Available *online* at: http://ajeeet.ft.unand.ac.id/

Andalas Journal of Electrical and Electronic Engineering Technology

ISSN 2777-0079

Remaining Useful Lifetime Prediction of Distribution Transformer Using Dynamic Multi-Scale Attention-based CNN-LSTM

Elvis Tamakloe, Benjamin Kommey, Jerry John Kponyo, Daniel Opoku, Francis Boafo Effah

Faculty of Electrical and Computer Engineering, College of Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, 00233, Ghana

ARTICLE INFORMATION

Received: February 17, 2026
Revised: March 15, 2026
Accepted: March 31, 2026
Available online: May 31, 2026

KEYWORDS

Transformer degradation, RUL, Dynamic Multi-Scale Attention, Data-driven predictive maintenance, Oil-immersion

CORRESPONDENCE

Phone: +00233507703286

E-mail: bkommey.coe@knust.edu.gh

A B S T R A C T

Oil-immersed transformers are critical assets in the energy industry linking most power utilities to end-users. Their failure results in prolong outages, leading to huge revenue loss incurred during downtimes and replacement cost. In extreme cases, transformers in an unhealthy state poses a significant threat to the safety of grid operators. Interestingly, traditional reactive and preventive methods have been inefficient in determining when legitimate maintenance actions are due, often leading to either early over-maintenance of healthy transformers or late under-maintenance of serviceable and unhealthy transformers. Predictive maintenance based on determining the remaining useful lifetime (RUL) acts as an actionable step that resolves these challenges by delivering exactly the most appropriate time to undertake maintenance whiles ensuring optimal utilization of resources which saves maintenance cost, reduces downtimes and ensures operator safety and grid reliability. This work proposed an advanced Dynamic Multi-Scale Attention (DMSA) model and leverages on multi-modal data fusion from electrical, mechanical, thermal, and environmental sources to provide an improved data-driven solution for accurate prediction of the RUL of distribution transformers. This technique addressed the drawbacks of employing single modality approaches in capturing complex operational interactions. In this work, dynamic scaling model is incorporated to adaptively adjust the attention weights based on the importance of the input features. For short term predictions, the proposed model experimentally achieved an enhanced performance of 0.2300 mean absolute error and 0.9872 coefficient of determination value. Additionally, the DMSA CNN-LSTM model demonstrated accurate prediction, evidenced by a concordance correlation coefficient value of 0.9936. These statistical gains were achieved in a computational time of 587.3387s, demonstrating superior scalability in the event of real time deployment. Furthermore, the long-term prediction was performed using Prophet to fit the data which predicted a RUL of 25 years at 95% confidence interval which corresponded with the reference standard in IEEE STD C57.91.

INTRODUCTION

Maintenance of distribution transformers have played a central role in their management and particularly in ensuring reliable power delivery at economical voltages [1,2]. Continuous operation and exposure of distribution transformers to stressors without proper maintenance reduces their performance and accelerates their deterioration over time especially in events of undetected failures [3,4]. This unpleasant outcome if left unchecked leads to severe service interruptions, high outages, maintenance downtime and cost. Thus, to make correct informed decisions regarding the proper management of distribution transformers, it is of utmost necessity to know their remaining useful lifetime (RUL). This is essential in order to facilitate decisions on transformer maintenance, injections and replacement [5,6]. Conventional reactive and preventive practices (periodic inspection, schedule maintenance and condition monitoring) have provided an incomplete understanding of the health of transformers in service. This shortcoming results in inaccuracies of knowing exactly how long a transformer can continue operating safely. RUL addresses this limitation by estimating the time left before the transformer reaches a failure point or an unacceptable level of operation. In view of this, math or physics-based models were developed to resolve this problem and predict the RUL of transformers using physical properties [7-9]. However, this method requires an in-depth domain knowledge to sufficiently model the complex interaction in distribution transformers. Furthermore, it is less adaptable to variable transformer operations and requires manual adjustments or calibration which limits the accuracy of their prediction [10,11]. The progress in ML and AI has reshaped the terrain of predicting the RUL more accurately compared to physics-based techniques by leveraging on historical operational data sampled from sensor measurements [12]. In the quest to achieve better performance, several DL models have employed hybrid techniques and incorporated a variety of attention mechanisms in their respective architecture to capture complex machine degradation. This is intended to capture the intricate dependencies and focus on relevant features that contribute to the deterioration of the transformer in order to make accurate predictions of their RUL

https://doi.org/10.25077/ajeeet.v6i1.210

[13-15]. The enhanced performance of neural networks models has been attributed to the integrating of an attention mechanism. However, several types of attentions mechanisms exist and their applications in neural networks vary based on the task requirement, model architecture, and its computational efficiency [16,17]. In reference to predicting the RUL of distribution transformers, most deep learning models are unable to captured complex deterioration across different time scales typically in situations where the feature importance varies. Although the adoption of multi-scale attention mechanism enables DL models focus on the relevant features contributing to transformer deterioration, their performance is suboptimal since they apply fixed weights and are unable to re-weigh vital features to identify sudden changes [18-20]. Considering the economic value and productive time that is lost as a result of damaged in-service power transformers, it is imperative to accurately predict the end of useful life in order to put in place the right contingency plans for maintenance action. This enables proper asset management by making informed decisions at the right time [21]. Due to aging and fault stress levels, the RUL of power transformers can be thoroughly mapped using features acquired non-intrusively in the data acquisition phase [22, 23]. This work proposed an innovative DMSA CNN-LSTM model to resolve the problem and improve the accuracy of predicting the RUL of distribution transformer for efficient health management.

[23] proposed an advanced CNN-LSTM model to tackle inaccurate RUL prediction. The model comprises a multi-level 1D CNN layer and a stacked multi-layer LSTM for extracting deep spatial and time-dependent features, respectively. That is, the 1D CNN is used to process convolutions along the two dimensions of the input data. With a time-window ranging from 15-30 and a convolution kernel size of 3 to 7, the input data is processed and transferred into the LSTM network. The stacked multilayer LSTM network employed dropout regularization to prevent overfitting during training. Thus, the identified feature vectors from the processed data are transferred to the fully connected layers for prediction. The model was subsequently trained on 60 iterations with a learning rate set at 0.01 and tested accordingly to verify its performance by evaluating both the scoring function and the RMSE value. A root mean squared error value of approximately 18.2084 was recorded on the test set. This implies that nearly 18.2% of error was identified between the actual and predicted values, which suggests a good performance. However, the scoring results were not provided, and the model's performance was not benchmarked. Although the authors claimed to have conducted a comparative assessment with other traditional data-driven models, no result was shown. Additionally, the type of dataset used in this experiment was unknown and the issue of data imbalance was not extensively addressed. Reducing the estimation error and time is critical for models developed as solutions for estimating the RUL of power transformers. In efforts to improve the model performance, [24] proposed a Bayesian neural network that evaluates power transformers based on a variety of features. In achieving this, actual data from about 500 transformers were collected at different periods and external conditions. The sampled data was preprocessed using wavelet transform (WT) to minimize errors inherent in the dataset. That is, a scale parameter is employed in this case without the need for a frequency parameter as observed in Fourier transforms. The resulting features are used as inputs to train and validate the Bayesian neural network. Principal Component Analysis (PCA) was primarily utilized to reduce data dimensions which is central for weighting to ensure the best performance. The proposed model achieved a classification accuracy of about 98.4%, implying that nearly 98% of instances were correctly classified by the model. This result indicated superiority as compared to other supervised learning models like the KNN, SVM, and multi-layer perceptron neural network (MLPNN) which achieved accuracies of approximately 92%, 93%, and 96.4% respectively. Thus, about 1.6% of the instances were incorrectly classified, which suggests a reduced error rate compared to the counterpart models. Although the model recorded a high accuracy, using this metric alone does not fully represent the model's overall performance. Thus, the F1 score, recall, precision, and hamming loss values were not accounted for and making this work inclined towards classification rather than regression. Furthermore, expressing RUL in the time domain rather than in percentage is more appropriate given the above task. In [25], two back propagation neural networks (BPNN) models were developed to forecast the degree of polymerization (DP) of a transformer cellulose paper and predict the loss of life (LOL) of the transformer. This approach utilized concentration data of 2-Furaldehyde (2FAL) collected from dissolve gas analysis (DGA) of the transformer oil. Based on the forecast result of the DP together with the 2FAL data obtained from the first BPNN algorithm, the LOL was then predicted using the second BPNN algorithm. The later model achieved a MSE value of 497.58 and correlation coefficient of 0.999. Considering this performance, the authors highlighted the success of the model. However, the high MSE value recorded at the 5 epochs suggests a high difference or variance between the actual and predicted value. Moreover, the correlation coefficient value of close to unity suggests a high performance between the actual and predicted values. These two metrics conflicts an in-depth assessment of the performance. Thus, more metrics are required to comprehensively evaluate the performance of this approach. Numerous studies have also focused on combining integral sources to predict the loss of life of transformers based on the oil insulations [26-28]. However, the models in these studies require extensive multi-modal data to capture intricate details to enhance its predictive performance. To improve the predictive performance, [29] proposed a multi-task model that combined LSTM and GRU models. After training on an entire dataset that included dissolve gases and oil quality of the transformer, the result revealed a MSE, MAE, R2-score and a MedAE of 2.543, 0.1346, 0.985, and 0.0284 which suggests a significant improvement over other combinations of DL architecture and conventional single regression and ensemble algorithms. Additionally, the work utilized Shapley Additive Explanations (SHAP), an explainable AI tool which provided a global and instance-level details on features that predominantly influenced the outcome of the model's prediction. Howbeit, numerous iterations over large number of epochs were required to train the model considering the computational resources used in the work. Furthermore, authors did not sufficiently address the case of data imbalance with regards to their data acquisition system and during preprocessing. To achieve an equal performance using optimal resources, [30] utilized a Support Vector Machine (SVM) to determine the condition of the cellulose insulation in order to estimate the remaining functional lifetime of the transformer. Although this approach achieved more than 95% accuracy in

classifying conditions of the insulation using a variety of SVM models, it nonetheless, did not explicitly describe how this was employed to determine the RUL of the transformer which is a regression study.

METHOD

DL based RUL Estimation Framework

Figure 1 presents a prognostic architecture that predicts the duration the transformer in service can continue to effectively operate before failure or breakdown threshold is reached. The objective of this approach is to estimate this duration based on the features extracted from the transformer data in order to know when to perform maintenance. Hence, the preprocessed data is used to train the DMSA-based CNN-LSTM model to analyze this regression task using Python.

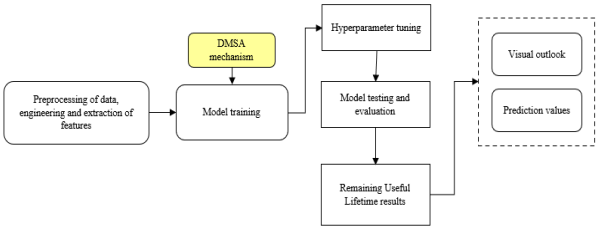

Figure 1. DL based RUL Estimation Framework.

After rigorous training, the model is evaluated and its hyperparameters are well-tuned to provide the most suitable result required to accurately predict the RUL of the power transformer. The results obtained from this stage are visualized to graphically determine the correlation between the predicted and true (actual) RUL values. The goal is to observe possible cases of under-or-over prediction patterns which gives insight into the model's performance and discrepancies. Numerical values of the prediction are then used to make informed decisions with regards to maintenance scheduling either on short-term or long-term bases. Hence, the integration of several features through multi-modal data fusion provides a comprehensive platform for accurately predicting transformer RUL. This is ultimately vital for extending transformer life and enabling reliable power supply. To a greater extent, by leveraging on the DMSA mechanism offered by the proposed model, the relevant features across the respective time scales (short and long) are adaptively prioritized to improve the interpretability. This implies that features with more weights in the model's predictions provide insights into the main features influencing the health of the transformer which depicts its RUL.

DL DMSA based CNN-LSTM Block

Oil-immersed distribution transformers are complex electrical machines and adopting predictive maintenance offers a huge opportunity to improve their reliability in the power system for a longer period. However, developing very accurate and robust predictive maintenance models to ensure a comprehensive diagnosis and prognosis of transformers require not only the use of multi-modal datasets but also on the algorithm that handles the associated complexities. In this work, a DMSA mechanism as presented in Figure 2 was introduced as an innovative and novel approach in an ensembled CNN-LSTM model to capture both the short and long-term dependencies within the complex multi-modal dataset.

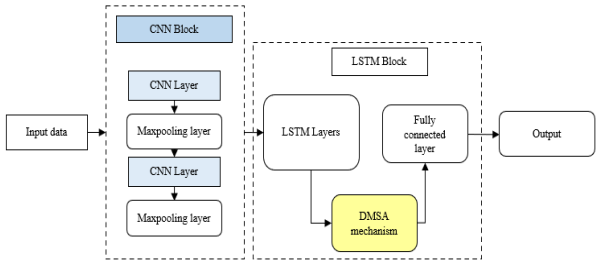

Figure 2. DL based DSMA CNN-LSTM Block.

In contrast to traditional attention mechanisms, the developed approach incorporates dynamic scaling, enabling the model to adaptively assign weights to both short and long-term temporal scales. This is done in response to changing transformer conditions. The design allows the model to selectively emphasize the most valuable time scale during anomaly detection while preserving a broader temporal context for accurate RUL prediction. A hybrid CNN-LSTM block is employed, where the CNN component extracts spatial features from the composite multimodal dataset. The LSTM component models temporal dependencies with the data given its sequential nature. This combination effectively learns both steady-state and transient behavioural patterns. The blocks are specifically structured to consist of two 1-D convolutional layers and two LSTM layers with each followed by a 1-D maxpooling layer. Utilizing 64 and 128 units, a progressive configuration was chosen for the respective CNN and LSTM blocks. The CNN block employs a kernel size of 3, while a pool size of 2 is used in the maxpooling layers. The output of the final CNN maxpooling layer serves as input to the LSTM block. The DL based DMSA mechanism is then applied to the LSTM output for adaptive temporal weighting and enhanced predictive capability across multiple timescales.

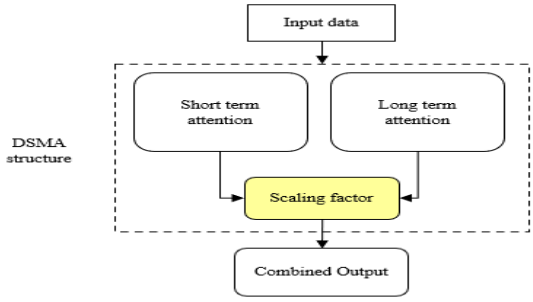

Figure 3. Developed DMSA structure

The developed DMSA structure shown in Figure 3 is designed to model both short and long-term temporal dependencies as established. This is accomplished through two complementary attention modules. One is precisely focused on short-term patterns and the other on long-term sequence data. In the short-term attention branch, the input is passed through a single-unit dense layer with a hyperbolic tangent (tanh) activation function. The resulting output is flattened, and a softmax activation is applied to determine the relative significance of each time step. The attention weights are repeated across the LSTM units and permuted to match the input dimension, generating an attention weighted representation that projects the short-term dynamics. The long-term attention mechanism is implemented whereby attention scores are found to take into account broader temporal dependencies in the data. The attention weights from both short and long-term branches are multiplied element-wise with the

 https://doi.org/10.25077/ajeeet.v6i1.210

corresponding inputs to enhance the most informative features. The dynamic scaling factor is implemented using two separate lightweight single-unit dense layers with each activated with a tanh-function. This is to regulate the contributions derived from the attention blocks based on computed scaling coefficients. The scaling coefficients are multiplied with the respective attention outputs. The dynamically scaled attention outputs are combined through element-wise addition to give the unified feature map for the two fully connected layers. Unlike standard multi-scale attention that applies static or fixed weights and lacks temporal state awareness, the innovation of dynamically scaled attention output enables the model to adaptively reweight the multi-scale temporal features in response to changing transformer condition. This is derived from the evolving LSTM hidden state which gives the DMSA continuous temporal awareness given the heterogenous nature of transformer data In this manner, accurate RUL prediction in the transformer is captured by emphasizing on the most relevant and informative temporal patterns.

Analytical Representation RUL framework

The proposed attention mechanism is applied to enable the model to focus on the significant parts of the transformer data. To do so, an attention score (δ_t), attention weights (ε_t) and context vector (τ) are computed using equations (1), (2), and (3) respectively. This provided the innovative base for this work's contribution.

$$\delta_t = v^T tanh(W_h h_t + b_h) \quad (1)$$

$$\varepsilon_t = \frac{exp(\delta_t)}{\sum_{k=1}^{T} exp(\delta_t)} \quad (2)$$

$$\tau = \sum_{t=1}^{T} (\varepsilon_t h_t) \quad (3)$$

$$s = tanh(W_s h_t + b_s) \quad (4)$$

The dynamic scaling factor (s) is computed accordingly and applied to the context vector to obtain a dynamically scaled result or effective context vector (τ_e). This incorporates the individual context vector for the short-term (τ_{short}) and long-term (τ_{long}) as expressed in the following equation:

$$\tau_e = s \cdot (\tau_{short} + \tau_{long}) \quad (5)$$

Thus, τ_e is then forwarded to the dense layer for final prediction as represented in the equation (6):

$$y' = (W_{out} \cdot \tau_e + b_{out}) \quad (6)$$

To predict the RUL which is a regression task, the output and the loss function ($L_\delta{}'$) were set accordingly. Here, linear activation function is employed in the output layer and Huber loss function is used to reduce the difference between predicted and actual target values as expressed in this Equation 7 by leveraging on the merits of MSE and MAE to address the impact outliers that influence the performance of the model [40, 41].

$$L_\delta{}'(y', \hat{y}) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2}(y_i' - \hat{y}_i)^2 & |y_i' - \hat{y}_i| \le \delta \\ \frac{1}{N}\sum_{i=1}^{N} \delta\left(|y_i' - \hat{y}_i| - \frac{1}{2}\delta\right), & |y_i' - \hat{y}_i| > \delta \end{cases} \quad (7)$$

Imperatively, δ is the threshold loss parameter controlling the transition between the MSE and MAE. Hence, the loss function transitions to MSE given that $|y_i' - \hat{y}_i|$ is small and MAE provided that $|y_i' - \hat{y}_i|$ is huge [42].

Process Flow Diagram of the RUL Framework

The process flow diagram in Figure 4 predicts the continuous RUL of the transformer based on the learned features from the data provided. This represents a structured regression task where the fused multi-modal data is preprocessed, and the relevant features are extracted accordingly via the data preprocessing and feature extraction stages.

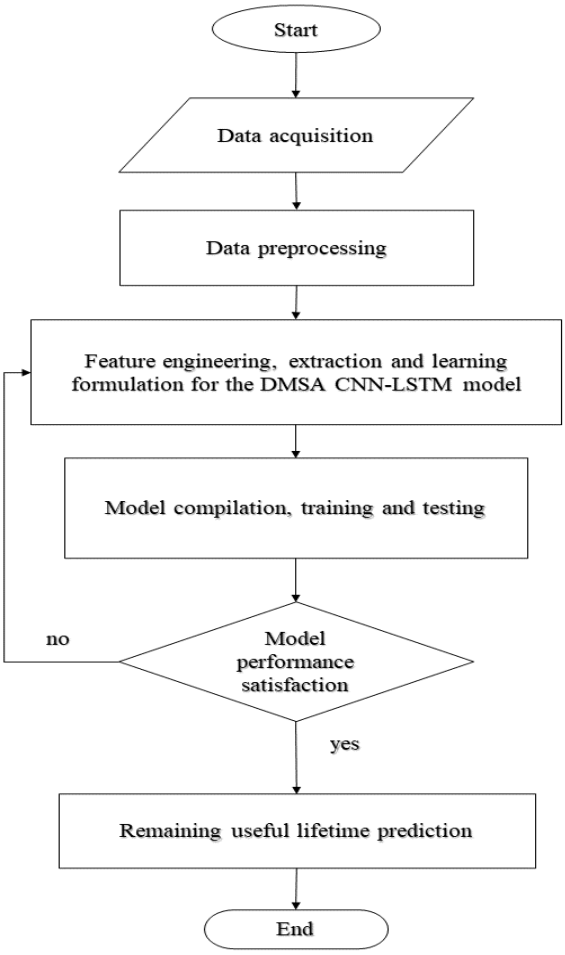

Figure 4. Process Flow of the RUL Prediction Framework

Here, degradation trends that impact the lifetime of the transformer are effectively identified with the innovative proposed DMSA mechanism. This is achieved through the model's adaptive focus on the short-term or long-term patterns. The model compilation and training are realized with regression-based loss function (mean squared error), and metric (mean absolute error). The decision block determines whether the trained model requires further adjustment based on its performance validation. A "yes" means further optimization is required and a "no" suggests additional retuning is not required. The final stage of the flowchart constitutes three different processes that emphasize model evaluation, prediction and error analysis of the predicted RUL. Therefore, the proposed prognostic approach does not only depict the accuracy of the predictions but also ensures interpretability. Thus, the attention score of the most vital features and time scale that immensely

contribute to the RUL are understood to aid in timely decision making in the phase of maintaining the transformer.

Data Description and Handling

To predict the RUL of a power transformer requires a thorough understanding of the trend, patterns and dependencies in the dataset. The dataset used in this work represents both electrical and environmental data sampled from a 100kVA rated ONAN distribution transformer at 1s and 15min interval respectively for nearly 10 months [31]. Thermal DGA data from [32, 33] mapped internal condition of the oil insulation. This means that similar degradation pattern of the oil insulation exist for transformers with same characteristic and operating conditions [34-36]. Featured engineered mechanical vibrations from the current data provided unique contribution highlighted in [37]. Early fusion was then applied to combine all the modalities into a unified dataset to inherently retain all the original information [38, 39]. Missing values and outlier were treated and a StandardScaler was then used to normalize the data. The unified dataset was then divided into 64% training, 16% validation and 20% testing set to ensure representation and robust evaluation. In the data exploration phase, all conditions were set as features including the health index and the RUL was dedicated as the target or label. A visual distribution of all the respective features is illustrated in a histogram plot in Figure 5 and a summary of the dataset provided in Table A1 (see appendix).

Data Investigation before RUL Prediction Assessment

In this Figure 5, each subplot represents the individual features plotted against the frequency and value on the y-axis and x-axis respectively. This is critical to unearth the underlying patterns in the data in order to improve the performance of the DMSA CNN-LSTM model for predicting the RUL of the transformer. The plot reveals that the individual voltage and current features were skewed towards the right with some high values although some values are concentrated on the lower axis. Other features including the power factor showed narrow distribution indicating that majority of the transformer's operation are within a normal range. Moreover, peaks in the fault gases imply a corresponding degradation of insulation materials. Furthermore, the visualization of the health index provided a holistic representation of the transformer's health. Based on this reference, targeted maintenance strategies can then be applied to facilitate in the identification of early signs of abnormalities. Before that, it is imperative to know that by analyzing how these features affect each other, the proposed model can then effectively and adaptively learn which duration or time scale is essential for predicting the RUL of the transformer. That is, the variation in the time-series data shows the significance of having both the short-term and long-term attention mechanisms in the model to deal with the complex patterns that results. Thus, abrupt fluctuations in some features (temperature or gas concentrations) are addressed with the short-term attentions whereas more slowly evolving changes (like gradual rise in water content and gradual depletion of the insulation) are captured with the long-term attention mechanism. Therefore, by visualizing these features, the most influential features that adversely affect the life expectation of the distribution transformer is identified. A line plot of the life expectation (RUL) is presented over the period (Year_Month). See Figure 6 in appendix. An initial increase in life expectation was observed between June 2019 to August 2019 signifying an initialization period due to initial commencement and adjustments in the operating conditions. From August 2019 onwards, no great variations were recorded as a steady RUL value stood at about 32 years. However, the slight dips and rises in the RUL under this stable period stemmed from the small variations in the operational features visualized in the composite dataset.

RESULTS AND DISCUSSION

Performance of RUL Prediction or Estimation Model.

The proposed DMSA CNN-LSTM RUL prediction model was trained, tested and validated over a stipulated number of epochs using the new preprocessed composite data. The paramount objective of the model is to accurately predict or estimate the RUL of the transformer based on the historical data in order to prioritize maintenance actions thereby facilitating efficient asset management. As discussed in the preceding subsection, RUL prediction is a prognostic task and in this case the performance of the model is evaluated using standard unique regression metrics [43]. Table 1 presents the outcome of the proposed model compared with other baseline DNN models.

Table 1. Initial result of the DMSA CNN-LSTM model against other DNN models on default parameters

Model	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
CNN	1.0236	1.7469	0.9806	4.4959	0.9896	220.6341
LSTM	0.6067	1.7546	0.9804	3.0430	0.9899	1159.2371
GRU	0.2694	1.2192	0.9905	1.6663	0.9952	1161.8729
CNN-LSTM	0.3289	1.8826	0.9774	2.7119	0.9888	410.1172
MSA CNN-LSTM	0.5166	1.6257	0.9832	3.2051	0.9913	607.0822
DMSA CNN-GRU	0.2263	1.6815	0.9820	1.4309	0.9910	646.2549
DMSA CNN-LSTM	**0.2520**	**1.4921**	**0.9858**	**1.5985**	**0.9930**	**595.2953**

The performance of these respective models was obtained given a batch size of 32, 50 epochs, a default learning rate of 0.001 with an Adam optimizer and a default threshold loss parameter (δ) of 1.0. Therefore, this table presented the best performance relating to the mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2-score), mean absolute percentage error (MAPE), concordance correlation coefficient (CCC) and computation time required by the model. The said metrics used in the evaluation process of the model were computed based on their individual unique analytical contributions as given in Equations (8), (9), (10), (11), (12) and (13).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i' - \hat{y}_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i' - \hat{y}_i\right)^2} \tag{9}$$

$$R2\ score = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\left(y_i' - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i' - \bar{y}\right)^2} \tag{10}$$

 https://doi.org/10.25077/ajeeet.v6i1.210

$$MAPE\ (\%) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|y_i - \hat{y}_i|}{y_i}\right) \times 100 \quad (11)$$

$$CCC = \frac{2\rho\sigma_{y_i'}\sigma_{\hat{y}_i}}{\sigma_{y_i'}^2 + \sigma_{\hat{y}_i}^2 + \left(\mu_{y_i'} - \mu_{\hat{y}_i}\right)^2} \quad (12)$$

$$Computational\ time\ (t)\ = t_{end} - t_{start} \quad (13)$$

In these equations, y_i', $\hat{y}_i$, $\bar{y}$, and N represent the actual, predicted, mean of the actual value, and the total number of samples present in the dataset respectively. Furthermore, the variance and means of the actual and predicted values are denoted by σ and μ respectively. Additionally, ρ indicates the Pearsons Correlation Coefficient (or PCC) which is a measure of the linear correlation between the two variables. Imperatively, the choice of these performance metrics is to provide a thorough evaluation of the model's ability to accurately predict the life expectation of the distribution transformer. The results of the individual models in Table 1 shows that using DL models achieves more than 97% R^2 score in predicting the RUL. Evaluating the performance in the table, the DMSA CNN-LSTM model recorded the second lowest overall MAE of 0.2520, and MAPE of 1.5985 surpassing all the baseline models except the DMSA CNN-GRU model which is used in this ablation study. A lower MAE value is desirable as it indicates minimum prediction errors. Additionally, it achieved the second lowest RMSE value of about 1.4921 after the GRU which obtained a value of 1.2192. This outcome suggests the proposed model's capability in effectively predicting the RUL of the transformer with less deviations from the actual values. Moreover, a computation time of 595.2953s indicates the proposed model's scalability in efficiently utilizing the available computational resource thus ensuring its real-time application in practical scenarios. Interestingly, aside the MAE, MAPE and computational time, the baseline GRU model performed slightly ahead of the DMSA CNN-LSTM model in terms of the R^2 score, and CCC performance metrics indicating slightly better correlation with the true or actual values of the RUL. Notably, the closer the metrics are to a value of 1, the better the performance of the model. The ensembled CNN-LSTM model achieved the highest RMSE value of 1.8826 which is undesirable since it produces the largest deviations between predicted and actual values. It also has the lowest R^2 score of 0.9774 suggesting the least correlation compared to its counterpart models. Though this result is below the proposed model, it is however achieved with a slightly lower computation time of 425.4025s. The MSA CNN-LSM, LSTM and CNN models obtained an R^2 score of 0.9832, 0.9804, and 0.9806 with comparatively good CCC values of 0.9913, 0.9899, and 0.9896 respectively. Evidently, the CNN model produced the minimum overall computation time of 220.6341s while recording the highest MAE and MAPE values of 1.0236 and 4.4959 respectively. Comparing the results between the DMSA CNN-LSTM and DMSA CNN-GRU architecture, it is noticeable that the former generally performed better on four critical evaluation metrics which includes the RMSE, CCC, R^2 score, and computational time than the latter. In this context, the statistical significance of the proposed model obtained about 11.3%, 0.39%, 0.20% and 7.9% improvement in terms of RMSE, R^2 score, CCC and computational time respectively over the DMSA CNN-GRU model. From this perspective, it is largely important to highlight that leveraging on the DMSA CNN-LSTM architecture resulted in a holistically better performance compared to any of the said DNN models.

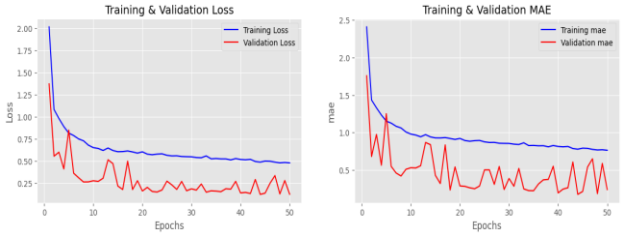

Figure 7. Initial Training and Validation loss and MAE curves of the proposed RUL model on default parameters

The learning outcome of the proposed RUL model was visualized over 50 epochs of the training and validation process as shown in Figure 7. In this figure, both the training and validation loss decreased substantially as the number epochs (iterations) increased. In other words, the training loss decreased smoothly from over 2.00 and eventually a value close to 0.48. The validation loss fluctuated in the early stages of the training process from 1.37 and finally stabilized to approximately 0.13. This graphical outcome shows that the model learned effectively on the training data and partly generalized well on the validation data despite some degree of recorded fluctuations. Similar to the training loss, the training MAE of the DMSA CNN-LSTM model reduced gradually and steadily as number of epochs increased indicating consistent learning for improved predictions. Unlike the training MAE, the validation MAE experienced an unsmooth downward trend with pronounced fluctuations over the epochs. To improve the performance of the proposed model, hyperparameter tuning was applied and the result was compared with the individually tuned baseline DNN models. Table 2 highlights the hyperparameters applied to achieve the improved performance of the DMSA CNN-LSTM model in Table 3.

Table 2. Tuned hyperparameter elements in the RUL prediction model

Hyperparameter	Value /Designation
Loss function r	Huber
Threshold parameter (δ)	9.5
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epoch	50

Table 3. Result of the DMSA CNN-LSTM model against other DNN models on hyperparameters

Model	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
CNN	1.2132	1.7750	0.9799	6.6913	0.9891	222.3042
LSTM	0.4415	1.5199	0.9853	4.6293	0.9924	1144.6522
GRU	0.3054	1.1937	0.9909	2.9675	0.9954	1312.2633
CNN-LSTM	0.3125	1.4913	0.9858	2.8965	0.9929	432.2007
MSA CNN-LSTM	0.4543	1.5459	0.9848	3.7617	0.9922	620.8878
DMSA CNN-GRU	0.2634	1.5994	0.9837	1.7885	0.9919	617.0923
DMSA CNN-LSTM	**0.2300**	**1.4175**	**0.9872**	**1.8205**	**0.9936**	**587.3387**

Analyzing the result of the various hyperparameter tuned models in Table 3 suggests that the DMSA CNN-LSTM model outperforms its counterpart by achieving the lowest error rate of 0.2300 MAE value whiles realizing a robust generalization with high R^2 score and CCC value of 0.9872 and 0.9936 respectively. Furthermore, the MAPE value of 1.8205% (the second lowest value) achieved by the model demonstrates minimum relative error which implies high accuracy. Considering a tradeoff between computational efficiency and overall model performance, although the DMSA CNN-LSTM model achieved a computational time of 587.3387s which is appreciably higher than 222.3042s and 432.2007s recorded from CNN and CNN-LSTM models respectively. However, it is significantly faster than the baseline LSTM and GRU models. Analytical study of the results on both the DMSA CNN-LSTM and DMSA CNN-GRU architecture shows that utilizing the LSTM as the backbone of the proposed model achieved the lowest errors values, and fit better with stronger agreement between predicted and actual RUL values. Moreover, the faster computational time recorded with the proposed model suggests approximately 5% edge over the GRU supported DMSA architecture. This outcome reveals that the simplified gating inherent in the GRU combined with the DMSA architecture made it less suitable for real-time deployment especially given the involvement of much longer sequences within the data.

A chronological approach was introduced to find the most suitable optimizer in the hyperparameter tuning stage. On this ground, the proposed model was trained on the Adam, AdamW, Nadam, Root Mean Squared RMSprop, and Adadelta optimizers with a default learning rate of 0.001 respectively. Table 4 presents the result of these five optimizers after training on 50 epochs and 32 batch sizes.

Table 4. Performance of proposed model on different optimizers

Optimizer	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
RMSprop	0.3651	1.7551	0.9804	2.1594	0.9900	491.0204
Adadelta	1.2513	3.5747	0.9186	7.8170	0.9575	625.4497
AdamW	0.3136	1.7073	0.9814	2.5231	0.9907	726.7736
Nadam	0.2382	1.7201	0.9812	1.9887	0.9905	608.8358
Adam	**0.2520**	**1.4921**	**0.9858**	**1.5985**	**0.9930**	**595.2953**

Examining the outcome in the Table 4, demonstrates that Adam optimizer produced the best overall result achieving the lowest MAE, RMSE, and MAPE of 0.2520, 1.4921 and 1.5985 respectively while maintaining high R^2-score and CCC of 0.9858 and 0.9930 respectively at an efficient computation time of 595.2953s compared to the recorded values by the other optimizers. Numerically, the low error rates (MAE, RMSE, and MAPE suggests that the Adam optimizer minimizes errors between predicted and actual values and explains inherent variance better than its counterparts. Hence, this provided an invaluable insight in the accuracy of predicting the RUL in distribution power transformers. Other variants of Adam specifically AdamW and Nadam performed equally well across the respective performance metrics. Additionally, the RMSprop optimizer being the fastest with a computational time of 491.0204 also recorded values close to AdamW optimizer although weaker than Adam and Nadam. On the contrary, the Adadelta optimizer performed the least across most of the performance metrics excluding the computational time. Hence, this practical analysis reiterates or affirms the Adam optimizer as the most suited optimizer for the proposed RUL prediction model. Furthermore, various learning rates were evaluated with the Adam whiles other parameters maintained accordingly. In Table 5, the default learning rate of 0.001 emerged as the best choice, showing higher accuracy and efficiency compared to other investigated learning rates.

Table 5. Performance of proposed model on different learning rate

Learning rate	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
0.01	0.6348	3.4300	0.9251	4.9323	0.9615	606.4158
0.001 (default)	**0.2520**	**1.4921**	**0.9858**	**1.5985**	**0.9930**	**595.2953**
0.0001	0.5861	1.8088	0.9792	3.2241	0.9894	676.7374
0.00001	0.7213	3.0056	0.9425	4.7534	0.9706	709.3019
0.000001	1.4207	3.7233	0.9117	9.6255	0.9539	546.1124

From the table, learning rates specifically 0.01, 0.00001, and 0.000001 resulted in high error rates (MAE, RMSE, and MAPE values) which practically makes them unsuitable for training the proposed model due to their poor convergence. Conversely, an optimal learning rate of 0.001 provided the lowest MAE, RMSE, and MAPE as highlighted in the table. Additionally, the highest R^2 score and CCC values recorded by this default learning rate suggests the best balance between accurate predictions and error minimization Moreover, the DMSA CNN-LSTM RUL prediction model was trained on four separate epochs (50, 100, 150 and 200 epochs) whiles keeping the other parameters constant. Therefore, their respective performance is presented accordingly in Table 6.

Table 6 Performance of proposed model on different epochs

Epoch	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
50	**0.2520**	**1.4921**	**0.9858**	**1.5985**	**0.9930**	**595.2953**
100	0.5160	1.8586	0.9780	3.5437	0.9887	1157.2039
150	0.3617	1.8075	0.9792	2.1919	0.9895	1638.8936
200	0.3933	1.7574	0.9803	2.5202	0.9900	2589.5685

Detailed evaluation of the result in Table 6 indicates that at 50 epochs, the proposed model performed best which is evident by the low MAE. RMSE, and MAPE together with the highest R^2-score and CCC values. This optimal performance was recorded at a desirable training time of 595.2953s showing computational efficiency. However, increasing the number of epochs beyond 50 significantly increased the computational time to approximately 1157.2039s, 1638.8936s and 2589.5685s for 100, 150 and 200 epochs respectively. Additional analysis of the result suggested that the performance of the proposed model on the error and correlation metrics slightly declined after 50 epochs suggesting that possible overfitting. Thus, considering this analysis, training the proposed model for 50 epochs gave the best overall performance, balancing predictive accuracy with computational efficiency. Various batch sizes were explored in the tuning stage of the proposed model's hyperparameters to enhance training efficiency, obtain better results, and optimize the use of computational resources. Table 7 summarized the performance of the proposed model on five different batch sizes.

 https://doi.org/10.25077/ajeeet.v6i1.210

Table 7 Performance of proposed model on different batch sizes

Batch sizes	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
16	0.4558	1.7182	0.9812	2.6323	0.9904	906.9590
32	**0.2520**	**1.4921**	**0.9858**	**1.5985**	**0.9930**	**595.2953**
64	0.8298	1.8366	0.9785	3.7000	0.9888	385.6871
128	0.3877	1.7661	0.9801	2.1206	0.9899	289.6951
256	0.4696	1.6840	0.9819	2.3060	0.9908	253.0175

Interestingly, as the batch size increases by factor of two, a corresponding decrease in the computation time was observed during the training process. In other words, the proposed model converged faster on larger batch sizes than for smaller batch sizes. Noticeably, the 32-batch size produced the minimum errors values across the MAE, RMSE, and MAPE. It also provided the highest correlations (R^2-score, and CCC) at an optimal computation time of 595.2953s. Evaluating the other respective batch sizes reveal that the batch size of 128 performed relatively well on key error metrics (MAE and MAPE) compared to 16, 64, and 256 batch sizes. Hence, the overall assessment immensely validated the 32-batch size as the most appropriate in providing a balance between the proposed model's accuracy and computational speed or efficiency. Furthermore, the threshold parameter (δ) of the Huber loss function was investigated to identify the most effective value suitable to enhance the performance of the proposed model as shown in Table 8. In view of this, different threshold parameter values were experimented at 0.5 interval starting from the default value 1.0.

Table 8 Performance of proposed model on different threshold parameters

Threshold parameter (δ)	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
1.0	0.2520	1.4921	0.9858	1.5985	0.9930	595.2953
1.5	0.3144	1.5635	0.9844	2.0917	0.9921	733.7020
2.0	0.3543	1.9129	0.9767	3.2228	0.9882	645.6112
2.5	0.3177	1.4491	0.9866	3.0872	0.9932	584.2833
3.0	0.4744	1.6009	0.9837	3.9953	0.9918	689.4466
3.5	0.3457	1.4862	0.9859	3.4082	0.9929	723.1224
4.0	0.3389	1.6119	0.9834	3.6273	0.9916	731.5702
4.5	0.4093	1.4397	0.9868	3.9847	0.9933	705.7537
5.0	0.2784	1.5511	0.9847	2.1038	0.9923	735.6569
5.5	0.3312	1.8358	0.9785	2.8167	0.9892	730.8444
6.0	0.4320	1.6732	0.9822	4.0810	0.9910	720.1319
6.5	0.3693	1.5357	0.9850	2.8123	0.9925	800.6139
7.0	0.3857	1.6543	0.9826	2.0701	0.9912	717.4889
7.5	0.2850	1.6700	0.9822	2.2425	0.9911	710.1202
8.0	0.2275	1.6266	0.9831	1.6323	0.9916	697.4387
8.5	0.2397	1.6091	0.9835	1.7483	0.9917	709.2849
9.0	0.2023	1.4292	0.9836	1.4410	0.9935	695.8626
9.5	**0.2300**	**1.4175**	**0.9872**	**1.8205**	**0.9936**	**587.3387**
10.0	0.2491	1.6699	0.9822	1.7198	0.9911	831.9207

The various threshold parameter values in the table denote how the tuning factors affect the proposed model's training and prediction of outputs. It is imperative to note that at $\delta = 9.5$, the proposed model achieves the best overall performance. This is evident by the lowest MAE, RMSE, and MAPE values of 0.2300, 1.4175, and 1.8205 which implies minimal prediction error, reduced deviations from actual RUL values and reliable overall predictions respectively. More so, the highest R^2-score and CCC values of 0.9872, and 0.9936 depicts significant correlation and prediction which aligns better with the actual or true data values. Achieving this outcome within an optimal computation time 0f 587.3387s indicates that the threshold value of 9.5 is most preferable for achieving improved result during hyperparameter tuning.

The overall performance of the hyperparameter tuned DMSA CNN-LSTM RUL prediction model is summarized in Table 9. Additionally, a comparative assessment of the proposed tuned DMSA CNN-LSTM model was performed to evaluate its performance against the hyperparameter tuned baseline GRU model.

Table 9 Overall performance of the best performing models after hyperparameter tuning

Tuned Model	MAE	RMSE	R^2-score	MAPE (%)	CCC	Compt. time (s)
GRU	0.3054	1.1937	0.9909	2.9675	0.9954	1312.2633
DMSA CNN-GRU	0.2634	1.5994	0.9837	1.7885	0.9919	617.0923
DMSA CNN-LSTM	**0.2300**	**1.4175**	**0.9872**	**1.8205**	**0.9936**	**587.3387**

This result show that the proposed DMSA CNN-LSTM model merged strong predictive accuracy with computational efficiency. That is, it achieved a lower MAE, and MAPE of 0.2300, and 1.8205 respectively. Even though, the RMSE is slightly higher than the tuned GRU model by a margin of 0.2238, the overall error minimization indicates that the proposed model dominated in this aspect. Furthermore, the high R^2-score and CCC value of 0.9872 and 0.9936 obtained by the proposed model suggests a close performance with the tuned baseline GRU model. This implies close similarity in terms of thoroughly identifying the underlining patterns inherent in the data. Remarkably, the proposed model achieved this result approximately 55% faster than the GRU model and 5% faster than the DMSA CNN-GRU model thus indicating its superiority and preferability for reliable prediction of distribution transformer maintenance periods.

The six subplots (see Appendix) in the respective Figures 8(a), 8(b), and 8(c) demonstrates how the proposed model's Huber loss evolved across different hyperparameter settings taking into consideration the learning rate, batch size and epochs. It is observed from the result in Figure 8(a) that applying a 0.001 learning rate produced a much faster and stable convergence of the loss curves in tuning 1-3 compared to those outside the bracket (specifically 0.01 and 0.0001) which resulted in much slower convergence. A batch size of 64 and 128 produced more stable curves within the 10th epoch in tuning 2 and 3 of Figure 8(a). However, considering faster convergence and computational efficiency, the batch size of 32 enabled the DMSA CNN-LSTM model achieve similar result using less memory requirements per iteration. The justification on 50 epochs enabled the proposed model to substantially reduce training time without overfitting compared to training on larger epochs. That is, in real-time RUL prediction, rapid convergence was selected over extended training which yields lower losses. Hence, unlike other tunings in Figure 8(a), 8(b), and 8(c), tuning 1 in Figure 8(a) provided the much-needed balance of speed and accuracy which is more suitable and in events where the model requires additional

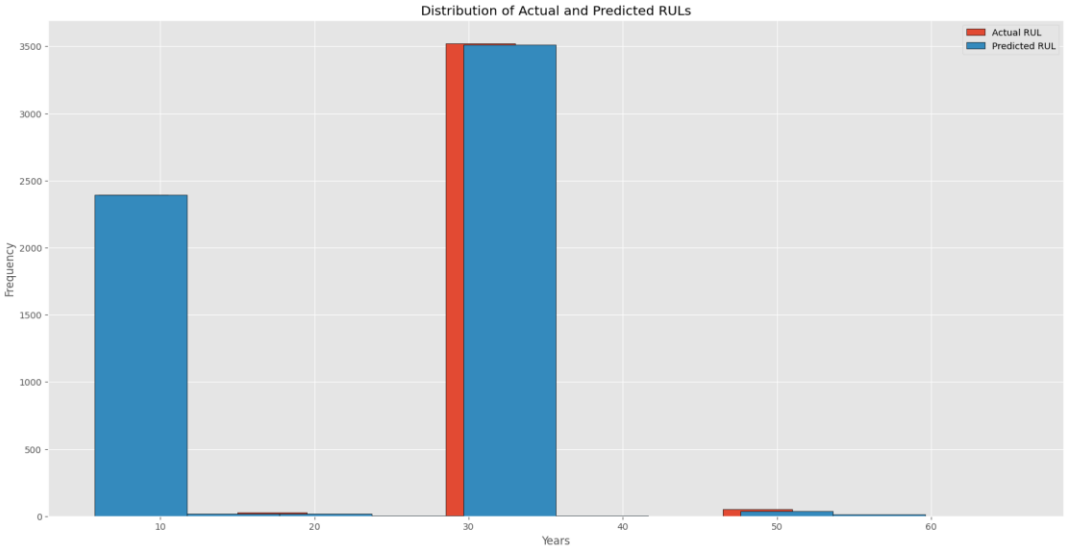

Figure 9. Histogram plot of the distribution of the actual and predicted RUL values

updates. In contrast to the results in Figure 8(a), very strong oscillations, unstable and divergent loss curves were obtained in Figure 8(c) particularly for tuning 1-4 given the high learning rate of 0.01. Furthermore, training the proposed model on lower learning rate typically 0.000001 resulted in very slow learning and severe underfitting as seen in hyperparameter tuning 5-6. This outcome made these sections of tuning parameters inappropriate for the proposed DMSA CNN-LSTM model. Interestingly, the loss curves in the individual subplot of Figure 8(b) showed stable characteristics but at the expense of slower convergence and risk of potential underfitting. Hence, this requires significant tradeoffs which made them unreliable for predicting transformer RULs in real-time sensitive scenarios.

The histogram plot in Figure 9 represents the distribution of the actual and predicted values in the test data. It shows the actual RUL (marked in orange) which denotes the ground-truth life expectancy distribution and the predicted RUL (marked in blue) which also represents the distribution of the predicted life expectancy values by the DMSA CNN-LSTM model. Based on the results in this figure, it is noticeable that the RUL values are continuously spread across three main categorizes (10, 30 and 50 years). However, the dominant distributions fell within 10 years and 30 years as indicated by their level of frequency in the test data. Imperatively, the predicted RUL values matched closely with the actual RUL values showing high degree of overlaps indicative of the strong performance by the proposed model as numerically validated in Table 10.

An outlook of the prediction error was visualized in Figure 10 to buttress the strong predictive performance of the DMSA CNN-LSTM RUL estimation model. In this figure, the error distribution is centered around a value of zero without any significant skewness towards either the extreme positive or negative values. This suggests that the model is unbiased and captures the relevant trend in the data without given preference to a particular error value. Therefore, based on this performance, it is evident that the proposed model does not significantly overpredicts or underpredicts the RUL values in the entire test data. Given the strength of the proposed model, the performance of its future RUL predictions was established for short term predictions only.

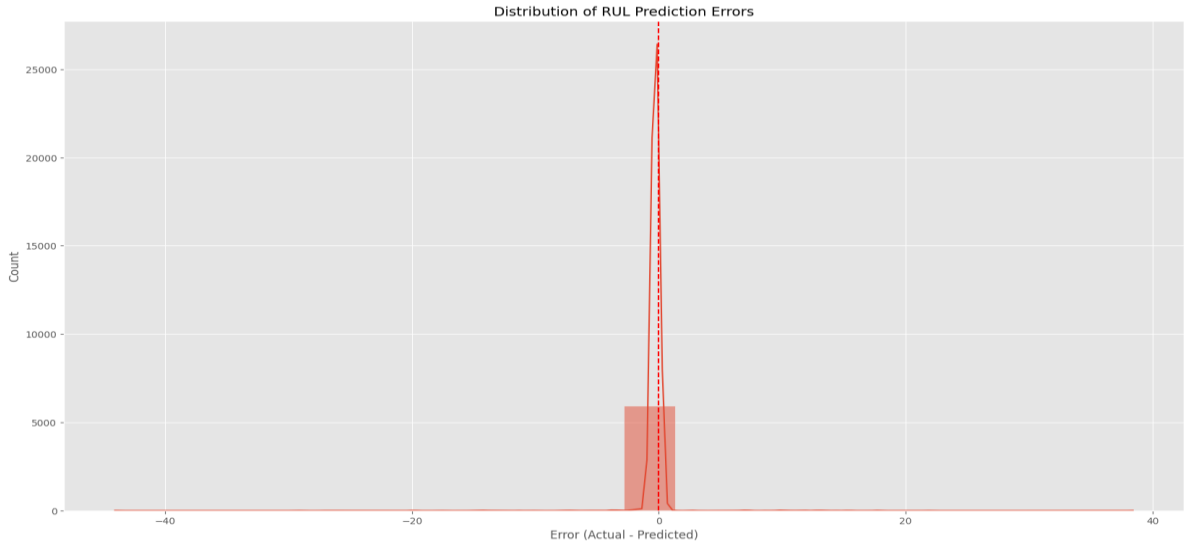

Figure 10. Distribution of RUL prediction errors between actual and predicted values

 https://doi.org/10.25077/ajeeet.v6i1.210

Long-Term RUL Predictions with Prophet

The significance of accurately predicting the RUL of the distribution transformer is an indispensable prognostic approach for reliable asset management. In this regard, a prediction of the RUL for the next 10 years was performed and illustrated in Figure 11 indicating trends from 2020 to 2030 using Prophet forecasting function.

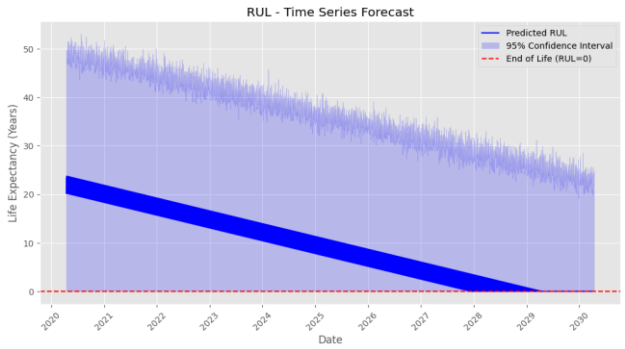

Figure 11. Prediction of the trend of the RUL of the distribution transformer.

The downwards trend in the RUL plot as seen in Figure 11 shows a gradual and continuous deterioration in the life expectancy of the distribution oil immersed transformer over the period on the condition that no maintenance interventions are carried out. This steady linear degradation implies that the deterioration pattern of the transformer starts from about 25 years and predictably declines to zero considering the impacts of all the stressors involved in the transformer's operation. Based on rate of decline, it is expected that the transformer reaches it end of life between 2028-2029. That is, at this point the life expectancy value reaches zero provided the operating conditions remain the same and no unexpected failures occur. Furthermore, in the early years of the forecast, the 95% confidence interval visibly appear large however, its outlook narrowed as the years progressed. This implies that in the early prediction of the RUL value, Prophet exhibited a sizeable uncertainty but this uncertainty reduced as its approached 2028, and 2029. The reduction in the uncertainty means more confidence in long term RUL predictions. The mean RUL values predicted from 2027 to the end of life in 2029 is presented in Table 11.

Table 10 Numerical difference between actual and predicted values in the test and validation data

Test samples	Actual values	Predicted values	Difference
1	32.0	32.139458	-0.139458
2	32.0	32.176796	-0.176796
…	…	…	…
5991	32.0	32.215897	-0.215897
5992	32.0	32.205292	0.205292

From the table, the predicted RUL of the distribution transformer indicates a generally steady degradation with respect to time. Notably, the RUL value of the transformer in the period of 2027 remained over a year until early 2028. This implies that the transformer is still operational despite the continuous deterioration pattern. This steady pattern was followed in the early period from January 2028 to April 2028 nonetheless, after May to December 2028 the RUL fell below 1 year suggesting an accelerated decline in the transformer useful lifetime. Moreover, in the early period of 2029 the RUL reached zero indicating that the transformer has failed, nonoperational and its end of predicted lifespan has been realized. Holistically, the decline in RUL values was initially gradual but became accelerated towards the end of the transformer's life expectancy.

Table 11 Long-term RUL predictions for 2027, 2028, and 2029

Date (m/yr)	RUL Predictions (yrs)	…	Date	RUL predictions (yrs)	…	Date	RUL predictions (yrs)
1/2027	4..26	…	*1/2028*	1.59	…	*1/2029*	0.12
2/2027	3.96	…	*2/2028*	1.36	…	*2/2029*	0.06
3/2027	3.62	…	*3/2028*	1.29	…	*3/2029*	0.04
4/2027	3.60	…	*4/2028*	1.14	…	*4/2029*	0.00
5/2027	3.33	…	*5/2028*	0.86	…	*5/2029*	0.00
6/2027	2.97	…	*6/2028*	0.89	…	*6/2029*	0.00
7/2027	3.00	…	*7/2028*	0.70	…	*7/2029*	0.00
8/2027	2.56	…	*8/2028*	0.59	…	*8/2029*	0.00
9/2027	2.40		*9/2028*	0.56		*9/2029*	0.00
10/2027	2.29		*10/2028*	0.38		*10/2029*	0.00
11/2027	1.90		*11/2028*	0.32		*11/2029*	0.00
12/2027	1.79		*12/2028*	0.25		*12/2029*	0.00

Noticeably, this characteristic trend is typical of distribution transformers given the cumulative stress experienced. Therefore, based on this result, the long-term prediction using Prophet suggests that identified maintenance activities including transformer replacement can be undertaken as the last resort before the year 2028. Thus, the RUL in 2029 is undesirable for reliable power distribution as the transformer is at risk of run-to-failure. Hence, taking into account the prediction made regarding the distribution oil-immersed transformer as evident in Figure 11, it is clear that the RUL of the transformer is expected to be approximated 25 years considering its current operational condition. Interestingly, this outcome aligned with the expected degradation pattern of distribution transformers as referenced in IEEE STD C57.91, which represents a standard that estimates the life expectancy of a transformer to be approximately 20.55 years [44]. Thus, considering the objective of this research, predicting both the short and long-term RUL of distribution transformers provides a comprehensive guide for improved decision making to enhance their effective management.

Comparative Assessment of the RUL Prediction or Estimation Model

The current industrial revolution has led to the development of innovative models aimed at improving the limitations of the existing one's whiles taking notice of the enormous implications that could occur in the event of failing to accurately predict the RUL of the transformer. In reference to this, the performance of the proposed model is benchmarked against current state-of-the-art models to understand its overall contribution and highlight its dominance. Imperatively, the complementary use of Prophet as a statistical tool in this domain of RUL prediction provides the baseline benchmark for better validation and assessment of results [45, 46]. Hence, the performance of the DMSA CNN-LSTM RUL prediction model was benchmarked against other recent works in this domain. In view of this, a comparative assessment and evaluation was presented in Table 12 and Table 13 to thoroughly underline the importance of this work.

Table 12. Comparative assessment of the proposed work with current benchmarks

Author	[47]	[48]	[49]	[50]	[51]	[52]
Algorithm(s)	Spatio-Temporal Complete Graph Convolution Network (STCGCN)	Extreme Gradient Boosting (XGB)	Back propagation neural network (BPNN)	Random Forest (RF)	Adaptive Network-Based Fuzzy Inference System (ANFIS)	Physics-Informed Neural Network (PINN)
Features	DGA data and Oil temperature	Electrical load and temperature	Electrical current and temperature	Ambient temperature, humidity, load, current, and degree of polymerization	Electrical load and temperature	Degree of polymerization (DP)
Purpose	Remaining Useful Lifetime Prediction	Remaining Useful Lifetime Prediction	Remaining Useful Lifetime Prediction	Useful Lifetime Estimation	Estimating Transformer Loss of Life	Remaining Useful Lifetime Prediction
Performance	MAE: 0.0125 RMSE: 0.0158 MAPE: 21.59	$RMSE_{(TOTavg)}$: e_{train}: 1.67±1.14 e_{test}: 6.23±2.17	(50_{trn} : 50_{tst}) & (100_{trn} : 100_{tst}) MSE_{avg}: 4.68 Acc_{avg}:93.92% (75_{trn} : 25_{tst}) MSE_{avg}: 11.03 Acc_{avg}:88.23%	Load factor prediction as inference (with HMM) RMSE:0.0172 and (no HMM) RMSE:0.0197	RMSE:2.946×10^{-10} R^2-score: 0.96 Compt. time: 25.7s	MSE: 26.8 RMSE: 5.18 MAE: 3.52 At (1.5% noise, 110˚C)
Dataset	Dissolve gas and oil temperature	Electrical load and temperature	Electrical current and temperature	Ambient temperature, humidity, load and current	Electrical load and temperature	Degree of polymerization (DP)
Data Availability	Open access	N/A	Simulation	N/A	N/A	N/A
Origin	Multi- sourced	Multi- sourced	Simulation (MATLAB)	Multi- sourced (Iran)	Simulation (MATLAB)	Synthetic generation
Duration	From 01-05-2012 to 13-01-2017. (>60months)	3years and 10months (Nov. 2012-Sep. 2016)	N/A	1 year	1 year	25years
Sampling frequency	1 day for DGA dataset and 1 hour for oil temperature.	1 hour	N/A	0.5hours	1 hour	1 day

Table 13. Comparative assessment of proposed work with current benchmarks (continued from table 12).

Author	[53]	[54]	[55]	This work
Algorithm	Nguyen. Widrow Neural Network	Time Series Decomposition	Decomposition-based Neural Controlled Differential Equation (DNCDE)	Dynamic Multi-Scale Attention CNN-LSTM (DMSA CNN-LSTM) and Prophet
Features	Electrical current, harmonics and temperature	Electrical load and ambient temperature	Vibrations	Electrical, DGA (thermal), Transformer vibrations (mechanical), and Ambient (external temperature)
Purpose	Distribution Transformer Lifetime Prediction	Distribution Transformer Remaining Useful Lifetime Estimation	Forecasting of Power Transformer Remaining Service Life	Fault localization (multi-label classification)
Performance	MAE: 0.024 (75_{trn}:25_{tst})^{H} MAE: 0.023189 (75_{trn}:25_{tst})^{M} MAE: 0.031132 (75_{trn}:25_{tst})^{D} H-Haar wavelet M-Meyer wavelet D-Daubechies wavelet	TSD (load kVA) MAE: 2.94 RMSE: 3.90 TSD (Ambient) MAE: 2.94 RMSE: 3.90 RUL_{err}:2.2062±0.0195(avg)	Accuracy: 92±1 R: 0.9721±0.0125 MAE: 0.0318±0.0111 Running time: 0.0115s	MAE: 0.2300 RMSE: 1.4175 R^2-score: 0.9872 MAPE: 1.8205 CCC: 0.9936 Training time: 587.3387
Dataset	Electrical data	Electrical load and ambient temperature	Vibration data	Fused multi-modal dataset (Electrical, environmental, DGA (thermal), and vibration signatures)
Data Availability	N/A	Private	Private (Hubei DC Company)	Open access
Origin	Multi- sourced (Surabaya city)	Multi- sourced	Single-sourced	Multi- sourced
Duration	12 hours	2014-2016	N/A	Approximately 10 months (electrical and environmental data) and others N/A
Sampling frequency	N/A	N/A	10kHz	1s-15min electrical and environmental data) and others N/A

 https://doi.org/10.25077/ajeeet.v6i1.210

The comparative assessment seen in both Table 12 and Table 13, indicates that the DMSA CNN-LSTM model with Prophet proved advantageous compared to the other baseline models evaluated as benchmarks. The value of thoroughly extracting both the temporal and spatial dependencies via the CNN and LSTM layers provided a huge lead in accurately predicting the RUL both in the short-term and long-term respectively. Thus, employing DMSA mechanism gave the required adaptability to enable the model focus on only the important features that contributes most to the transformer's life expectancy. Therefore, fusing the rich multi-modal dataset ensured an efficient prediction of both the short term and long term of the remaining useful lifetime.

CONCLUSIONS

With reference to the outcome achieved in this work, it is noteworthy that using the proposed DMSA CNN-LSTM model combined with multi-modal fusion is a highly effective data-driven approach for short-term prediction of the RUL in predictive maintenance of distribution power transformers. Importantly, this research analysis highlights the need for combining diverse data sources in predictive maintenance since single-modality techniques lacks efficiency in terms of capturing complex interactions between different operational conditions necessary for accurate RUL prediction. The fusion of multi-modal datasets enabled the DMSA CNN-LSTM to correlate several sensor readings, facilitating accurate short-term predictions with minimal errors. That is, it achieved a performance of 0.2300 MAE, 1.4175 RMSE, 0.9872 R2-score and 1.8205 MAPE. Moreover, the proposed DMSA CNN-LSTM model demonstrated accurate prediction, evidenced by a CCC value of 0.9936. which is preferably closer to unity and achieved a computational time of 587.3387s. Furthermore, the long-term prediction performed with Prophet predicted a RUL of 25 years at 95% confidence interval which agrees with the reference standard in IEEE STD C57.91 as expected. This sets a new benchmark for future research and development in this domain. Therefore, the use of the proposed solution together with Prophet library tool in predictive maintenance supports fast and proactive interventions which facilitates informed decision making, and ensures effective management of transformers in energy sector leading to reliable power distribution.

REFERENCES

[1]. Q.T. Tran, K. Davies, L. Roose, P. Wiriyakitikun, J. Janjampop, E.R. Sanseverino, and G. Zizzo, "A Review of Health Assessment Techniques for Distribution Transformers in Smart Distribution Grids." *Applied Sciences*, vol. 10, pp: 1-20. 2020. doi: 10.3390/app10228115.

[2]. L. Jin, D. Kim, and A. Abu-Siada, "State-of-the-art review on asset management methodologies for oil-immersed power transformers." *Electric Power Systems Research*, vol. 218, pp: 1-18. 2023. doi: 10.1016/j.epsr.2023.109194.

[3]. L. Jin, D. Kim, A. Abu-Siada, and S. Kumar, "Oil-Immersed Power Transformer Condition Monitoring Methodologies: A Review." *Energies*, vol. 15, pp: 1-32. 2022. doi: 10.3390/en15093379.

[4]. L.R. Chandran, G.S.A. Badu, M.G. Nair, and K. Ilango, "A review on status monitoring techniques of transformer and a case study of loss of life calculation of distribution transformers." *Materials Today: Proceedings*, vol. 46, no. 10, pp: 4659-4666. 2021. doi: 10.1016/j.matpr.2020.10.290.

[5]. J.Z. Balanta, S. Rivera, A.A. Romero, and G. Coria, "Planning and Optimizing the Replacement Strategies of Power Transformers: Literature Review." *Energies*, vol. 16, no. 11, pp: 1-16. 2023. doi: 10.3390/en16114448.

[6]. G.S. Rêma, B.D. Bonatto, A.C.S. De Lima, and A.T. De Carvalho, "Emerging Trends in Power Transformer Maintenance and Diagnostics: A Scoping Review of Asset Management Methodologies, Condition Assessment Techniques, and Oil Analysis." *IEEE Access*, vol. 12, pp: 111451-111467. 2024. doi: 10.1109/ACCESS.2024.3441523.

[7]. H. Zhao, J. Chang, and Y. Qu, "Prediction method of residual life of transformer oil-paper insulation based on Wiener random process improved by strong tracking filter." *IET Generation, Transmission and Distribution*, vol. 16, pp: 4007-4016. 2022. doi: 10.1049/gtd2.12581.

[8]. X. Chen, X. Sun, X. Si, and G. Li, "Remaining Useful Life Prediction Based on an Adaptive Inverse Gaussian Degradation Process With Measurement Errors." *IEEE Access*, vol. 18, pp: 3498-3510. 2020. doi: 10.1109/ACCESS.2019.2961951.

[9]. M. Zhang, J. Liu, L. Liao, Q. Chen, P. Qi, and X. Chen, "Method for predicting the remaining life of oil-paper insulation system based on stochastic degradation process." *IET Science, Measurement & Technology*, vol. 13, no. 4. 2019. doi: 10.1049/iet-smt.2018.5041.

[10]. H. Li, Z. Zhang, T. Li, and X. Si, "A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities." *Mechanical Systems and Signal Processing*, vol. 209, pp: 1-32. 2024. doi: 10.1016/j.ymssp.2024.111120.

[11]. J.J.M. Jimenez, S. Schwartz, R. Vigerhoeds, B. Grabot, and M. Salaün, "Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics." *Journal of Manufacturing Systems*, vol. 56, pp: 539-557. 2020. doi: 10.1016/j.jmsy.2020.07.008.

[12]. C. Ferreira, and G. Gonçalves, "Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods." *Journal of Manufacturing Systems*, vol. 63, pp: 550-562. 2022. doi: 10.1016/j.jmsy.2020.05.010.

[13]. F. Wang, A. Liu, C. Qu, R. Xiong, and L. Chen, "A Deep-Learning Method for Remaining Useful Life Prediction of Power Machinery via Dual- Attention Mechanism." *Sensors*, vol. 25, no. 2, pp: 1-19. 2025. doi: 10.3390/s25020497.

[14]. A. Wahid, M. Yahya, J.G: Breslin, and M.A. Intizar, "Self-Attention Transformer-Based Architecture for Remaining Useful Life Estimation of Complex Machines." *Procedia Computer Science*, vol. 217, pp: 456-464. 2023. doi: 10.1016/j.procs.2022.12.241.

[15]. G.S. Chadha, S.R. Bin Shah, A. Schwung, and S.X. Ding, "Shared Temporal Attention Transformer for Remaining Useful Lifetime Estimation." *IEEE Access*, vol. 10, pp: 74244-74258. 2022. doi: 10.1109/ACCESS.2022.3187702.

[16]. X. Xu, X. Li, W. Ming, and M. Chen, "A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction." *Computer &Industrial Engineering*, vol. 169, pp: 1-14. 2022. doi: 10.1016/j.cie.2022.108204.

[17]. Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning." *Neurocomputing*, vol. 452, pp: 48-62. 2021. doi: 10.1016/j.neucom.2021.03.091.

[18]. A. Kara, "Multi-scale deep neural network approach with attention mechanism for remaining useful life estimation." *Computers & Industrial Engineering*, vol. 169, pp: 1-9. 2022. doi: 10.1016/j.cie.2022.108211.

[19]. H. Lv, J. Chen, T. Pan, T. Zhang, Y. Feng, and S. Liu, "Attention mechanism in intelligent fault diagnosis of

machinery: A review of technique and application." *Measurement*, vol. 199, pp: 1-18. 2022. doi: 10.1016/j.measurement.2022.111594.

[20]. Z. Lu, B. Li, C. Fu, J. Wu, L. Xu, S. Jia, and H. Zhang, "Remaining Useful Life Prediction Method Based on Dual-Path Interaction Network with Multiscale Feature Fusion and Dynamic Weight Adaptation." *Actuators*, vol. 13, no. 10, pp: 1-25. 2024. doi: 10.3390/act13100413.

[21]. E.T. Mharakurwa, "In-Service Power Transformer Life Time Prospects: Review and Prospects." *Journal of Electrical and Computer Engineering-Hindawi*, vol. 2022, Art. ID: 9519032, pp: 1-20. 2022. doi: 10.1155/2022/9519032

[22]. E.T. Mharakurwa, G.N. Nyakoe, and A.O. Akumu, "Transformer Remnant Life Estimation and Asset Management model based on Insulation Stress Assessment." *2019 Electrical Insulation Conference (EIC)*, pp: 325-329. 2019. doi: 10.1109/EIC43217.2019.9046526.

[23]. S. Xuntao, H. Ran, O. Mingyu, Y. Lei, K. Qingpai, W. Weiwei and L. Kairan, "A Method for Remaining Useful Life Prediction of Transformer Based on the CNN-LSTM Network." *2022 IEEE 5th International Electrical and Energy Conference (CIEEC)*, pp: 64-69. 2022. doi: 10.1109/CIEEC54735.2022.9845984.

[24]. M.K.K. Alabdullh, M. Joorabian, S.G. Seifossadat, and M. Saniei, "A New Model for Predicting the Remaining Lifetime of Transformer Based on Data Obtained Using Machine Learning." *Journal of Operation and Automation in Power Engineering*, vol. 12, no. 3, pp: 224-232. 2024. doi: 10.22098/JOAPE.2023.11093.1830.

[25]. M.K. Ngwenyama, and M.N. Gitau, "Application of back propagation neural network in complex diagnostics and forecasting loss of life of cellulose paper insulation in oil-immersed transformers." *Scientific Reports*, vol. 14, no. 6080, pp: 1-28. 2024. doi: 10.1038/s41598-024-56598-x.

[26]. A. Mário et.al., "Artificial Neural Networks Application for Top Oil Temperature and Loss of Life Prediction in Power Transformers." *Electric Power Components and Systems*, vol. 50, pp: 546-560. 2022. doi: 10.1080/15325008.2022.2137599.

[27]. M.K.K. Alabdullh, M. Joorabian, S.G. Seifossadat, M. Saniei, and M. Abasi, "A novel method to estimate the lifetime of mineral oil-type power transformers based on the analysis of chemical and physical indicators using artificial intelligence." *Heliyon*, vol. 10, pp: 1-17. 2024. doi: 10.1016/j.heliyon.2024.e40447.

[28]. Z. Liang, Y. Fang, H. Cheng, Y. Sun, B. Li, K. Li, W. Zhao, Z. Sun, and Y. Zhang, "Innovative Transformer Life Assessment Considering Moisture and Oil Circulation." *Energies*, vol. 17, pp: 1-21. 2024. doi: 10.3390/en17020429.

[29]. N. El-Rashidy, Y.A. Sultan, and Z.H. Ali, "Predecting power transformer health index and life expectation based on digital twins and multitask LSTM-GRU model." *Scientific Reports*, vol. 15, no. 1359, pp: 1-29. 2025. doi: 10.1038/s41598-024-83220-x.

[30]. A.-M. Aciu, M.-C. Nitu, C.-I. Nicola, and M. Nicola, "Determining the Remaining Functional Life of Power Transformers Using Multiple Methods of Diagnosing the Operating Condition Based on SVM Classification Algorithms." *Machines*, vol. 12, no. 37, pp: 1-34. 2024. doi: 10.3390/machines12010037.

[31]. S. Putchala, "Distributed Transformer Monitoring." https://www.kaggle.com/datasets/sreshta140/ai-transformer-monitoring. (Accessed Feb 2, 2025).

[32]. R.M.A. Velásquez and J.V.M. Lara, "Root cause analysis improved with machine learning for failure analysis in power transformer," *Engineering Failure Analysis*, vol. 115, pp: 1-20. 2020. doi: 10.1016/j.engfailanal.2020.104684.

[33]. E. Li, L. Wang, and B. Song, "Fault Diagnosis of Power Transformers With Membership Degree." *IEEE Access*, vol. 7, pp: 1-8. 2019. doi: 10.1109/ACCESS.2019.2902299.

[34]. O.E. Gouda, and A.Z. El Dein, "Prediction of Aged Transformer Oil and Paper Insulation." *Electric Power Components and Systems*, vol. 47, no. 4-5, pp: 406-419. 2019. doi: 10.1080/15325008.2019.1604848.

[35]. M. S. Ali, A.H. Abu Bakar, A. Omar, A.S. Abdul Jaafar, and S.H. Mohamed, "Conventional methods of dissolved gas analysis using oil-immersed power transformers for fault diagnosis: A review" *Electric Power Systems Research*, vol. 216, pp: 1-16. 2022. doi: 10.1016/j.epsr.2022.109064.

[36]. A. Wajid, A.U. Rehman, S. Iqbal, M. Pushkarna, S.M. Hussain, H. Kotb, M. Alharbi, and I. Zaitsev, "Comparative Performance Study of Dissolved Gas Analysis (DGA) Methods for Identification of Faults in Power Transformer." *International Journal of Energy Research (Hindawi)*, vol. 2023, Art. ID: 9960743, pp: 1-14. 2023. doi: 10.1155/2023/9960743.

[37]. J. Gielniak, and M. Czerniak, "Investigation of Distribution Transformers Vibrations in Terms of Core and Winding Condition Assessment." *Energies*, vol, 15, no. 13, pp: 1-18. 2022. doi: 10.3390/en15010013.

[38]. M. Hussain, M. O'Nils, J. Lundgren, and S.J. Mousavirad, "A Comprehensive Review on Deep Learning-Based Data Fusion." *IEEE Access*, vol. 12, pp: 180093-180124. 2024. doi: 10.1109/ACCESS.2024.3508271.

[39]. M. Pawłowski, A. Wróblewska, and S. Sysko-Romańczuk, "Effective Techniques for Multimodal Data Fusion: A Comparative Analysis." *Sensors*, vol. 23, pp: 1-16. 2023. doi: 10.3390/s23052381.

[40]. Z. Wang, D. Li, S. Wu, Y. Huang, Z. Yang, and W. Nai, "Huber Loss Function Based on Cockroach Swarm Algorithm with T-Distribution Parameter." *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp: 2490-2493. 2021. doi: 10.1109/IAEAC50856.2021.9390734.

[41]. M. Hu, M. Li, and L. Kong, "Trustworthy regularized huber regression for outlier detection." *Journal of Statistical Computation and Simulation*, vol. 94, no. 5, pp: 1121-1137. 2024. doi: 10.1080/00949655.2023.2281643.

[42]. Y. Feng, and Q. Wu, "A statistical learning assessment of Huber regression." *Journal of Approximation Theory*, vol. 273, pp: 1-21. 2022. doi: 10.1016/j.jat.2021.105660.

[43]. A.A. Imam, A. Abusorrah, M.M.A. Seedahmed, and M. Marzband, "Accurate Forecasting of Global Horizontal Irradiance in Saudi Arabia: A Comparative Study of Machine Learning Predictive Models and Feature Selection Techniques." *Mathematics*, vol. 12, no. 16, pp: 1-24. 2024. doi: 10.3390/math12162600.

[44]. "IEEE Guide for Loading Mineral-Oil-Immersed Transformer and Step-Voltage Regulators: Redlines:" *IEEE Std C57.91-2011*, pp: 1-123. 2012. doi: 10.1109/IEEESTD.2012.6166928.

[45]. L. Shao, Y. Zhang, X. Zheng, X. He, Y. Zheng, and Z. Liu, "A Review of Remaining Useful Life Prediction for Energy Storage Components Based on Stochastic Filtering Methods." *Energies*, vol. 6, no. 3, pp: 1-22. 2023. doi: 10.3390/en16031469.

[46]. V. Alchakov and V. Pisarev, "Application of Machine Learning Methods and Hybrid Modeling for Predicting the Remaining Useful Life of Equipment." *2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, pp:1600-1605. 2025. doi: 10.1109/EDM65517.2025.11096662.

[47]. M. Xing, W. Ding, T. Zhang, and H. Li, "STCGCN: a spatio-temporal complete graph convolution network for remaining useful life prediction of power transformer." *International Journal of Web Information Systems*, vol. 19, no. 2, pp: 102-117. 2023. doi: 10.1108/IJWIS-02-2023-0023.

[48]. J.I. Aizpurua, S.D.J. McArthur, B.G. Stewart, B. Lambert, J.G. Cross, and V.M. Catterson, "Adaptive Power Transformer Lifetime Predictions Through Machine

Learning and Uncertainty Modeling in Nuclear Power Plants." *IEEE Transactions on Industrial Electronics*, vol. 66, no.6, pp: 4726-4737. 2019. doi: 10.1109/TIE.2018.2860532.
[49]. Rosmaliati, N. Elok, R.I. Putri, A. Priyadi, Taufik, M.H. Purnomo, "The Remaining Life of Distribution Transformer Prediction by Using Neuro-Wavelet Method." *Przegląd Elecktrotechniczny*, pp: 114-122. 2023. doi: 10.15199/48.2023.02.19.
[50]. H. Gorginpour, H. Ghimatgar, and M.S. Toulabi, "Lifetime Estimation and Optimal Maintenance Scheduling of Urban Oil-Immersed Distribution-Transformers Considering Weather-Dependent Intelligent Load Model and Unbalanced Loading." *IEEE Transactions on Power Delivery*, vol. 37, no. 5, pp: 4154-4165. 2022. doi: 10.1109/TPWRD.2022.3146154.
[51]. A. Majzoobi, M. Mahoor, and A. Khodaei, "Machine learning applications in estimating transformer loss of life." *2017 IEEE Power and Energy Society General Meeting*, Chichago-USA, pp: 1-5. 2017. doi: 10.1109/PESGM.2017.8274564
[52]. M.A. Saleh et.al., "A PINN-Based Lifetime Predictor for Oil-Impregnated Paper Insulation Degradation in Power Transformers Using Degree of Polymerization." *2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, Gdynia-Poland, pp: 1-6. 2024. doi: 10.1109/CPE-POWERENG60842.2024.10604393.
[53]. Rosmaliati, N.E. Setiawati, M.H. Purnomo, and A. Priyadi, "Nguyen-Widrow Neural Network for Distribution Transformer Lifetime Prediction." *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, Surabaya-Indonesia, pp: 305-310. 2018. doi: 10.1109/CENIM.2018.8710815.
[54]. H.M. Usman, R. ElShatshat, and A.H. El-Hag, "Distribution Transformer Remaining Useful Life Estimation Considering Electric Vehicle Penetration." *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp: 3130-3141. 2023. doi: 10.1109/TPWRD.2023.3265671.
[55]. L. Lei, Y. He, and Z. Xing, "Neural-controlled differential equation with decomposition for online forecasting of power transformer remaining service life." *Electric Power Systems Research*, vol. 242, pp: 1-11. 2025. doi: 10.1016/j.epsr.2025.111466.

NOMENCLATURE

2FAL	2.Furaldehyde
C_2H_2	Acetylene
Adam	Adaptive Moment Estimation
Nadam	Adaptive Moment Estimation with nesterov momentum
AdamW	Adaptive Moment Estimation with weight decay
ANFIS	Adaptive Network-Based Fuzzy Inference System
avg	average
BPNN	Back Propagation Neural Network
b	Bias term
CO_2	Carbon dioxide
CO	Carbon monoxide
R2-score	Coefficient of determination
Compt.	Computation
CCC	Concordance Correlation Coefficient
CNN	Convolution Neural Network
R	Correlation coefficient
D	Daubechies wavelet
DNCDE	Decomposition-based Neural Controlled Differential Equations
DL	Deep Learning
DP	Degree of Polymerization
DGA	Dissolved Gases Analysis
DMSA	Dynamic Multi-Scale Attention
C_2H_6	Ethane
C_2H_4	Ethylene
XGB	Extreme Gradient Boosting
GRU	Gated Recurrent Unit
H	Haar wavelet
h	Hidden state
H_2	Hydrogen
tanh	Hyperbolic tangent
KNN	K-Nearest Neighbour
LSTM	Long Short-Term Memory
LOL	Loss of Life
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MedAE	Median Absolute Error
CH_4	Methane
M	Meyer wavelet
MLPNN	Multi-Layer Perceptron Neural Network
MSA	Multi-Scale Attention
1D	One-Dimensional
PD	Partial Discharge
PCC	Pearson Correlation Coefficient
PINN	Physics-Informed Neural Network
PCA	Principal Component Analysis
RF	Random Forest
ReLU	Rectified Linear Unit
RUL	Remaining Useful Lifetime Estimation
RMSE	Root Mean Square Error
RMSprop	Root Mean Squared Propagation
SHAP	Shapley Additive Explanations
STCGCN	Spatio-Temporal Complete Graph Convolution Network
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
tst	testing
TSD	Time Series Decomposition
TOT	Top Oil Temperature
trn	training
v^T	Weight vector
WT	Wavelet Transform
W	Weights

APPENDICES

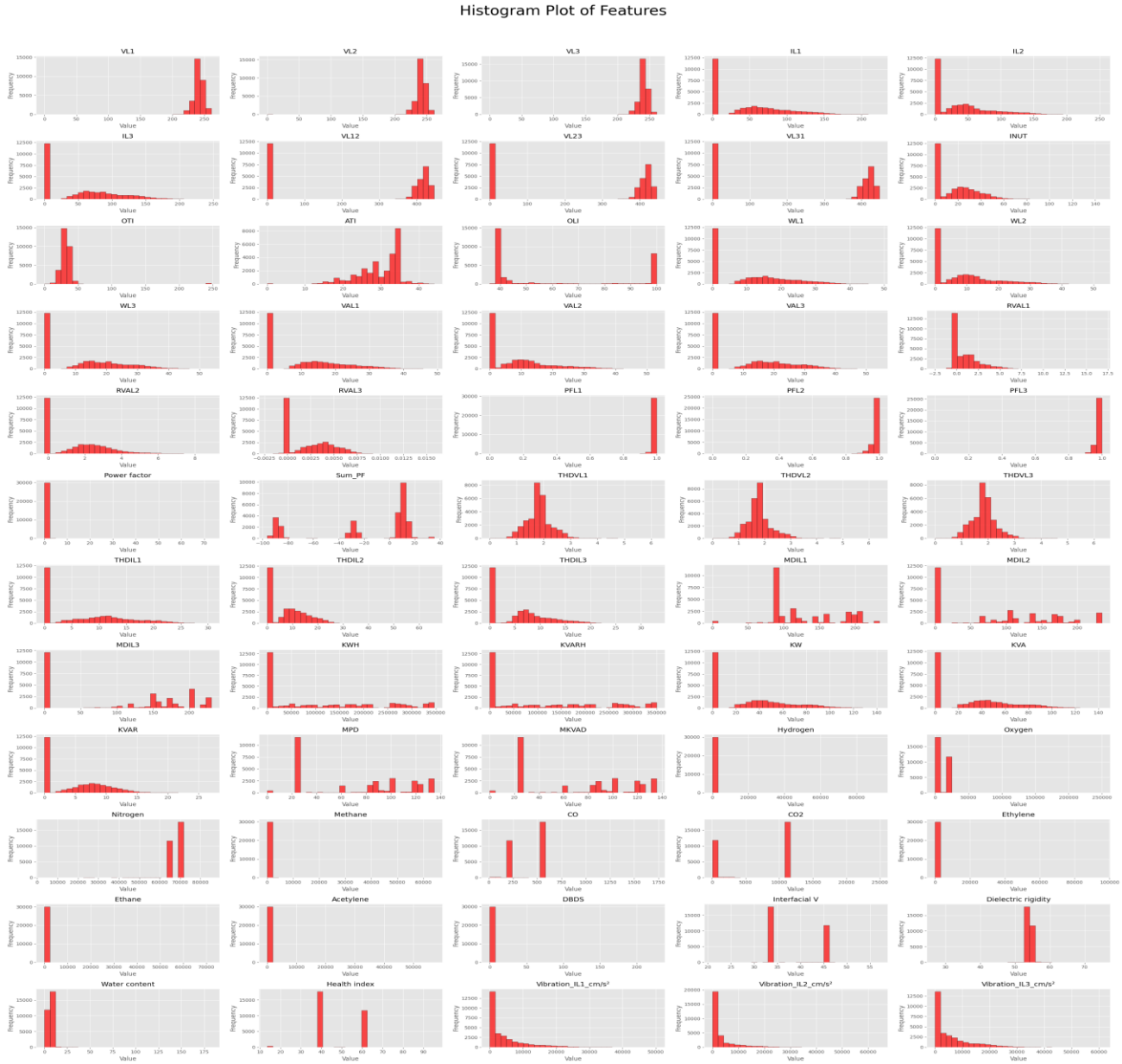

Figure 5. Distribution analysis of the features in Histogram plot

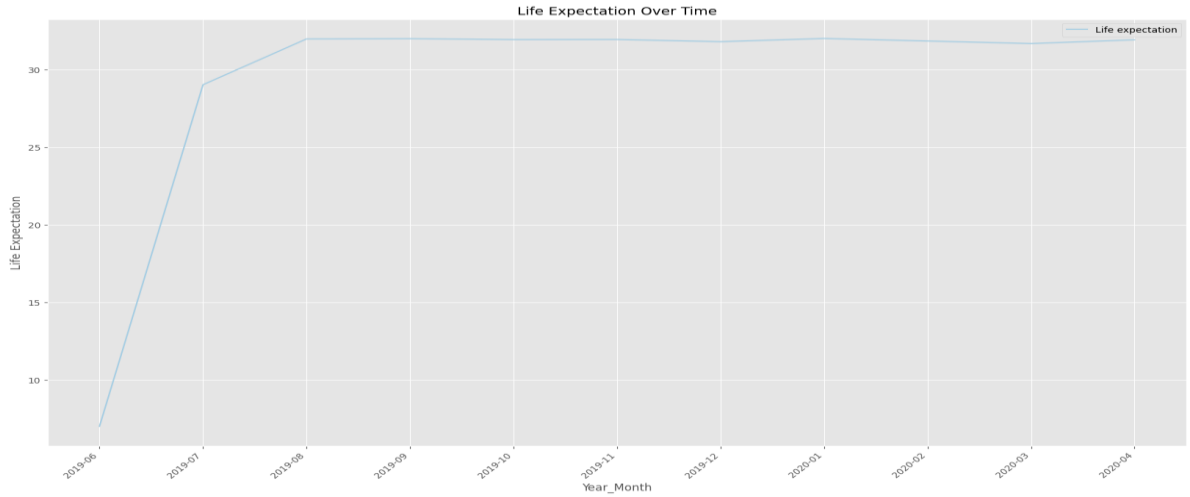

Figure 6. Plot of the RUL of the transformer over time (Year_Month)

 https://doi.org/10.25077/ajeeet.v6i1.210

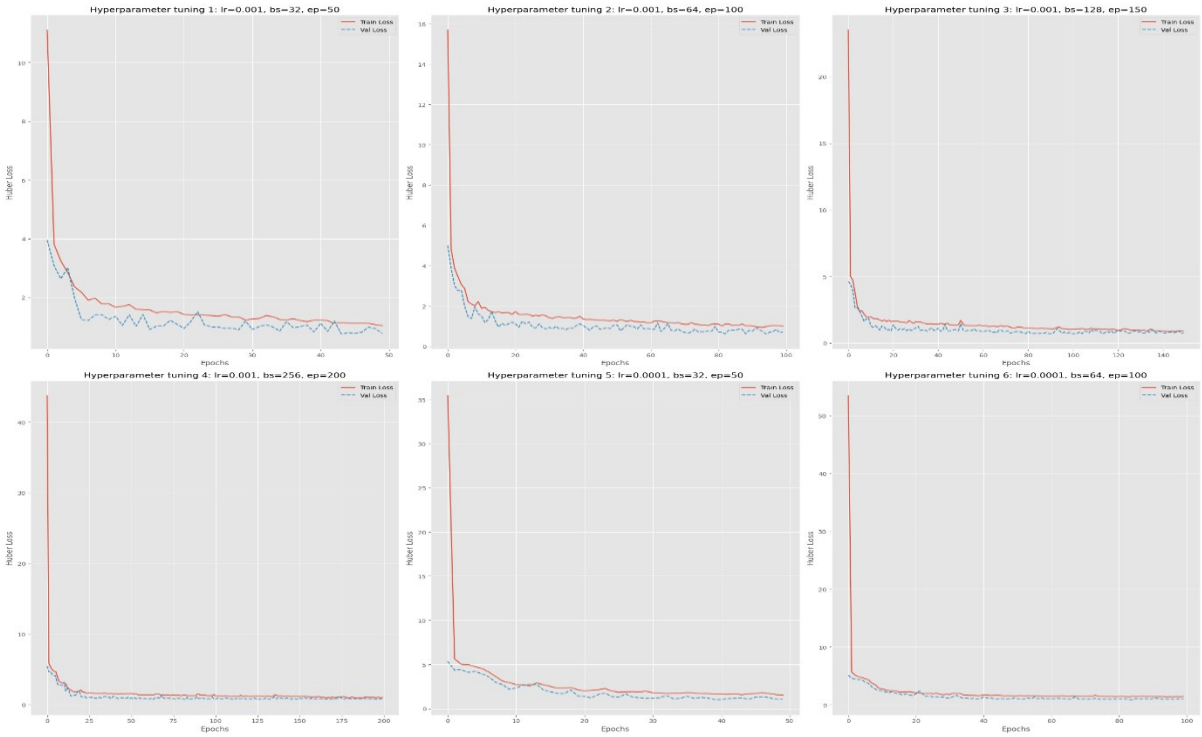

Figure 8(a). Training and validation loss curves of the proposed RUL model on respective hyperparameters

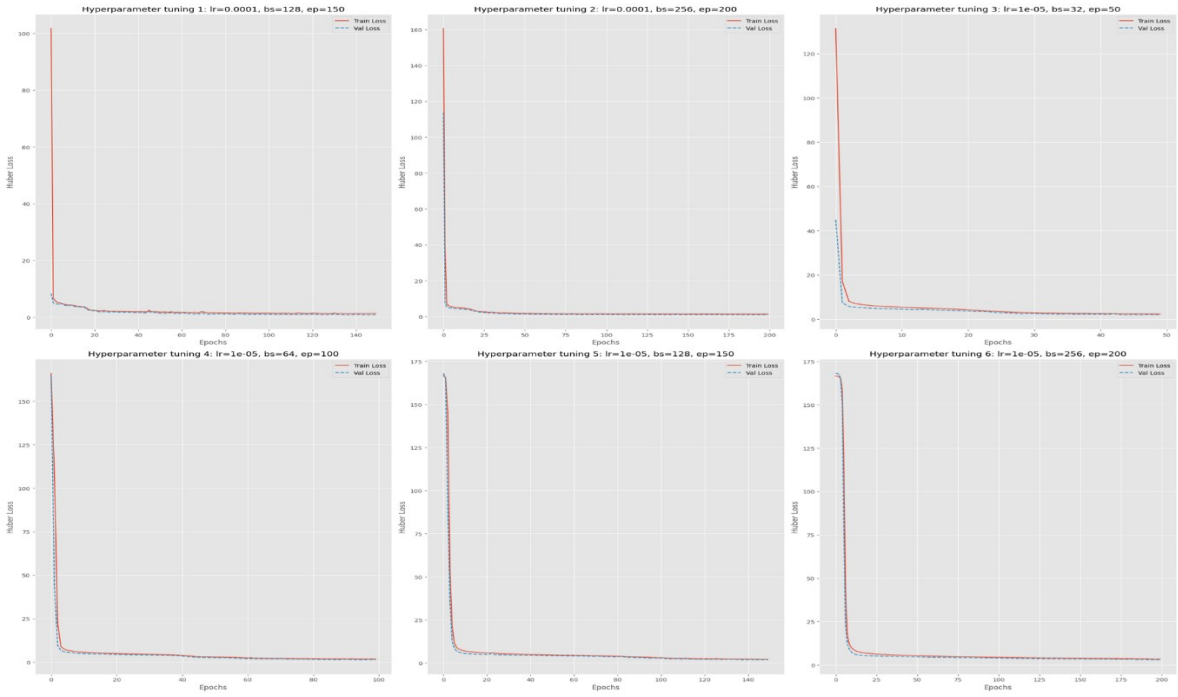

Figure 8(b). Training and validation loss curves of the proposed RUL model on respective hyperparameters

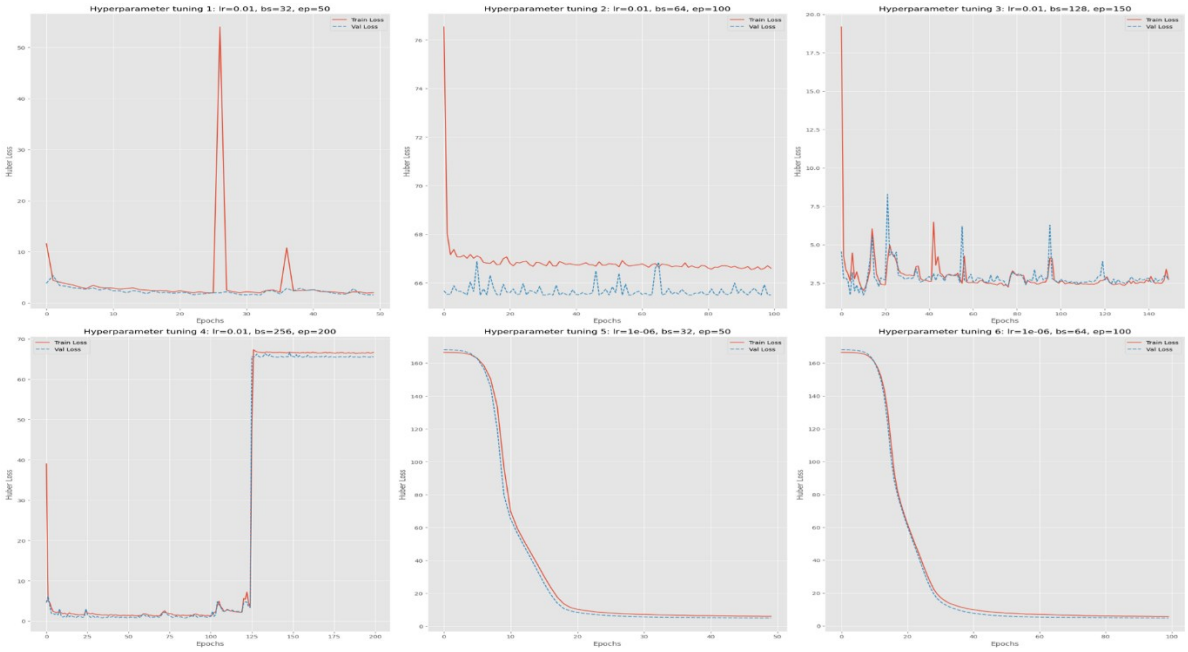

Figure 8(c). Training and validation loss curves of the proposed RUL model on respective hyperparameters

Table A1. Mapping data modalities to features

Data modality	**Measure Quantity**	**Features**
Electrical	Phase and phase-phase voltages	VL1, VL2, VL3, VL12, VL23, VL31.
	Phase and neutral currents	IL1, IL2, IL3, INUT.
	Active and apparent power	KW; WL1, WL2, WL3, KW
	Reactive power and energy	KVAR; KVARH, VAL1, VAL2, VAL3, RVAL1, RVAL2, RVAL3, KWH
	Power factor	Power factor, Sum_PF
	Voltage and current harmonic distortion	THDVL1, THDVL2, THDVL3, THDIL1, THDIL2, THDIL3
Thermal (Oil DGA)	Dissolved gas concentration in transformer oil	H_2, O_2, N_2, CH_4, CO, CO_2, C_2H_4 (ethylene), C_2H_6 (ethane), C_2H_2 (acetylene)
	Oil temperature, level, insulation and degradation.	OTI, OLI, Dielectric Rigidity, Interfacial Voltage, Water Content
Environmental	Ambient temperature	ATI
Mechanical	Winding vibration acceleration derived from phase currents.	Vibration_IL1_cm/s², Vibration_IL2_cm/s², Vibration_IL3_cm/s²

 https://doi.org/10.25077/ajeeet.v6i1.210