

Impact of Diagnostic Parameter Group Availability on Transformer Health Index Estimation using Real-World Oil Diagnostic Data

Omar Ahmed Mohammed Al-Shaikh
Faculty of Electrical Technology and Engineering (FTKE),
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia

Arfah Ahmad
Faculty of Electrical Technology and Engineering (FTKE),
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia

Muhammad Sharil Yahaya
Faculty of Electrical Technology and Engineering (FTKE),
Universiti Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia

Ahmad Jazlan Haja Mohideen
Faculty of Engineering, International Islamic University
Malaysia (IIUM)
Kuala Lumpur, Malaysia

Abstract - Transformer Health Index estimation commonly depends on diagnostic information obtained from dissolved gas analysis, oil quality analysis, and furan analysis. In practical utility operation, however, not all diagnostic groups are always available because testing frequency, cost, and data completeness differ across monitoring programs. This paper evaluates the impact of diagnostic parameter group availability on Transformer Health Index estimation using 3,392 real-world transformer oil diagnostic records. Complete records were used to establish the reference health-index target, and seven reduced and complete input scenarios were then developed using dissolved gas analysis, oil quality analysis, furan analysis, and their combinations. The same preprocessing pipeline, transformer-level 80:20 train-test split, and XGBoost regression model were applied to all scenarios to ensure a fair comparison and to prevent records from the same transformer appearing in both training and testing subsets. The complete diagnostic scenario achieved the best performance, with root mean square error of 3.7517, mean absolute error of 2.8735, coefficient of determination of 0.9753, and condition-class agreement of 84.90 percent. Among reduced scenarios, the combination of dissolved gas analysis and oil quality analysis provided the strongest practical estimation alternative. The findings show that combining fault-gas, oil-quality, and paper-ageing indicators substantially improves Transformer Health Index estimation, while reduced diagnostic groups can only approximate the reference health index when complete diagnostic information is unavailable.

Keywords - Transformer Health Index; dissolved gas analysis; oil quality analysis; furan analysis; diagnostic parameter availability; XGBoost; condition monitoring.

I. INTRODUCTION

Power transformers are critical assets in transmission and distribution networks because unexpected failures may cause service interruption, expensive replacement, equipment damage, and safety risk. Modern transformer asset management therefore relies on condition-based maintenance supported by laboratory oil diagnostics, online monitoring, and quantitative health indicators [1-7].

The Transformer Health Index integrates heterogeneous diagnostic evidence into a single condition score that can be used for asset ranking, maintenance planning, and risk-informed decision-making [8], [9]. Conventional health-index approaches are attractive because they are transparent and interpretable for utility engineers. However, these approaches depend on the availability of several diagnostic groups, including dissolved gas analysis, oil quality analysis, and furan analysis [10-13].

Machine learning has been increasingly applied to transformer condition assessment because it can learn nonlinear relationships between diagnostic variables and reference health-index values [14-18]. Tree-based ensemble models are particularly suitable for tabular diagnostic data because they can capture threshold-like behaviour, nonlinear interactions, and mixed-scale feature effects [19-25]. Nevertheless, many published studies mainly report model accuracy under a complete-feature setting and give less attention to the practical problem of diagnostic parameter availability.

In practice, complete diagnostic information is not always available. Dissolved gas analysis may be performed more frequently than furan analysis, while some oil-quality parameters may be missing due to testing constraints or incomplete historical records. This creates an important question for utilities: how much estimation performance is lost when one or more diagnostic groups are unavailable? This paper addresses this question through a diagnostic parameter group availability study using complete historical records to define the reference health-index target and reduced input scenarios to evaluate practical estimation reliability.

The main contributions of this paper are as follows: first, seven diagnostic input scenarios are formulated using dissolved gas analysis, oil quality analysis, furan analysis, and their combinations; second, each scenario is evaluated using the same preprocessing and XGBoost modelling pipeline under a transformer-level train-test split; third, numerical prediction accuracy and maintenance-class agreement are compared; and fourth, practical guidance is provided on which diagnostic

groups should be prioritized for reliable Transformer Health Index estimation when complete diagnostic data are not available.

II. RELATED WORK

A. Transformer health index and oil diagnostics

Dissolved gas analysis is a well-established diagnostic technique for identifying electrical and thermal faults in mineral-oil-filled transformers. IEEE and IEC guides describe the interpretation of gases generated under fault and ageing mechanisms, including hydrogen, methane, ethane, ethylene, acetylene, carbon monoxide, and carbon dioxide [1], [2], [10]. Oil quality analysis provides complementary information on dielectric and chemical oil condition through breakdown voltage, moisture, acidity, interfacial tension, and colour [4]. Furan analysis is commonly used to support assessment of cellulose paper ageing and insulation degradation [7], [11].

Health-index frameworks combine these diagnostic channels using scoring tables, weighting factors, and aggregation functions [8], [9], [17], [18]. Their main strength is interpretability, because the final score can be traced to engineering thresholds and diagnostic factors. However, fixed scoring rules may not fully capture nonlinear interactions among gases, oil-quality variables, and paper-ageing indicators [14-16].

B. Machine learning for transformer health estimation

Machine learning methods have been used to estimate health-index values by learning from historical diagnostic data. Neural-network and general regression approaches have been reported for transformer health-index calculation [12], [14]. Random forests, support vector regression, AdaBoost, gradient boosting, and XGBoost have also been used because they are able to model nonlinear feature interactions in structured engineering datasets [19-25].

XGBoost is adopted in this work as a common learner for all diagnostic scenarios. It builds an additive ensemble of regression trees and includes regularization to reduce overfitting [21]. Using one learner across all feature scenarios ensures that observed performance changes are mainly due to diagnostic group availability rather than changes in model family. The evaluation procedure follows standard predictive-modelling practice, including training-only preprocessing, fixed train-test evaluation, and regression metrics [26-31].

C. Diagnostic parameter availability gap

Data completeness is a major practical challenge in transformer monitoring. A full laboratory test may contain dissolved gases, oil-quality parameters, and furan content, but older maintenance records or limited testing programs may provide only a subset of these variables. Without a systematic availability analysis, utilities may not know whether reduced diagnostic information is sufficient for health-index estimation.

This study therefore treats diagnostic availability as an experimental variable. Instead of comparing many algorithms on the same complete dataset, the proposed framework compares how the same model behaves when different diagnostic groups are available. This design provides direct evidence on the value of each diagnostic group for Transformer Health Index estimation.

III. MATERIALS AND METHODS

A. Dataset and feature groups

The dataset contains 3,392 real-world transformer oil diagnostic records from 564 transformer serial numbers. Thirteen diagnostic input variables were used and grouped into dissolved gas analysis, oil quality analysis, and furan analysis. The supervised target was Real_HI%, which was calculated from a conventional transformer health-index procedure using complete diagnostic information and used as the reference output for model development. Identifier fields were used only for transformer-level train-test splitting and were not used as prediction features.

The estimation problem is formulated in (1), where the predicted health index is a function of the available diagnostic feature groups.

$$\widehat{THI} = f(X_{DGA}, X_{OQA}, X_{FA}) \quad (1)$$

Table I summarizes the diagnostic parameter groups used in this study

TABLE I. DIAGNOSTIC PARAMETER GROUPS USED IN THE STUDY

Group	Variables	Diagnostic role
DGA gases	H ₂ , CH ₄ , CO, CO ₂ , C ₂ H ₄ , C ₂ H ₆ , C ₂ H ₂	Internal electrical and thermal fault indicators
Oil quality	Dielectric breakdown, moisture, neutralization value, interfacial tension, colour	Insulating-oil dielectric and chemical condition
Furan analysis	2FAL	Cellulose paper ageing indicator
Target	Real_HI%	Reference Transformer Health Index output

B. Diagnostic input scenarios

Seven scenarios were prepared to represent different levels of diagnostic parameter availability. The complete scenario used all groups, while the reduced scenarios were created by selectively excluding one or more diagnostic groups from the model inputs. These reduced scenarios do not recalculate a new full conventional THI; instead, they test how accurately limited diagnostic groups can approximate the reference Real_HI% when some diagnostic information is unavailable. The scenario design is summarized in Table II and illustrated in Fig. 1.

TABLE II. DIAGNOSTIC PARAMETER AVAILABILITY SCENARIOS

Scenario	Input groups	Purpose
S1	DGA only	Gas-based fault diagnostic information only
S2	OQA only	Oil-quality information only
S3	FA only	Paper-ageing indicator only
S4	DGA + OQA	Fault gas and oil-quality information
S5	DGA + FA	Fault gas and paper-ageing information
S6	OQA + FA	Oil-quality and paper-ageing information
S7	DGA + OQA + FA	Complete diagnostic information

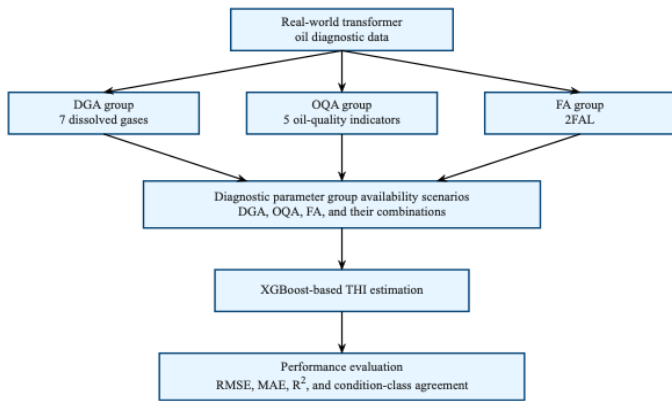


Fig. 1. Proposed diagnostic parameter group availability framework.

C. Data preprocessing and model development

For each input scenario, only the selected diagnostic features were retained. Missing numerical values were handled using median imputation fitted on the training subset only. Min-max normalization was applied after imputation to maintain consistent numerical scaling across scenarios, as expressed in (2). This training-only preprocessing strategy was used to avoid information leakage from the testing subset.

The dataset was divided using a transformer-level 80:20 split based on Serial No., so that records from the same transformer were assigned only to either training or testing. The original dataset contained 3,392 records. Since transformer-level splitting requires a valid Serial No. identifier, three records without valid transformer identifiers were excluded from the group-aware model evaluation, resulting in 3,389 records used for training and testing. The resulting group-aware evaluation used 2,720 training records from 449 transformers and 669 testing records from 115 transformers. XGBoost regression was trained independently for each diagnostic availability scenario using the same random seed, preprocessing steps, and evaluation set to ensure a fair comparison.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

D. Evaluation metrics

Performance was evaluated using root mean square error, mean absolute error, coefficient of determination, and condition-class agreement. Lower error values indicate better numerical prediction, while higher coefficient of determination and class agreement indicate stronger health-index estimation capability.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

In (3)-(5), y_i is the reference health-index value, \hat{y}_i is the predicted value, \bar{y} is the mean reference health-index value, and n is the number of test samples. Condition-class agreement was calculated after mapping numerical health-index values into maintenance-oriented classes.

IV. RESULTS AND DISCUSSION

A. Dataset summary

The final dataset contained 3,392 records, 564 unique transformer serial numbers, and 13 diagnostic input variables. The reference health-index values covered a broad range of transformer conditions, with a mean of 58.91 percent, standard deviation of 23.20 percent, minimum value of 0.00 percent, and maximum value of 100.00 percent. This distribution provides a meaningful basis for testing the sensitivity of Transformer Health Index estimation to diagnostic group availability. Table III presents the dataset summary used in the analysis

TABLE III. DATASET SUMMARY

Item	Value
Number of records	3,392
Unique Serial No.	564
Diagnostic input variables	13
Target variable	Real_HI%
Mean THI	58.91%
Standard deviation	23.20%
Minimum/Maximum THI	0.00% / 100.00%

B. Effect of diagnostic group availability

Table IV presents the comparative prediction results for all diagnostic availability scenarios. The complete scenario, which combined dissolved gas analysis, oil quality analysis, and furan analysis, produced the strongest overall performance. This confirms that Transformer Health Index estimation benefits from combining fault-gas, oil-quality, and paper-ageing information, as also shown by the error, explained-variance, and class-agreement comparisons in Figs. 2-4.

TABLE IV. PREDICTION PERFORMANCE UNDER DIAGNOSTIC AVAILABILITY SCENARIOS USING TRANSFORMER-LEVEL SPLIT

Scenario	Input groups	RMSE	MAE	R2	Class agreement %
S1	DGA only	13.9525	11.1347	0.6581	48.58
S2	OQA only	10.3881	8.2241	0.8105	60.54
S3	FA only	15.0032	12.2226	0.6047	46.19
S4	DGA + OQA	6.1469	4.5487	0.9336	79.07
S5	DGA + FA	7.7280	6.2092	0.8951	76.83
S6	OQA + FA	9.6818	7.7799	0.8354	62.63
S7	DGA + OQA + FA	3.7517	2.8735	0.9753	84.90

The complete diagnostic model achieved RMSE of 3.7517, MAE of 2.8735, R2 of 0.9753, and condition-class agreement of 84.90 percent. This result indicates that the full feature group explains most of the variation in the reference health-index target and gives the most reliable classification support under the transformer-level evaluation setting.

Among the reduced multi-group scenarios, DGA + OQA achieved the best practical performance with RMSE of 6.1469, MAE of 4.5487, R2 of 0.9336, and condition-class agreement of 79.07 percent. This suggests that combining fault-gas indicators with oil-quality variables provides a strong estimation alternative when furan analysis is unavailable. However, the

error remains higher than the complete scenario, indicating that 2FAL contributes important information about paper-insulation ageing.

The single-group scenarios performed substantially worse. DGA-only and FA-only scenarios produced high error and low class agreement, while OQA-only performed better but remained clearly below the combined scenarios. These results show that no single diagnostic group is sufficient to fully represent transformer health across the available dataset.

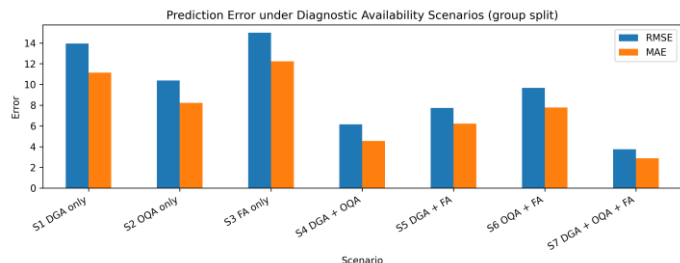


Fig. 2. RMSE and MAE comparison under diagnostic availability scenarios using transformer-level split.

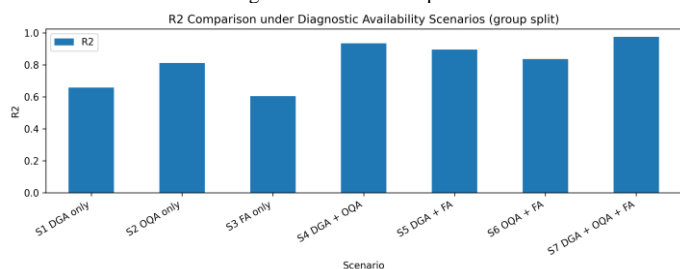


Fig. 3. Coefficient of determination comparison across diagnostic availability scenarios using transformer-level split.

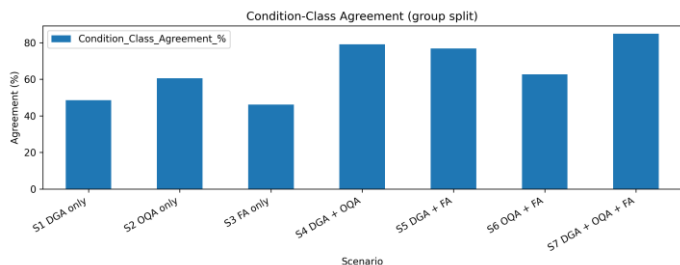


Fig. 4. Condition-class agreement under diagnostic availability scenarios using transformer-level split.

C. Condition-class agreement and interpretation

Condition-class agreement provides a maintenance-oriented view of the results because utilities often use health-index classes for inspection scheduling, refurbishment planning, and risk ranking. The complete diagnostic scenario achieved the highest agreement of 84.90 percent, followed by the DGA + OQA scenario with 79.07 percent. The prediction plots in Figs. 5 and 6 show that the complete model has tighter alignment with the reference THI than the reduced DGA + OQA model, while Fig. 7 indicates that most full-model residuals are distributed around zero. Fig. 8 summarizes the relative contribution of each diagnostic group in the complete XGBoost model.

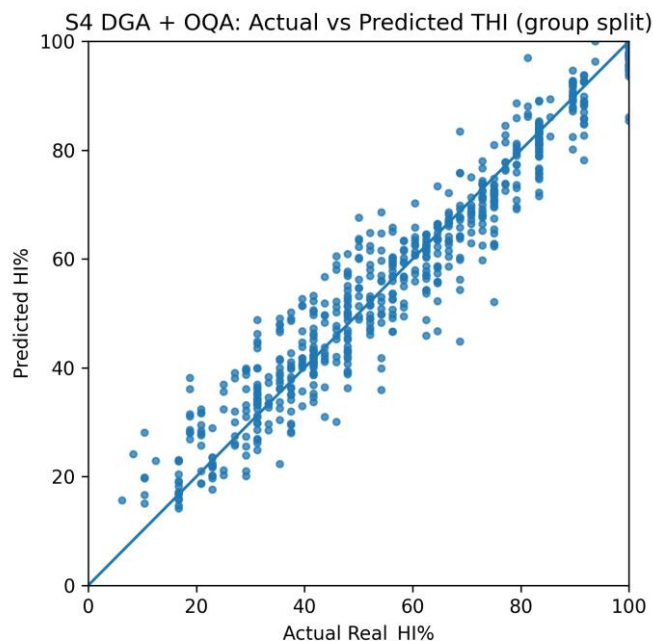


Fig. 5. Actual and predicted THI values for the reduced DGA + OQA scenario using transformer-level split.

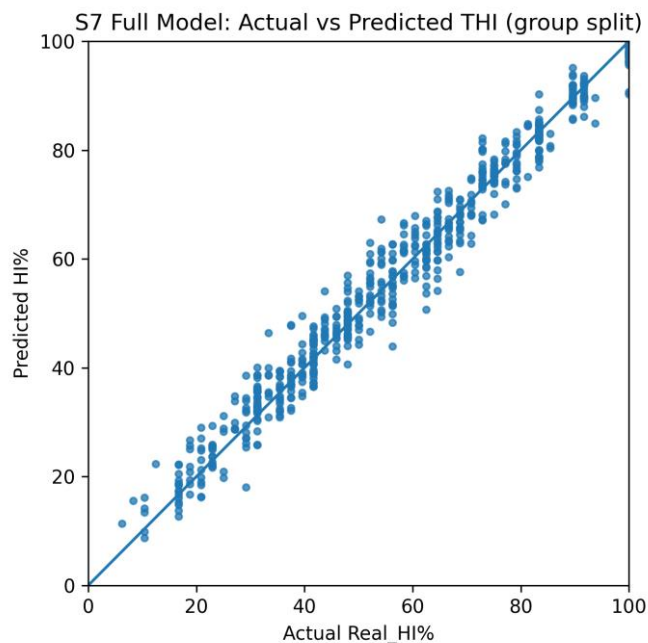


Fig. 6. Actual and predicted THI values for the complete DGA + OQA + FA scenario using transformer-level split.

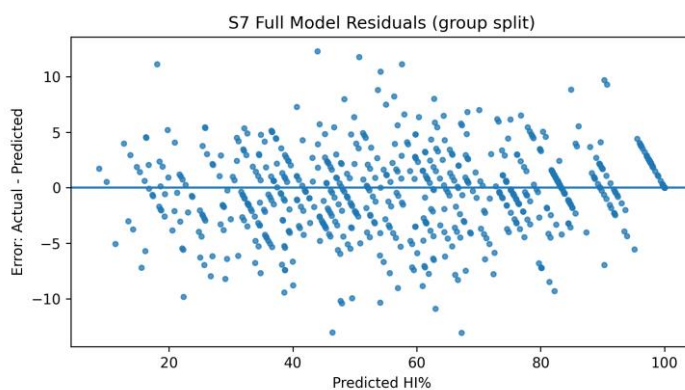


Fig. 7. Residual distribution for the complete DGA + OQA + FA model using transformer-level split.

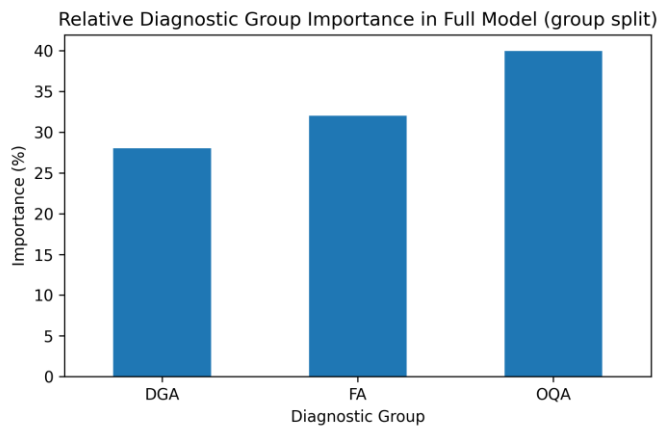


Fig. 8. Relative diagnostic group importance from the complete XGBoost model using transformer-level split.

D. Practical implications

The findings provide practical guidance for utilities managing incomplete transformer diagnostic records. When full DGA, OQA, and FA information is available, it should be preferred because it produces the lowest prediction error and highest class agreement. When furan analysis is unavailable, DGA + OQA is the strongest reduced estimation alternative, but it should be interpreted as an approximation of the reference THI rather than a full replacement for complete conventional DGA + OQA + FA assessment. The group-importance result in Fig. 8 further supports the need to consider all diagnostic groups when making high-risk maintenance decisions.

The weak performance of DGA-only indicates that gas data alone should not be treated as a complete health-index predictor when oil-quality and paper-ageing information are relevant to the condition score. DGA remains essential for fault diagnosis, but it should be complemented by OQA and FA for holistic health-index estimation whenever these tests are available.

V. CONCLUSION

This paper presented a diagnostic parameter group availability analysis for Transformer Health Index estimation using 3,392 real-world transformer oil diagnostic records. Seven input scenarios were evaluated using dissolved gas analysis, oil quality analysis, furan analysis, and their combinations under a common XGBoost modelling framework and transformer-level train-test split.

The complete diagnostic scenario achieved the best performance, with RMSE of 3.7517, MAE of 2.8735, R2 of 0.9753, and condition-class agreement of 84.90 percent. Among reduced scenarios, DGA + OQA provided the strongest practical estimation alternative, while single-group scenarios produced higher prediction errors and lower class agreement. The results demonstrate that integrated diagnostic information is important for reliable Transformer Health Index estimation and maintenance-class decision support.

The reduced scenarios should be understood as approximations of a reference health-index target derived from complete diagnostic information, not as replacements for complete conventional THI calculation. Future work may extend this analysis using multi-utility datasets, uncertainty quantification, explainable artificial intelligence, temporal oil-

test histories, and cost-sensitive diagnostic planning. These extensions can further support practical transformer asset management when complete diagnostic information is not always available.

ACKNOWLEDGMENT

The authors acknowledge the Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka (UTeM), for academic and research support. This work was supported by the Ministry of Higher Education (MOHE), Malaysia, under the Fundamental Research Grant Scheme (FRGS), Grant No. FRGS/1/2024/FTKE/F00579.

REFERENCES

- [1] P. T. Committee, "IEEE guide for the interpretation of gases generated in mineral oil-immersed transformers," *IEEE Std C*, vol. 57, pp. 40–42.
- [2] I. IEC, "60599—Mineral Oil-Impregnated Electrical Equipment in Service—Guide to the Interpretation of Dissolved and Free Gases Analysis," *International Electrotechnical Commission: Geneva, Switzerland*, 2015.
- [3] I. Std, "Mineral insulating oils in electrical equipment—Supervision and maintenance guidance," *Mineral insulating oils in electrical equipment—Supervision and maintenance guidance*, 2013.
- [4] I. C. 106-, "IEEE guide for acceptance and maintenance of insulating mineral oil in electrical equipment," ed: Institution of Electrical and Electronics Engineers Piscataway, 2015.
- [5] W. Cigré, "A2. 37; Transformer Reliability Survey," *Technical Brochure*, vol. 642, 2015.
- [6] M. Wang, A. J. Vandermaar, and K. D. Srivastava, "Review of condition assessment of power transformers in service," *IEEE Electrical insulation magazine*, vol. 18, no. 6, pp. 12–25, 2003.
- [7] T. K. Saha, "Review of modern diagnostic techniques for assessing insulation condition in aged transformers," *IEEE transactions on dielectrics and electrical insulation*, vol. 10, no. 5, pp. 903–917, 2003.
- [8] A. Jahromi, R. Piercy, S. Cress, J. Service, and W. Fan, "An approach to power transformer asset management using health index," *IEEE Electrical Insulation Magazine*, vol. 25, no. 2, pp. 20–34, 2009.
- [9] A. Naderian, S. Cress, R. Piercy, F. Wang, and J. Service, "An approach to determine the health index of power transformers," in *Conference Record of the 2008 IEEE International Symposium on Electrical Insulation*, 2008: IEEE, pp. 192–196.
- [10] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE electrical Insulation magazine*, vol. 18, no. 3, pp. 8–17, 2002.
- [11] D. Shroff and A. Stannett, "A review of paper aging in power transformers," in *IEE Proceedings C (Generation, Transmission and Distribution)*, 1985, vol. 132, no. 6: IET, pp. 312–319.
- [12] W. H. Tang and Q. H. Wu, *Condition monitoring and assessment of power transformers using computational intelligence*. Springer Science & Business Media, 2011.

- [13] R. E. James and Q. Su, *Condition assessment of high voltage insulation in power system equipment*. IET, 2008.
- [14] M. M. Islam, G. Lee, and S. N. Hettiwatte, "Application of a general regression neural network for health index calculation of power transformers," *International Journal of Electrical Power & Energy Systems*, vol. 93, pp. 308–315, 2017.
- [15] A. Alqudsi and A. El-Hag, "Application of machine learning in transformer health index prediction," *Energies*, vol. 12, no. 14, p. 2694, 2019.
- [16] D. Rediansyah, R. A. Prasajo, and A. Abu-Siada, "Artificial intelligence-based power transformer health index for handling data uncertainty," *IEEE Access*, vol. 9, pp. 150637–150648, 2021.
- [17] A. Azmi, J. Jasni, N. Azis, and M. A. Kadir, "Evolution of transformer health index in the form of mathematical equation," *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 687–700, 2017.
- [18] J. Jasni, A. Azmi, N. Azis, M. Yahaya, and M. Talib, "Assessment of transformer Health Index using different model," *Pertanika Journal of Science and Technology*, vol. 25, no. 1, pp. 143–150, 2017.
- [19] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [20] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.
- [21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [22] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.
- [23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.
- [24] V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [25] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [26] T. Hastie, "The elements of statistical learning: data mining, inference, and prediction," ed: springer, 2009.
- [27] M. Kuhn and K. Johnson, *Applied predictive modeling*. Springer, 2013.
- [28] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [29] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, 1995, vol. 14, no. 2: Montreal, Canada, pp. 1137–1145.
- [30] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.
- [31] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.