Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 22, Iss. 2, pp. 19-29, 2024

ENHANCING PREDICTIVE MAINTENANCE OF POWER TRANSFORMERS THROUGH MACHINE LEARNING APPROACHES *

Carla Pacheco[1], Vagner Paes[1], Marcelo de Carvalho[1], Felipe Lopes[1], Gabriel Machado[1]

Alex Garcia[1], Edmilson Neto[1], Jefferson Santos[2], Edward Haeusler[1], Ana Marotti[3]

[1] *Informatics Department, PUC-Rio, Brazil*

Emails: {cpacheco, vpaes, mdcarvalho, flopes, gresende, agarcia, eneto, hermann}@inf.puc-rio.br

[2] *FGV EBAPE, Rio de Janeiro, Brazil*. Email: jefferson.santos@fgv.br

[3] *Eletrobras, Brazil*. Email: amarotti@eletrobras.com

Abstract – This paper focuses on the use of machine learning algorithms to assist in predictive maintenance, aiming to reduce downtime and associated costs for electrical equipment. It introduces two machine learning predictive indicators: the Chromatographic Assay Indicator (CAI) and the Electrical Failure Risk Indicator (EFRI), which leverage chromatographic and sensor data, respectively. The CAI indicator was trained on external data, and its generalization capability was assessed using company-specific chromatographic data with expert assistance. The evaluation involved two heuristics, both focusing on keyword identification in free-text diagnoses. Results showed over 90% accuracy in predicting failures in reactors and power transformers, and an AUC of nearly 85% for reactors. Additionally, CAI outperformed classical Dissolved Gas Analysis (DGA) methods. On the other hand, the EFRI indicator was trained using company's internal data from a monitoring system of operational sensors and maintaining power transformer data. The classification model achieved over 95% accuracy and an AUC of almost 90% on the test set. These results indicate that both indicators can be integrated into a solution to support maintenance specialists in their decision-making processes.

Keywords – predictive maintenance, machine learning, power transformers, random forest

1 Introduction

The downtime of transmission assets such as power transformers leads to high costs for energy transmission companies, in addition to disturbances in the delivery of electricity to people. According to ONS (National Operator of the Electric System), temporary interruptions of energy supply may have several causes, such as failures in equipment or in protection and control systems [1].

It is undoubtedly important to avoid interruptions in the energy supply, as these interruptions can lead to significant economic losses and operational inefficiencies. Research on maintenance effectiveness reveals that one-third of all maintenance costs is wasted due to unnecessary or improperly conducted maintenance activities [2]. This inefficiency not only wastes financial resources but also diverts manpower and materials that could be better utilized elsewhere. The primary reason behind such ineffective maintenance management is the insufficient availability of accurate and real-time data [2]. Without this crucial information, it becomes challenging to accurately assess the condition and performance of plant equipment and systems, leading to either over-maintenance, where resources are spent on equipment that does not need attention, or under-maintenance, where issues are overlooked until they cause serious problems. Consequently, implementing robust data collection and analysis practices is essential to optimize maintenance schedules, ensure the reliability of energy supplies, and maximize the return on investment in maintenance operations.

Traditional methods of maintenance are reactive and preventive [2]. The former is based on actual equipment failure and is the most expensive method for run-to-fail management. The latter is based on statistical trends and predetermined time intervals or operating hours to reduce the probability of failure or loss of performance. Predictive methods are at the vanguard [3] because they determine the scheduling of maintenance actions in an adaptive and flexible manner, according to the need of the equipment instead of at fixed intervals as in preventive maintenance.

This work aims to improve the maintenance of power transformers and thereby reduce the costs associated with their downtime. In this paper, we present two ML predictive indicators: CAI (Chromatographic Assay Indicator) and EFRI (Electrical Failure Risk Indicator), which use chromatographic and sensor data, respectively, both from the company's internal systems. The developed solution integrates these two ML indicators with several classical and traditional data analytics approaches to support maintenance specialists. After several experiments, the Random Forest algorithm was chosen to predict data from the operation

*ANEEL funds this work as R&D project PD-00394-1907/2019.

and maintenance of the equipment. This resulted in enhanced reliability, safety, and cost-effectiveness of power transformer and reactor operations.

The CAI indicator was assembled using chromatographic data from the company, as described in Section 4. After being trained on external data, its generalization capability was evaluated on cromatographic data from the company with the assistance of the specialists. The evaluation involved two heuristics, both of which focused on identifying specific keywords in the free-text diagnoses. Results obtained in both heuristics show accuracy above 90% for predicting failures in reactors and power transformers, and an AUC reaching nearly 85% for reactors. In addition, the results of CAI were also benchmarked against classical Dissolved Gas Analysis (DGA) methods, where it presented greater performance. Regarding the EFRI indicator, it was trained on data coming from internal systems of the company responsible for monitoring operational data of sensors and maintaining information of power transformers. According to Section 6.4, it is noticeable the corresponding classification model achieved an accuracy over 95%, as well as AUC of almost 90% on Test set.

This paper is organized into six more Sections. Section 2 describes the main related works. Section 3 describes the use of chromatography and Supervisory Control and Data Acquisition (SCADA) systems in power transformers monitoring and maintenance. Section 4 describes the chromatographic dataset used to develop the CAI indicator whereas Section 5 details the monitoring dataset used by the EFRI indicator. Section 6 presents the results of experimental evaluation of CAI and EFRI. In the end, Section 7 concludes and presents future works.

2 Related Work

In the first part of this Section, we are going to address the state-of-the-art articles related to predictive maintenance of transmission power transformers in the context of DGA. In order to find the articles that best match the current work, we performed a systematic search at SCOPUS by using the query string: ("predictive" AND "maintenance") AND ("dissolved" AND "gas" AND "analysis") AND ("transmission") AND ("machine" AND "learning") AND ("power" AND "transformer"). We obtained five results, comprising journals and conference proceedings. Afterwards, we ranked them based on their relevance. After that, we selected the top 3 articles and the references therein that best matched the current work, which we are going to describe in details as follows.

In [4], the authors introduced a novel methodology for incipient fault diagnosis in power transformers by using artificial neural networks (ANN). They used a dataset available in [5], which contains fault types and the corresponding concentration of the dissolved gases in the insulating oil of the power transformers, such as Hydrogen (H_2) and Methane (CH_4). Moreover, in their proposed approach, the authors used a cascade structure of Multi-Layer Perceptron networks (CMLP) for fault classification and the goal was to improve the network's performance by simplifying the number of relationships required to be learned by the ANN, which resulted in 85% accuracy in the test set. In another work [6], the authors proposed a robust multilayer framework for online condition assessment of power transformers. They handled the measurement of uncertainties and fused the results of independent DGA methods without losing their fault diagnosis outcome. Moreover, the authors used an ANN to intelligently assign the weight of each independent method in the fusion procedure, depending on its fault type for a given range of input gases concentration. The models were tested on a combination of IEC TC 10 [5] and IEEE standard [7] datasets, and the best obtained model had an overall accuracy of 96%. In another interesting work [8], the authors used a Convolutional Neural Network (CNN) to predict power transformer fault types under different noise levels in measurements. In this work, the authors used different categories of input ratios concerning DGA: conventional ratios, such as four ratios of Rogers [9], new ratios, which were conveniently created, and hybrid ratios, which are a combination of the former and the latter ratios. The datasets used for training and testing were collected from 16 sources, such as in [10, 11]. The CNN model had an accuracy of 98.5 % for 0% noise level and 96.6% for $\pm$ 20% noise level in the test data, which shows the power of CNNs to model faulty power transformers when data is noisy.

In the second part of this Section, we are going to address the state-of-the-art articles related to maintenance of power transformers in the context of modeling by using SCADA data. In order to find the articles that best matched the current work, we performed a systematic search at SCOPUS by using the following query string: ("maintenance") AND ("machine" AND "learning") and ("scada") AND ("power" AND "transformer"), which resulted in two articles. We ranked them by relevance, and found one article that matched this work. Moreover, in order to get more articles related to this work, in the aforementioned query string, we replaced "machine learning" by "artificial intelligence" and found three articles. After that, we ranked them by relevance and found out that two articles are related to the current work. In the following, we are going to describe in details such articles and the references therein. In another interesting work that used manual inspection of power transformer to confirm the fault diagnosis [12], the authors proposed a system called SADTRAFOS in order to support maintenance's decision with respect to power transformers fault prediction. In that work, the authors used a fuzzy inference module for fault diagnosis, and a decision support, which provided recommendations to the managers. The final model had an overall accuracy of 80% on real data concerning fault diagnosis and made correct recommendations for the maintenance team.

In [13], the authors applied ML to fault prediction in wind turbines for generating corrective maintenance strategies. In this work, SCADA data was used to capture the operational status of turbines, and by using a dual transformer model, comprised of two stages, they obtained an accuracy of up to 96.75% for alarm prediction in the dataset obtained from [14]. In another work [15], the authors proposed a model to detect temperature anomalies in key components of wind turbines, such as gearbox and transformers. They used ANN to address the challenge of the limited pre-classified data and then categorized the operating conditions into: (1) the Normal Behavior (NB) module, (2) the Expected Time To Failure (ETTF) module, and (3) Anomaly

Detection (AD) module. Then, they applied the model in data from an offshore wind farm in Germany, and obtained an accuracy of 94% concerning transformers with 7 hours ahead prediction.

3 Current maintenance practice

Maintenance plays a major role in the industrial sector, as it greatly affects costs and reliability, thereby playing a vital role in a company's competitiveness in the market. Unforeseen interruptions or failures in equipment can severely hamper a company's primary operations, potentially leading to substantial penalties and immeasurable damage to its reputation. Consequently, it is crucial to detect and address any faults in equipments to prevent disruptions in the production processes [16]. There are several practices related to equipment's maintenance. These practices can be grouped in three main categories [17]:

- Reactive or Run-to-Failure (R2F) maintenance;
- Preventive Maintenance (PvM);
- Predictive Maintenance (PdM).

The advancements in modern techniques like the IoT, sensing technology, and artificial intelligence have brought a shift in maintenance strategies from R2F to PvM to PdM. Reactive Maintenance is a simplistic approach where interventions are carried out only after failures happen. It is carried out solely to restore equipment to its functioning state after a failure has already occurred, leading to delays and high costs associated with reactive repairs. On the other hand, Preventive Maintenance follows a predetermined schedule based on time or process iterations to prevent breakdowns. While it aims to prevent failures, it may result in unnecessary maintenance and incur high prevention costs [18].

To strike a balance between the two approaches, PdM is employed. PdM relies on real-time estimation of equipment "health" to identify potential failures and intervenes in a timely manner before their occurrence. This is the best of both worlds, allowing maintenance to be performed with the lowest possible frequency to prevent unplanned reactive maintenance, while avoiding the costs associated with excessive preventive maintenance [18].

Predictive maintenance has four levels of maturity [19]. The highest level is supposed to use data analytics and real-time monitoring of equipment. Data analytical techniques include ML algorithms to uncover hidden relationships and identify meaningful patterns in large amounts of high-dimensional and multivariate data, presented in complex and dynamic environments like industrial settings.

3.1 Chromatography

Chromatography is a technique that allows the separation of a mixture of species in separate compounds [20]. In this context, "species" is a generic term used to describe the different compounds, molecules, or elements that are mixed together. It comprises various separation techniques depending on the material to be analyzed after separation. Physical and chemical characteristics such as mass, density, and type of intermolecular bonding are taken into consideration in the analysis. In [21], whose authors are the pioneers of chromatography, there is an allusion to the similarity between the chromatogram and distillation columns. Therefore, new methods and techniques for obtaining quantities of specific elements in a sample gained prominence in various areas of study.

In the electrical sector, chromatography enables the analysis of the quality of certain transmission assets, by identifying the concentration of gases dissolved in the insulation oil. This analysis is often based on the the concentration of the following nine gases: H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2, O_2, and N_2. This data allows for several inferences be made about the condition of the transmission asset, as detecting equipment defects, as the proportion of these compounds can indicate electrical and thermal defects.

3.2 SCADA/EMS Systems

A SCADA/EMS system is a comprehensive solution for managing a utility's electricity grid, from data collection and analysis to real-time control of equipment and optimization of system performance. The concept means it is a Energy Management System (EMS) implemented using Supervisory Control and Data Acquisition (SCADA) architecture.

- SCADA: This is an industrial control system that uses computers, networked data communications, and graphical user interfaces for high-level process supervisory management. It also uses other peripheral devices, such as programmable logic controllers and discrete proportional-integral-derivative controllers to interface with process plant or machinery. The operator interfaces, which enable monitoring and issuing of process commands, are where specialists monitor and control processes through SCADA. This kind of system consolidates, distributes and monitors real-time data, allowing operators to directly interact with devices such as sensors, valves, pumps, motors, and more through human-machine interface (HMI) software.
- EMS: An Energy Management System is a system of computer-aided tools used by operators of electric utility grids to monitor, control, and optimize the performance of power generation and transmission. The monitor and control functions are known as SCADA, whereas optimization packages are often referred to as "advanced applications". The EMS assists

in the process of ensuring that the power system is operating efficiently, with real-time adjustments to changes in the load or power supply. It can also help plan and schedule maintenance activities.

In Brazil, one of the most important SCADA/EMS systems is SAGE (acronym in Portuguese for "Open System for Energy Management"). SAGE is a large-scale and high-performance SCADA/EMS system, developed and constantly updated by Eletrobras Research Center (CEPEL). It is used by dozens of electric power generation, transmission, and distribution agents in Brazil, particularly Eletrobras and ONS. Its modular architecture allows for proper customization, enabling it to be used as a communication gateway, a data concentrator for a distribution system, a local or regional supervisory system, an operation center for a system, or even a "multi-site" system composed of multiple synchronized and redundant control centers [22].

4 Chromatographic Dataset

It is common practice in the electrical sector to monitor the condition of transformers through DGA. Our objective in creating the CAI risk indicator was to use Machine Learning to develop a classifier that assists maintenance specialists in their decision-making process. Upon receiving the company's data, it was evident that there were few data points and, furthermore, they were not labeled with the type of failure present. Therefore, they were not suitable for training a ML algorithm. We opted to combine several public datasets [23–28] to train the algorithm. These datasets have reliable data, since they were validated in these scientific works and some transformers were manually inspected. To use these databases, it was necessary to work with a sample universe of only 5 types of gases, as some of the databases considered only these gases.

The labeled classes were identified from the aforementioned external datasets, as follows: NF - (No Faults); PD - (Partial Discharge); D1 - (Low Energy/Spark Discharge); D2 - (High Energy/Arc Discharge); T1 - Low Temperature Fault (t $< 300°$C); T2 - Middle Temperature Fault ($300°$C $\leq$ t $\leq 700°$C); T3 - High Temperature Fault (t $> 700°$C). The gases considered in this work in predictive maintenance modeling were five out of the nine presented in Section 3.1, namely, H_2, CH_4, C_2H_2, C_2H_4, and C_2H_6. After joining the datasets, we started the data preprocessing step, which consisted in two steps: (i) data cleaning, and (ii) data normalization. The data cleaning process consisted in two main steps: (1) replacement of null or invalid values and (2) removal of duplicated samples or blank values.

After data joining and cleaning, we calculated four concentration ratios, which were calculated and inserted as new columns in the unified dataset, and we adjusted the non-detectable value by adding a constant value, as suggested in [29]: $R_1 = \frac{CH_4}{H_2} + 0.4$, $R_2 = \frac{C_2H_2}{C_2H_4} + 0.4$, $R_4 = \frac{C_2H_6}{CH_4} + 0.4$, and $R_5 = \frac{C_2H_4}{C_2H_6} + 0.4$. For normalization, we used the IQR (Interquartile Range) method, since it is robust to outliers, and it is given by Equation 1:

$$Norm_{IQR} = \frac{x - Q_{2/4}}{Q_{3/4} - Q_{1/4}} \quad (1)$$

where Q_i is the i^{th} quartile. The above Equation was applied to the following attributes of the unified dataset: (i) H_2, (ii) CH_4, (iii) C_2H_4, (iv) C_2H_6, (v) R_1, (vi) R_4 and (vii) R_5. However, the attributes corresponding to C_2H_2 gas and to the ratio R_2 needed a particular normalization process, due to the distribution of their values containing several outliers and zero values. Therefore, in order to normalize these columns, Equations 2 and 3 were applied, respectively.

$$Norm_{C_2H_2} = \frac{\log(x+1)}{max(log(C_2H_2 + 1))} \quad (2)$$

$$Norm_{R_2} = \frac{x}{max(R_2)} \quad (3)$$

5 Monitoring Dataset

The EFRI uses analogical monitoring data from the SAGE system and relates it to the maintenance data from the Enterprise Resource Planning (ERP) system. This indicator may suggest a high risk of an electrical failure based on analogical measurement data in a six-day time window of the failure event. A major benefit is the maintenance planning can be done in advance, assisting the decision making of the operator on the maintenance of the asset.

A big electric energy company provided the data from SAGE and SAP Plant Maintenance (SAP PM) systems. The data from SAP has the maintenance history of power transformers and it was used in the categorization of electric failures of SAGE data. Such SCADA system reads the sensors of equipment and records that in a daily 5-minute log file. We calculated some daily statistics from each sensor, including mean, quartiles and standard deviation. Therefore, we could calculate the mean of each attribute in a six-day sliding window. The sensors have related alarms (digital data) which were computed to extract the mean and maximum statistics.

The processed data from SAGE was matched with the categorized SAP data. We analyzed the description of maintenance data and grouped it into categories according to some keywords. Such keywords and categories were manually chosen through a detailed study of data and related works. The categorization process also used the Levenshtein algorithm [30] in order to find the similarity in the free-text description. Thus, the conducted experiments that showed the best results belonged to the electric failure category. Only the attributes of electric nature were considered in the development of the model.

5.1 Methodology

We elaborated a methodology that covers the whole workflow of the ML algorithms and preprocessing. Each step can be briefly pointed as: (1) Data cleansing; (2) Dataset partition and feature selection; (3) Data balancing and normalization; (4) Model development and optimization; and (5) Model assessment. The following Subsections contain the description of each step.

5.1.1 Data balancing and normalization

There is a huge discrepancy in the amount of failures and non-failures in the dataset. To overcome this problem, we chose to balance the Training data trough the combination of two techniques, respectively: (i) reduction of non-failures in 3.7 times and then (ii) data augmentation of failures in 143 times. Both techniques performed better than only one. We applied the RandomUnderSampler algorithm from Imblearn library [31] to reduce the majority class instances. Then, we applied the SMOTE (Synthetic Minority oversampling Technique) algorithm [32], from the same library, to oversample the minority class, based on the parameter of the nearest neighbors, which was chosen to be 10. This augmentation technique creates synthetic data based on the existent data, increasing the size and variety of the minority class. Both undersampling and oversampling techniques used the same number for reproducibility. The balanced dataset had 10,142 instances of each class and was normalized based on the Gaussian function. This choice was based on comparison tests against Min-Max and IQR-based normalization methods. After that, we applied the normalization in the Test set with the parameters of the Training set in order to avoid data leakage, according to the Equation 4. With this in mind, the missing data of Test set was replaced by the mean of the respective attribute of the Training set.

$$Norm_{x_{Test}} = \frac{x_{Test} - \overline{x}_{Treino}}{\sigma_{Treino}} \tag{4}$$

5.1.2 Model development and optimization

We conducted experimental tests with the Random Forest algorithm with the max depth of trees parameter instantiated with the value 13, based on empirical analysis. Thereafter, this algorithm was used as a part of the exhaustive method of optimization GridSearchCV that uses 10-fold CV with the same parameters used by GridSearchCV in the feature selection step.

5.1.3 Model assessment

As the main metric, we chose the recall of the minority class because it evaluates the total amount of failures captured by the model. Furthermore, we analyzed the metric Relative Risk (RR) [33] that indicates the number of times there is an increased risk of occurring an electrical failure. RR is the rate of two metrics: False Omission Rate (FOR) and Predicted Positive Value (PPV). The former is the rate of the electric failures classified as non-failures, according to Equation 5. The latter is the rate of failures correctly classified over the total amount of instances predicted as failures, as shown in Equation 6. Hence, Equation 7 shows the RR metric which is dimensionless.

$$FOR = \frac{FN}{FN + TN} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

$$RR = \frac{PPV}{FOR} \tag{7}$$

The minority class was considered as to be the positive class, in other words, failures. Furthermore, we analyzed the classic metrics as follows: accuracy, AUC, F1-score, recall and recall of minority class. The best results are presented in Section 6.

6 Results and Discussion

This Section presents the results of the both indicators: CAI and EFRI. They indicate an increased risk of occurring an electric failure. The combination of these indicators provide decision-making support to the specialist with the most recent monitored data. In the first Subsections, CAI results will be discussed in details. Specifically, Subsection 6.1 will show how the best ML algorithm for CAI was determined, Subsection 6.2 will show CAI compared with classical approaches, and then Subsection 6.3 addresses the CAI model evaluated on internal company's data by assuming two different heuristics. Finally, EFRI results will be shown in training and test sets.

6.1 Assessing various ML algorithms to determine the optimal CAI using external data

We ran a benchmark against some algorithms and possible scenarios. A Vapnik-Chervonenkis dimension [34] study was conducted which, due to the limited amount of data records available, excluded the use of neural networks. The experiments were executed with the following four algorithms: (i) Support Vector Machine (LibSVM) [35]; (ii) Random Forest [36]; (iii) Fuzzy Unordered Rule Induction Algorithm (FURIA) [37]; and (iv) Random Trees [38]. We performed the following experiments:

- with and without label class aggregation;
- selection of three different set of concentration ratios (used in addition to all five gas concentrations);
- five or ten CV folds;

We tested all the possible scenarios, and the best result is presented in the following scenario:

- Three label classes:
  i. PD, D1 and D2 labels grouped into Electric Faults;
  ii. T1, T2 and T3 labels grouped into Thermal Faults;
  iii. Normal.
- All four concentration ratios (R1, R2, R4 and R5);
- Ten CV folds.

We split the data into training and test sets, with 80% and 20% of the original data, respectively. In order to find the best hyperparameters, we used the RandomizedSearchCV method from Sklearn [39] library in the training set by using ten folds.

Table 1: Experiment results obtained in the best scenario

Algorithms	LibSVM	Random Forest	FURIA	Random Tree
Accuracy	74.2%	**92.7%**	89.2%	81.1%

6.2 Comparison between CAI and classical chromatographic indicators

In order to validate the Random Forest model obtained from the last Subsection, we compared such a model with classical DGA methods, for instance: (i) Rogers, (ii) Doernenburg, (iii) NBR 7274, (iv) IEC 599 and (v) Duval's Triangle [40–44]. The comparison results are displayed in Table 2, from which we can see that the Doernenburg's method had the highest accuracy among the classical DGA methods. The hybrid DIEC-R method, which consists of a combination of Doernenburg's and IEC Ibrahim's method [26], presented the best result among all classic/hybrid methods. From the same Table, we can clearly notice that the Random Forest model outperformed all the other classical methods in almost 19 percentage points, presenting 92.2% accuracy and 92.1% F1-score in the test set.

Table 2: Classical DGA methods and Random Forest model performance in the Test set

Method	**Accuracy**	**F1-score**
Rogers	35.1%	24.2%
Rogers (refined)	46.8%	27.8%
Doernenburg	13.6%	14.5%
NBR 7274	51.7%	43.3%
IEC Ratio	51.2%	41.0%
IEC (refined)	64.4%	56.6%
Duval's Triangle	60.5%	38.8%
Doernenburg + Durval	69.7%	52.3%
Doernenburg + IEC (Ibrahim)	73.2%	71.3%
Random Forest	**92.2%**	**92.1%**

6.3 CAI results on internal data of the company

After training the CAI's model on external data, it became necessary to verify its generalization capability on data from a specific company's system, which contains chromatographic samples followed by their corresponding diagnoses provided by a specialist. To accomplish this, a dataset comprising chromatographic samples from this company's system was used, including the same five gases described in Section 4, along with a novel feature representing the specialists' diagnoses in a free-text format instead of discrete labels. To obtain target labels from this free-text format, two different heuristics were applied. Essentially, both heuristics constrained the classification problem to a binary scope, and they are detailed more extensively in the following subsections.

6.3.1 First Heuristic

In the first heuristic, the labels '0' (representing the Non-Failure class) and '1'(representing the Failure class) were assigned according to the following rule: if the expert report field is blank then assign label '0', otherwise assign label '1'. According to experts, blank fields actually represent non-failure scenarios. Following this labeling heuristic, the CAI model was evaluated on both transformers and reactors data. The confusion matrices, as well as the main evaluation metrics, are presented in Tables 3 and 4, respectively.

Table 3: Confusion matrices of CAI evaluated with the First Heuristic.

	Transformers		**Reactors**	
Class	**(predicted) Positive**	**(predicted) Negative**	**(predicted) Positive**	**(predicted) Negative**
(real) Positive	9	9	10	4
(real) Negative	52	646	9	300

Table 4: Evaluation metrics of CAI evaluated with the First Heuristic.

Metric	**Transformers**	**Reactors**
Accuracy	91.48%	95.98%
AUC	71.28%	84.26%
F1-score	59.14%	79.24%
Recall (failure class)	50.00%	71.43%
Relative Risk	10.74	40.00

In Table 4, "Relative Risk" indicates how many times there is an increased risk of an abnormality in a equipment whenever the model assigns a chromatographic sample to the Failure class [45]. Also in Table 4, it is noteworthy that the recall was only 0.5 (in case of transformers). By analyzing the Failure Class records, it was found that some reports are inconclusive and they are filled indicating the need of getting a new chromatographic sample. Therefore a new heuristic was implemented to deal with the report field by searching keywords.

6.3.2 Second Heuristic

After evaluating the CAI indicator results from the first heuristic, another experiment was conducted using internal data of the company based on a second heuristic. The reports provided by the specialists for the chromatographic data of both transformers and reactors from the company were reviewed, and labels '1' for Failure or '0' for Non-Failure were assigned according to the presence of certain keywords. This new heuristic works according to the following algorithm:

1. If the expert report field contains a non-failure keyword then assign label '0'.
2. Otherwise, if the expert report field contains a failure keyword then assign label '1'.
3. Otherwise assign label '0'.

For blank fields, the result is the same as the previous heuristic. However, for fields containing some report, results may differ. The keywords used in the labeling process are listed below:

- **Non-Failure keyword:** resampling;
- **Failure keywords:**: increase, elevation, acetylene, generation, out, shutdown, monitoring, inspection/inspected.

The CAI model was re-evaluated using the new heuristic. The resulting confusion matrices, as well as the main evaluation metrics, are shown in Tables 5 and 6, respectively.

In comparison to the first heuristic, there was a significant improvement in the results, particularly in the failure class recall metric, which for transformers increased from 50% to 61.54%. Concerning reactors, the results also point out to a relevant improvement in the model's generalization capability, improving AUC, recall and Relative Risk metrics.

Table 5: Confusion matrices of CAI evaluated with the Second Heuristic.

	Transformers		**Reactors**	
Class	**(predicted) Positive**	**(predicted) Negative**	**(predicted) Positive**	**(predicted) Negative**
(real) Positive	8	5	8	1
(real) Negative	53	650	11	303

Table 6: Evaluation metrics of CAI evaluated with the Second Heuristic.

Metric	**Transformers**	**Reactors**
Accuracy	91.90%	96.28%
AUC	77.00%	92.69%
F1-score	58.68%	77.60%
Recall (failure class)	61.54%	88.89%
Relative Risk	17.18	128.00

6.4 EFRI results

6.4.1 Dataset partition and feature selection

The dataset was split into two disjoint subsets: Training (80%, being 71 failure and 37,401 non-failure) and Test (20%, being 18 failure and 9,350 non-failure). In both sets, the proportion of the classes is the same: 1/527. Missing values in the Training set were replaced by the mean value of the attribute. Afterwards, we conducted preliminary tests with the decision tree based algorithm DecisionTreeClassifier from Python's Sklearn library [46]. Such algorithm is deterministic and was applied to the Training set in order to rank the most important attributes. To help this choice, we picked some hyperparameters of this algorithm, such as the minimum amount of samples required to split a node between 2 and 403 with an interval of 20, with the support of a method of exhaustive optimization of estimators, named GridSearchCV [47], from the same library. This method used refitting of the estimator with the best parameters found in the Training set based on the recall metric of the minority class. In addition, we chose the strategy to evaluate the performance of the Cross-Validation (CV) algorithm on the test fold using a list of the following multiple metrics: F1 of minority class, recall of minority class, AUC (Area Under the ROC Curve). After that, we ranked the most important attributes and selected the top 19 analogical in addition to the remaining 42 digital attributes to use in the modeling from now on, totaling 61 attributes.

The best results indicated that there was an increased risk of a power transform to fail in 99.7 times (indicated by the RR metric). The Random Forest performance is shown in confusion matrix of the Test set in Table 7. The evaluation metrics of Training and Test sets are present in Table 8.

Table 7: Confusion matrix of EFRI on Test set

Class	**(predicted) Positive**	**(predicted) Negative**
(real) Positive	15	3
(real) Negative	432	8918

Table 8: Evaluation metrics on Training and Test sets

Metric	**Training**	**Test**
Accuracy	97%	95.4%
AUC	97%	89%
F1-score	97%	52%
F1-score (failure class)	98%	6.5%
Recall	97%	89.4%
Recall (failure class)	99.9%	**83.3%**
Relative Risk	3061	**99.7**

The algorithm parameters were chosen using recall metric evaluated on the failure class. The results show that the model can be a helpful aid to the maintenance team on pinpointing the equipment in need of servicing.

7 Conclusions and Future Work

The exploration of predictive maintenance technologies is of great importance for the electrical sector, which is typically highly regulated and complex. Expanding predictive maintenance as applied in this article, leveraging the methodology in the maintenance of power transformers, and generalizing it to other transmission assets that allow the use of CAI and EFRI indicators for decision support can be a good path for future related work.

In an era of smart, data-driven, and cost-effective operations, the electrical sector can benefit from the development of a cloud platform to support the creation of other machine learning models, which is an important step towards predictive maintenance of power substation equipment. Cloud platforms offer scalability, data accessibility, and computing power that boost the development and deployment of complex ML models such as the ones presented in this work. The two ML indicators presented in this paper show that ML can be a useful addition to classic data analytics tools in supporting maintenance specialists in their decision-making process.

In summary, we used ML-based approaches to predict failures of power transformers by using chromatographic data (CAI) and data from SAGE (EFRI). In the case of CAI, we outperformed the classical DGA methods by using an RF model, as described in Section 6, obtaining an improvement in the accuracy of predicting failures on the test set by 19% compared to the best classical method. We tested the CAI model on the company's internal data, achieving AUC metrics close to 0.77 and 0.92 for transformers and reactors, respectively, evidencing the great generalization capability of the CAI model for the company's purpose. In the case of EFRI, by using a Random Forest model, we obtained an accuracy of 95.4% and a recall of 89.4% on the test set, showing that this method is quite helpful in predicting failures. Using the aforementioned indicators of failure, the maintenance planning of power plants can be performed in a data-driven way, leading to more efficient maintenance strategies. Moreover, the modeling process shown here opens up possibilities for modeling the downtime of power plants and wind turbines with reliability and reducing associated costs. In the initial phase of our prediction project, the estimation of the minimum size of the learning dataset required to achieve good predictions, based on the PAC-learning agnostic model, indicated that Support Vector Machines (SVM) and Random Forest, with a slight preference towards the latter, would be the most effective predictors. This is due to their relatively low Vapnik-Chervonenkis dimension, which ensures they can achieve good generalization performance even with smaller datasets.

References

[1] ONS. "Qualidade do Suprimento". https://www.ons.org.br/paginas/resultados-da-operacao/qualidade-do-suprimento, July 2024.

[2] R. K. Mobley. *An introduction to predictive maintenance*. Elsevier, 2002.

[3] J. LaRiviere, P. McAfee, J. Rao, V. K. Narayanan and W. Sun. "Where predictive analytics is having the biggest impact". *Harvard business review*, vol. 25, 2016.

[4] S. M. de Andrade Lopes and R. A. Flauzino. "A Novel Approach for Incipient Fault Diagnosis in Power Transformers by Artificial Neural Networks". In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pp. 1–5, 2021.

[5] E. Li. "Dissolved gas data in transformer oil—Fault Diagnosis of Power Transformers with Membership Degree", 2019.

[6] S.-A. Ahmadi and M. Sanaye-Pasand. "A Robust Multi-Layer Framework for Online Condition Assessment of Power Transformers". *IEEE Transactions on Power Delivery*, vol. 37, no. 2, pp. 947–954, 2022.

[7] M. Duval and A. dePabla. "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases". *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001.

[8] I. B. M. Taha, S. Ibrahim and D.-E. A. Mansour. "Power Transformer Fault Diagnosis Based on DGA Using a Convolutional Neural Network With Noise in Measurements". *IEEE Access*, vol. 9, pp. 111162–111170, 2021.

[9] I. E. Commission. "Mineral oil-impregnated electrical equipment in service – guide to the interpretation of dissolved and free gases analysis". IEC Standard 60599, IEC, 1999.

[10] Dept. Sci. Technol., Govt of India. "Technology Information Forecasting and Assessment Council (TIFAC) Laboratory", 2020.

[11] I. B. Taha. "DGALab: an extensible software implementation for DGA". *IET Generation, Transmission & Distribution*, vol. 12, pp. 4117–4124(7), October 2018.

[12] L. T. B. Santos, M. M. B. R. Vellasco and R. Tanscheit. "Decision Support System for Diagnosis of Power Transformers". In *2009 15th International Conference on Intelligent System Applications to Power Systems*, pp. 1–6, 2009.

[13] J. Chatterjee and N. Dethlefs. "A Dual Transformer Model for Intelligent Decision Support for Maintenance of Wind Turbines". In *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10, 2020.

[14] O. Catapult. "Platform for Operational Data (POD) Disseminated by ORE Catapult". `https://pod.ore.catapult.org.uk`, accessed on 9th May, 2023.

[15] B. Alzawaideh, P. T. Baboli, D. Babazadeh, S. Horodyvskyy, I. Koprek and S. Lehnhoff. "Wind Turbine Failure Prediction Model using SCADA-based Condition Monitoring System". In *2021 IEEE Madrid PowerTech*, pp. 1–6, 2021.

[16] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto and S. G. Alcalá. "A systematic literature review of machine learning methods applied to predictive maintenance". *Computers & Industrial Engineering*, vol. 137, pp. 106024, 2019.

[17] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone and A. Beghi. "Machine learning for predictive maintenance: A multiple classifier approach". *IEEE transactions on industrial informatics*, vol. 11, no. 3, pp. 812–820, 2014.

[18] Y. Ran, X. Zhou, P. Lin, Y. Wen and R. Deng. "A survey of predictive maintenance: Systems, purposes and approaches". *arXiv preprint arXiv:1912.07383*, 2019.

[19] M. Haarman, M. Mulders and C. Vassiliadis. "Predictive maintenance 4.0: predict the unpredictable". Technical report, PwC and Mainnovation, 2017.

[20] W. De Malsche. *Solving advanced micromachining problems for ultra-rapid and ultra-high resolution on-chip liquid chromatography*. Twente University Press, 2008.

[21] A. J. Martin and R. L. Synge. "A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins". *Biochemical Journal*, vol. 35, no. 12, pp. 1358, 1941.

[22] L. A. Pereira, L. C. Lima, A. J. Silva, P. A. Machado, M. F. Amorim, L. Ayru Filho, G. P. Azevedo, N. Lambert, P. D. Zarur, V. V. Tavares *et al.*. "SAGE-Um Sistema Aberto para a Evolução", 2014.

[23] E. Li, L. Wang and B. Song. "Fault Diagnosis of Power Transformers With Membership Degree". *IEEE Access*, vol. 7, pp. 28791–28798, 2019.

[24] E. Li. "Dataset directly extracted from Table 9 of Enwen Li (2019)", 2019. Table 9 available in the article: `https://ieee-dataport.org/documents/dissolved-gas-data-transformer-oil-fault-diagnosis-power\protect\penalty-\@M-transformers-membership-degree`.

[25] E. Li. "Dissolved gas data in transformer oil - Fault Diagnosis of Power Transformers with Membership Degree", 2019.

[26] S. I. Ibrahim, S. S. Ghoneim and I. B. Taha. "DGALab: an extensible software implementation for DGA". *IET Generation Transmission Distribution*, vol. 12, no. 18, pp. 4117–4124, 2018.

[27] M. Duval and A. dePabla. "Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases". *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001.

[28] D. R. Morais *et al.*. "Ferramenta inteligente para detecção de falhas incipientes em transformadores baseada na análise de gases dissolvidos no óleo isolante". Master's thesis, Universidade Federal de Santa Catarina, 2004.

[29] J. Kreps. "Questioning the lambda architecture". https://www.oreilly.com/radar/questioning-the-lambda-architecture/, July 2014.

[30] V. I. Levenshtein. "Binary codes capable of correcting deletions, insertions and reversals." *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, February 1966. Doklady Akademii Nauk SSSR, V163 No4 845-848 1965.

[31] The imbalanced-learn developers. "Under-sampling". `https://imbalanced-learn.org/stable/under_sampling.html`, accessed on 17th May, 2023.

[32] The imbalanced learn developers. "Over-sampling". `https://imbalanced-learn.org/stable/over_sampling.html`, accessed on 17th May, 2023.

[33] D. Katz, J. Baptista, S. P. Azen and M. C. Pike. "Obtaining Confidence Intervals for the relative risk in Cohort Studies". *Biometrics*, vol. 34, no. 3, pp. 469–474, 1978.

[34] A. Ehrenfeucht, D. Haussler, M. Kearns and L. Valiant. "A general lower bound on the number of examples needed for learning". *Information and Computation*, vol. 82, no. 3, pp. 247–261, 1989.

[35] C.-C. Chang and C.-J. Lin. "LIBSVM: A Library for Support Vector Machines". *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, 2011.

[36] L. Breiman. "Random forests". *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[37] J. Hühn and E. Hüllermeier. “FURIA: An algorithm for unordered fuzzy rule induction”. *Data Mining and Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, 2009.

[38] B. Pfahringer. “Random Model Trees: An Effective and Scalable Regression Method”. Working paper. University of Waikato, Department of Computer Science. Hamilton, New Zealand. Available at `https://researchcommons.waikato.ac.nz/handle/10289/4056`.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[40] I. E. Commission. “Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service”. Technical report, IEC, Geneva, Switzerland, 1978.

[41] R. Hooshmand and M. Banejad. “Fuzzy Logic Application in Fault Diagnosis of Transformers Using Dissolved Gases”. *Journal of Electrical Engineering and Technology*, vol. 3, no. 3, pp. 293–299, 2008.

[42] 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.

[43] M. Duval and A. dePabla. “Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases”. *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001.

[44] A. Abu-Siada and S. Islam. “A new approach to identify power transformer criticality and asset management decision based on dissolved gas-in-oil analysis”. *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 3, pp. 1007–1012, 2012.

[45] C. L. Sistrom and C. W. Garvan. “Proportions, odds, and risk”. *Radiology*, vol. 230, no. 1, pp. 12–19, 2004.

[46] scikit-learn developers. “sklearn.tree.DecisionTreeClassifier”. `https://scikitlearn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html`, accessed on 17th May, 2023.

[47] scikit-learn developers. “scikit-learn: Machine Learning in Python”. `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html`, accessed on 17th May, 2023.