Advanced Health Index Prediction for Transformers Using CatBoost Algorithm

Kanaka Raju Kalla 1, B.Srinivasa Rao 2*, Gorriparthi Tagore Prasanth 3, Rama Harshavardhini 4 , Pallela Keertana 5, Sahithi Palle 6

[1,2,3,4,5,6] Aditya Institute of Technology And Management, Tekkali, Srikakulam, Andhra Pradesh 532201, India.

Abstract:

The reliability of power transformers is critical for ensuring stable and efficient energy distribution. Traditional transformer maintenance methods often rely on reactive strategies, leading to unexpected failures and costly downtime. This study proposes an advanced health index prediction model using the CatBoost algorithm to enhance predictive maintenance strategies. By leveraging key diagnostic indicators such as dissolved gas analysis (DGA) parameters, oil quality indicators, and electrical properties, the model effectively predicts transformer health status. The research explores various train-test split ratios and hyperparameter tuning to optimize model performance. Results indicate that a 75-25 train-test split yields the best predictive accuracy, with an R^2 of 0.759 and the lowest RMSE of 9.089. The findings highlight the effectiveness of CatBoost in handling categorical data, improving model interpretability, and reducing overfitting. This approach enables utilities to make proactive maintenance decisions, minimizing unexpected failures and extending the lifespan of transformers.

Key Words: Transformer Health Index, Predictive Maintenance, CatBoost Algorithm, Machine Learning, Dissolved Gas Analysis (DGA).

1. Introduction:

The Advanced Health Index Prediction for Transformers aims to leverage the power of CatBoost algorithm, which is known for their efficiency and accuracy in handling categorical data, to predict the health status and potential failures of transformer equipment [1]. By integrating machine learning techniques with domain-specific knowledge, this approach seeks to enhance predictive maintenance strategies and minimize downtime in electrical systems [2]. The implementation of this predictive model will not only improve the reliability of transformer operations but also facilitate proactive decision-making by providing insights into maintenance needs before failures occur [3]. This innovative model is expected to transform how utilities manage their assets, ultimately leading to cost savings and improved service continuity for consumers [4]. On transformer maintenance reveals that traditional methods often rely on reactive strategies, which can result in unexpected failures and costly repairs [5]. Transitioning to a predictive maintenance framework allows for continuous monitoring of transformer health, enabling utilities to address potential issues before they escalate into serious problems [6]. This shift not only enhances operational efficiency but also fosters a culture of safety and accountability within the utility sector, as stakeholders become more informed about the condition of their assets [7]. Implementing such a proactive approach can significantly reduce downtime and extend the lifespan of transformers, ensuring that utilities can deliver reliable power to their customers without interruption [8]. By leveraging advanced analytics and real-time data collection, utilities can gain deeper insights into the performance trends of their transformers [9], ultimately leading to more informed decision-making and resource allocation. This data-driven approach enables utilities to prioritize maintenance activities based on the health status of each transformer, thereby optimizing operational costs and enhancing overall system reliability. This not only minimizes the risk of unexpected failures but also fosters a culture of continuous improvement within the utility sector, as organizations adapt to evolving technologies and customer expectations [10]. Transformers play a crucial role in the electrical grid by stepping up or stepping down voltage levels to ensure efficient power distribution across long distances, thus maintaining stability and reliability in energy supply. Understanding the different types of transformers, such as distribution and power transformers, is essential for utilities to effectively manage their infrastructure and meet the demands of a growing energy landscape [11]. The ongoing advancements in transformer technology, including smart transformers and renewable integration, are set to further improve the efficiency and adaptability of power systems in response to evolving energy needs [12]. Objectives of this research include exploring the effectiveness of CatBoost algorithms in improving prediction accuracy of transformer health, assessing their impact on operational efficiency, and identifying best practices for integrating these advanced analytical tools into existing energy management systems [13]. This research aims to provide insights into the practical applications of CatBoost, highlighting its potential to transform energy management strategies and enhance sustainability efforts within the industry. The findings from this research could pave the way for innovative solutions that not only optimize resource allocation but also contribute to reduce downtime and extend the lifespan of transformers [14].

2. Literature Review:

The maintenance of transformers has evolved significantly over the years, transitioning from traditional reactive strategies to more proactive predictive maintenance frameworks. This shift is largely due to advancements in machine learning and data analytic, which have enabled utilities to monitor transformer health in real-time and make informed decisions regarding maintenance and resource allocation. Historically, transformer maintenance practices have relied on reactive strategies, where utilities respond to failures after they occur. This approach often leads to unexpected outages, costly repairs, and prolonged downtime, which can severely impact the reliability of power supply. Studies have shown that reactive maintenance can result in significant financial losses for utilities, highlighting the need for a more proactive approach [15]. In recent years, the assessment and prediction of transformer health have gained significant attention due to their critical role in power system reliability. Various methodologies have been explored, ranging from traditional statistical models to advanced AI-driven techniques. This review presents a comparative analysis of recent studies focusing on transformer fault diagnosis and health index prediction, highlighting their methodologies, results, and identified research gaps.

Author	Method	Results	Research Gap
[16]	SVM optimized by firefly algorithm with RBF kernel for transformer fault diagnosis.	Improved accuracy and fault prediction through optimized hyperparameters, enabling early fault detection.	Lack of integration with broader power system stability strategies.
[17]	Fuzzy logic-based health index using diagnostic test results.	Achieved 92% accuracy with data from 200 transformers, outperforming expert models.	Challenges in real-world implementation and integration with predictive maintenance systems.
[18]	Failure probability-based life prediction using health index and polynomial regression.	Established a mapping between health index and failure probability for life prediction.	Limited consideration of external factors influencing degradation.
[19]	SVM and ANN for transformer health detection using maintenance data.	SVM with nonlinear kernels showed high performance but overfitting issues.	Need for better generalization and optimal model selection strategies.
[20]	Regression and classification for health index prediction with missing data scenarios.	Gradient Boosting achieved 97.87% accuracy, addressing data unavailability.	Limited exploration of cumulative effects of missing data.
[21]	Fuzzy inference-based health assessment using multi-dimensional lifecycle data.	Accurate health ratings and maintenance guidance validated by field tests.	Incomplete coverage of condition indicators and reliance on expert input.
[22]	Health index integrating operational history and test data.	Improved maintenance planning and remaining life prediction.	Need for validation across diverse transformer types and conditions.
[23]	Comparing WSM, FIS, and combined methods for health index calculation.	Combined approach with K-means clustering provided consistent maintenance insights.	Lack of standardization in health index methods and integration strategies.
[24]	AI-based health index prediction using decision tree classifiers.	Achieved 96.3% accuracy, with insulation resistance as the key factor.	Limited exploration of other algorithms or hybrid approaches for improvement.

While substantial progress has been made in transformer health assessment through machine learning, fuzzy logic, and hybrid approaches, several challenges remain. Key research gaps include the need for better integration with predictive maintenance strategies, improved generalization of models, consideration of external influencing factors, and standardization of health index methodologies. Future research should focus on developing robust, scalable, and interpretable models that enhance real-world implementation and ensure comprehensive transformer health monitoring.

3. Proposed Approach

The Figure 1 visually represents the key components of transformer health prediction using the CatBoost algorithm. It highlights the essential aspects, including data sources, feature engineering, hyperparameter tuning, model performance metrics, and strategies for improvement**.** These elements collectively contribute to developing a robust predictive model for transformer health assessment. By leveraging CatBoost, the framework ensures effective handling of categorical data, reducing overfitting, and improving model interpretability. This structured approach enhances the accuracy and reliability of transformer health prediction.

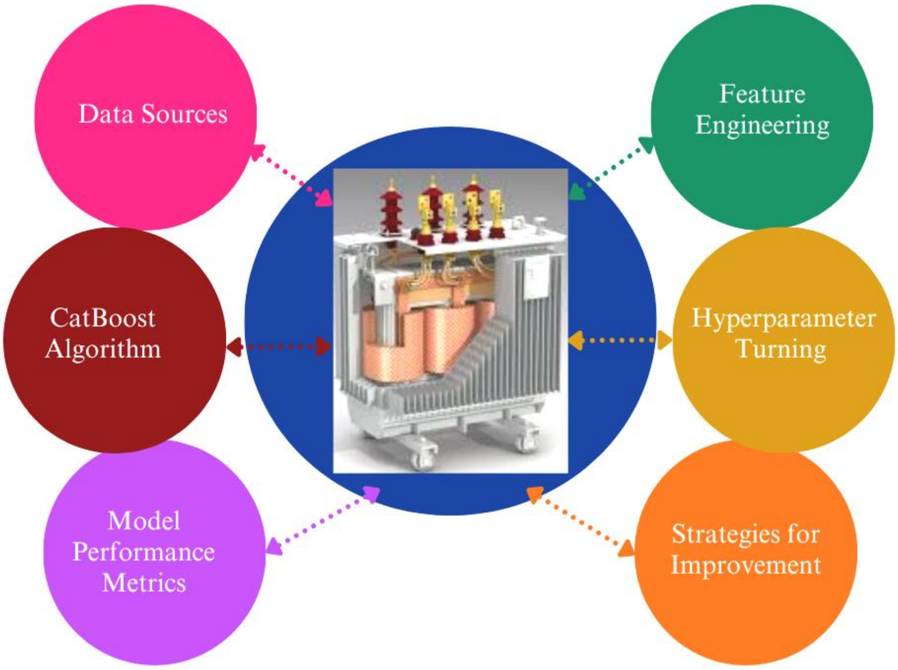

Figure 1. CatBoost-Based Transformer Health Diagnosis Framework

3.1. Data Source

The dataset used for this study is sourced from [25], which provides failure analysis data for power transformers. This dataset, available on Mendeley Data, contains crucial diagnostic information that aids in identifying root causes of transformer failures. By leveraging this data, the proposed CatBoost-based transformer health diagnosis framework enhances predictive accuracy through effective feature engineering, hyperparameter tuning, and performance evaluation.

3.2 Data Specification

The dataset used for transformer fault diagnosis consists of 470 records with 15 attributes, covering key diagnostic indicators for transformer health assessment. The attributes include:

1. Dissolved Gas Analysis (DGA) Parameters:

- Hydrogen, Oxygen, Nitrogen, Methane, Carbon Monoxide (CO), Carbon Dioxide (CO_2), Ethylene, Ethane, Acetylene

2. Oil Quality Indicators:

- Dibenzyl Disulfide (DBDS), Power Factor, Interfacial Voltage, Dielectric

Rigidity, Water Content

3. Target Variable:

- Health Index – A numerical value representing the overall condition of the transformer.

These features are crucial in identifying early signs of transformer degradation, enabling predictive maintenance through machine learning model CatBoost. The dataset ensures a comprehensive evaluation of transformer health by integrating gas concentrations, electrical properties, and oil quality parameters.

3.3 Flow chart

The Figure 2 represents flowchart of machine learning workflow for transformer health prediction using CatBoost. The process begins with data collection and preprocessing, including data normalization and correlation analysis. Various train-test splits (60-40 to 90-10) are explored to ensure robustness. The CatBoost model is then trained, and hyperparameter tuning is performed using GridSearchCV. Model performance is evaluated based on accuracy (R-squared) and error metrics (MAE, MSE, RMSE, RMSLE, MAPE).

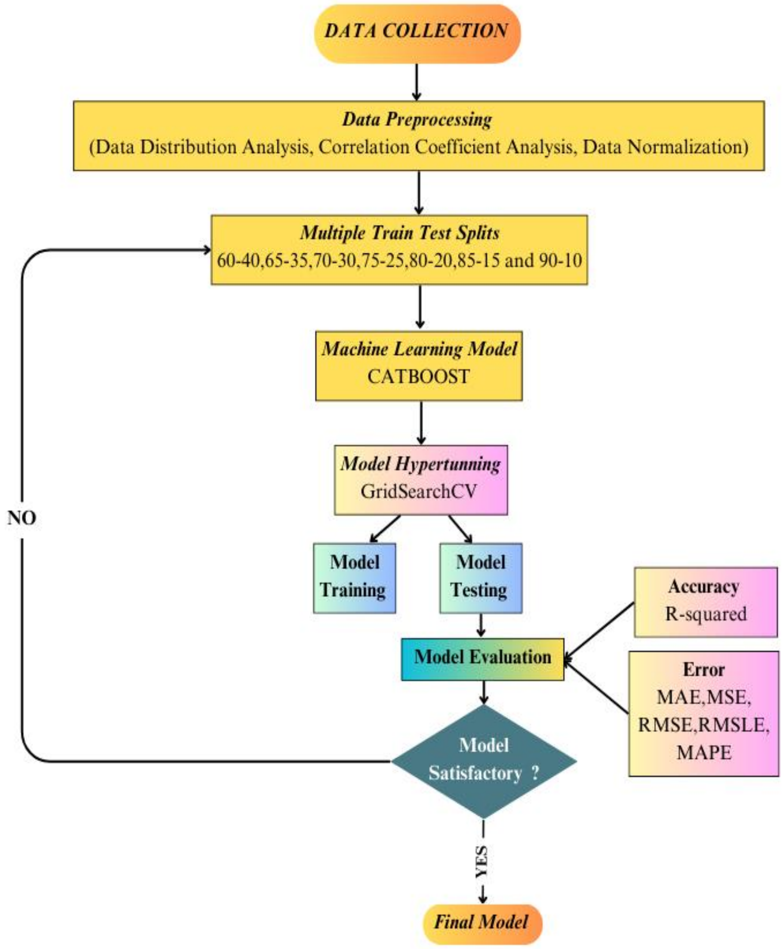

Figure 2. Flow Chart

3.4 Data Preprocessing

The dataset has no missing values. The Figure 3 is a pair plot visualization, which presents scatter plots of pair wise relationships between multiple variables along with their univariate distributions. This is useful for identifying correlations, patterns, and potential outliers in the dataset.

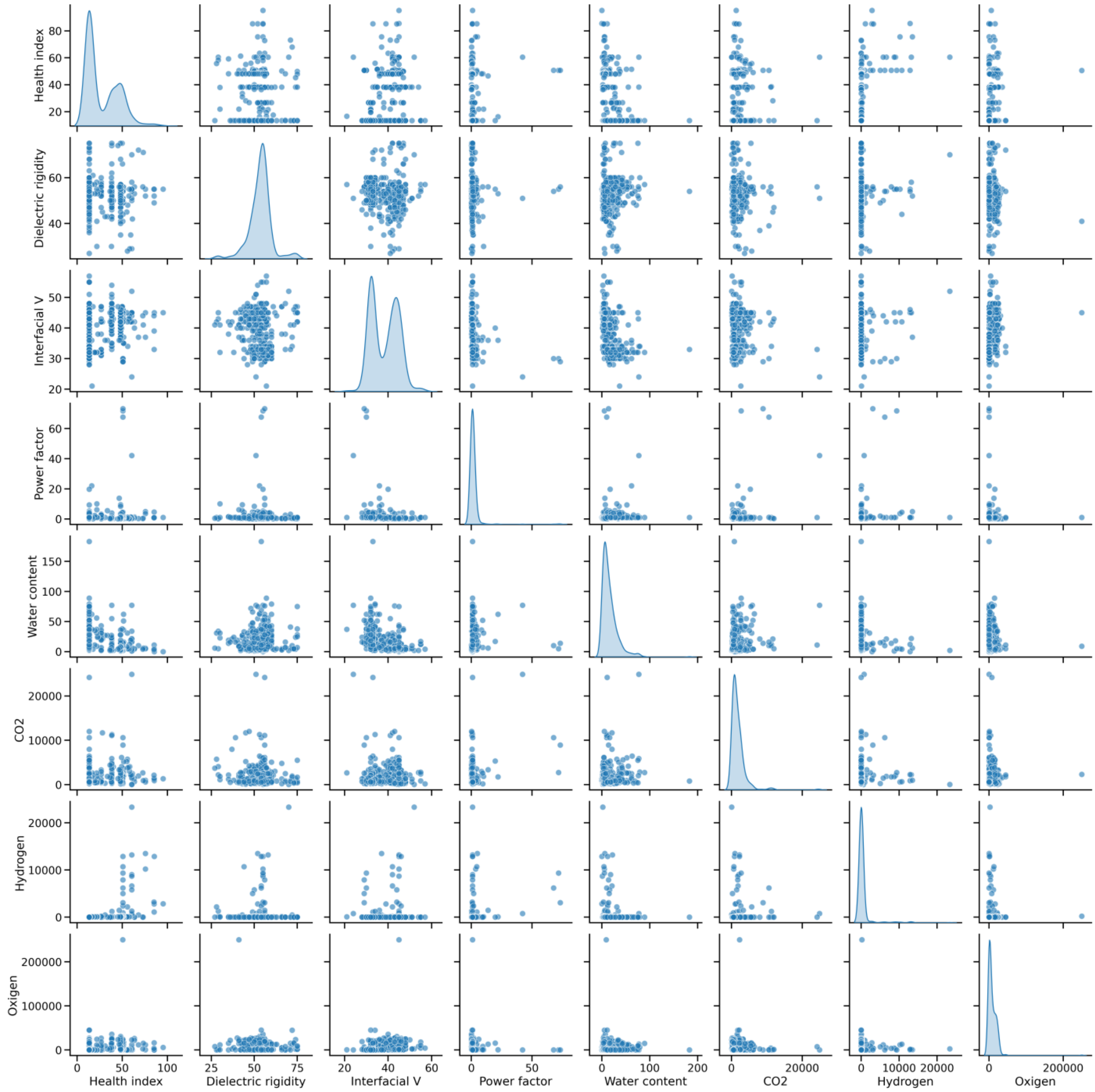

Figure 3. Pair Plot Visualization

The Figure 4 presents Distribution Analysis of Key Features, showcasing their distributions. This visualization is crucial for understanding the spread, skewness, and presence of outliers in the dataset. Notably, several features, such as Hydrogen, Oxygen, Methane, CO, CO_2, Ethylene, and Ethane, exhibit highly skewed distributions with most values concentrated near zero. In contrast, Nitrogen and Dielectric Rigidity follow a more normal-like distribution. These insights help in preprocessing steps like normalization or transformation to enhance model performance.

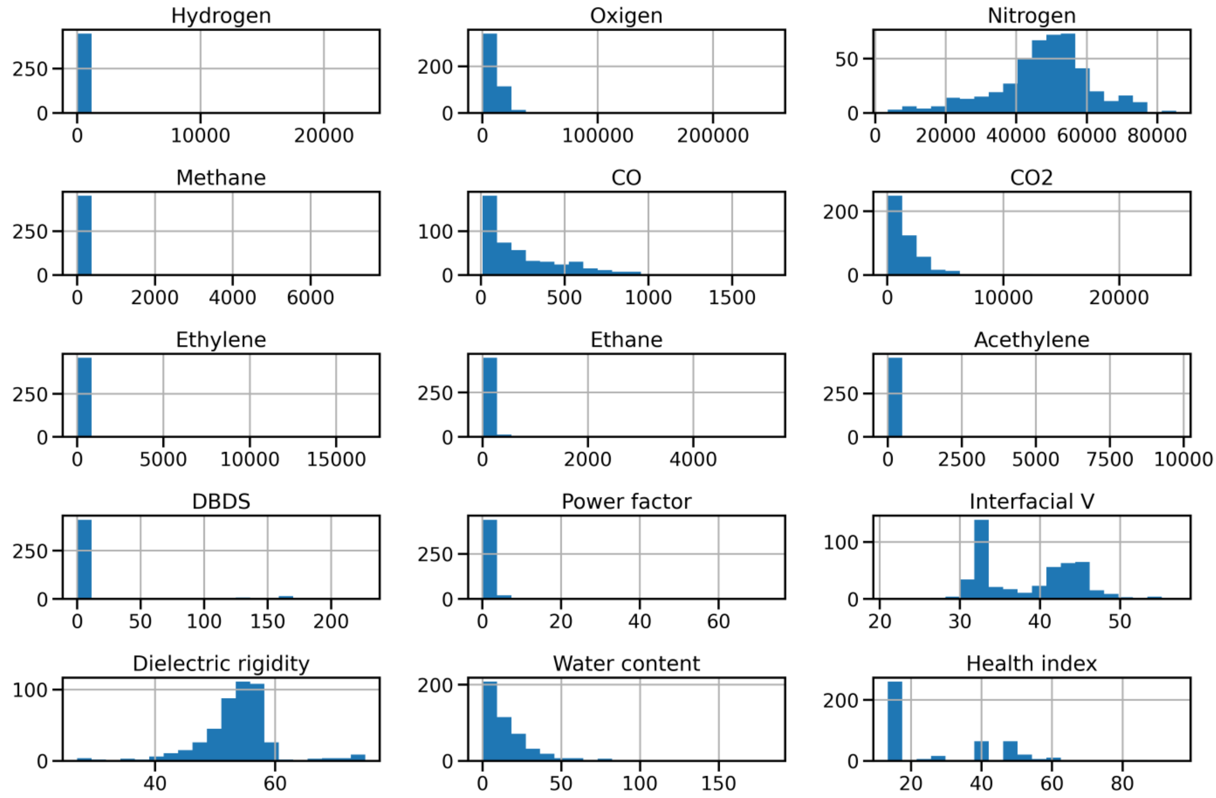

Figure 4. Distribution Analysis of Key Features

The Figure 5 represents Feature Correlation Heatmap, which illustrates the pair wise correlation coefficients between different features. From the heatmap:Methane and Ethane (0.91), Methane and Ethylene (0.80), and Ethylene and Ethane (0.76) show strong positive correlations, suggesting they may carry redundant information. DBDS and Health Index (0.47) suggest that DBDS might be an important feature for predicting health index. Interfacial V and Health Index (-0.40) and Water Content and Health Index (-0.28) show moderate negative correlations, indicating that higher interfacial voltage and water content might degrade the health index. This analysis helps in feature selection, reducing multicollinearity, and improving model interpretability.

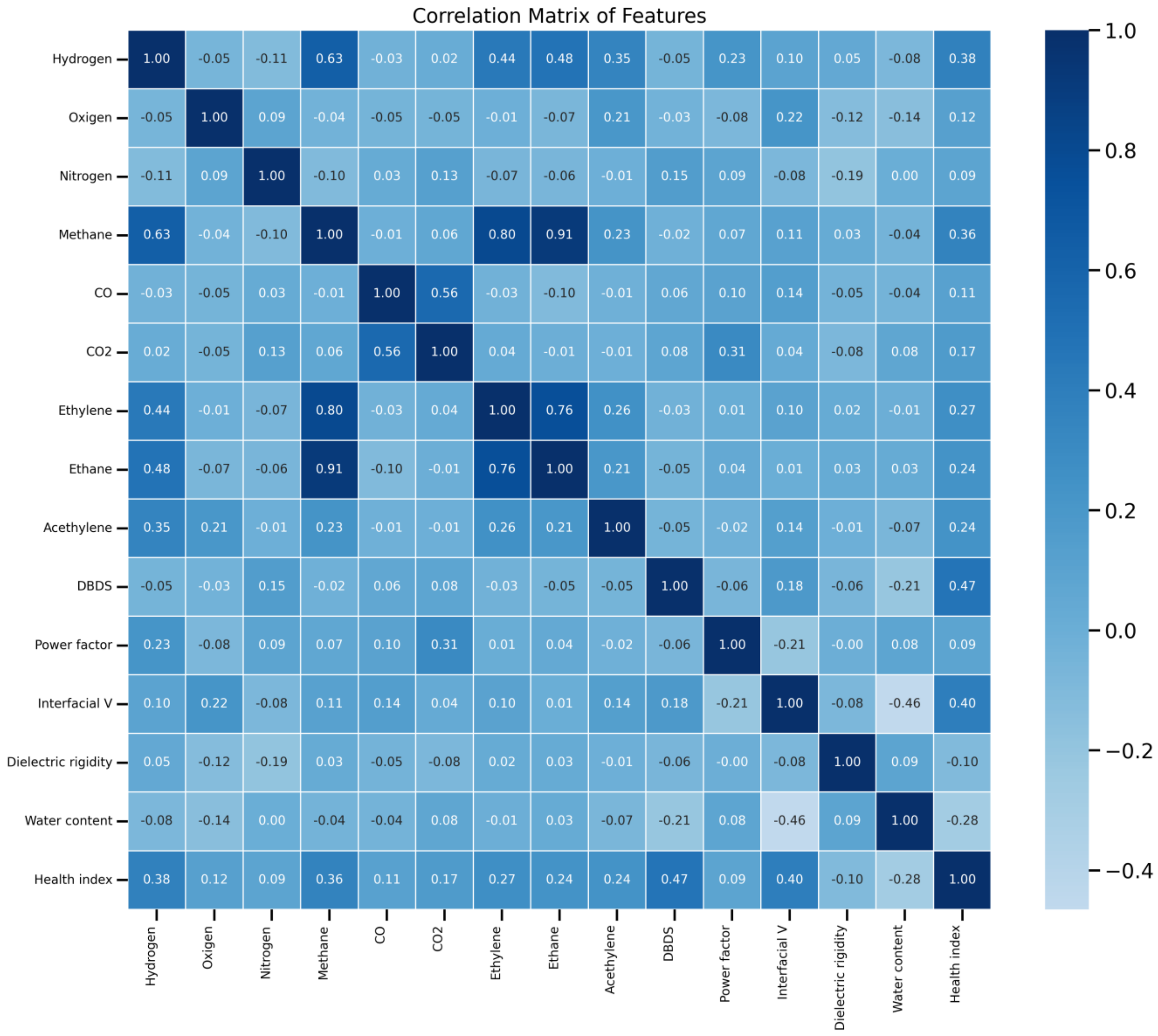

Figure 5. Feature Correlation Heatmap

4. Results and Discussion

The Table 1 presents the cross-validation scores for different train-test split ratios, with the mean cross-validation (CV) score serving as a measure of model performance stability across multiple training subsets. Observations include: The 60-40 split has the lowest mean CV score (0.6148), suggesting that using less training data leads to lower model generalizability. The 85-15 (0.6599) and 90-10 (0.6608) splits show the highest mean CV scores, indicating that increasing training data improves model performance. 80-20 (0.6351) and 75-25 (0.6184) splits demonstrate balanced performance. Variability in individual CV scores across different splits highlights fluctuations in model performance, emphasizing the need for a well-optimized train-test ratio to achieve stable predictions. These insights are critical for determining the optimal train-test split in transformer health prediction, ensuring that the model learns effectively while maintaining strong generalization.

Table 1 Cross-Validation Scores For Different Train-Test Split Ratios

Train Test Splits	Cross-Validation Scores	Mean CV Score
60-40	[0.762, 0.5685, 0.6987, 0.5168, 0.7386, 0.7342, 0.6015, 0.7037, 0.33, 0.4936]	0.6148
65-35	[0.6956, 0.5426, 0.804, 0.6385, 0.6902, 0.4849, 0.3823, 0.5958, 0.7286, 0.6144]	0.6177
70-30	[0.5914, 0.6363, 0.7013, 0.7059, 0.4718, 0.6578, 0.7447, 0.5444, 0.5378, 0.6259]	0.6217
75-25	[0.511, 0.474, 0.6391, 0.7378, 0.7939, 0.6006, 0.665, 0.6162, 0.6157, 0.531]	0.6184
80-20	[0.5043, 0.6695, 0.6931, 0.3378, 0.6277, 0.8055, 0.627, 0.7189, 0.6648, 0.7025]	0.6351
85-15	[0.6563, 0.6054, 0.5371, 0.6858, 0.6663, 0.8705, 0.7087, 0.6901, 0.7234, 0.4554]	0.6599
90-10	[0.7364, 0.6781, 0.6447, 0.7093, 0.668, 0.3763, 0.644, 0.7282, 0.6265, 0.7967]	0.6608

The Table 2 presents Optimal CatBoost Hyperparameters For Different Train-Test Split Ratios and corresponding best cross-validation (CV) scores for different train-test splits in training a CatBoost model. Key observations include: The best CV scores improve as the training data increases, with the 90-10 split achieving the highest score (0.6984), indicating that more training data enhances model performance. Across most splits, the optimal depth is 4, suggesting that a shallow decision tree structure effectively balances model complexity and generalization. The learning rate varies, with 0.1 being optimal for larger train sets (60-40, 65-35, 70-30, 85-15), while 0.05 is preferred for 75-25, 80-20, and 90-10, indicating that smaller train sets benefit from a lower learning rate to improve stability. The number of iterations increases (200) for 80-20 and 90-10, suggesting that more training data allows for deeper model learning. The L2 regularization parameter (l2_leaf_reg) shifts between 1 and 3, with higher values (3) appearing in 80-20 and 85-15, indicating a need for stronger regularization as training data increases. These insights help determine the optimal train-test split and hyperparameter tuning for maximizing CatBoost model performance in transformer health prediction.

Table 2 Optimal CatBoost Hyperparameters For Different Train-Test Split Ratios

Train Test Splits	Best Parameters	Best CV Score
60-40	{'depth': 4, 'iterations': 100, 'l2_leaf_reg': 1, 'learning_rate': 0.1}	0.64366241
65-35	{'depth': 4, 'iterations': 100, 'l2_leaf_reg': 1, 'learning_rate': 0.1}	0.650501917
70-30	{'depth': 4, 'iterations': 100, 'l2_leaf_reg': 1, 'learning_rate': 0.1}	0.683980674
75-25	{'depth': 4, 'iterations': 100, 'l2_leaf_reg': 1, 'learning_rate': 0.05}	0.659914026
80-20	{'depth': 4, 'iterations': 200, 'l2_leaf_reg': 3, 'learning_rate': 0.05}	0.674304996
85-15	{'depth': 4, 'iterations': 100, 'l2_leaf_reg': 3, 'learning_rate': 0.1}	0.673916667
90-10	{'depth': 4, 'iterations': 200, 'l2_leaf_reg': 1, 'learning_rate': 0.05}	0.698387144

The Table 3 compares the best cross-validation (CV) scores with the mean CV scores across different train-test splits to analyze model improvement. Key observations include: All train-test splits show improvement when using optimal hyperparameters, with the highest increase (10.02%) observed for the 70-30 split, indicating that hyperparameter tuning significantly enhances model performance at this ratio. The difference between the best and mean CV scores varies, with the largest gain (0.0623) occurring at 70-30, suggesting this split benefits the most from hyperparameter optimization. The lowest percentage improvement (2.12%) is at 85-15, implying that model performance is already near optimal and less sensitive to further tuning. The 90-10 split achieves the highest best CV score (0.6984), but its improvement percentage (5.69%) is moderate compared to 70-30. These findings support the selection of an optimal train-test ratio and emphasize the impact of hyperparameter tuning in maximizing the predictive accuracy of transformer health prediction.

Table 3 Comparison of the Best Cross-Validation (CV) Scores with the Mean CV Scores

Train-Test Splits	Best CV Score	Mean CV Score	Difference	% Improvement
60-40	0.6436	0.6148	0.0288	4.68
65-35	0.6505	0.6177	0.0328	5.31
70-30	0.6841	0.6217	0.0623	10.02
75-25	0.6599	0.6184	0.0415	6.71
80-20	0.6743	0.6351	0.0392	6.17
85-15	0.6739	0.6599	0.0141	2.12
90-10	0.6984	0.6608	0.0376	5.69

Table 4 Performance Metrics across Different Train-Test Split Ratios

Train-Test Splits	R^2	MAE	MSE	RMSE	RMSLE	MAPE	Comments
60-40	0.694	6.928	104.968	10.245	0.35	29.325	Low R^2 and high error metrics. Not optimal.
65-35	0.709	6.987	104.892	10.242	0.34	28.671	Slightly better R^2, but high MAE/MAPE.
70-30	0.746	6.242	93.512	9.67	0.32	24.084	Good overall balance; best MAPE and low errors.
75-25	0.759	6.315	82.602	9.089	0.31	26.512	Best R^2 and RMSE; very competitive.
80-20	0.755	6.064	83.989	9.165	0.31	25.564	Very close to Set 4; better MAE/MAPE.
85-15	0.73	6.469	95.07	9.75	0.31	26.106	Decent R^2, but higher errors than Sets 70-30,75-25.
90-10	0.744	6.62	103.997	10.198	0.32	24.504	Competitive MAPE, but lower R^2 and high errors.

Table 4 shows Performance Metrics across Different Train-Test Split Ratios, to determine the best set, we evaluate each metric with the following criteria: R^2: Higher is better. MAE, MSE, RMSE, RMSLE, MAPE: Lower is better. Best split is 75-25: Highest R^2 (0.759), lowest RMSE (9.089), and competitive RMSLE (0.31). This set performs best overall, particularly if R^2 and RMSE are prioritized. Split 70-30: Best for MAPE (24.084) and low errors overall, but slightly lower R^2. Split 80-20: Very close to Split 75-25, with better MAE and MAPE but slightly lower R^2 and RMSE. Train Test Split 75-25 is the best choice for overall performance. Figure 6. is showing Performance Metrics for Different Train Test Splits. Figure 7-13 shows scatter plots of different train test splits.

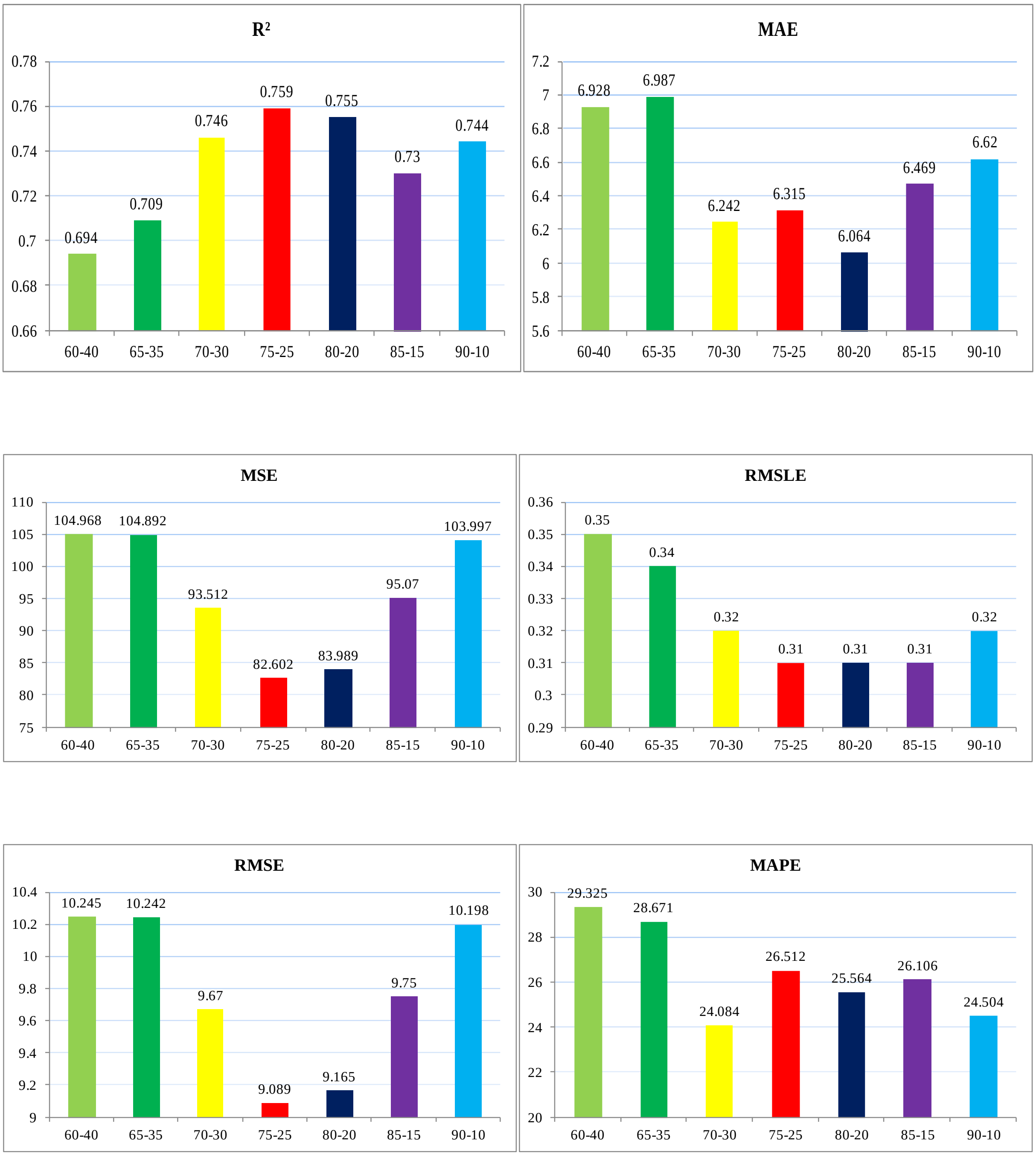

Figure 6. Performance Metrics for Different Train Test Splits

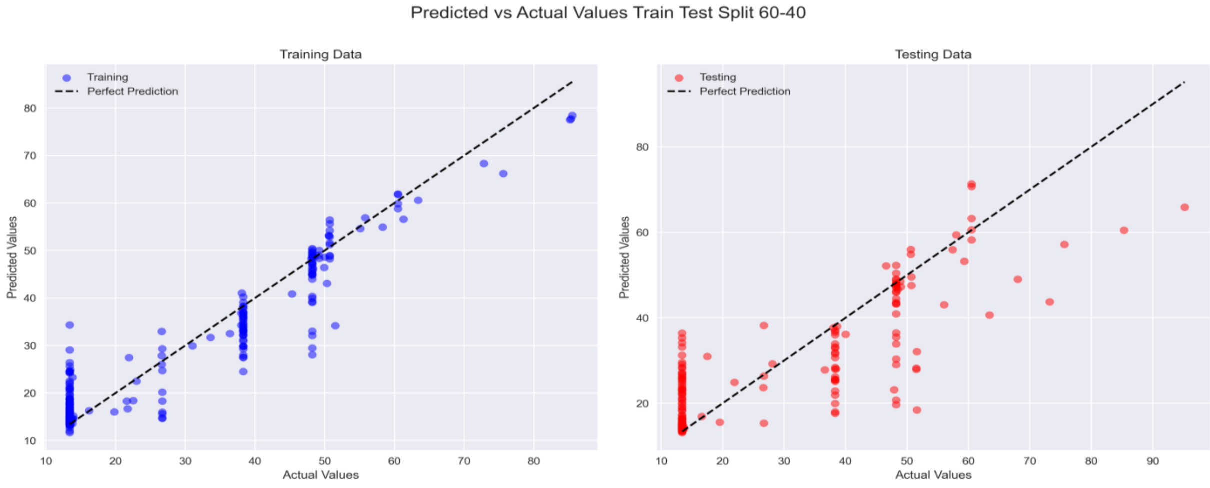

Figure 7. Scatter plot of Train Test Split 60-40

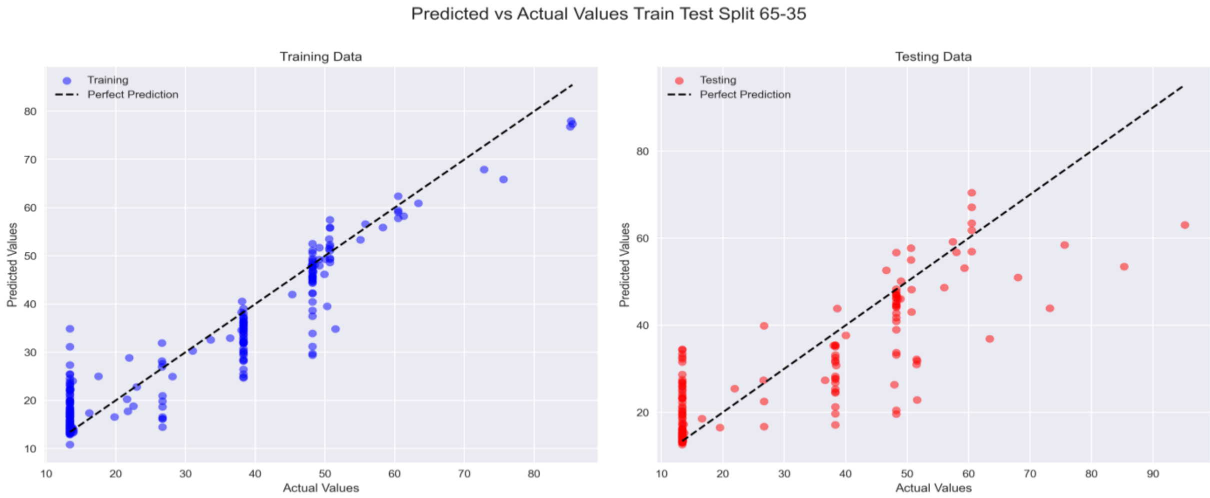

Figure 8. Scatter plot of Train Test Split 65-35

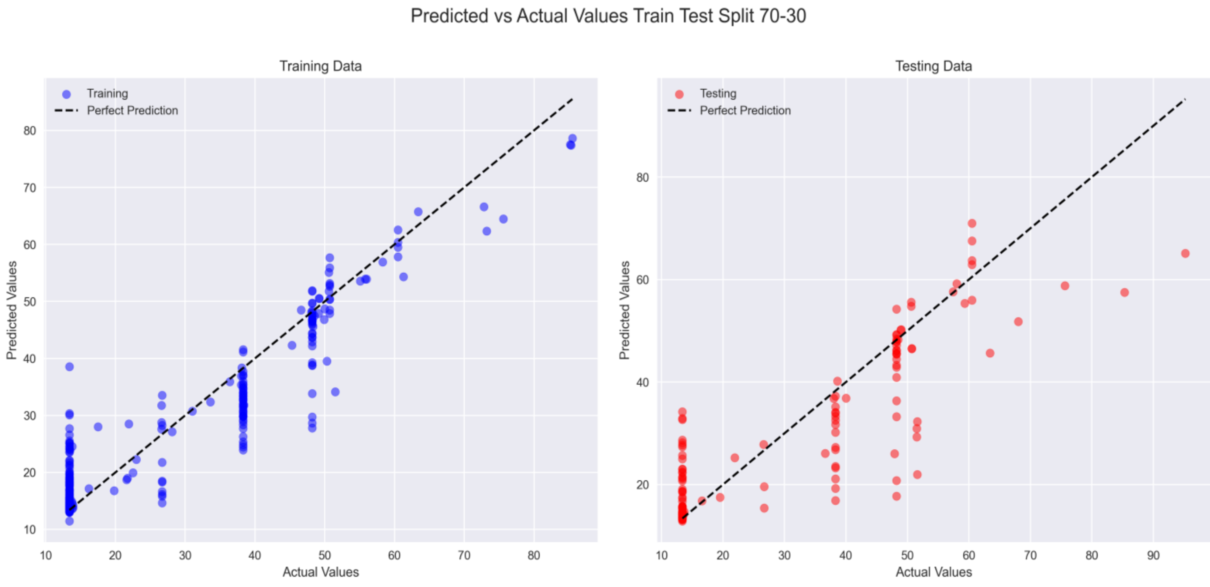

Figure 9. Scatter plot of Train Test Split 70-30

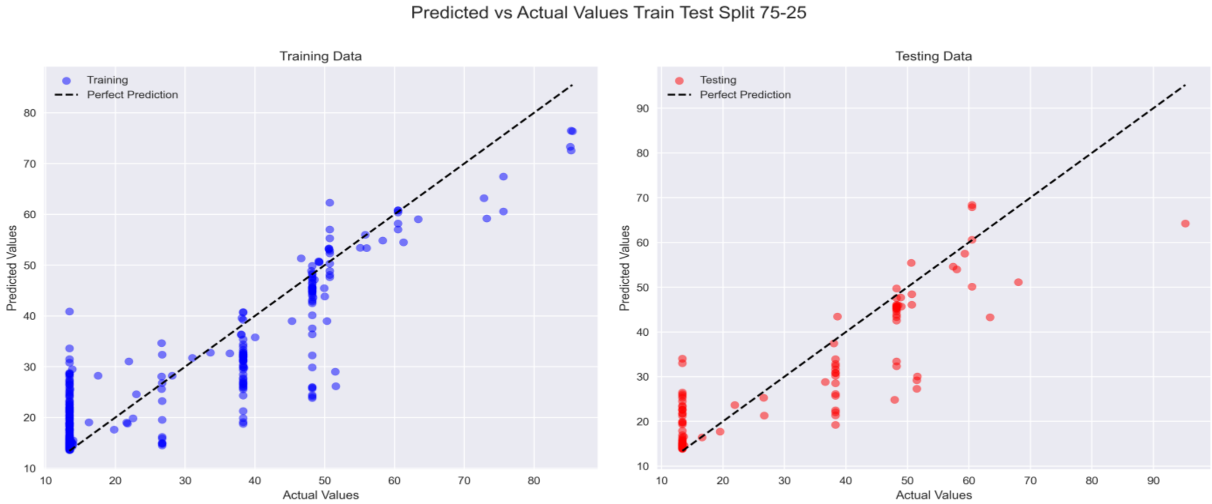

Figure 10. Scatter plot of Train Test Split 75-25

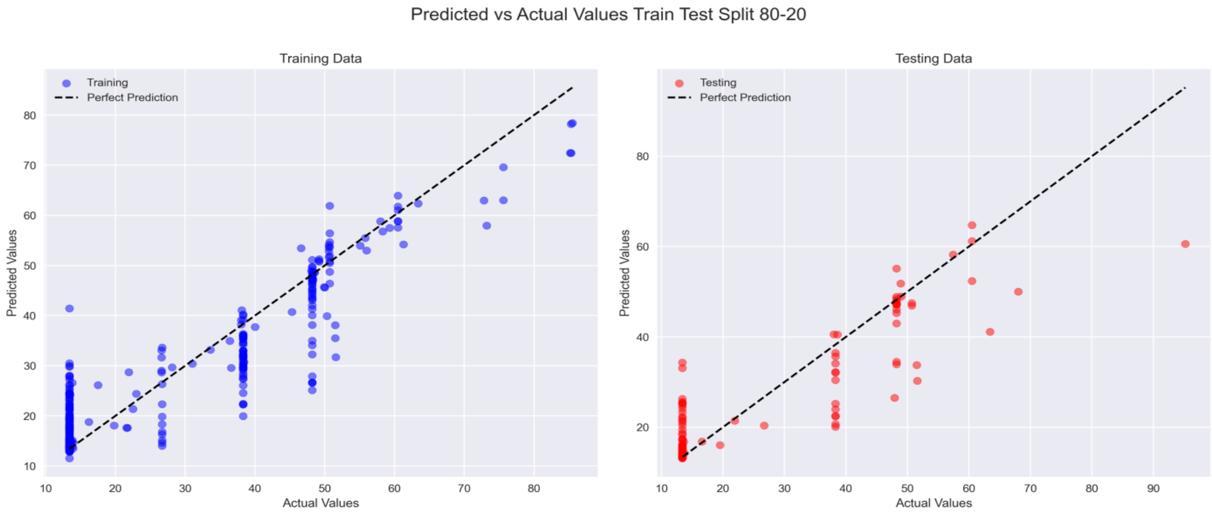

Figure 11. Scatter plot of Train Test Split 80-20

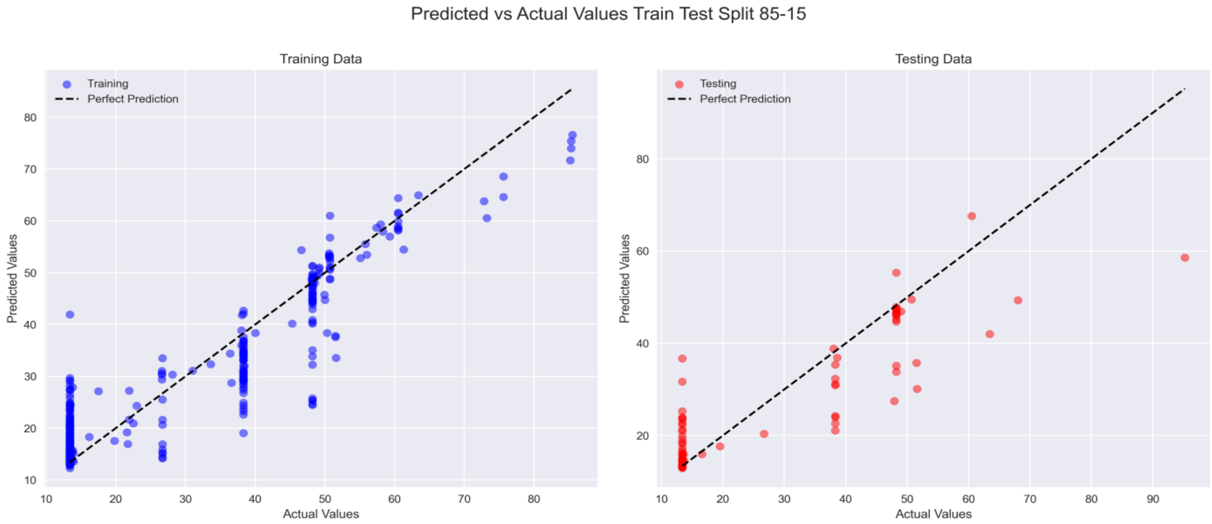

Figure 12. Scatter plot of Train Test Split 85-25

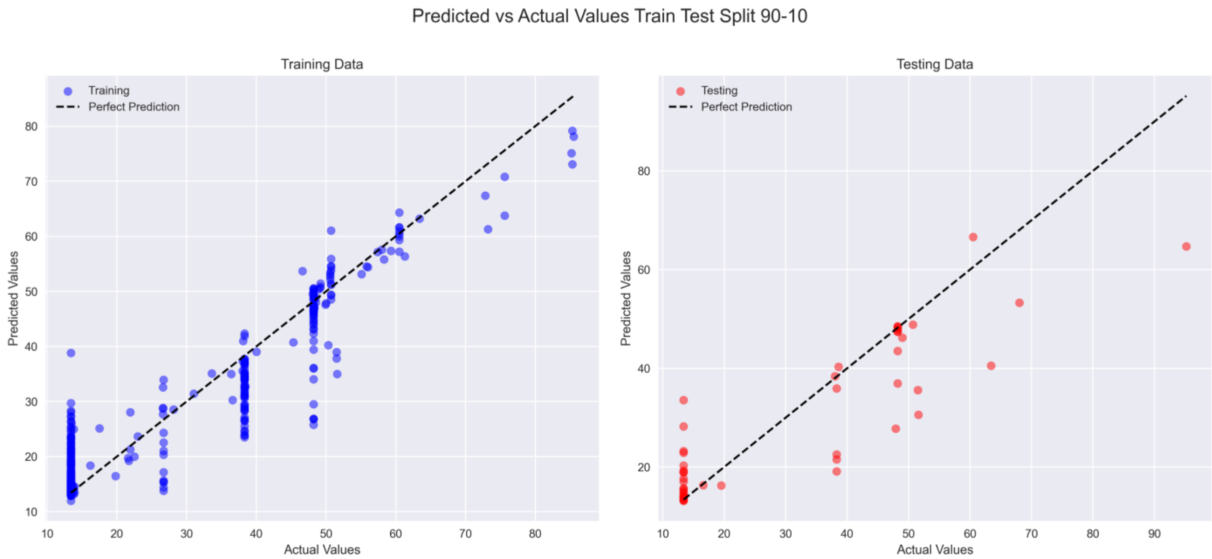

Figure 13. Scatter plot of Train Test Split 90-10

Conclusion:

The Advanced Health Index Prediction model for transformers demonstrates significant promise in revolutionizing maintenance strategies within the utility sector. Our analysis reveals that the optimal train-test split ratio of 75-25 provides the most reliable predictive performance, achieving an R^2 value of 0.759 and an RMSE of 9.089. The study shows varying levels of improvement across different split ratios, with the 70-30 split showing the highest percentage improvement (10.02%) after hyperparameter optimization, while the 90-10 split achieved the highest absolute CV score (0.6984). These findings underscore the importance of proper data partitioning and model tuning in developing robust predictive maintenance systems. The implementation of this CatBoost-based model represents a significant step forward in transitioning from reactive to predictive maintenance strategies, potentially leading to reduced downtime, extended transformer lifespan, and improved service reliability for utility providers. Future work could focus on incorporating real-time monitoring capabilities and expanding the model's applicability to diverse transformer types and operating conditions.

References

[1] M. Islam, G. Lee, S. N. Hettiwatte, and K. Williams, "Calculating a Health Index for Power Transformers Using a Subsystem-Based GRNN Approach," IEEE Transactions on Power Delivery, Aug. 2018, doi: 10.1109/TPWRD.2017.2770166.

[2] A. J. Alvares and R. R. Gudwin, "Integrated System of Predictive Maintenance and Operation of Eletronorte Based on Expert System," IEEE Latin America Transactions, Sep. 2019, doi: 10.1109/TLA.2019.8826707.

[3] A. J. C. Trappey, C. V. Trappey, L. Ma, and J. C. M. Chang, "Integrating Real-Time Monitoring and Asset Health Prediction for Power Transformer Intelligent Maintenance and Decision Support," 2015, Engineering asset management,doi: 10.1007/978-3-319-09507-3_46.

[4] Nordman, M.M. & Nordström, Lars. "Considerations on maintenance strategies for data communication assets in power distribution utilities" (2006). 7 pp.. 10.1109/PES.2006.1709504.

[5] Badawi, Mohamed & Ibrahim, Shimaa & Mansour, Diaa-Eldin & El-Faraskoury, Adel & Ward, Sayed & Mahmoud, Karar & Lehtonen, Matti & Darwish, M. M. F.. "Reliable Estimation for Health Index of Transformer Oil Based on Novel Combined Predictive Maintenance Techniques". IEEE Access (2022). 10. 25954-25972. 10.1109/ACCESS.2022.3156102.

[6] Kampourakis, Manthos & Fiotakis, Georgios & Livieris, Ioannis & Fourakis, Spyridon & Meintanis, Ioannis & Kokkorikos, Stefanos & Kyriakoulis, Nikolaos & Chondronasios, Apostolos.." A decision support system for online predictive maintenance of distribution transformers" (2022). 10.1109/ICECET55527.2022.9872875.

[7] D. V. Joao, P. Z. Lodetti, A. B. dos Santos, M. A. I. Martins, S. de Francisci, and J. F. B. Almeida, "A Smart Badge Implementation on Electrical Power Sector for Safety Improvement for Workforce - A Study Case," in IEEE PES Innovative Smart Grid Technologies Conference, Feb. 2021. doi: 10.1109/ISGT49243.2021.9372220.

[8] H. Ayar, H. Nhaila and M. E. Khaili, "Modeling and simulation of power transformer breakdowns to implement predictive maintenance," 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mohammedia, Morocco, 2023, pp. 1-8, doi: 10.1109/IRASET57153.2023.10152941.

[9] Youssef, Mohamed & Ibrahim, Rania & Desouki, Hussein & Moustafa, Mohamed. "An Overview on Condition Monitoring & Health Assessment Techniques for Distribution Transformers" (2022).. 187-192. 10.1109/ICGEA54406.2022.9791900.

[10] A. Arnaiz, E. Konde, and J. Alarcón, "Continuous Improvement on Information and On-line Maintenance Technologies for Increased Cost-effectiveness," Procedia CIRP, Jan. 2013, doi: 10.1016/J.PROCIR.2013.07.038.

[11] M.-S. Kang, P. Enjeti, and I. J. Pitel, "Analysis and design of electronic transformers for electric power distribution system," in IEEE Industry Applications Society Annual Meeting, Oct. 1997. doi: 10.1109/IAS.1997.629077

[12] G. Singh and I. Kaur, "Designing of a Smart Transformer," International Journal of Technology Enhancements and Emerging Engineering Research, Jun. 2015.

[13] A. Aziz, M. A. Talib, A. F. Abidin, and S. A. M. Al Junid, "Development of Power Transformer Health Index Assessment Using Feedforward Neural Network," Journal of Advanced Research in Applied Sciences and Engineering Technology, May 2023, doi: 10.37934/araset.30.3.276289

[14] L. O. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in Neural Information Processing Systems, Dec. 2018.

[15] H. Nhaila and M. El Khaili, "Modeling and simulation of power transformer breakdowns to implement predictive maintenance," May 2023. doi: 10.1109/IRASET57153.2023.10152941

[16] S. Abitha, M. Yuvarani, B. Hirija, and B. Vigneshwaran, "Intelligent Techniques based Power Transformer Health Monitoring Index using Dissolved Gas Analyser," pp. 373–377, Sep. 2024, doi: 10.1109/icosec61587.2024.10722324.

[17] T. Manoj and C. Ranga, "Multi-parameter-based fuzzy logic diagnostic system for health assessment of power transformer," Australian journal of electrical and electronics engineering, May 2024, doi: 10.1080/1448837x.2024.2356436.

[18] Z. Wang, W. Ma, Z. Ma, W. Chai, and J. Yu, "Transformer Life Prediction Based on Failure Probability and Health Index," Apr. 2024, doi: 10.1109/ceepe62022.2024.10586550.

[19] A. Er, W. K. Wong, F. H. Juwono, I. M. Chew, S. Sivakumar, and A. P. Gurusamy, "Automated Transformer Health Prediction: Evaluation of Complexity and Linearity of Models for Prediction," Springer Science+Business Media, 2024, pp. 21–33. doi: 10.1007/978-981-99-9833-3_3.

[20] N. R. Arsya, S. Suwarno, R. A. Prasojo, G. A. Sudrajad, and A. Abu-Siada, "A Holistic Integration of Conventional and Machine Learning Techniques to Enhance the Analysis of Power Transformer Health Index Considering Data Unavailability," IEEE Access, vol. 12, pp. 124549–124561, Jan. 2024, doi: 10.1109/access.2024.3450110.

[21] L. Xie, Z. Jiang, L. Feng, C. Chu, Z. Huang, and X. Liu, "Transformer Health Condition Assessment Method Based on Full Life Cycle Data," Springer Science+Business Media, 2023, pp. 317–325. doi: 10.1007/978-981-99-7413-9_29.

[22] C. Barriga, "Health index for power transformer condition assessment based on operation history and test data," Energy Reports, vol. 8, pp. 9038–9045, Nov. 2022, doi: 10.1016/j.egyr.2022.07.041. Available: https://doi.org/10.1016/j.egyr.2022.07.041.

[23] D. A. Zaldivar and A. A. Romero, "Health Index for Power Transformer Condition Assessment: A Comparison of Three Different Techniques.," Journal of applied research and technology,vol.20, no. 5, pp. 536–545, Oct. 2022, doi: 10.22201/icat.24486736e.2022.20.5.1606.

[24] I. Ahmad, M. U. Jamil, and K. N. Paracha, "Artificial Intelligence-Based Approach for Prediction of Power Transformer Health Index," pp. 1–4, Oct. 2022, doi: 10.1109/ICPEI55293.2022.9986545.

[25] Arias, Ricardo; Mejia Lara, Jennifer (2020), "Data for: Root cause analysis improved with machine learning for failure analysis in power transformers", Mendeley Data, V1, doi: 10.17632/rz75w3fkxy.1.