Health Assessment and Lifespan Estimation for Power Transformer using Machine Learning

Ajaan Anubhav Borah*, Ritu Nazneen Ara Begum*, Barnali Goswami*

* Department of Electrical Engineering, Assam Engineering College

DOI: 10.29322/IJSRP.14.04.2024.p14812
10.29322/IJSRP.14.04.2023.p14812

Paper Received Date: 15th February 2024
Paper Acceptance Date: 25th March 2024
Paper Publication Date: 6th April 2024

Abstract- This paper investigates the application of various machine learning models for predicting the health index and remaining lifespan of power transformers. We evaluate the performance of individual models including Random Forest, Support Vector Regression (SVR), Gradient Boosting, and Elastic Net in estimating the health index of power transformers. Additionally, a stacking ensemble approach is implemented by combining the predictions of these individual models. The performance of all models is compared using the R-squared metric to identify the most accurate approach. The model with the highest R-squared value is then employed to predict the remaining lifespan of power transformers based on their health index estimations. This research contributes to the development of reliable and efficient methods for power transformer health assessment and lifespan prediction, promoting preventative maintenance strategies and ensuring grid stability.

Index Terms- Elastic Net, Gradient Boosting, Health Index, Life Expectation, Machine Learning, Power Transformer, Random Forest, Stacking Ensemble

I. INTRODUCTION

Power transformers are the backbone of the electrical grid, efficiently transferring power at various voltage levels. Their reliable operation is crucial for grid stability and consistent power delivery. However, these complex assets degrade over time due to factors like thermal stress, electrical loading, and material aging. This degradation can lead to catastrophic failures, causing economic losses, outages, and safety risks [1].

Predictive maintenance strategies are essential for mitigating these risks. By proactively assessing transformer health and estimating remaining life, utility companies can implement targeted maintenance before failures occur. Traditional methods rely on periodic offline tests and inspections, which are time-consuming and may not fully capture the transformer's condition.

Machine learning (ML) has emerged as a game-changer in various industries [2], including the power sector. ML algorithms can analyze vast amounts of data from transformers, including operational data, sensor measurements, and historical maintenance records. This data allows ML models to learn complex relationships between various parameters and the health state of a transformer. This enables the development of non-intrusive, continuous health monitoring systems, providing valuable insights for predictive maintenance.

This research explores the application of several ML algorithms for predicting the health index and remaining lifespan of power transformers. We focus on two key areas:

a. *Individual Model Evaluation*: The research investigates the performance of individual ML models like Random Forest, Support Vector Regression (SVR), Gradient Boosting, and Elastic Net. These models will be trained on historical data containing operational parameters, transformer characteristics, and measured health indicators. We will evaluate the effectiveness of each model in estimating the health index using statistical metrics like R-squared [3].
b. *Ensemble Learning for Improved Accuracy*: The research will explore the potential benefits of ensemble learning by implementing a stacking ensemble approach. This technique combines predictions from individual models to create a more robust and accurate health index estimation. The stacking ensemble leverages the strengths of each base model, potentially leading to improved performance compared to individual models.
c. *Remaining Lifespan Prediction:* The model with the highest R-squared value, signifying the best performance in health index estimation, will then be used for predicting the remaining lifespan of power transformers. The estimated health index serves as a key indicator of the transformer's current condition, allowing the model to extrapolate its remaining useful life. This information is invaluable for utility companies to make informed decisions regarding maintenance scheduling and potential transformer

replacement strategies.

This research contributes to the development of reliable and efficient methods for power transformer health assessment and lifespan prediction. By utilizing machine learning algorithms, we aim to move beyond traditional methods and provide a more comprehensive understanding of a transformer's health. This approach can empower utility companies to implement proactive maintenance strategies, promoting grid stability, reducing outage risks, and ensuring reliable power delivery.

II. LITERATURE REVIEW

Maintaining the health and longevity of power transformers is paramount for guaranteeing grid stability and efficient power transmission. Traditionally, periodic inspections and offline diagnostic tests have been the mainstay of this process. However, these methods have limitations. They can be time-consuming, resource-intensive, and often provide a one-time snapshot rather than continuous health monitoring [4].

The emergence of machine learning (ML) offers a transformative opportunity to address these limitations and establish a more proactive, data-driven approach to transformer health assessment [5]. This review explores current research on applying ML to transformer health assessment and lifespan estimation, with a specific focus on algorithms relevant to this project: Support Vector Regression (SVR), Random Forest, and Gradient Boosting.

Current Applications of Machine Learning in Transformer Health Assessment

a. *Dissolved Gas Analysis (DGA):* A widely used technique for identifying internal faults in transformers by analyzing dissolved gases within the insulating oil. Studies have shown promise in integrating ML algorithms with DGA data to improve fault diagnosis accuracy and predict remaining lifespan. For instance, research by previous authors employed SVR to analyze DGA data and achieved enhanced fault classification compared to traditional methods [6] [7].
b. *Partial Discharge (PD) Detection:* PD events within transformers indicate the beginning of insulation degradation. Research efforts have explored using Random Forest to classify PD patterns and assess the severity of internal faults, such as the work done by other researchers [8].
c. *Operational Data Analysis:* Analyzing operational data like load profiles, vibration measurements, and environmental parameters can provide valuable insights into transformer health. Studies have utilized Gradient Boosting to predict winding deformation in transformers based on operational data, contributing to lifespan estimation, as demonstrated in previous research [9].

Unveiling Opportunities and Addressing Gaps

a. While existing research demonstrates the potential of ML for transformer health assessment, there is a need for further exploration of comparisons between multiple algorithms and feature selection techniques. This paper's comparative analysis of SVR, Random Forest, Elastic Net and Gradient Boosting using R-squared as a metric directly addresses this gap.
b. Integrating diverse data sources such as DGA, operational, and environmental data into a unified health index framework presents an opportunity for comprehensive health assessment. This project's focus on developing a dynamic health index based on multiple features aligns with this promising direction.
c. Visualizing the health index profile over time for individual transformers can provide valuable insights and facilitate trend analysis. This research's objective to create graphical representations of the health index addresses this need for visual interpretation.

This review underscores the promising potential of ML for enhancing power transformer health assessment and lifespan estimation. This project, by utilizing SVR, Random Forest, and Gradient Boosting to develop a dynamic health index with visual representations, addresses key gaps in existing research and contributes to a more data-driven approach to transformer management.

III. METHODOLGY

Machine learning offers a powerful approach to assess power transformer health and predict remaining lifespan. This methodology leverages data from Kaggle [10]. Here's the key workflow:

a. During data pre-processing Anaconda environment is set up with pandas for data manipulation, matplotlib and seaborn for visualization, and scikit-learn for machine learning. We load the Kaggle dataset into a pandas Data Frame and explore its structure, data types, and summary statistics. Techniques like removing rows or implementing custom functions address missing values and outliers.
b. In Exploratory data analysis (EDA) heatmaps are created to visualize correlations between features and using scatterplots to understand how individual features influence transformer health. Feature scaling is then performed using standardization to improve model training.
c. After that splitting the pre-processed data into training and testing sets, We define and evaluate machine learning models like Elastic Net, random forest regression, and others. The model with the highest R-squared score, indicating superior health index prediction, is chosen.
d. Finally, relevant literature helps identify the established relationship between the predicted health index and remaining lifespan. This relationship, often involving formulas or degradation curves, is used with the predicted health index to estimate the remaining lifespan for transformers.

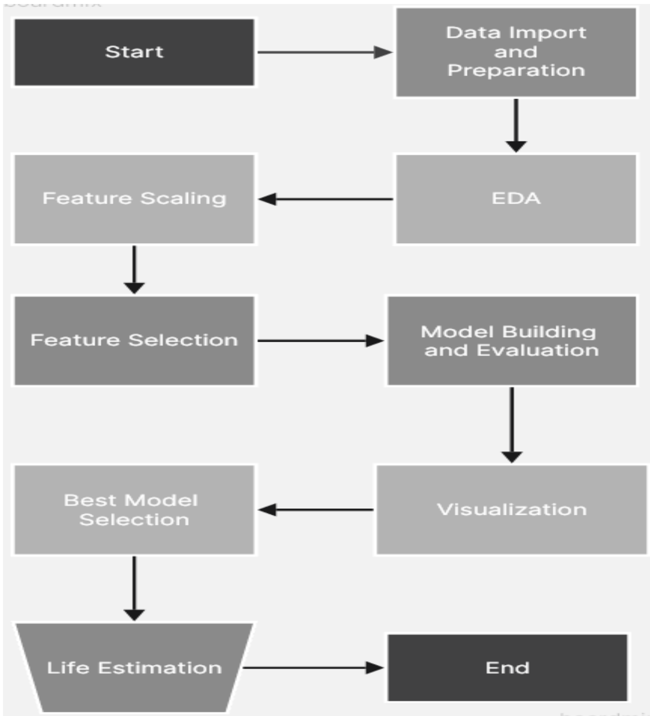

Figure 1: Flowchart of the process

Several machine learning models were created to assess their effectiveness in predicting power transformer health. These models include:

a. Elastic Net (model1) [11]:
   - Combines linear regression with regularization techniques (L1 and L2)
   - .L1 regularization encourages sparsity, potentially removing irrelevant features.
   - The chosen parameters (l1_ratio=0.8 and alpha=0.5) prioritize feature selection while maintaining some regularization strength.

b. Random Forest Regressor (model2) [12]:
   - An ensemble method that builds multiple decision trees and combines their predictions.
   - Robust to outliers and handles high-dimensional data effectively.
   - The configuration (n_estimators=150 and max_depth=5) balances model complexity with generalization.
   - Setting random state ensures reproducibility of results.

c. Support Vector Regression (SVR) (model3) [12]:
   - Uses a kernel function (here, radial basis function) to map data into a higher-dimensional space, enabling the capture of non-linear relationships.
   - Regularization parameter (C=15.0) controls the trade-off between fitting the data and model complexity.

d. Gradient Boosting Regressor (model4) [13]:
   - Another ensemble method that builds decision trees sequentially, each correcting the errors of the previous trees.
   - The configuration (n_estimators=100, max_depth=3, and learning_rate=0.1) focuses on reducing variance while controlling model complexity.

e. Stacking Regressor (model5) [14]:
   - An ensemble method that combines predictions from the models (model1, model2, model3, and model4).
   - Aims to leverage the strengths of each base model to potentially achieve better performance.
   - Utilizes linear regression as the final estimator for simplicity and interpretability.

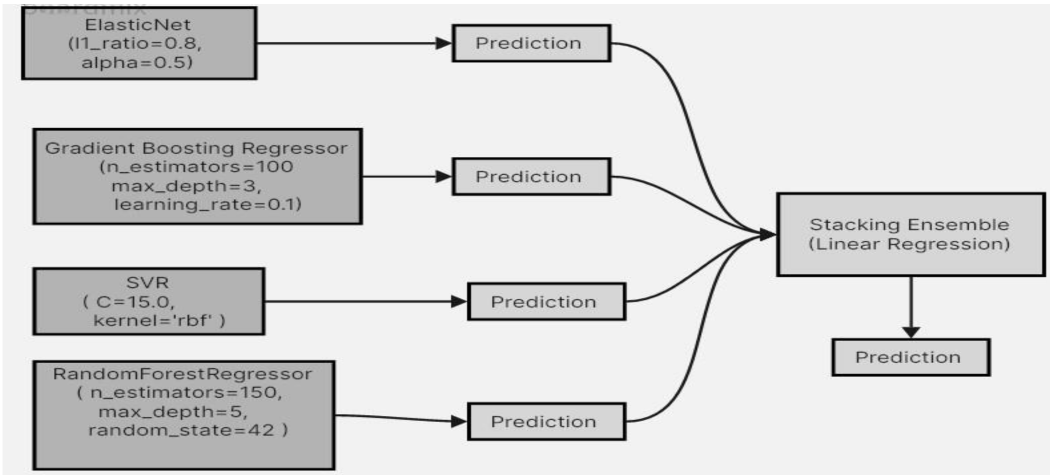

Figure 2: Machine Learning Models

The R-squared metric was used to evaluate each model's performance. The model with the highest R-squared score indicates the best fit between the predicted health indices and the actual health data. After selecting the best model, the predicted health indices can be used in conjunction with a pre-established relationship [15] to estimate the remaining lifespan of each transformer.

IV. RESULT AND DISCUSSION

A. Health Index Determination:

The results provided assess the performance of several models in predicting health index for power transformer. The R-squared score, ranging from 0 to 1, indicates how well a model explains the data's variance. Here, Model 2 (Random Forest) and a tie between Model 5 and 6 achieved the strongest individual fits, explaining over 71% of the variance. Model 1 (Elastic Net) and Model 4 (Gradient Boosting) showed moderate performance, while Model 3 (SVR) underperformed significantly.

Interestingly, the best overall performer was a Stacking Regressor. This technique combines predictions from multiple models (including Model 1, 2, 4, and potentially 5 and 6) to create a more robust final model. Since Models 5 and 6 achieved the highest individual R-squared scores, they likely played a significant role in the Stacking Regressor's success.

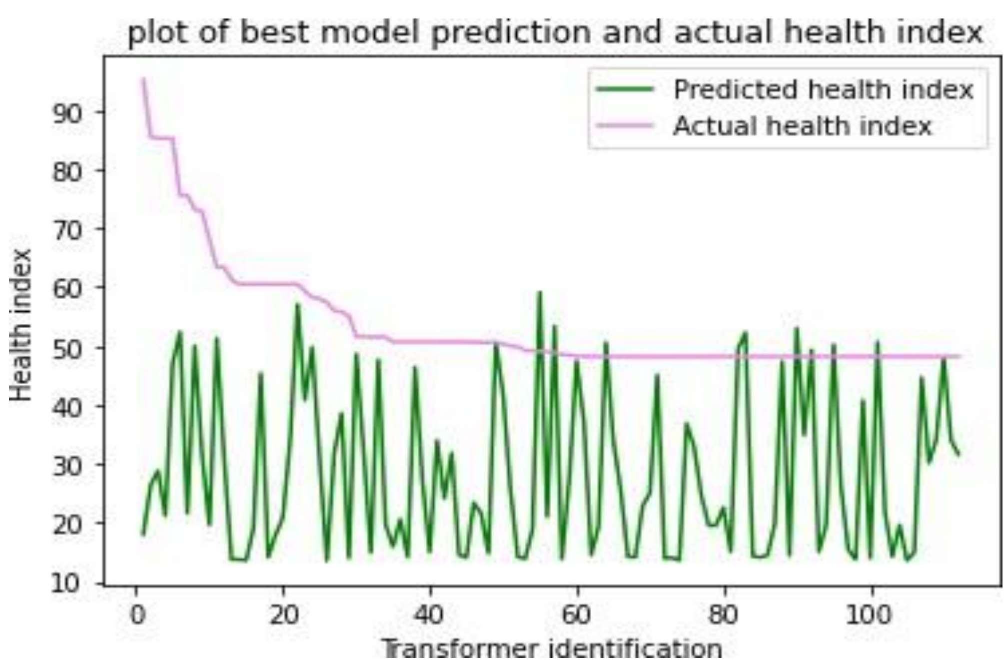

Figure 3 : Comparison of best predicted health index with already available health index

While several individual models showed promise, the Stacking Regressor emerged as the best option by leveraging the strengths of multiple models. Further analysis could explore how each model contributes to the Stacking Regressor's predictions and investigate ways to improve Model 3's performance.

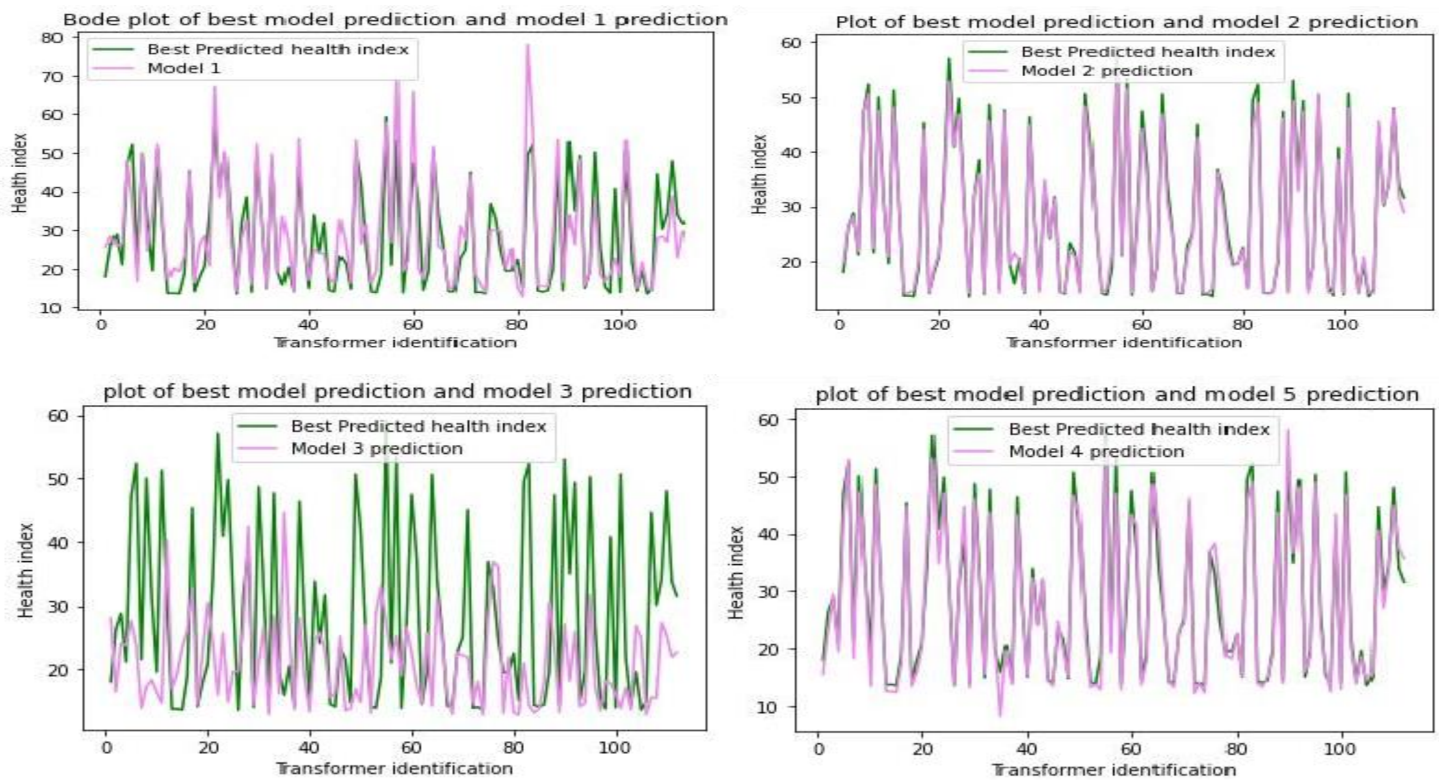

Figure 4: Comparison of the best predicted result with other results

B. Life Condition

Table 1 outlines a health grading system (A-E) for power transformers based on their estimated remaining lifespan. This system provides a clear and concise way to assess transformer condition and guide maintenance decisions. [15]

- **Category A: Minor Deterioration** - Transformers in this category exhibit minimal signs of wear and tear. They are expected to function reliably for an extended period with potentially only routine maintenance required.
- **Category B: Significant Deterioration** - These transformers show a more pronounced decline in their condition compared to Category A. While they may still be operational, closer monitoring and potentially some corrective maintenance might be necessary to ensure continued reliable operation.
- **Category C: Widespread Deterioration** - Transformers in Category C exhibit significant degradation across various components. Their remaining lifespan is considerably reduced, and preventative maintenance or refurbishment may be necessary to avoid unexpected failures.
- **Category D: Widespread Very Serious Deterioration** - This category signifies a critical stage. The transformer experiences severe degradation throughout its system, and its functionality is highly compromised. Immediate mitigation strategies, like replacement or emergency repairs, are likely required to prevent catastrophic failure.
- **Category E: Extensive Deterioration** - Transformers in Category E have suffered extensive damage and are deemed inoperable. Replacing the transformer is the most probable course of action.

Table 1: Transformer Grading System

Grade	Health Index	Remaining Life (in years)
A	85-100	>15
B	70-85	>10
C	50-70	>10
D	30-50	>3
E	0-30	>1

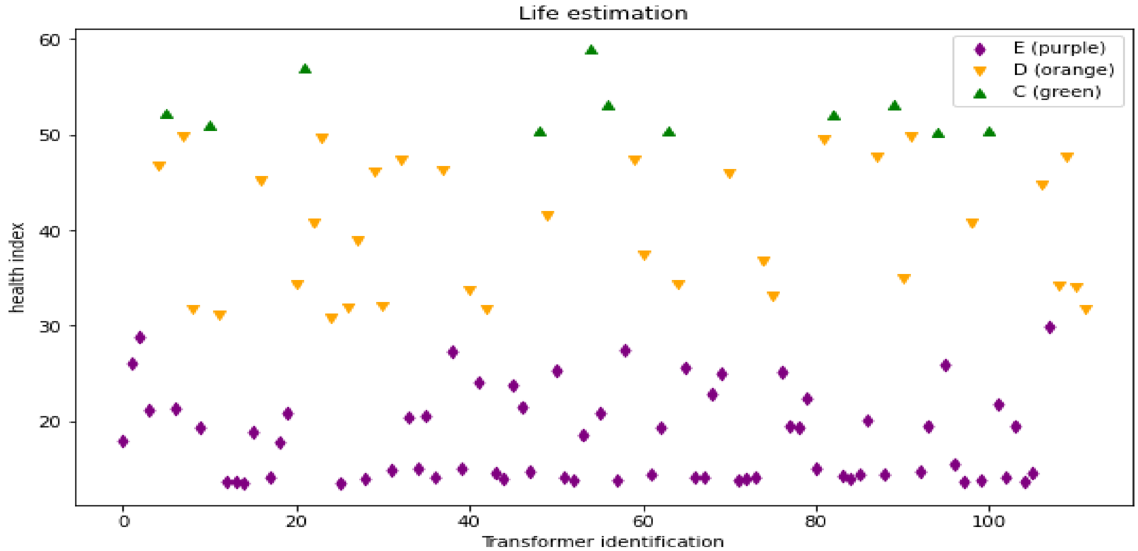

Figure5: Life Estimation of the Transformers

V. CONCLUSION

Our investigation aimed to assess the effectiveness of various machine learning models in predicting the health index of power transformers. The results revealed that the Stacking Regressor achieved the strongest performance, exceeding a fit of 71% (R-squared) by combining the strengths of individual models like Random Forest and potentially the top individual performers (Model 5 and 6). While these individual models also demonstrated promising capabilities, the Stacking Regressor's ability to leverage their combined strengths underscores the potential of ensemble techniques for power transformer health index prediction.

This study paves the way for further exploration in several key areas. Firstly, a more in-depth analysis could be conducted to elucidate how each individual model contributes to the Stacking Regressor's predictions. This would provide valuable insights into which models offer the most crucial information and potentially lead to optimizing the ensemble for even better performance. Secondly, investigating the reasons behind Model 3's (SVR) underperformance is crucial. Exploring alternative parameter tuning strategies or potentially using a different kernel function could significantly improve its accuracy and potentially make it a valuable contributor to the ensemble in future iterations.

Finally, integrating the Stacking Regressor into a real-world transformer health monitoring system would be a significant step forward. This would allow for real-time health index prediction and remaining lifespan estimation, facilitating proactive maintenance strategies and potentially preventing costly transformer failures. By addressing these future research avenues, researchers can further refine the model's accuracy and pave the way for its practical application in power grid management, ultimately enhancing grid reliability and efficiency.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to the Department of Electrical Engineering at Assam Engineering College. The department's unwavering support and access to a wealth of journal publications proved invaluable throughout this project.

REFERENCES

[1] J. H. Harlo “Electric Power Transformer Engineering”. 1st edition, 2003

[2] M. Bertolini, D. Mezzogori , M. Neroni, F. Zammori ,” *Machine Learning for industrial applications: A comprehensive literature review*” Expert Systems with Applications .Volume 175, 1 August 2021, 114820

[3] A. C. Cameron , F. A.G. Windmeijer “*An R-squared measure of goodness of fit for some common nonlinear regression models* “ Journal of Econometrics Volume 77, Issue 2, April 1997, Pages 329-342

[4] D. Glover ,T. Overbye , M. S. Sarma. “Power System Analysis and Design”J. . ISBN: 9789353502089. Edition: 6th, 2017

[5] A. Alqudsi and A. EI-Hag, “*Application of Machine Learning in Transformer Health Index Prediction*” *Energies* 2019, *12*(14), 2694

[6] H. Sun , Y. Huang , C. Huang, “*A Review of Dissolved Gas Analysis in Power Transformers”* Energy Procedia Volume 14, 2012, Pages 1220-1225

[7] K. I. Annapoorani, K. Udayakumar, B. Umamaheswari “*Diagnosis of fault in power transformer by dissolved gas analysis using support vector machine*” june-2017

[8] S. Wang, C. Ping,G. Xue , “Transformer Partial Discharge Pattern Recognition Based on Random Forest” March 2019, Journal of Physics Conference Series 1176(6):062025

[9] O. Núñez-Mata, F. Acuña, **C.** González, G. A. Gómez-Ramírez, “Assessment of Power Transformers using a Methodology Based on Health Indices” March 2021, Conference: IEEE ICA/ACCA2021 - International Conference on Automation

[10] https://www.kaggle.com/code/fatemehpanahandeh/power-transformer2

[11] J. K. Tay, B. Narasimhan, and T. Hastie; “*Elastic Net Regularization Paths for All Generalized Linear Models”* Published in final edited form as:J Stat Softw. 2023;

106: 1.Published online 2023 Mar 23

[12] K. Leauprasert, T. Suwanasri, C. Suwanasri and N. Poonnoy, "*Intelligent Machine Learning Techniques for Condition Assessment of Power Transformers*," 2020 International Conference on Power, Energy and Innovations (ICPEI), Chiangmai, Thailand, 2020, pp. 65-68

[13] A. I. Khalyasmaa, M. D. Senyuk and S. A. Eroshenko, "*Analysis of the State of High-Voltage Current Transformers Based on Gradient Boosting on Decision Trees*," in IEEE Transactions on Power Delivery, vol. 36, no. 4, pp. 2154-2163, Aug. 2021, doi: 10.1109/TPWRD.2020.3021702.

[14] H. Wau and D, Levinson, "The ensemble approach to forecasting: A review and synthesis" Tansportation Research Part C: Emerging Technologies Volume 132, November 2021, 103357

[15] A. N. Jahromi, R. Piercy, S. Cress. J. R. R. Service and W. Fan, "An Approch to Power Transformer Asset Management Using Health Index" IEEE Electrical Insulation Magazine-2009

ISSN 2250-3153

AUTHORS

First Author – Ajaan Anubhav Borah, M. Tech, Assam Engineering College, ajaananubhav523g@gmail.com

Second Author – Ritu Nazneen Ara Begum, Assistant Professor, Assam Engineering College ,aecritu7@gmail.com

. **Third Author** – Dr Barnali Goswami, Professor, Assam Engineering College and bgoswami.ele@aec.ac.in .

Correspondence Author – Ajaan Anubhav Borah, M. Tech, Assam Engineering College,