ENHANCING POWER TRANSFORMER INSULATION ASSESSMENT USING GRADIENT BOOSTED DECISION TREES AND 2-FAL ANALYSIS

Huy Vu Tran, Kim Anh Nguyen*, Dinh Duong Le

The University of Danang – University of Science and Technology, Vietnam

*Corresponding author: nkanh@dut.udn.vn

(Received: May 12, 2025; Revised: June 12, 2025; Accepted: June 20, 2025)

DOI: 10.31130/ud-jst.2025.23(9B).503E

Abstract - This study presents an empirical model for predicting the degree of polymerization (DP) and estimating the remaining useful life (RUL) of transformer insulation based on 2-furfuraldehyde (2-FAL) concentrations. Using 125 samples collected from multiple credible sources, a Gradient Boosted Decision Trees (GBDT) model was developed to improve prediction accuracy. Compared with existing models such as Chengdong, Burton, and Heisler, the proposed GBDT achieved superior performance, with lower Mean Absolute Error (MAE = 38.65) and Root Mean Squared Error (RMSE = 63.04). Graphical analyses confirmed a strong agreement between predicted and measured DP values, effectively capturing the nonlinear relationship between 2-FAL and DP. Sensitivity analysis showed that the model responds notably to small variations in 2-FAL at early degradation stages. The results enhance transformer diagnostics and enable proactive asset management through accurate, non-invasive, and data-driven monitoring of insulation aging.

Key words - Power transformers; Dissolved Gas Analysis; insulation aging; health index; 2-furfuraldehyde; Gradient Boosted Decision Trees; remaining useful life

1. Introduction

The dependable performance of power transformers is vital for ensuring the stability and efficiency of electrical power networks, making the evaluation of insulation condition a fundamental aspect of asset management in power systems [1]. Transformer insulation, primarily made of cellulose-based materials, undergoes degradation over time due to thermal, oxidative, and mechanical stresses, resulting in a decline in the degree of polymerization (DP) [2]. The DP serves as a critical measure of insulation degradation, with lower values indicating more advanced deterioration [3]. Accurate DP estimation is essential for predicting the remaining useful life (RUL), supporting proactive maintenance, and avoiding unexpected failures that may result in substantial financial losses and safety risks. However, traditional DP assessment methods, such as chemical analysis of oil samples, are often invasive and time-intensive, underscoring the necessity for non-invasive, data-driven solutions.

Approaches like regression analysis and empirical modeling have been used to link DP with measurable indicators, particularly the concentration of 2-FAL in transformer oil, a dependable marker of cellulose breakdown. Previous models, such as those developed by Chengdong *et al.* [4], Heisler *et al.* [5], De Pablo [6, 7], Dong *et al.* [8], Stebbins *et al.* [9], Pahlavanpour *et al.* [10], Burton [11], Vauchex *et al.* [12], and Lin *et al.* [13] have applied various regression and neural network techniques, but their predictive accuracy varies, with some showing significant inaccuracies due to oversimplified assumptions or poor handling of non-linear degradation patterns.

This research introduces an innovative empirical model to improve DP estimation and RUL prediction, utilizing a robust dataset of 125 samples drawn from four significant studies [14–17]. The model adopts a logarithmic regression technique to effectively capture the exponential characteristics of insulation degradation, overcoming shortcomings of prior methods. By incorporating advanced regression methods and validating the model using established metrics, R-square (R^2), mean absolute error (MAE), and root mean square error (RMSE), this study seeks to deliver a reliable tool for assessing transformer condition. Graphical and numerical evaluations across a broad spectrum of 2-FAL concentrations (0–40 ppm) will further assess the model's effectiveness compared to existing methods. The results are anticipated to enhance maintenance strategies, streamline asset management, and bolster the reliability of power transformers in contemporary electrical grids.

2. Methodology

2.1. Furan compound

Furan compounds found in transformer oils have been widely researched as markers of cellulose insulation degradation, with theoretical models established to estimate the degree of polymerization (DP) of paper insulation based on their concentrations. When transformer cellulose paper is subjected to temperatures of 100°C or higher, various degradation processes produce byproducts that dissolve into the oil, allowing furan concentration analysis to evaluate insulation condition. Recent research highlights that furan compounds are primarily generated from electrical discharges within the cellulose insulation, though in lesser amounts compared to those produced by thermal stress. Higher quantities of furanic compounds are formed when insulation undergoes substantial thermal stress, such as temperatures exceeding 120°C. The formation rate of these compounds is influenced by factors such as the oil's water content and oxygen levels. Typically, five primary furanic compounds are identified: 2-furfuraldehyde (2-FAL), 2-acetylfuran (2-ACF), 5-methyl-2-furaldehyde (5-MEF), 2-furfuryl alcohol (2-FOL), and 5-hydroxymethyl-2-furaldehyde (5-H2F) as Figure 1. Of these, 2-FAL stands out due to its greater production rate and stability in the transformer environment.

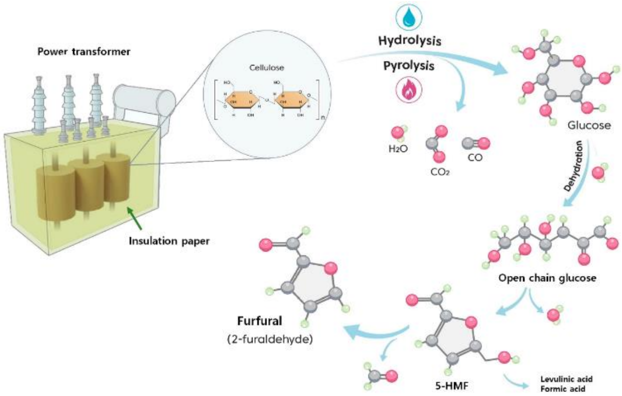

Figure 1. *Schematic diagram of a power transformer (oil/paper) and the degradation of the insulation paper (cellulose) to furfural through 5-HMF [18]*

Nevertheless, the stability of furanic compounds is a crucial consideration for their effectiveness as diagnostic indicators. Compounds that maintain stability over extended periods offer more reliable insights into insulation aging, whereas those that are unstable under fluctuating conditions, such as prolonged exposure or extreme temperatures, may yield inaccurate conclusions and are less suitable for diagnostic applications.

2.2. The degree of polymerization

Cellulose is the main component of transformer insulation systems, and its condition is evaluated using the degree of polymerization (DP). This metric measures changes in the molecular structure of cellulose, which is composed of extended chains of glucose units and other associated polymers. As cellulose undergoes degradation due to aging, the DP indicates the average number of glucose rings in the polymer chains, offering a window into the paper's condition. The DP is a vital indicator that affects the mechanical stability of the insulation system during manufacturing.

Table 1. *Condition of the transfomer according to the 2-FAL content*

Mechanical Strength	DP	2-FAL (ppm)	Extent of degradation
Highest (new paper)	1200–700	0–0.1	Healthy
Regular (normal operation)	700–450	0.1–1.0	Moderate deterioration
Endangered (obsolete)	450–250	1–10	Extensive deterioration
Completely lost the strength (end use)	<250	>10	End-of-life criteria

Freshly manufactured Kraft paper typically displays an average DP ranging from 700 to 1200, with a tensile strength of approximately 1200. As the DP drops from 450 to 700, the paper undergoes moderate deterioration, yet its strength remains relatively stable. However, when the DP falls within the range of 450 to 250, the paper undergoes significant mechanical weakening and turns dark brown, indicating severe degradation. Once the DP decreases below 250, the paper loses all mechanical strength, becomes fully decomposed, and marks the end of the transformer's operational life [19–21] as Table 1.

2.3. The relationship between 2-FAL and DP

Evaluating the condition of transformer insulation paper through direct sampling during operation or downtime is often challenging, prompting the use of furan derivatives, such as furans, to monitor aging driven by thermal or electric stress. Measuring furans in oil is relatively simple, although the underlying aging mechanisms are intricate due to the interaction of multiple degradation pathways. Research indicates that electrical discharges within cellulose generate furans, though in smaller amounts compared to those produced under thermal stress. When insulation is subjected to significant thermal stress, such as temperatures above 120 °C [22], furan production increases substantially, with formation rates affected by factors like the oil's water and oxygen content. Among the furans, 2-FAL is the most critical indicator of the degree of polymerization (DP) of cellulose insulation, reflecting the severity of degradation. The relationship between 2-FAL concentration and DP underscores the strength and nature of their connection, often measured by the correlation coefficient. This coefficient, ranging from -1 to +1, quantifies the extent of association between these variables, offering valuable insights into the insulation's state. Various studies have developed models to characterize the link between 2-FAL levels and DP, confirming a robust correlation that enhances understanding of cellulose aging in transformer oil. This correlation is a key foundation for assessing insulation health and estimating transformer lifespan [23].

2.4. Current approaches for DP prediction

Directly sampling paper insulation from a transformer poses significant challenges, particularly when the transformer remains operational. Improper execution of this method can result in severe damage or failure. As a result, an indirect approach involving the analysis of furanic compounds in the insulating oil has been developed. These compounds are byproducts of the aging process of paper insulation. Studies indicate that at lower temperatures, the primary byproducts include carbon oxides and moisture, whereas at intermediate temperatures, furanic compounds become predominant. However, at elevated temperatures, these compounds tend to become unstable. Over the years, multiple studies have sought to establish a correlation between furan concentrations and DP values, leading to the creation of mathematical models for predicting DP. These efforts have facilitated the use of 2-FAL concentrations to assess transformer insulation aging. Common models for this purpose include exponential decay models, polynomial regression models, and logarithmic models, which are widely utilized for insulation assessment.

Several mathematical models have been developed to estimate the DP using 2-FAL concentrations, each with its own strengths and limitations. The model by Burton [11] uses the equation:

$$\mathrm{DP} = \frac{2.5 - \lg(\text{2-FAL})}{0.005} \tag{1}$$

It utilizes logarithms to better capture the nonlinear relationship between 2-FAL and DP, but it fails to adequately reflect the strong nonlinearity of the aging process. Chendong *et al.* [4] proposed:

$$\mathrm{DP} = \frac{1.51 - \lg(\text{2-FAL})}{0.0035} \tag{2}$$

which also uses logarithms to describe the nonlinear relationship and includes a coefficient of 0.0035 in the denominator to adjust for sensitivity to changes in 2-FAL. However, developed in 1991, it may not align with modern standards and materials. De Pablo [6] introduced:

$$\mathrm{DP} = \frac{1850}{\left(\text{2-FAL} + 2.3\right)} \tag{3}$$

where the coefficient 2.3 in the denominator helps adjust for low DP values when 2-FAL levels are small, making it more stable at low 2-FAL concentrations. Nonetheless, it does not use logarithms, so it may not accurately depict the rapid DP decrease in early aging stages. De Pablo [7] suggested:

$$\mathrm{DP} = \frac{7100}{\left(8.84 + \text{2-FAL}\right)} \tag{4}$$

with the coefficient 8.84 making the formula more sensitive to large 2-FAL values and suitable for predicting DP in later aging stages. However, logarithms are not utilized, and the coefficient 7100 may overestimate DP in some cases compared to actual values.

The model by Vauchex *et al.* [12] is given by:

$$\mathrm{DP} = \frac{2.6 - \lg(\text{2-FAL})}{0.0049} \tag{5}$$

It leverages logarithms to capture the nonlinear relationship between 2-FAL and DP, with a coefficient of 0.0049 in the denominator allowing for sensitivity adjustments that can be tailored to specific operating conditions. However, developed in 2002, it may not comply with current standards and materials. Pahlavanpour *et al.* [10] proposed:

$$\mathrm{DP} = \frac{800}{\left[0.186(\text{2-FAL}) + 1\right]} \tag{6}$$

which shares similarities with De Pablo's approach and thus inherits similar strengths and weaknesses. Heisler et al. Heisler *et al.* [5] presented:

$$\mathrm{DP} = 325\left(\frac{19}{3} - \lg\left(\text{2-FAL}\right)\right) \tag{7}$$

where the coefficient 19/3 helps adjust for small 2-FAL values, but the coefficient 325 may not be suitable for all transformer types, leading to potential inaccuracies in DP prediction. Stebbins *et al.* [9] introduced:

$$\mathrm{DP} = \frac{\lg(\text{2-FAL} \times 0.88) - 4.51}{-0.0035} \tag{8}$$

incorporating a coefficient of 0.0035 in the denominator to adjust sensitivity to changes in DP, making the formula more adaptable. However, it may not be accurate for physically reasonable DP values in certain cases.

Dong *et al.* [8] developed:

$$\mathrm{DP} = 402.47 - 220.87\lg\left(\text{2-FAL}\right) \tag{9}$$

with coefficients 220.87 and 402.47 that can be adjusted for specific conditions, but it may yield large errors at high 2-FAL values. Lastly, Lin *et al.* [13] proposed:

$$\mathrm{DP} = 405.25 - 347.22\lg\left(\text{2-FAL}\right) \tag{10}$$

with coefficients 405.25 and 347.22 that can be tailored to specific conditions, sharing similarities with Dong's model and its associated limitations. These studies highlight the importance of indirect methods in evaluating the aging condition of transformer insulation, contributing to enhanced reliability and safety in transformer maintenance without the need for direct sampling.

The models discussed above aim to establish a mathematical relationship between DP decline and 2-FAL concentrations to facilitate transformer condition assessment. They include a variety of approaches such as exponential decay, polynomial regression, and logarithmic models, reflecting the diverse strategies used to predict insulation aging over time. These indirect methods are crucial for improving the reliability and safety of transformer maintenance by eliminating the need for invasive sampling procedures.

2.5. Remaining useful life of transformer

The remaining useful lifetime (RUL) of a power transformer can be computed by subtracting the elapsed service time from its expected total operational lifespan. Typically, a transformer's operational lifespan is estimated to be around 35 years. To evaluate this lifetime, the elapsed life (EL) is estimated based on the insulation paper's degree of polymerization (DP), assuming DP values degrade from an initial value of 1200 (new condition) down to 200 (end-of-life condition).

One established method by Pradhan calculates EL using a logarithmic formula, expressed as:

$$EL = 20.5\ln\left(\frac{1200}{DP}\right)\ \text{(years)} \tag{11}$$

Additionally, Kanumuri introduced a formula to determine the remaining service life percentage (%RSL), given by:

$$\%RSL = \frac{\lg\left(DP\right) - 2.903}{-0.006021} \tag{12}$$

Combining these formulas allows for precise assessment of transformer health, enabling accurate predictions of their remaining operational lifespan $RUL = 35 - EL$ (years) and facilitating proactive maintenance decisions to enhance reliability and performance in power systems [17].

2.6. Gradient Boosted Decision Trees

Gradient Boosted Decision Trees (GBDT) represents a robust ensemble learning technique that integrates multiple decision trees to enhance predictive accuracy. This method constructs decision trees sequentially, with each subsequent tree designed to correct errors made by its predecessors. This iterative refinement process gradually improves the model's forecasting precision [24].

The core principle of GBDT involves minimizing a loss function, with the initial prediction typically set as the mean of the target variable. In each iteration, a new decision tree is trained to predict the negative gradient of the loss function relative to the current prediction. This approach identifies the minimum loss function value. The predictions of each new tree are scaled by a learning rate and incorporated into the ensemble's output, effectively reducing errors. Ultimately, the GBDT algorithm can be outlined as follows [25].

- Step 1: Initialize the model begin with an initial model prediction using a constant value $F_0(x)$, determined by minimizing the total loss across n samples:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma) \tag{13}$$

where, $L(y_i, \gamma)$ is the loss function; y_i is the actual value for the ith sample; and γ is the initial constant prediction.

- Step 2: Iterative optimization. For each iteration m ($m = 1, \ldots, M$),

○ compute the loss function's negative gradient of ith sample at mth iteration:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \tag{14}$$

○ fit a decision tree $h_m(x)$ to the negative gradients r_{im} ; and compute the optimal step size r_{im} :

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{15}$$

○ update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{16}$$

- Step 3: Final prediction computes the final prediction as the sum of outputs from all trees:

$$F_M(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{17}$$

This study utilized the GBDT model, a widely recognized gradient boosting framework, to analyze data. The number of trees, learning rate, and tree depth were optimized using cross-validation to boost the model's performance on the DGA dataset. GBDT's ability to unravel complex nonlinear relationships in the data, yielded superior prognostic accuracy compared to traditional methods.

2.7. Performance evaluation

Researchers have devised various algorithms and metrics to assess the performance of models predicting transformer insulation health. Commonly utilized metrics include mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). MSE quantifies the average squared difference between predicted and actual values, while RMSE, the square root of MSE, provides better interpretability. MAE determines the average absolute difference between predicted and actual values [26]. These metrics are expressed through Equations (18–20):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{18}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{19}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{y}_i - y_i \right| \tag{20}$$

In these equations, N represents the number of samples, y_i is the actual value, and $\hat{y}_i$ is the predicted value for the *i*-th sample. Together, these metrics offer a robust framework for evaluating the accuracy and reliability of models predicting insulation degradation.

3. Proposed approach

3.1. Data collection and preprocessing

This study utilizes a comprehensive dataset of 125 samples gathered from diverse sources: [14] contributed 74 samples through a methanol-based DP estimation approach, [27] provided 9 samples using feedforward backpropagation neural networks, [16] supplied 32 samples focused on DP prediction models, and [17] added 10 samples for DP-based transformer lifespan evaluation. The relationship between DP and 2-FAL follows a clear trend, where an increase in FA corresponds to a decrease in DP, as depicted in Figure 2.

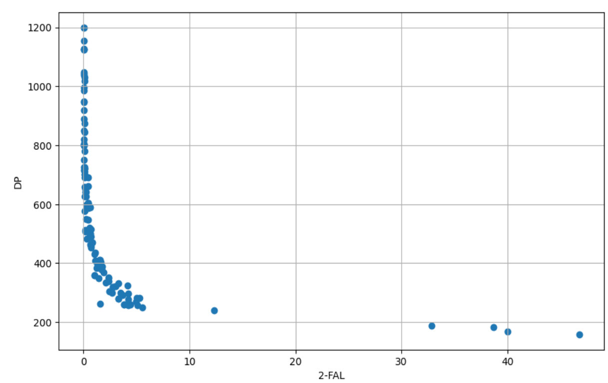

Figure 2. *Correlation between 2-FAL and DP*

3.2. Development of the GBDT model

In the development of the Gradient Boosted Decision Tree (GBDT) model, the process begins with model selection and development. The GBDT algorithm is selected due to its proven capability to effectively handle non-linear data and provide outstanding classification performance. To optimize its accuracy and efficiency, key parameters such as the number of trees, tree depth, and learning rate are carefully adjusted. Following this, the model undergoes training and evaluation using pre-processed data. The performance of the trained GBDT model is evaluated through established metrics, including MSE, RMSE, and MAE to ensure its reliability. Once training is complete, the fully developed DP prediction model is stored for practical deployment, facilitating seamless integration into real-world applications while ensuring usability and scalability of the system.

4. Results and discussion

4.1. Comparative results

This section evaluates the effectiveness of a novel GBDT model in estimating the DP based on 2-FAL concentrations for predicting the RUL of transformer insulation. The model's performance is compared against established models by [4–13]. The evaluation is conducted using three essential metrics (MAE, RMSE, and R^2) as detailed in Table 2 and illustrated in Figure 3.

Table 2. *Performance comparison of models for DP estimation in transformer insulation using MAE and RMSE, metrics*

	Stebbins	De Pablo 2	Pahlavanpour	De Pablo 1	Vaurchex	Dong	Chaouhui	Burton	Heisler	Chengdong	GBDT Model
MAE	849.29	205.52	173.05	105.66	92.81	86.06	86.36	80.75	75.80	60.26	38.65
RMSE	853.80	230.38	193.86	135.12	114.53	125.03	113.75	108.61	105.44	90.75	63.04

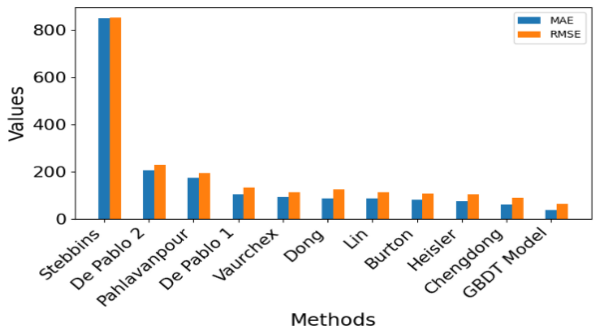

Figure 3. *Comparison of MAE, and RMSE for different methods*

The GBDT model demonstrates superior performance with an MAE of 38.65 and an RMSE of 63.04, outperforming all other models evaluated. Among the established models, three authors achieve results that are nearest to those of the proposed GBDT model. Chendong *et al.* [4] take the lead with an MAE of 60.26 and an RMSE of 90.75, as indicated in Table 2, indicating the smallest deviation from the new GBDT model. Heisler *et al.* [5] follows in second place, recording an MAE of 75.80 and an RMSE of 105.44, as depicted in Figure 3, suggesting strong performance despite marginally elevated errors. Burton [11] ranks third with an MAE of 80.75 and an RMSE of 108.61, as noted in Table 2, although its errors exceed those of Chendong and Heisler. Conversely, models such as Stebbins (MAE = 849.29, RMSE = 853.80), De Pablo 2 (MAE = 205.52, RMSE = 230.38), and Pahalavanpour (MAE = 173.05, RMSE = 193.86) exhibit considerably higher errors (see Figures 3 and 4), indicating reduced precision in DP estimation compared to the proposed model and the top three authors.

Figure 4 presents the correlation between 2-FAL concentrations (ranging from 0 to 5 ppm) and DP values (spanning 200 to 1200) for transformer insulation, comparing both measured and predicted DP values across various models. The graph includes five distinct lines: DP_measured, the GBDT model, Chendong, Heisler, and Burton. The DP_measured line exhibits a steep drop from approximately 1100 at 0 ppm to under 400 at 5 ppm, with evident fluctuations that reflect the non-linear nature of insulation degradation. The GBDT model closely mirrors this pattern, consistently aligning with the measured data across the entire range. Both Chendong and Heisler models approximate the measured DP values but show slight deviations at higher 2-FAL levels, tending to underestimate the DP. In contrast, Burton's model displays the most significant divergence from the measured data, particularly at 2-FAL concentrations exceeding 2 ppm, underscoring its reduced precision in tracking the degradation trend.

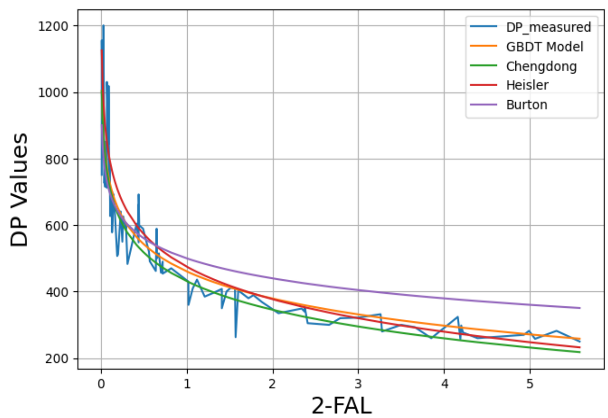

Figure 4. *Relationship between 2FAL and DP values of high-performing models*

To demonstrate the practical applicability and reliability of the proposed GBDT model, an independent validation was conducted using data from a set of 10 transformers not included in the model training phase. The detailed results, presented comprehensively in Table 3, underscore the model's strong predictive capabilities. Predictions of the degree of polymerization (GBDT DP) align closely with actual measured DP values, accurately reflecting the insulation conditions of each transformer.

For example, transformer TR1 exhibited extensive deterioration, with an actual DP of 430, closely matched by the GBDT prediction of 427.087. Its calculated EL using GBDT DP was 21 years, with a total transformer lifetime operation (TRL) of 35 years, resulting in an actual RUL of 14 years. Transformer TR7, categorized as healthy, showed minimal deviation between actual DP (727) and predicted DP (742.366), corresponding to an elapsed life of 10 years and a substantial remaining life of 25 years.

Table 3. *Test data of 10 independent transformers*

Trans.	2-FAL	Actual DP	GBDT DP	Extent of Degradation	EL-GBDT DP (year) [Prahan]	TRL operation (TRL) (year)	RUL-Actual DP (year)
TR1	1.02	430	427.087	Extensive deterioration	21	35	14
TR2	0.82	470	466.631	Moderate deterioration	19	35	16
TR3	3.66	293	290.059	Extensive deterioration	29	35	6
TR4	1.41	408	393.343	Extensive deterioration	23	35	12
TR5	0.56	520	513.278	Moderate deterioration	17	35	18
TR6	3.5	300	290.059	Extensive deterioration	29	35	6
TR7	0.1	727	742.366	Healthy	10	35	25
TR8	0.72	490	470.48	Moderate deterioration	19	35	16
TR9	4.22	277	275.632	Extensive deterioration	30	35	5
TR10	1.12	436	401.764	Extensive deterioration	22	35	13

Transformers TR2, TR5, and TR8 demonstrated moderate deterioration, with respective remaining lifespans ranging from 15 to 17 years, whereas transformers like TR3, TR4, TR6, TR9, and TR10 experienced extensive deterioration, reflecting significantly reduced RULs between 5 and 13 years.

These evaluations clearly illustrate the model's effectiveness in accurately classifying degradation extents (healthy, moderate, extensive) and calculating elapsed and remaining lifetimes. Consequently, the GBDT model not only ensures precise diagnostics but also supports effective condition-based maintenance strategies, enabling early interventions to maximize transformer reliability, performance, and safety within power systems.

4.2. Discussion

In terms of prediction accuracy, the GBDT model stands out, achieving the lowest MAE of 38.65 and RMSE of 63.04, surpassing all competing models. Among the alternatives, Chengdong recorded the nearest performance with an MAE of 60.26 and RMSE of 90.76, while Stebbins exhibited the poorest results, with an MAE of 849.29 and RMSE of 853.80. The notably low MAE and RMSE values of the GBDT model underscore its enhanced precision and capacity to manage significant errors, making it a valuable asset for estimating the RUL of transformer insulation with reliability.

When compared to Chengdong, the GBDT model demonstrates superior accuracy and robustness, evidenced by its lower MAE (38.65 versus 60.26) and RMSE (63.04 versus 90.76). In contrast, Dong's model, with an MAE of 86.06 and RMSE of 125.03, falls considerably behind, reinforcing the practical advantage of the proposed approach. Stebbins' model, marked by an MAE exceeding 849, performed the least effectively, likely due to limitations in its modeling framework.

The GBDT model demonstrates outstanding performance, achieving the lowest error metrics (MAE of 38.65 and RMSE of 63.04), supported by both numerical analyses and graphical representations. Notably, the proposed model closely aligns with the measured DP trend illustrated in Figure 4, confirming its ability to effectively capture complex, non-linear degradation patterns of transformer insulation. Future studies could extend the analysis by examining additional influencing factors on insulation degradation, such as the concentrations of CO_2, CO, and the CO_2/CO ratio, which may further enhance predictive accuracy and reliability. Moreover, further refinements of the GBDT approach could be explored to optimize error reduction. Collectively, these advancements position the proposed model as a robust and dependable tool for accurately evaluating transformer insulation conditions and reliably predicting their remaining useful life.

Figure 5 illustrates that the prediction errors of the proposed model are generally minimal. The majority of samples are concentrated within the 0–50 MAE range, with the peak frequency approaching 40 samples, reflecting high prediction accuracy. A smaller number of samples show errors surpassing 100, and only a few rare cases exceed 300. This error distribution underscores the model's robustness and dependability, particularly for practical transformer diagnostics, where minimizing error margins is essential for accurately evaluating insulation condition and estimating remaining useful life.

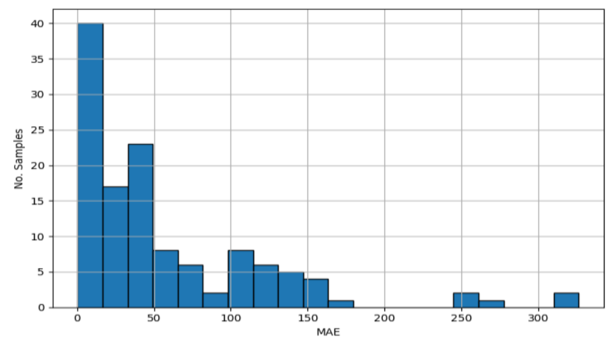

Figure 5. *Distribution of MAE*

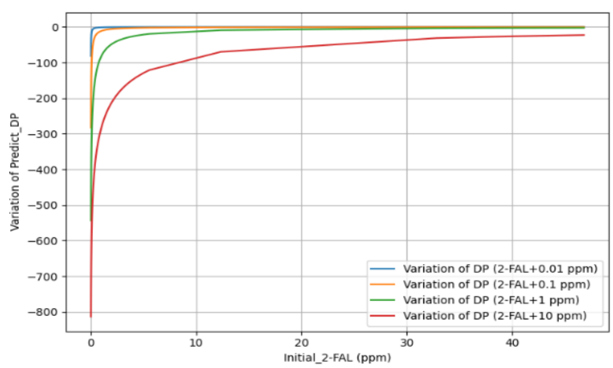

Figure 6. *Sensitivity analysis of predicted DP in response to changes in 2-FAL concentration*

Figure 6 depicts the variation in predicted DP as the concentration of 2-FAL increases incrementally across four different scenarios: 2-FAL+0.01, 2-FAL+0.1, 2-FAL+1, and 2-FAL+10 ppm. The data reveals that DP is highly responsive to minor changes in 2-FAL, particularly at lower concentrations. At the outset, with minimal 2-FAL levels, even a small increment leads to a significant reduction in predicted DP, most notably in the 2-FAL+10 ppm case, where the drop surpasses -800. As the initial 2-FAL concentration rises, the sensitivity diminishes, and the curves tend to level off. This pattern underscores the model's ability to detect early-stage degradation, positioning it as a useful tool for identifying potential faults in transformers at an initial stage.

By addressing these challenges, the suggested approach offers substantial benefits for evaluating transformer conditions and implementing predictive maintenance strategies, thereby minimizing operational risks and enhancing asset management efficiency within power systems.

5. Conclusion

This research successfully introduced an advanced GBDT model for predicting the DP and estimating the RUL of transformer insulation based on 2-FAL concentrations. Utilizing a robust dataset and validated through comprehensive comparative analyses, the GBDT model outperformed traditional approaches, achieving significantly lower error metrics. Its accuracy in capturing complex non-linear degradation patterns was evidenced by closely aligning predictions with actual DP measurements across varying 2-FAL concentrations. The model's sensitivity to incremental 2-FAL changes further highlights its capability for early detection of insulation deterioration, an essential aspect for timely maintenance interventions. Consequently, the proposed method provides a reliable, efficient, and non-invasive diagnostic tool, significantly contributing to the improvement of transformer asset management strategies. Future work may involve further refinement of the model to enhance accuracy and exploring its integration into broader condition-based maintenance frameworks within electrical power networks.

REFERENCES

[1] H. V. Tran, K. A. Nguyen, and D. D. Le, "Impact of data balancing algorithms on the accuracy of power transformer fault diagnosis based on dissolved gas analysis", in *Proc. 2nd Int. Conf. on Green Solutions for Emerging Technologies and Sustainability (GSETS 2025)*, Ho Chi Minh City, Vietnam, Apr. 2025, pp. 851–863.

[2] D. Feng, Z. Wang, and P. Jarman, "Transmission power transformer assessment using furan measurement with the aid of thermal model", in *Proc. 2012 IEEE Int. Conf. Condition Monitoring and Diagnosis (CMD)*, Sep. 2012, pp. 521–524.

[3] Y. Wang, S. Gong, and S. Grzybowski, "Reliability evaluation method for oil–paper insulation in power transformers", *Energies*, vol. 4, no. 9, pp. 1362–1377, 2011. https://doi.org/10.3390/en4091362

[4] X. Chendong, F. Qiming, and X. Shiheng, "To estimate the ageing status of transformers by furfural concentration in the oil", in *Proc. of the CIGRE Study Committee 33 Colloquium,* Leningrad, Moscow, 1991.

[5] A. Heisler, A. Banzer, and R. Ilgen, "Zustandsbeurteilung von Transformatoren mit Furfurol-Bestimmung", *EW – Magazin für die Energiewirtschaft*, vol. 102, no. 16, pp. 58–59, 2003.

[6] A. D. Pablo, "Interpretation of furanic compounds analysis–degradation models", CIGRÉ Working Group D1, Tech. Brochure 15-01, 1997.

[7] A. D. Pablo, "Furfural and ageing: How are they related", in *IEE Colloquium on Insulating Liquids (Ref. No. 1999/119)*, May 1999, pp. 5–1. IET.

[8] M. Dong, Z. Yan, and G. J. Zhang, "Comprehensive diagnostic and aging assessment method of solid insulation in transformer", in *Proc. 2003 Annu. Rep. Conf. Electrical Insulation and Dielectric Phenomena (CEIDP)*, Oct. 2003, pp. 137–140. IEEE.

[9] R. D. Stebbins, D. S. Myers, and A. B. Shkolnik, "Furanic compounds in dielectric liquid samples: Review and update of diagnostic interpretation and estimation of insulation ageing", in *Proc. 7th Int. Conf. Properties and Applications of Dielectric Materials (ICPADM)*, 2003, pp. 921–926.

[10] B. Pahlavanpour, M. Eklund, and N. Naphthenics, "Thermal ageing of mineral insulating oil and kraft paper", in *Proc. TechCon*, Oct. 2003.

[11] P. J. Burton, "Applications of liquid chromatography to the analysis of electrical insulating materials", *CIGRÉ Technical Brochure 15-8*, 1988.

[12] H. Vaurchex, I. Höhlein, and A. J. Kachler, "Transformer aging research on furanic compounds dissolved in insulating oil", in *Proc. Large High Voltage Electrical Systems Conf.*, Paris, France, Aug. 2002, pp. 25–30.

[13] C. Lin, B. Zhang, and Y. Yuan, "The aging diagnosis of solid insulation for oil-immersed power transformers and its remaining life prediction", in *Proc. 2010 Asia-Pacific Power and Energy Engineering Conf. (APPEEC)*, Mar. 2010, pp. 1–3. IEEE.

[14] A. Teymouri, B. Vahidi, and P. van der Wielen, "A novel methanol-based DP estimation method with a new methanol peak detector index for aging assessment of power transformer insulation paper", *IEEE Trans. Dielectr. Electr. Insul.*, vol. 29, no. 4, pp. 1506–1513, 2022.

[15] B. A. Thango, J. A. Jordaan, and A. F. Nnachi, "Assessment of transformer cellulose insulation life expectancy based on oil furan analysis (case study: South African transformers)", *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 6, pp. 29–33, 2021.

[16] S. Ghoneim, "The degree of polymerization in a prediction model of insulating paper and the remaining life of power transformers", *Energies*, vol. 14, art. no. 670, 2021.

[17] P. C. K. Guri and K. Najdenkoski, "Transformer remaining life calculation based on degree of polymerization (DP)", *Journal of Multidisciplinary Engineering Science and Technology*, vol. 4, no. 11, pp. 16779–16786, 2024.

[18] H. Park, E. Kim, B. S. Kwak, T. Jun, R. Kawano, and S. H. Pyo, "Selective aqueous extraction and green spectral analysis of furfural as an aging indicator in power transformer insulating fluid", *Separations*, vol. 10, no. 7, pp. 381, 2023.

[19] International Electrotechnical Commission (IEC), "Insulating liquids – Determination of acidity – Part 2: Colourimetric titration", IEC 62021-2, 2007.

[20] A. C. I. U. Ancuța-Mihaela, M. C. NIȚU, M. Nicola, C. I. Nicola, and F. Lăzărescu, "Complementary analysis of the degree of polymerization based on chemical markers 2-furaldehyde and methanol using fuzzy logic", in *Proc. 21st Int. Symp. Electrical Apparatus & Technologies (SIELA)*, Jun. 2020, pp. 1–6. IEEE.

[21] R. Soni and B. Mehta, "Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies", *Engineering Failure Analysis*, vol. 128, no. 2, pp. 105634, 2021.

[22] D. Kanumuri, V. Sharma, and O. P. Rahi, "Analysis using various approaches for residual life estimation of power transformers", *International Journal on Electrical Engineering & Informatics*, vol. 11, no. 2, pp. 389–407, 2019.

[23] K. Guri, "Correlation between degree of polymerization and data from transformer oil analysis", *Electrotechnica & Electronica (E+E)*, vol. 59, no. 1, pp. 24-32, 2024.

[24] H. V. Tran, K. A. Nguyen, D. D. Le, and D. H. Dinh, "Improving the reliability of oil-immersed power transformer fault diagnosis based on the evaluation of dissolved gas component input vectors", in *Proc. 2025 8th Int. Conf. on Circuits, Systems and Simulation (ICCSS 2025)*, Ho Chi Minh City, Vietnam, May 2025, pp. 109-114. https://doi.org/10.1109/ICCSS65911.2025.11081708

[25] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms", *Artificial Intelligence Review*, vol. 54, pp. 1937–1967, 2021. https://doi.org/10.1007/s10462-020-09896-5

[26] A. A. Adekunle, I. Fofana, P. Picher, E. M. Rodriguez-Celis, and O. H. Arroyo-Fernandez, "Analyzing transformer insulation paper prognostics and health management: A modeling framework perspective", *IEEE Access*, vol. 12, pp. 58349–58377, 2024.

[27] B. A. Thango and P. N. Bokoro, "Prediction of the degree of polymerization in transformer cellulose insulation using the feedforward backpropagation artificial neural network", *Energies*, vol. 15, no. 12, pp. 4209, 2022.