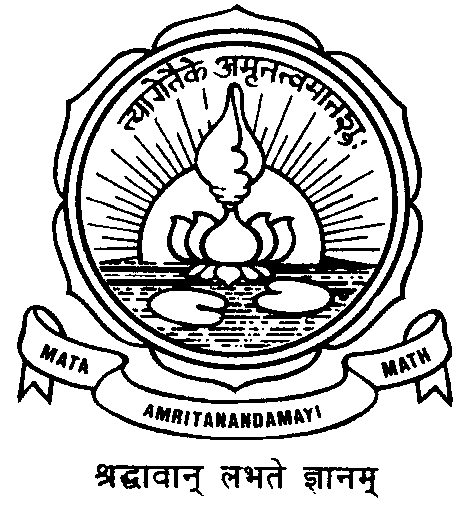
**AMRITA VISHWA VIDHYAPEETHAM**

(University Established U/S 3 of the UGC Act 1956)

**AMRITA SCHOOL OF ENGINEERING, BENGALURU**

**PROJECT REPORT**

Academic Year 2025 [Jan-May]

Department of Electrical and Electronics Engineering

**Title:** Predictive Transformer Health Assessment using a Machine Learning Approach for Condition Monitoring via Oil Composition Analysis

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Team Members** | **Roll Number** |
| 1. | Adina Sree Venkat Utham Kumar | [BL.SC.U4AIE24102] |
| 2. | Ankur De | [BL.SC.U4AIE24105] |
| 3. | Jiya Sachdeva | [BL.SC.U4AIE24123] |
| 4. | Mamidi Sai Tanushka | [BL.SC.U4AIE24130] |

**ABSTRACT**

This research introduces a novel approach to power transformer condition monitoring through the application of machine learning techniques on dissolved gas analysis (DGA) and oil composition data. Transformer failures can lead to substantial economic losses and power grid disruptions, necessitating reliable prognostic methodologies. Our study develops a comprehensive machine learning framework analysing a diverse dataset of 472 transformer oil samples. The model evaluates multiple compositional parameters including dibenzyl disulfide (DBDS), interfacial tension, dielectric constant, dissolved gases (hydrogen, oxygen, acetylene, methane, carbon dioxide), moisture content, acidity, and viscosity to predict two critical outputs: transformer health index values and remaining life expectancy. We employed ensemble machine learning methods to capture complex parameter interactions and identify deterioration patterns across different transformer designs and operational environments. The resulting software solution demonstrates significant potential for asset management application. Future development will include real-time monitoring capabilities through IoT sensor integration, a mobile application interface for field technicians, automated report generation with customizable thresholds, a comparative analytics dashboard for fleet-wide assessment, integration with maintenance management systems, and a recommendation engine for optimized maintenance scheduling. Additionally, the software architecture is being designed to incorporate federated learning capabilities to enable knowledge sharing across utilities while maintaining data privacy. This scalable software project aims to transform traditional transformer maintenance practices by providing a comprehensive digital solution for condition-based maintenance strategies.

**KEYWORDS**

Power transformers, condition monitoring, machine learning, dissolved gas analysis, oil composition, predictive maintenance, asset health management, life expectancy prediction, software development, IoT integration

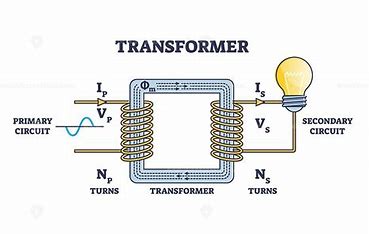
**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Topic discussed** | **Page Numbers** |
| 1. | Introduction | 4-7 |
| 2. | Literature Review | 8-12 |
| 3. | Machine Learning | 13-22 |
| 4. | Mathematical Ideation | 23-26 |
| 5. | Final Output | 27 |
| 6. | Future Scope | 28 |
| 7. | Conclusion | 29 |
| 8. | Acknowledgment | 30 |
| 9. | Bibliography | 31 |
| 10. | Appendix | 32-34 |

INTRODUCTION

# What are Transformers?

Transformers are electrical devices that transfer energy between circuits through electromagnetic induction. They consist of two or more coils (windings) wrapped around a core, usually made of laminated steel. When alternating current flows through the primary winding, it creates a changing magnetic field in the core, inducing voltage in the secondary winding. The voltage ratio between primary and secondary is determined by the turns ratio of the windings, allowing transformers to step voltage up or down while maintaining the same frequency.



# Turns Ratio Formula

N₁ = Number of turns in the primary winding

N₂ = Number of turns in the secondary winding

V₁ = Primary voltage

V₂ = Secondary voltage

I₁ = Primary current

I₂ = Secondary current

This equation demonstrates that the ratio of primary to secondary voltage is directly proportional to the ratio of primary to secondary turns, while the current ratio is inversely proportional to maintain power conservation (ignoring losses).

## For a step-up transformer

## For a step-down transformer

# Types of Transformers

## By Function:

* **Power Transformers:** Used in transmission networks for bulk power transfer
* **Distribution Transformers:** Convert distribution voltage to utilization voltage
* **Instrument Transformers**: Include current (CT) and potential (PT) transformers for measurement

## By Cooling Method:

* **Oil-Immersed:** ONAN (natural cooling), ONAF (forced air), OFAF (forced oil and air)
* **Dry-Type:** Air-cooled without oil, used in indoor environments

## By Phases:

* **Single-Phase:** Used in residential applications
* **Three-Phase:** Used in industrial applications and power transmission

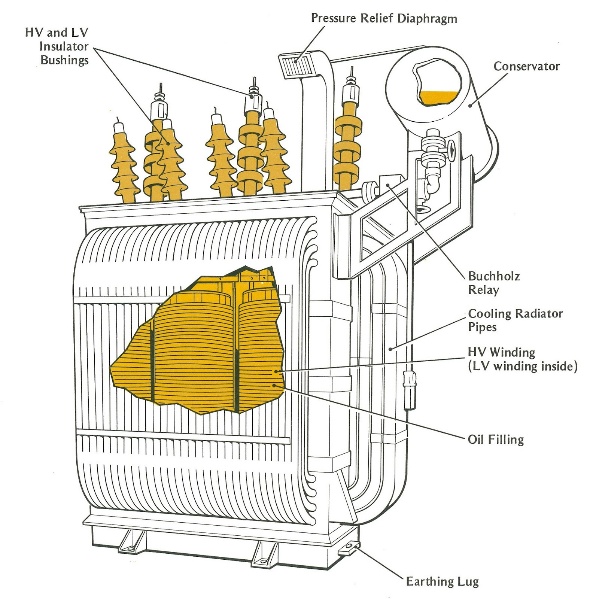
# Oil-Based Transformer Operation

Oil-immersed transformers use insulating oil that serves dual purposes: electrical insulation and cooling. The oil fills the tank containing the core and windings, providing high dielectric strength while absorbing and transferring heat.

Key components include:

* **Conservator:** Accommodates oil expansion/contraction
* **Buchholz Relay:** Detects gas formation from internal faults
* **Radiators:** Increase surface area for cooling
* **Breather:** Filters incoming air to prevent moisture

Oil circulates either naturally (thermosiphon effect) or through forced circulation (pumps), carrying heat from windings to the cooling surfaces.



**Real life model**   **Internal Structure**

# Causes of Transformer Failures / Blasts

1. **Electrical Faults:** Short circuits, insulation breakdown, or lightning strikes
2. **Oil Degradation:** Moisture contamination, oxidation creating sludge, or gas formation
3. **Mechanical Issues:** Core/winding movement or gasket failures
4. **Operational Problems:** Overloading, cooling system failures, or external fires

These issues can lead to catastrophic failures including explosions when arcing occurs in oil, rapidly decomposing it into explosive gases.

# Prevention Measures

1. **Regular Testing:** Dissolved gas analysis (DGA), oil quality tests, electrical tests
2. **Protective Equipment:** Differential protection, Buchholz relays, pressure relief devices
3. **Monitoring Systems:** Temperature, oil level, and gas monitoring
4. **Proper Loading:** Maintaining operation within rated capacity
5. **Oil Treatment:** Regular filtering and reconditioning

Oil-based transformers remain the standard for high-power applications and outdoor installations despite their risks, due to their reliability, efficiency, and cost-effectiveness. Proper maintenance and monitoring are essential to prevent failures and extend service life.

LITERATURE REVIEW

This serves as the review for all papers referred for the making of this project

# Paper 1

* Proposes an edge computing-based real-time transformer health monitoring system.
* Focuses on oil colour detection as a critical indicator of insulation degradation using ESP32 microcontroller and TensorFlow for image analysis.
* Shifts computational load to edge devices, reducing latency and reliance on centralized servers.
* Enhances scalability, minimizes network congestion, and improves maintenance, especially in remote locations.
* Promotes cost-effective and efficient monitoring for improved grid reliability.

**Technologies Used**: TensorFlow, Edge Computing, ESP32, Machine Learning, Deep Learning, Google Cloud, IoT Applications, Microgrid, Colour Detection.

# Paper 2

* Highlights the need for improved methodologies in transformer health index (HI) evaluation.
* Reviews limitations of traditional oil sample analysis methods, often reliant on small datasets and expert interpretation.
* Advocates for a hybrid AI-based approach combined with Subject Matter Expertise (SME) for better accuracy.
* Proposes a solution leveraging 65,600 oil samples with an accuracy >95% in predictive maintenance.
* Recommends integrating real-time analyses for advanced monitoring capabilities.

**Technologies Used**: Machine Learning Algorithms, Random Forest Classifier, Anomaly Detection, Hotelling's T-squared Test, Feature Importance Analysis, Risk Management.

# Paper 3

* Focuses on integrating LM35 sensors for transformer temperature monitoring with diagnostic tools like dissolved gas analysis (DGA).
* Highlights LM35's simplicity, cost-effectiveness, and suitability for real-time monitoring.
* Explores correlation between temperature variations and dissolved gas concentrations in transformers.
* Identifies challenges such as sensor noise/drift and integration costs but advocates machine learning (e.g., ANN, SVM) for predictive diagnostics.
* Emphasizes combining multiple parameters for enhanced transformer health assessment.

**Technologies Used**: IoT, LM35 Sensors, Arduino Microcontroller, Data Processing, GSM, Wi-Fi, Bluetooth, Machine-to-Machine Communication.

# Paper 4

* Highlights the critical role of power transformers in power systems and their vulnerability to degradation over time.
* Traditional diagnostics rely on infrequent testing (6–12 months), which may not provide timely fault detection.
* Emphasizes the potential of integrating sensor data (e.g., PMUs, DGA monitors) for advanced diagnostics.
* Stresses the need for AI and ML-based methods for accurate health assessment and timely fault prediction.
* Identifies a lag in the adoption of AI/ML in transformer diagnostics, calling for improved techniques.

**Technologies Used**: Machine Learning, Probabilistic Logic, Reliability Index, Phasor Measurement Units (PMUs), Dissolved Gas Analysis (DGA), Thermal Models, Dynamic Bayesian Network, Long Short-Term Memory (LSTM), Gas Concentration, Temperature Prediction.

# Paper 5

* Focuses on challenges faced by distribution transformers in smart grids due to overloading and faults.
* Proposes an IoT-based real-time monitoring system for anomaly detection in transformer loads.
* Utilizes low-cost IoT gateways and sensors to collect load current and oil level data.
* Employs a cloud-based anomaly detection algorithm, Isolation Forest, to detect faults within a 24-hour window.
* Includes a mobile application for real-time transformer monitoring and geographic tracking.

**Technologies Used**: IoT, Big Data, Cloud Computing, Smart Grid, Load Monitoring, Deep Learning, Anomaly Detection.

# Paper 6

* Integration of renewable energy sources affects transformer stability and grid reliability.
* Proposes a hybrid online system for real-time monitoring of transformer health using diagnostic factors (insulation quality, oil condition, dissolved gas analysis).
* Emphasizes the importance of indicators like power factor and dissolved gas analysis.
* Aims to improve preventive maintenance, reliability, and transformer lifespan.

# Paper 7

* Introduces a cost-effective, AI-driven approach for transformer insulation health assessment.
* Uses Artificial Neural Networks (ANN) for Health Index (HI) prediction, reducing the need for multiple expensive condition tests.
* Improves large-scale feasibility of monitoring through AI and feature-based exhaustive techniques.
* Focuses on reducing costs while ensuring efficient and accurate monitoring.

# Paper 8

* IoT-based health monitoring system uses sensors for data collection (temperature, oil level, vibration).
* Employs AI and machine learning for real-time fault detection and alerts.
* Data is transmitted via Wireless Sensor Networks (WSNs) and Fiber Optic Sensors (FOSs).
* Provides remote monitoring and reduces downtime with cloud-based applications.
* Highlights challenges like sensor reliability and secure communication.

# Paper 9

* Focuses on risks of large transformer failures and their global impact on power systems.
* Proposes an advanced real-time monitoring tool with predictive analytics.
* Utilizes Principal Component Analysis (PCA) and Back-Propagation Artificial Neural Network (BP-ANN) for health prediction.
* Historical data improves accuracy, with results significantly outperforming heuristic methods.
* Enhances preventive maintenance and asset management strategies.

# Paper 10

* Highlights limitations of traditional transformer health monitoring methods.
* Explores AI-driven solutions for Transformer Prognostic Health Management (TPHM).
* AI techniques enhance dissolved gas analysis (DGA), frequency response analysis (FRA), and partial discharge (PD) automation.
* Emphasizes predictive maintenance strategies to improve grid reliability and reduce risks.
* Advocates for a comprehensive AI-powered prognostic diagnosis ecosystem.

MACHINE LEARNING

# What is Machine Learning?

Machine Learning (ML) is a subfield of artificial intelligence that gives computers the ability to learn and improve from experience without being explicitly programmed. It focuses on developing algorithms that can access data, learn from it autonomously, and make predictions or decisions based on that learning. The primary aim is to allow computers to learn automatically without human intervention and adjust their actions accordingly.

# Types of Machine Learning

Machine learning can be broadly classified into the following categories:

1. **Supervised Learning:** The algorithm learns from labelled training data, making predictions based on that data. The system is provided with input-output pairs, and it learns to map the relationship between them.
2. **Unsupervised Learning:** The algorithm learns from unlabelled data by finding patterns and relationships within the data without external guidance. It identifies inherent groupings in data based on statistical properties.
3. **Semi-supervised Learning:** A hybrid approach where the algorithm learns from a combination of labelled and unlabelled data, typically a small amount of labelled data and a large amount of unlabelled data.
4. **Reinforcement Learning:** The algorithm learns by interacting with an environment, receiving feedback in the form of rewards or penalties based on its actions, and adjusting its behaviour to maximize rewards.
5. **Deep Learning:** A subset of ML involving neural networks with multiple layers (deep neural networks) that can learn and make intelligent decisions on their own.

# Common terms in Machine Learning

**Feature:** An individual measurable property or characteristic of a phenomenon being observed. Features are the input variables used in ML models.

**Label:** The output variable that a model is designed to predict in supervised learning.

**Training Data:** The dataset used to train a model by establishing relationships between features and labels.

**Validation Data:** A dataset used to tune hyperparameters and evaluate a model's performance during development.

**Test Data:** A separate dataset used to assess the final performance of a trained model.

**Overfitting:** When a model learns the training data too well, including its noise and outliers, performing poorly on new, unseen data.

**Underfitting:** When a model is too simple to capture the underlying pattern of the data, resulting in poor performance on both training and new data.

**Hyperparameters:** Parameters set before the learning process begins, as opposed to parameters that are derived via training.

**Cross-validation:** A technique to evaluate a model by training it on different subsets of the data and testing it on the remaining subset.

**Bias-Variance Trade-off:** The conflict in trying to simultaneously minimize two sources of error that prevent supervised learning algorithms from generalizing beyond their training set.

# How does it work?

**Data Collection:** Gathering relevant, high-quality data that represents the problem domain.

**Data Preprocessing:** Cleaning the data by handling missing values, outliers, and normalizing or standardizing features to improve model performance.

**Feature Engineering**: Selecting, transforming, or creating features to improve model performance.

**Model Selection:** Choosing an appropriate ML algorithm based on the problem type, data characteristics, and desired outcomes.

**Training:** The model learns patterns from the training data by minimizing a predefined loss function.

**Evaluation:** Assessing the model's performance using validation or test data with appropriate metrics.

**Tuning:** Adjusting hyperparameters to optimize the model's performance.

**Deployment:** Implementing the trained model in a production environment to make predictions on new data.

**Monitoring & Maintenance:** Continuously monitoring the model's performance and retraining as needed when data patterns change.

# Applications of Machine Learning

**Healthcare:** Disease prediction, medical image analysis, patient monitoring, drug discovery, and personalized medicine.

**Finance:** Fraud detection, algorithmic trading, credit scoring, portfolio management, and customer segmentation.

**Retail:** Product recommendations, inventory management, price optimization, customer behavior analysis, and demand forecasting.

**Manufacturing:** Predictive maintenance, quality control, supply chain optimization, and process automation.

**Transportation:** Autonomous vehicles, traffic prediction, route optimization, and logistics management.

**Energy:** Load forecasting, grid management, fault detection, and energy consumption optimization.

**Security:** Intrusion detection, biometric authentication, surveillance, and cybersecurity threat detection.

**Natural Language Processing:** Translation, sentiment analysis, chatbots, and content generation.

**Computer Vision:** Image recognition, object detection, facial recognition, and video analysis.

**Power Systems:** Transformer health monitoring, load prediction, fault diagnosis, and maintenance scheduling.

# Types of Machine Learning Models

**1. Linear Models**

**Linear Regression:** Predicts a continuous output variable based on one or more input variables by establishing a linear relationship.

**Logistic Regression:** Despite its name, it's a classification algorithm that predicts the probability of an instance belonging to a particular class

**2. Tree-Based Models**

**Decision Trees:** Models that make decisions based on a series of questions, resembling a flowchart structure.

**Random Forests:** An ensemble method that builds multiple decision trees and merges their predictions.

**Gradient Boosting Machines (GBM):** Builds trees sequentially, with each tree correcting errors of the previous ones.

**3. Support Vector Machines (SVM)**

Finds the optimal hyperplane that best separates data points of different classes in a high-dimensional space.

**4. Neural Networks**

**Feedforward Neural Networks:** The basic type where information moves in one direction from input to output.

**Convolutional Neural Networks (CNN):** Specialized for processing grid-like data, such as images.

**Recurrent Neural Networks (RNN):** Designed for sequential data by maintaining a memory of previous inputs.

**Transformers:** Architecture with self-attention mechanisms, primarily used for natural language processing tasks.

**5. Clustering Algorithms**

**K-means:** Partitions data into k clusters based on similarity.

**Hierarchical Clustering:** Creates a tree of clusters by recursively merging or splitting them.

**DBSCAN:** Density-based clustering that identifies areas of high density separated by areas of low density.

**6. Dimensionality Reduction**

**Principal Component Analysis (PCA):** Reduces dimensionality by projecting data onto principal components.

**t-SNE:** Non-linear technique for visualizing high-dimensional data in lower dimensions.

**7. Ensemble Methods**

**Bagging:** Builds multiple models in parallel and combines their predictions.

**Boosting:** Builds models sequentially, with each model correcting errors of the previous ones.

**Stacking:** Combines predictions from multiple models using another model.

# Data of all Models used

|  |  |  |
| --- | --- | --- |
| **Model** | **Test Health Index**  **(R2)** | **Test Life Expectancy (R2)** |
| **Multi Task Elastic Net** | 0.735071 | 0.719207 |
| **Random Forest Regressor** | 0.892221 | 0.917003 |
| **Extra Trees Regressor** | 0.916855 | 0.936815 |
| **Gradient Boosting Regressor** | 0.882263 | 0.928580 |
| **XGBoost** | 0.881960 | 0.922573 |
| **CatBoost** | 0.857018 | 0.937276 |
| **K-Nearest Neighbours Regressor** | 0.855033 | 0.883082 |
| **SVR (Multioutput Regressor)** | 0.745656 | 0.831921 |
| **GaussianProcessRegressor** | 0.202147 | 0.326832 |
| **MLP Regressor** | 0.514013 | 0.605714 |
| **RegressorChain** | 0.476155 | 0.409127 |
| **StackingRegressor** | 0.453056 | 0.574112 |
| **LightGBM Multi Output Regressor** | 0.887229 | 0.924203 |

# Models we have selected

## CatBoostRegressor

**CatBoostRegressor** is an implementation of gradient boosting on decision trees developed by Yandex. It's specifically designed to handle categorical features automatically without extensive preprocessing.

**Key Features:**

**Handling Categorical Features:** Automatically processes categorical variables without the need for one-hot encoding.

**Robust to Overfitting:** Implements symmetric trees and a novel gradient-based approach to reduce overfitting.

**GPU Acceleration:** Supports training on GPUs for faster computation.

**Built-in Regularization:** Uses multiple techniques to prevent overfitting, including a novel ordered boosting approach.

**Missing Value Handling:** Automatically deals with missing values in the data.

**Working Process:**

**Data Preparation:** The algorithm first converts categorical features into numerical ones using techniques like mean encoding, but with a preprocessing step to avoid target leakage.

**Ordered** **Boosting:** Unlike traditional gradient boosting, CatBoost changes the order of the training examples in each iteration, which helps to reduce overfitting.

**Feature Combinations:** Automatically generates and evaluates combinations of categorical features.

**Tree Building:** Constructs decision trees where each split is based on a single feature.

**Prediction:** For regression tasks, the final prediction is the sum of all tree predictions multiplied by the learning rate.

**Mathematical Foundation:**

CatBoost uses a gradient boosting framework that minimizes the objective function:

Where:

* is the loss function measuring the difference between the predicted and actual values.
* is the regularization term for the jth tree.
* n is the number of training examples.
* M is the number of trees.

For handling categorical features, CatBoost employs an ordered target statistics technique:

Where:

* is the numerical representation of the categorical feature.
* is 1 if the condition is true and 0 otherwise.
* is the target value.
* a is a smoothing parameter.
* p is a prior value, typically the mean of the target.

## ExtraTreesRegressor

ExtraTreesRegressor is an ensemble method that builds multiple randomized decision trees (extra-trees) and averages their predictions. It's an extension of the random forest algorithm with additional randomization.

**Key Features:**

* **High Randomization**: Introduces more randomness than Random Forests by selecting split points randomly.
* **Reduced Variance**: The high level of randomization helps reduce variance and overfitting.
* **Feature Importance**: Provides a measure of feature importance based on how much each feature contributes to decreasing the weighted impurity.
* **Parallelizable**: Trees can be built in parallel, making it efficient for large datasets.
* **Robust to Noisy Data**: The randomization makes it less sensitive to noise in the data.

**Working Process:**

1. **Bootstrap Sampling**: For each tree, a bootstrap sample is drawn from the training data (can be disabled by setting bootstrap=False).
2. **Random Feature Selection**: For each node in a tree, a random subset of features is considered for splitting.
3. **Random Split Selection**: Unlike Random Forests, Extra Trees selects a split point randomly from the range of values for the selected feature, rather than searching for the best split.
4. **Tree Construction**: Each tree is grown to its maximum depth (or until other stopping criteria are met).
5. **Prediction**: For regression tasks, the final prediction is the average of all tree predictions.

**Mathematical Foundation:**

The ExtraTreesRegressor algorithm uses the following approach for prediction:

Where:

* is the final prediction.
* *B* is the number of trees in the ensemble.
* is the prediction of the bth tree.

For each node in a tree, a feature j and a split point s are selected to minimize the impurity criterion:

Where:

* is the impurity measure (e.g., mean squared error for regression) of the set S
* and are the left and right subsets created by the split.
* |S|, and | are the sizes of the respective sets.

For regression, the impurity measure is typically the variance:

Where:

* is the target value for instance s.
* is the mean target value in set s.

MATHEMATICAL IDEATION

# Principal Component Analysis

PCA is a dimensionality reduction technique that transforms a dataset with potentially correlated variables into a set of linearly uncorrelated variables called principal components.

## The PCA Process:

1. **Standardization**: Normalize the data so that each feature has a mean of 0 and a standard deviation of 1.
2. **Covariance Matrix Computation**: Calculate the covariance matrix of the standardized data.
3. **Eigendecomposition**: Compute the eigenvectors and eigenvalues of the covariance matrix.
4. **Principal Components Selection**: Rank the eigenvectors by their corresponding eigenvalues and select the top k eigenvectors.
5. **Projection**: Project the original data onto the selected eigenvectors to get the principal components.

## Mathematical Foundation:

For a data matrix X with n samples and d features, the steps are:

1. **Standardization**:

Where is the mean and is the standard deviation of each feature.

1. **Covariance Matrix**:
2. **Eigendecomposition**:

Solve for eigenvectors v and eigenvalues such that:

1. **Principal Components**:

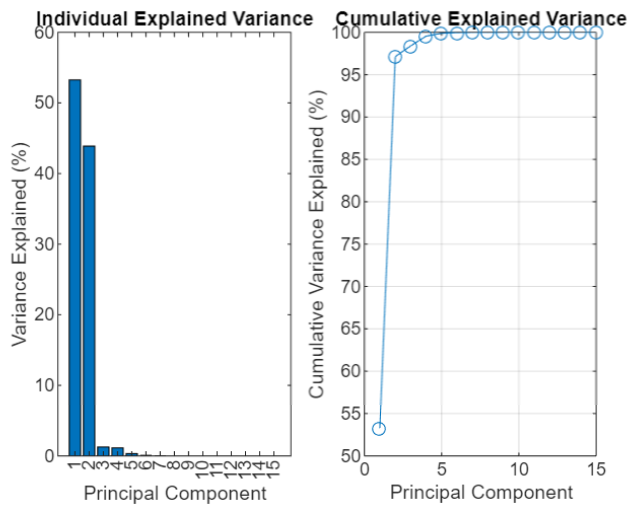
The principal components are given by:

1. **Variance Explained**:

The proportion of variance explained by the ith principal component is

## Benefits of PCA:

* Reduces dimensionality while preserving as much variance as possible.
* Removes multicollinearity by creating orthogonal features.
* Helps visualize high-dimensional data.
* Can improve the performance of machine learning algorithms by reducing overfitting.

****

# Data Normalisation

Data normalization is the process of rescaling numerical features to a standard range to ensure that no feature dominates the learning process due to its scale.

## Common Normalization Techniques:

1. **Min-Max Scaling (Normalization)**:

Rescales the data to a fixed range, typically [0, 1].

1. **Z-score Standardization**:

Rescales data to have a mean of 0 and a standard deviation of 1.

1. **Robust Scaling**:

Uses statistics that are robust to outliers, such as median and interquartile range.

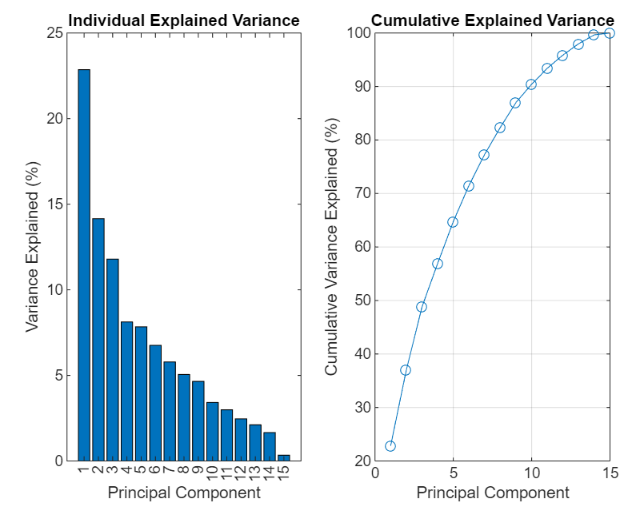
1. **Unit Vector Scaling**:

Scales each sample to have a unit norm (length of 1).

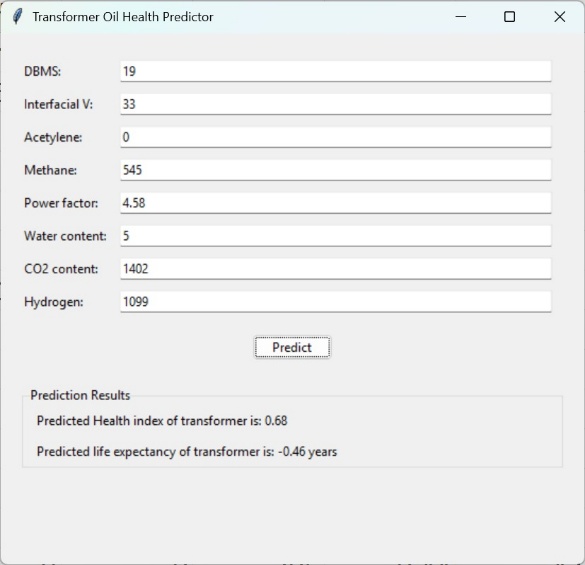
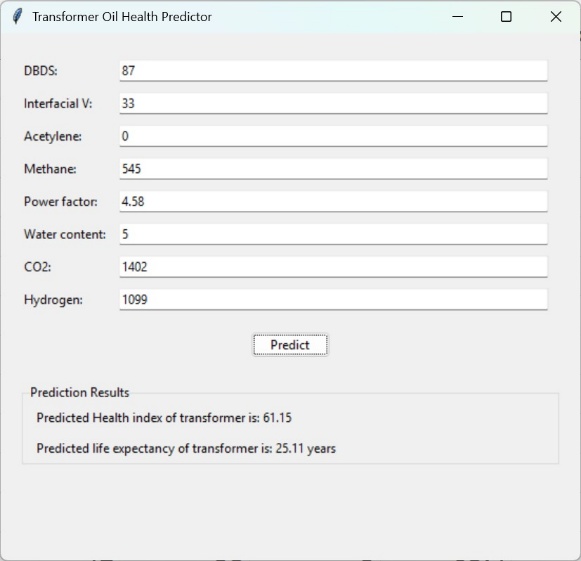
​

## Benefits of Normalization:

* Ensures all features contribute equally to the model.
* Improves convergence speed for many optimization algorithms.
* Required for distance-based algorithms like KNN and SVM.
* Facilitates comparison between features with different scales.

****

OUTPUT

** **

If output is negative, If output is positive,

Life expectancy decreases Life expectancy is a constant

FUTURE WORK

The integration of IoT devices with infrared sensors represents a promising advancement for transformer health monitoring. These sensors, strategically placed within transformers, continuously monitor oil color by analyzing infrared radiation reflection, with real-time data transmission to monitoring systems. Machine learning algorithms then compare oil color patterns against baseline data to detect anomalies and generate early warnings when parameters deviate from normal. This non-intrusive approach eliminates manual sampling, provides continuous assessment rather than periodic testing, enables early detection of issues, and integrates with existing health indices for comprehensive assessment.

As transformer monitoring systems become increasingly connected, robust cybersecurity measures are essential. Comprehensive protection should include secure communication protocols with end-to-end encryption and certificate-based authentication, alongside multi-factor authentication and role-based access control systems. Additional safeguards include AI-powered anomaly detection for network monitoring, regular security audits, secure coding practices, redundant systems with regular data backups, and integrated physical security measures to protect hardware from tampering. These cybersecurity measures, combined with hardware enhancements, ensure both reliability and security of transformer monitoring systems while protecting the integrity of electrical grid infrastructure.

**CONCLUSION**

This research study has been successful in demonstrating the capability of machine learning techniques to offer power transformer prognostics and health monitoring. By developing an ensemble machine learning model that forecasts transformer health index scores and life expectancy using an extensive database of 472 transformer oil samples, we have been able to develop a reliable means of forecasting transformer health index scores and life expectancy. The ability of the model to accommodate numerous parameters like DBDS levels, interfacial tension, dielectric characteristics, dissolved gases, moisture levels, acidity, and viscosity makes the model capable of recognizing complex patterns of degradation for different transformer designs and operating conditions.

The significance of this study is in the capacity to transform traditional maintenance methods into data-backed, condition-based maintenance. The software solution designed provides utilities with actionable data to optimize maintenance planning, reduce surprise failures, and optimize transformer life. Its integration with IoT sensors, mobile devices, automated reporting systems, and maintenance management systems, as envisioned, will also enhance its value of application in field operations.

As the power grids grapple with increasing loads and aging infrastructure challenges, this machine learning solution offers a scalable and adaptable solution to asset management. The inclusion of federated learning functionality will allow utilities to share experience with one another while maintaining the confidentiality of data, resulting in increased grid reliability and operational efficiency. This project represents a significant milestone in the use of advanced analytics to manage critical infrastructure and a stepping stone to future innovation in smart grid technology.

**ACKNOWLEDGMENT**

We would like to express our sincere gratitude to

**Dr. Praveen Kumar Mishra**

from the

**Department of Electrical and Electronics Engineering**

for his invaluable guidance, expertise, and continuous support throughout this research endeavour. His insightful suggestions, technical knowledge, and constructive feedback have been instrumental in shaping this work. Dr. Mishra's dedication to academic excellence and his willingness to share his profound knowledge have greatly influenced the quality of our research. We are deeply indebted to him for his patience and encouragement.

**BIBLIOGRAPHY**

[1] N. M. Lindsay and K. A. Nawas, "Design of Transformer Health Monitoring System Using TensorFlow Architecture," 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10823358

[2] A. Cathignol, V. Thuillie-Demont, L. Baldi, L. Micheau, J.-P. Petitpretre, and A. Ouberehil, "Hybrid AI-Subject Matter Expert Solution for Evaluating the Health Index of Oil Distribution Transformers," 2024. [Online]. Available: https://papers.phmsociety.org/index.php/phme/article/download/4102/2433

[3] S. Swetha, "Smart Transformer Health Monitoring System," 2022. [Online]. Available: https://sist.sathyabama.ac.in/sist\_naac/documents/1.3.4/1822-b.e-eee-batchno-30.pd

[4] S. Cvijic, N. Gupta, and S. Lux, "Need for AI in Transformer Diagnostics and Prognostics," 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10088199

[5] J. Ramesh, S. Shahriar, A. R. Al-Ali, A. Osman, and M. F. Shaaban, "Machine Learning Approach for Smart Distribution Transformers Load Monitoring and Management System," Energies, vol. 15, no. 21, p. 7981, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/21/7981

[6] O. Laayati, H. El Hadraoui, A. El Magharaoui, N. El-Bazi, M. Bouzi, A. Chebak, and J. M. Guerrero, "An AI-Layered with Multi-Agent Systems Architecture for Prognostics Health Management of Smart Transformers: A Novel Approach for Smart Grid-Ready Energy Management Systems," Energies, vol. 15, no. 19, p. 7217, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/19/7217

[7] A. Y. Alqudsi and A. H. El-Hag, "A cost-effective artificial intelligence-based transformer insulation health index," 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8280194

[8] M. Ragavan, K. Y. Kumar, A. Kiran, M. S. Saini, and R. A. Choudhury, "Transformer Health Condition Monitoring System Using the Internet of Things (IoT)," 2022. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4490230

[9] 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. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-09507-3\_46

[10] A. Vatsa, A. S. Hati, and A. K. Rathore, "Enhancing Transformer Health Monitoring With AI-Driven Prognostic Diagnosis Trends: Overcoming Traditional Methodology's Computational Limitations," 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10318187>

**APPENDIX**

**DATA SET EXPLAINED**

The utilised data set contains the following parameters as inputs for the model.

**Dissolved Gas Analysis (DGA) Parameters**

1. **Hydrogen (H₂):** A key gas generated by partial discharges, overheating of oil, and various fault conditions. High levels (>100 ppm) often indicate partial discharge activity occurring within the transformer.
2. **Oxygen (O₂):** Present due to air ingress or oil oxidation. Normal levels are typically 3,000-40,000 ppm. Significant decreases can indicate accelerated aging reactions consuming oxygen.
3. **Nitrogen (N₂):** Primarily present due to contact with air. Normal range is 10,000-100,000 ppm. The N₂/O₂ ratio helps identify air ingress issues.
4. **Methane (CH₄):** Produced during thermal decomposition of oil. Normal levels are <120 ppm. Higher concentrations indicate thermal faults at moderate temperatures (300-500°C).
5. **Carbon Monoxide (CO):** Generated primarily from thermal degradation of cellulose insulation. Normal levels are <500 ppm. Elevated levels indicate paper insulation deterioration.
6. **Carbon Dioxide (CO₂):** Also produced from degradation of cellulose materials. Normal levels are <5,000 ppm. The CO₂/CO ratio helps assess the rate of paper degradation.
7. **Ethylene (C₂H₄):** Forms during oil thermal decomposition at high temperatures. Normal levels are <50 ppm. Higher concentrations indicate serious thermal faults (500-800°C).
8. **Ethane (C₂H₆):** Generated during thermal oil decomposition at lower temperatures. Normal levels are <65 ppm. Indicates thermal faults at 300-500°C.
9. **Acetylene (C₂H₂):** Produced under extremely high temperatures or electrical arcing. Normal levels are <2 ppm. Presence above this threshold strongly indicates arcing within the transformer.
10. **DBDS (Dibenzyl Disulfide):** A corrosive sulfur compound that can form copper sulfide, leading to insulation failure. Its presence in oil (typically measured in mg/kg) indicates potential for copper corrosion.

**Oil Quality Parameters**

1. **Power Factor:** Measures the dielectric losses in the insulating oil. It is the ratio of resistive current to total current through the oil. Higher values (>0.5%) indicate contamination, moisture, or oxidation products in the oil.
2. **Interfacial Tension:** Measures the force required to rupture the oil-water interface, expressed in mN/m. New oil typically has values >40 mN/m. Lower values (<25 mN/m) indicate presence of polar contaminants and oxidation byproducts.
3. **Dielectric Rigidity (Breakdown Voltage):** The voltage at which electrical breakdown occurs across a standard gap in the oil, typically measured in kV. Higher values (>30 kV) indicate good insulating properties, while lower values suggest contamination or moisture.
4. **Water Content:** The amount of dissolved water in transformer oil, measured in ppm. Acceptable levels depend on the oil type and transformer voltage class, but generally should be <20 ppm for high-voltage transformers. Higher water content accelerates aging and reduces dielectric strength.

**Transformer Assessment Metrics**

1. **Health Index:** A comprehensive numerical indicator (typically 0-100 or 0-10) that aggregates multiple condition parameters to represent the overall health of a transformer. It incorporates electrical, mechanical, and chemical test results weighted according to their importance.
2. **Life Expectancy:** An estimate of the transformer's remaining useful life, typically expressed in years. It is calculated based on the health index, historical loading patterns, and aging acceleration factors. The normal design life of a power transformer is 25-40 years, but actual lifetime depends heavily on operating conditions and maintenance practices.