transactions on Power Delivery*, vol. 32, no. 3, pp. 1314–1320, 2017.

979-8-3315-8138-1