On the use of Machine Learning for predictive maintenance of power transformers

1st Carla Pacheco, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, cpacheco@inf.puc-rio.br

2nd Vagner Paes, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, vpaes@inf.puc-rio.br

3rd Marcelo de Carvalho, *Furnas, Eletrobras*, Rio de Janeiro, Brazil, marcarv@furnas.com.br

4th Felipe Lopes, *Furnas, Eletrobras*, Rio de Janeiro, Brazil, feliperl@furnas.com.br

5th Gabriel Machado, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, gresende@inf.puc-rio.br

6th Alex Garcia, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, agarcia@inf.puc-rio.br

7th Edmilson Neto, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, eneto@inf.puc-rio.br

8th Jefferson Santos, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, jsantos@inf.puc-rio.br

9th Edward Haeusler, *Informatics Department, PUC-Rio*, Rio de Janeiro, Brazil, hermann@inf.puc-rio.br

10th Ana Marotti, *Furnas, Eletrobras*, Rio de Janeiro, Brazil, amarotti@furnas.com.br

***Abstract*—This paper focuses on the use of machine learning algorithms to assist predictive maintenance aiming at reducing downtime and maintenance costs associated with power transformers. The paper presents two ML predictive indicators, Chromatographic Assay Indicator (CAI) and Electrical Failure Risk Indicator (EFRI), which use chromatographic and sensors data, respectively. The CAI evaluation showed a significant improvement in predicting failures compared with classical methods, whereas the EFRI tests showed it can be helpful to help maintenance team in identifying potential problems. The proposed solution integrates classical chromatographic analysis techniques with these ML indicators and aims at supporting maintenance specialists in decision-making processes, leading to more efficient maintenance management and reduced costs associated with equipment downtime.**

***Index Terms*—predictive maintenance, machine learning, power transformers, random forest algorithm, energy transmission**

I. 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 (Operador Nacional do Sistema Elétrico), the temporary interruptions of energy supply may have several causes, such as failures in equipment or in protection and control systems.

It is undoubtedly important to avoid energy supply interruption. Researches on maintenance effectiveness indicate that a third of all maintenance costs is wasted as a result of unnecessary or incorrectly conducted maintenance [1]. The predominant reason for this ineffective maintenance management is the lack of factual data to quantify the actual need for repair of plant equipment and systems [1]. The traditional methods of maintenance are *reactive* and *preventive* [2]. The former is based on the 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. The *predictive methods* are in the vanguard 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*.

The *predictive maintenance* uses direct monitoring of mechanical condition, system efficiency, and other indicators to determine the actual MTTF (mean-time-to-failure) or efficiency loss for each piece of plant equipment. Such predictive method aims to ensure the maximum interval between maintenance services and minimize the amount as well as the cost of unplanned downtime caused by failures. The monitoring may use disruptive technologies such as Internet of Things (IoT), Cloud Computing and Machine Learning (ML). In the lights of [3], maintenance has been one of the areas with the largest number of applications of modern Predictive Analysis techniques. IoT allows real-time telemetry to provide data on the operation of systems and equipment to be submitted to analytical procedures in order to make their failures more predictable.

This work aims to improve the power transformers maintenance, and thus to 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). The solution being developed exhibits these two ML indicators together with several classical indicators and traditional data analytics tools, in order to support maintenance specialists.

The CAI indicator uses chromatographic data, whereas the EFRI indicator uses sensors and maintenance data. Both CAI and EFRI models used the Random Forest algorithm. The former model had accuracy and F1-score metrics greater than 92% when evaluated on the test set. The latter model had 95% accuracy and 52% F1-score. The results of CAI were

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

benchmarked against classical DGA methods, presenting a much superior performance.

This paper is organized into six more sections. Section II describes the main related works. Section III describes the use of chromatography and SCADA systems in power transformers monitoring and maintenance. Section IV addresses the failure risk prediction based on chromatographic data whereas Section V details the workflow of the Electrical Failure Risk Indicator based on both maintenance and monitoring data. Section VI presents the results of experimental evaluation and the discussion conducted in this paper. In the end, Section VII concludes and presents future works.

II. 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 dissolved gas analysis (*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 gasses 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 MLP 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 gasses 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 Roger [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.

III. Current maintenance practice

Maintenance plays a major role in the industrial sector, as it greatly affects expenses 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 the equipment 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 is a simplistic approach where interventions are carried out only after failures happen;
- Preventive Maintenance (PvM) involves maintenance actions executed based on a predetermined schedule, either time-based or process-based;
- Predictive Maintenance (PdM) involves maintenance performed based on estimated equipment health status. PdM systems utilize prediction tools, historical data, health factors, statistical inference methods, and engineering approaches to detect impending failures in advance and enable timely interventions before failure occurs.

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 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, Predictive Maintenance (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.

A. *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, referred to as the analyte, 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, it is possible to highlight the analysis of the quality of certain transmission assets. There is a standardization for gas in oil samples that serves as a reference in Brazil, the NBR7070. Power transformers use oil in their closed circulation system to cool the equipment. However, as a secondary function, this oil can be used to interpret the amount of dissolved gas in the oil [22]. Such an indicator can be based on the analysis of 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. Its function is to support decision-making in predictive maintenance of transmission assets by comparing the composition of the new insulating oil with the sample to be analyzed. Based on this data, several inferences can be made about the condition of the transmission asset, as well as prevention of equipment defects, as the proportion of these compounds can indicate electrical and thermal defects in the analyzed asset.

B. *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 is a combination of two types of control systems: Supervisory Control and Data Acquisition (SCADA) and Energy Management System (EMS).

- 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 CEPEL [23]. It is used by dozens of electric power generation, transmission, and distribution agents in Brazil, particularly the founding companies of CEPEL (Chesf, Furnas, Eletronorte, and Eletrosul), as well as the National Electric System Operator (ONS), in all of their control centers. 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 [24].

IV. Evaluating Failure Risk from Chromatographic Data

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 ML to develop a classifier that assists maintenance specialists in their decision-making process. Upon receiving the company's data, it became apparent 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 [25]–[30] 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. Despite this limitation, the results were very satisfactory. In the testing dataset, the risk indicator outperformed classical chromatographic methods, as we will describe in Section VI-A. The risk indicator was also tested on the company's data, yielding encouraging results that demonstrate its ability to generalize the classification capability to the company's data.

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 III-A, namely, H_2, CH_4, C_2H_2, C_2H_4, and C_2H_6. After joining the datasets, we started the data pre-processing 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 nondetectable value by adding a constant value, as suggested in [31]: $R_1 = \frac{CH_4}{H_2} + 0.4$, $R_2 = \frac{C_2H_2}{C_2H_4} + 0.4$, $R_4 = \frac{C_2H_6}{CH_4} + 0.4$, 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:

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

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.

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

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

V. Evaluating Failure Risk from Monitoring Data

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 [32] algorithm 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.

A. Methodology

We elaborated a methodology that covers the whole workflow of the ML algorithms and pre-processing as depicted in Fig. 1. 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.

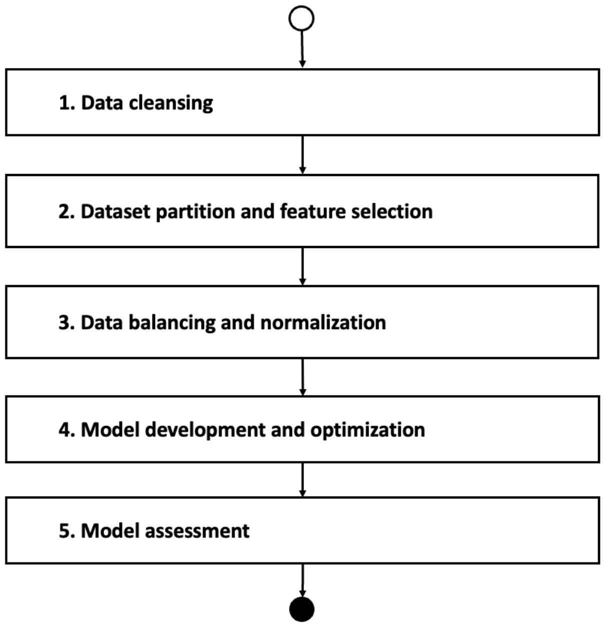

Fig. 1. Workflow for power transformer failure prediction

1) Data cleansing: As a premise, we looked at the digital data with measurements from more than one power transformer, representing the half of the base. Afterwards, we discarded analogical and digital attributes which variance was equal or less than one, considering the amount of missing values in this removal and keeping an attribute of temperature of auxiliary winding. In total, we worked with 147 attributes.

2) Dataset partition and feature selection: The remaining data was split into two disjoint subsets: Training (80%, being 71 *failure* and 37,401 *nonfailure*) and Test (20%, being 18 *failure* and 9,350 *nonfailure*). In both sets, the proportion of *nonfailures* is 520 times bigger than *failures*. The missing values at Training set were replaced by the mean of the analogical attribute. Digital attributes did not present missing values. Afterwards, we conducted preliminary tests with the decision tree based algorithm *DecisionTreeClassifier* from Python's sklearn library [33]. 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* [34], 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, AUROC (Area Under ROC curve). After that, we ranked the most important attributes and selected the top 19 analogical plus the remaining 42 digital attributes to use in the modeling from now on, totaling 61 attributes.

3) Data balancing and normalization: There is a huge discrepancy in the amount of *failures* and *nonfailures* in the dataset. To overcome this problem, we chose to balance the Training data trough the combination of two techniques, respectively: (i) reduction of *nonfailures* 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 [35] to reduce the majority class instances. Then, we applied the SMOTE (Synthetic Minority Over-sampling Technique) algorithm [36], from the same library, to over-sample the minority class, based on the parameter of the nearest neighbors, which was chosen to be 10. Both under- and over-sample 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 $Norm_{x_{Test}} = \frac{x_{Test}-\overline{x}_{Treino}}{\sigma_{Treino}}$. With this in mind, the missing data of Test set was replaced by the mean of the respective attribute of the Training set.

4) 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) 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) [37] 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 *nonfailures*, $FOR = \frac{FN}{FN+TN}$. The latter is the rate of failures correctly classified over the total amount of instances predicted as failures, $PPV = \frac{TP}{TP+FP}$. Hence, $RR = \frac{PPV}{FOR}$ is dimensionless.

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 VI.

VI. 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.

A. CAI results

We ran a benchmark against some algorithms and possible scenarios. The experiments were executed with the following four algorithms: (i) Support Vector Machine (LibSVM) [38]; (ii) Random Forest [39]; (iii) Fuzzy Unordered Rule Induction Algorithm (FURIA) [40]; and (iv) Random Trees [41]. We performed the experiments in the following ways:

- 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 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 scikit-learn [42] in the training set by using ten folds.

TABLE I
EXPERIMENT RESULTS OBTAINED IN THE BEST SCENARIO

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

In order to validate the Random Forest model, we compared it with classical *DGA* methods, such as: (i) Rogers, (ii) Doernenburg, (iii) NBR 7274, (iv) IEC 599 and (v) Duval's Triangle [22], [43]–[46]. The comparison results are displayed in Table II, from which we can see that the Doernenburg's method had the highest accuracy within the classical *DGA* methods. The hybrid DIEC-R method, which consists of a combination of Doernenburg's and IEC Ibrahim's method [28], presented the best result among all classic/hybrid methods. From the same Table, we can clearly notice that the *Random Forest* model outperformed the other classical methods in almost 19 percentage points, presenting 92.2% accuracy and 92.1% F1-score in the test set.

TABLE II
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%**

B. EFRI results

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 III. The evaluation metrics of Training and Test sets are present in Table IV.

TABLE III
CONFUSION MATRIX OF EFRI ON TEST SET

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

TABLE IV
EVALUATION METRICS ON TRAINING AND TEST SETS

Metric	**Training**	**Test**
Accuracy	97%	95.4%
AUROC	97%	89%
F1-score	97%	52%
F1-score of *failure*	98%	6.5%
Recall	97%	89.4%
Recall of *failure*	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.

VII. 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 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 works.

In a era of smart, data-driven and cost-effective operations, 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 substations equipment. Cloud platforms offer scalability, data accessibility and computing power that boosts 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 experts in their decision making process.

In summary, we have 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 have outperformed the classical *DGA* methods by using a *RF* model, as described in the *Results and Discussion section*, obtaining an improvement of the accuracy to predict failures on the test set by 19% compared to the best classical method. In the case of *EFRI*, by using a *RF* model, we obtained on the test set an accuracy of 95.4% and recall of 89.4%, which show that this method is quite helpful in predicting failures. Using the aforementioned indicators of failure, the maintenance planning's decision of the 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 instance, for modeling the downtime of power plants, as well as wind turbines, with reliability and reducing associated costs.

REFERENCES

[1] S. Yacoub, B. Cukic, and H. H. Ammar, "A scenario-based reliability analysis approach for component-based software," *IEEE transactions on reliability*, vol. 53, no. 4, pp. 465–480, 2004.

[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)*, 2021, pp. 1–5.

[5] E. Li, "Dissolved gas data in transformer oil—fault diagnosis of power transformers with membership degree," 2019. [Online]. Available: https://dx.doi.org/10.21227/h8g0-8z59

[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. 111 162–111 170, 2021.

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

[10] Dept. Sci. Technol., Govt of India, "Technology information forecasting and assessment council (tifac) laboratory," New Delhi, India, 2020.

[11] I. B. Taha, "Dgalab: an extensible software implementation for dga," *IET Generation, Transmission & Distribution*, vol. 12, pp. 4117–4124(7), October 2018, n/a. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2018.5564

[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*, 2009, pp. 1–6.

[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)*, 2020, pp. 1–10.

[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*, 2021, pp. 1–6.

[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, p. 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] "Predictive maintenance 4.0: predict the unpredictable," PwC and Mainnovation, Tech. Rep., 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, p. 1358, 1941.

[22] 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.

[23] "Cepel description," https://sage.cepel.br/index.php/pt/sage/visao-geral, accessed: 2023-05-14.

[24] 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.

[25] E. Li, L. Wang, and B. Song, "Fault diagnosis of power transformers with membership degree," *IEEE Access*, vol. 7, pp. 28 791–28 798, 2019.

[26] 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\break-transformers-membership-degree.

[27] "Dissolved gas data in transformer oil - fault diagnosis of power transformers with membership degree," 2019. [Online]. Available: https://dx.doi.org/10.21227/h8g0-8z59

[28] 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.

[29] 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.

[30] 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.

[31] J. Kreps, "Questioning the lambda architecture," *Online Article*, p. 205, July 2014. [Online]. Available: https://www.oreilly.com/radar/questioning-the-lambda-architecture/

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

[33] scikit-learn developers, "sklearn.tree.decisiontreeclassifier," https://scikitlearn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html, accessed on 17th May, 2023.

[34] 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.

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

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

[37] 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, p. 469–474, 1978.

[38] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," vol. 2, no. 3, 2011.

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

[40] 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.

[41] B. Pfahringer, "Random model trees: An effective and scalable regression method," 2010, working paper. University of Waikato, Department of Computer Science. Hamilton, New Zealand. Available at https://researchcommons.waikato.ac.nz/handle/10289/4056.

[42] 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.

[43] I. E. Commission, *Interpretation of the analysis of gases in transformers and other oil-filled electrical equipment in service*, 1st ed. Geneva, Switzerland: International Electrotechnical Commission, 1978, no. 599, iEC Publication.

[44] 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.
[45] 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.
[46] 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.