FEDERAL UNIVERSITY OF ITAJUBÁ

GRADUATE PROGRAM

IN ELECTRICAL ENGINEERING

DATA ANALYSIS METHODOLOGY UTILIZING THE STATISTICAL METRICS WEIGHT OF EVIDENCE (WoE) AND INFORMATION VALUE (IV) TO ASSIST IN ASSET MANAGEMENT OF POWER TRANSFORMERS

Gabriela Sampaio Rêma

December 2024

Itajubá - MG

FEDERAL UNIVERSITY OF ITAJUBÁ

GRADUATE PROGRAM
IN ELECTRICAL ENGINEERING

Gabriela Sampaio Rêma

DATA ANALYSIS METHODOLOGY UTILIZING THE STATISTICAL METRICS WEIGHT OF EVIDENCE (WoE) AND INFORMATION VALUE (IV) TO ASSIST IN ASSET MANAGEMENT OF POWER TRANSFORMERS

Thesis submitted to the Graduate Program in Electrical Engineering as part of the requirements to obtain the title of Doctor of Science in Electrical Engineering.

Concentration Area: Electrical Power Systems

Advisor: Ph.D. Benedito Donizeti Bonatto
Co-advisor: Ph.D. Antonio Carlos Siqueira de Lima

December 2024

Itajubá - MG

ACKNOWLEDGMENT

I acknowledge God for the health, wisdom and strength to complete this cycle.

I acknowledge advisor Benedito Donizeti Bonatto for his guidance and support in carrying out this work and to co-advisor Antonio Carlos Siqueira de Lima for his co-supervision.

I acknowledge Rafael Marinho Soares and André Tomaz de Carvalho for all the the teachings and encouragement.

I acknowledge the financial support provided in part from UNIFEI – Federal University of Itajuba; UNICAMP – Universidade Estadual de Campinas; CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil –finance code 001; CNPq – National Council for Scientific and Technological Development - Brazil; FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais, under grant APQ-03609-17; and INERGE – Instituto Nacional de Energia Elétrica. One author was funded by grant #2021/11380-5 from CPTEn – São Paulo Center for Energy Transition Studies; FAPESP – São Paulo Research Foundation; and FAPEMIG (grant APQ-02845-2021).

I acknowledge ISA CTEEP for providing the data used in the case study and for granting permission for its disclosure. The development was carried out within the scope of the Research, Development and Innovation Program of the ANEEL, reference PD-00068-0067/2023.

I acknowledge my family and friends for all their support and my little dog Lelê for her companionship throughout the development of this work.

ABSTRACT

The Brazilian electric power transmission sector faces a significant challenge involving the end of the regulatory useful life of several of equipment. Given the technical and economic infeasibility of renewing all depreciated assets from a regulatory standpoint, the need for an assertive risk management analysis and a reliable assessment of the physical useful life of assets is emphasized, especially power transformers, the main asset in the electrical energy transmission sector. In view of this scenario, the objective of the proposed thesis is to add value to asset management through the development of a data analysis methodology to assist in decision-making regarding the direction of maintenance investment in power transformers. Due to their similarity, reactors are also evaluated. To this end, data on moisture in the insulating oil of the equipment were used and the following categorical variables: voltage class, installation region (Regional), criticality, type, and age of the equipment. It is noteworthy that these variables are technical registry data of the assets, and the water content is an essential characteristic for determining the operational condition of the insulating oil, being one of the properties measured in the physicochemical tests. The original contribution of the thesis is the selection of categories with greater weight and categorical variables with higher predictive power using the statistical metrics Weight of Evidence (WoE) and Information Value (IV). Analyzing the predictive importance of a variable before developing a predictor can lead to better performing models. Furthermore, data based decisions lead to more assertive and proactive actions, and the prioritization of variables for evaluation is an important contribution, especially considering large equipment parks. The methodology was applied to a dataset of almost 10 thousand oil samples from 795 power transformers and reactors from the ISA CTEEP, electrical energy transmission company in Brazil, responsible for approximately 95% of the energy transmitted in the state of São Paulo and about 30% of all energy in Brazil.

Keywords – Power transformers; Asset management; Preventive actions; Insulating oil; Moisture in the oil; Variables prioritization; Information Value; Weight of Evidence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADASYN	Adaptive Synthetic Sampling Approach
AI	Artificial Intelligence
ANEEL	*Agência Nacional de Energia Elétrica* – Brazilian Electricity Regulatory Agency
CF	Consequence Factor
CNN	Convolution Neural Network
df	Data frame
DGA	Dissolved Gas Analysis
DIT	*Demais Instalações de Transmissão* – Other Transmission Installations
DWT	Discrete Wavelets Transform
EDM	Extended Debye Model
EPE	*Empresa de Pesquisa Energética* – Energy Research Company
EW	Entropy-Weighted
FA	Furan Analysis
FCM	Fuzzy C-Means
FDS	Frequency Domain Spectroscopy
FIS	Fuzzy Inference System
FLC	Fuzzy Logic Controller
FR	Frequency Ratio
FRA	Frequency Response Analysis
FSM	Failure-Sensitive Matrix
FTIR	Fourier Transform Infrared
GEP	Gene Expression Programming
GRA	Grey Relational Analysis
GWO	Gray Wolf Optimizer
HI	Health Index
HPLC	High-Performance Liquid Chromatography
IRT	Infrared Thermography
IV	Information Value
KL	Kullback–Leibler
kNN	k-Nearest Neighbor

MCPSE	*Manual de Controle Patrimonial do Setor Elétrico* – Asset Control Manual of the Electric Sector
MDPI	Multidisciplinary Digital Publishing Institute
ML	Machine Learning
MLP	Multilayer Perceptron
MME	*Ministério de Minas e Energia* – Ministry of Mines and Energy
MSC	Multi-State Condition
MUSIC	Multiple Signal Classification
NIR	Near-Infrared Spectroscopy
PCA	Principal Component Analysis
PD	Partial Discharge
PDC	Polarization and Depolarization Current
PDE	*Plano Decenal de Expansão de Energia* – Ten-Year Energy Expansion Plan
PNN	Probabilistic Neural Network
PRISMA-ScR	Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
PSO	Particle Swarm Optimization
RB	*Rede Básica* – Basic Grid
RVM	Recovery Voltage Monitoring
SMOTE	Synthetic Minority Over-Sampling Technique
SRC	Sweep Reflection Coefficient
SVM	Support Vector Machine
TB	*Bauru*
TC	*Cabreúva*
TE	*Expansão Nacional*
TS	*São Paulo*
TT	*Taubaté*
VF	Vector Fitting
WoE	Weight of Evidence

1. INTRODUCTION

Asset management plays an essential role when it comes to services that require high availability and high reliability, as is the case in the electrical sector. Considering the entire lifecycle of the asset, the Operation and Maintenance stage represents the longest period, and in this stage, asset management includes maintenance management.

The efficient planning and scheduling of maintenance activities depend on a reliable assessment of the condition of the equipment. This is particularly critical in the transmission sector, where business sustainability is directly linked to the availability of assets and facilities.

Failures in electrical network equipment result in substantial expenses for power utilities. Therefore, employing assessment techniques is essential for effectively diagnosing and estimating the true operational status of such equipment. Efficient diagnoses enable the management of this asset chain, aiming to strike the optimal balance between investments, maintenance costs, and operational performance [1].

This thesis proposes an original contribution to assist in the condition assessment of power transformers, the main asset in the transmission sector. Due to their similarity, reactors are also evaluated. The following sections will present the main motivations, objectives, and contributions of this work.

1.1. MOTIVATION

At a global level, in order to meet the growing demand for electrical energy, transmission companies made significant investments with a certain temporal concentration for the physical implementation of transmission electrical systems and currently, they are facing a major challenge, which involves the end of the regulatory lifespan of various equipment [2].

In Brazil, the Energy Research Company (*Empresa de Pesquisa Energética* – EPE) and the Ministry of Mines and Energy (*Ministério de Minas e Energia* – MME) point out, in the Ten-Year Energy Expansion Plan (*Plano Decenal de Expansão de Energia* – PDE) [3], that by 2031, the amount of fully depreciated transmission assets could reach, approximately, US$ 219 billion (considering US$ 1 equal to R$ 6.08). This context can bring risks to the system, impact sectoral planning, and affect the capacity to support the energy transition.

It is worth highlighting that in Brazil, in general, transmission equipment is operating satisfactorily, even with time exceeding regulatory expectations, which implies pressure to obtain spare parts and qualified labor to carry out maintenance. This need increases the risks of future failures for the transmission system, in addition to being able to cause penalties and losses for the transmission companies that operate the assets. In these terms, decision-making on how to proceed, that is, maintain, modernize or replace assets at the end of their useful life, is one of the biggest challenges for network operators and regulators, with repercussions also for the energy production industry. transmission system components [2].

Another important issue to highlight is that the management of transmission assets, on an international level, is characterized by the diversity of metrics used to evaluate performance. There is no consensus on the best practices to be adopted, which limits the exchange of international experiences. It is noteworthy that technical standards should consider the entire lifecycle, from design to decommissioning of equipment, including installation, operation, maintenance, modernization, and diagnostics. In this sense, there is a need and an opportunity for the development of standards and regulations for asset management in transmission networks [2].

1.2. Objectives and original contribution

Given the technical and economic infeasibility of renewing all depreciated assets from a regulatory standpoint, the need for an assertive risk management analysis and a reliable assessment of the physical useful life of assets is emphasized, especially power transformers, the main asset in the electrical energy transmission sector.

Considering that dielectric failures are the main failure mode in transmission substation transformers in Brazil and worldwide [4], data analysis methodologies related to such failures can assist in preventive actions.

Thus, the general objective of the proposed thesis is to add value to asset management through the development of a data analysis methodology to assist in decision-making regarding the direction of maintenance investment in power transformers. Due to their similarity, reactors are also evaluated.

To achieve the general objective of the thesis, the following specific objectives were developed:

- Scoping review for analysis and understanding of what has been developed worldwide on the proposed topic;
- Analysis of tests for the evaluation of the dielectric of power transformers;
- Descriptive analysis of technical data and physicochemical test data on the oil of power transformers and reactors in operation;
- Application of statistical metrics that evaluate the predictive importance of a variable in relation to the output;
- Development of an analysis methodology to assist in the management of power transformers and reactors through the prioritization of categorical variables, based on real data from equipment in operation at ISA CTEEP, an electrical energy transmission company in Brazil.

The scoping review carried out was published in the article "Emerging Trends in Power Transformer Maintenance and Diagnostics: A Scoping Review of Asset Management Methodologies, Condition Assessment Techniques, and Oil Analysis" [5] in IEEE Access.

The methodology and results of this thesis are presented in the article “Data analysis methodology utilizing the statistical metrics Weight of Evidence (WoE) and Information Value (IV) to assist in asset management of power transformers” also published in IEEE Access [6].

Figure 1.1 illustrates the object of study of this thesis, a power transformer in operation, installed in an ISA CTEEP substation.

Figure 1.1 – Power transformer in operation, installed in an ISA CTEEP substation [5].

The original contribution of the thesis is the selection of categories with greater weight and categorical variables with higher predictive power using the statistical metrics Weight of Evidence (WoE) and Information Value (IV). Analyzing the predictive importance of a variable before developing a predictor can lead to better performing models. Furthermore, data based decisions lead to more assertive and proactive actions, and the prioritization of variables for evaluation is an important contribution, especially considering large equipment parks.

The proposed methodology is based on data on moisture in the insulating oil of the equipment and the following categorical variables: voltage class, installation region (Regional), criticality, type, and age of the equipment. The feasibility of applying the methodology is highlighted, as these categorical variables consist of technical registration data of transformers and reactors, and the moisture content is an essential characteristic for determining the operational condition of the insulating oil, being one of the properties measured in the physicochemical tests performed on the oil.

1.3. THESIS STRUCTURE

The document is structured as follows: this chapter presents an introduction to the topic, the motivation, objectives, and contributions of the thesis.

Chapter 2 provides a foundation on failures in transformers, the assessment of the condition of insulating oil and the metrics Weight of Evidence (WoE) and Information Value (IV). Chapter 3 presents the scoping review conducted.

Chapter 4 presents the proposed methodology and a comparative analysis of the application of the proposed metrics in predictive models from different areas. The results and discussions are presented in Chapter 5.

Finally, Chapter 6 presents the conclusions and include publications and possibilities for future work. In the APPENDIX are complementary analysis presentations.

2. THEORETICAL FOUNDATION

This chapter provides a foundation on transformer failures, assessment of the condition of insulating oil and the metrics Weight of Evidence (WoE) and Information Value (IV).

2.1. FAILURE MODES IN POWER TRANSFORMERS

There is limited literature available in the public domain discussing transformer failure statistics [7], both in relation to failure numbers and modes [4].

For this reason, in 2008, the Cigrè formed Working Group A2.37 Transformer Reliability and published the resulting Transformer Reliability Survey material [4], with contributions from 58 concessionaires from 21 countries. A total of 964 failures occurring in transformers were examined, with 799 occurring in transmission substation transformers and 165 occurring in power plant step-up transformers.

The mode describes the nature of the failure. The failure modes in the study were classified as follows: (a) dielectric – primarily related to PD and flashover; (b) electrical – open circuit, short circuit, and failure in electrical contacts; (c) thermal – overheating and hotspot; (d) chemical – oil contamination and corrosion; and (e) mechanical – bending, breakage, displacement, and loosening.

Figure 2.1 shows the classification of the failure modes of the 964 failures collected. Dielectric-origin failures are the majority, representing more than 36%.

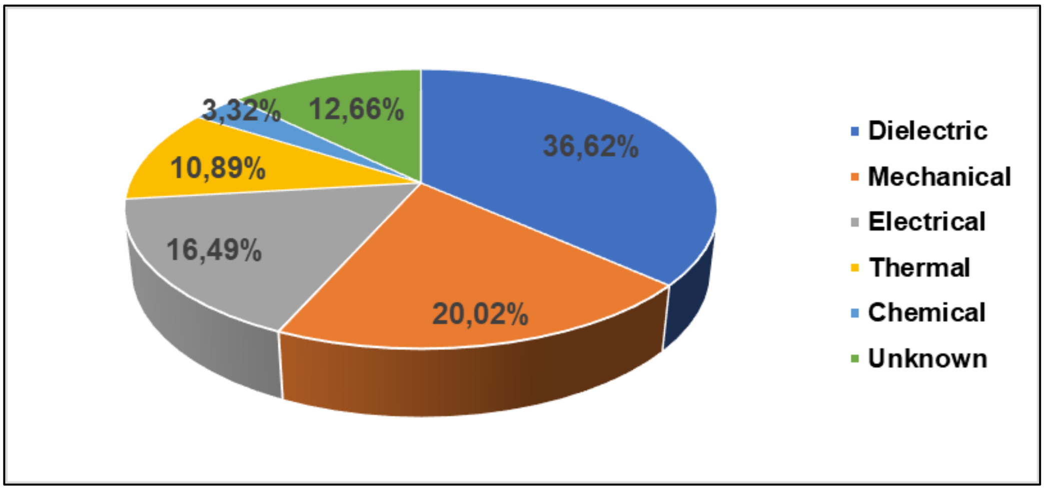


Figure 2.1 – Classification of the failure modes of the 964 power transformer failures from concessionaires in 21 countries [4].

Figure 2.2 shows the failure mode classification analysis for transmission substation and power plant transformers. In the case of power plant transformers, thermal-origin failures surpassed dielectric failures.

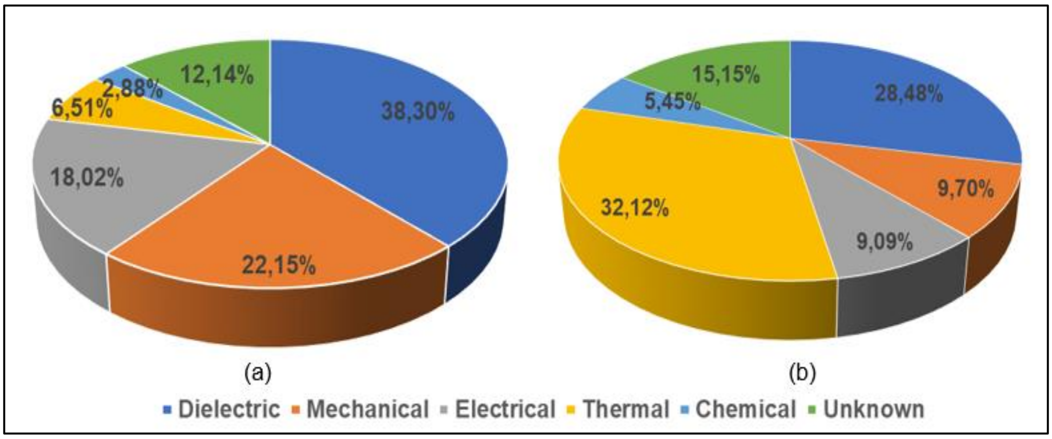


Figure 2.2 – Classification of the failure modes of (a) 799 failures in transmission substation transformers and (b) 165 failures in power plant transformers [4].

Within the Brazilian electrical sector, the study was conducted with data from 3198 transformers and reactors from voltage classes starting at 138 kV. The information came from thirteen concessionaires in the country, representing more than 70% of the installed capacity in Brazil. The classification of failure modes is shown in Figure 2.3. It was found that dielectric failures accounted for more than 45% of the total failures.

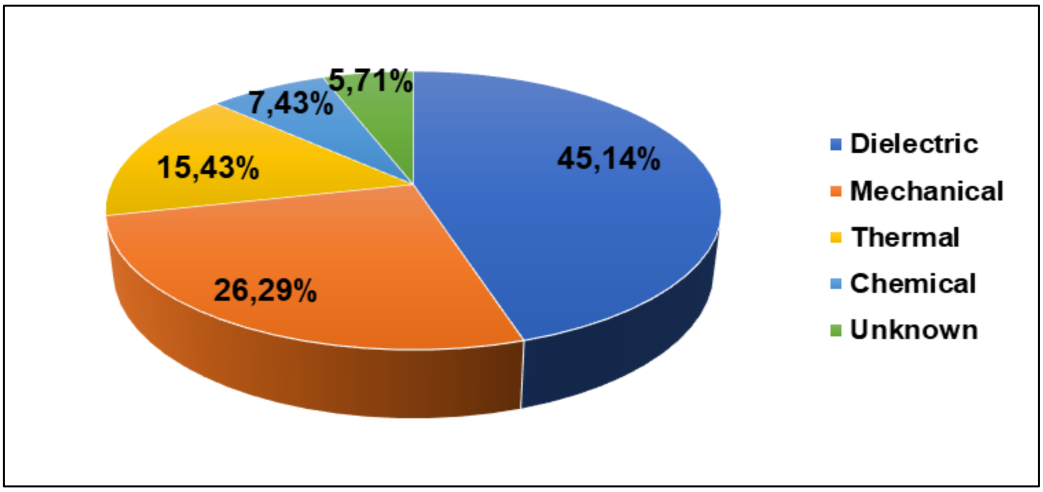


Figure 2.3 – Classification of failure modes of transformers and reactors with voltage classes starting from 138 kV from concessionaires in Brazil [4].

2.2. Assessment of the condition of insulating oil

Insulating oil serves as both a dielectric and cooling agent in high-voltage electrical equipment. ABNT NBR 10576:2017 [8] based on IEC 60422:2013 – Mineral insulating oils in electrical equipment - Supervision and maintenance guidance – standardizes guidelines for supervision and maintenance related to mineral insulating oil in equipment in Brazil. According to the standard, oil is defined as a dielectric, heat transfer, and arc extinction medium, with its performance dependent on certain basic characteristics that can affect the overall performance of electrical equipment.

Transformers are most significant apparatus in interconnected power systems with consideration of several aspects like reliability, efficiency, economy, wide applications and operation [9]. Therefore, condition assessment of these worthy components has been the main subject of many researches. Solid and oil insulation of transformers are the two dominant factors limiting transformers lifetime [10].

Physical-chemical and gas chromatography tests are conducted on the insulating oil of power transformers, respectively, to verify proper operating conditions and to analyze dissolved gases.

The physical-chemical tests of the oil, classified as Group 1 - routine tests - evaluate the following properties: a) color and appearance; b) dielectric loss factor; c) neutralization index (acidity); d) dielectric strength; e) interfacial tension; f) moisture; and aim to determine if the oil conditions are suitable for continuous operation and suggest the type of corrective action necessary, if applicable [8].

2.2.1. PERIODICITY

According to ABNT NBR 10576:2017 [8], it is not possible to establish a general rule for the periodicity of oil analysis for equipment in service. For power transformers and reactors, the following sampling periodicity is suggested:

a) Before energization;

b) 24 to 72 hours after energization;

c) One month after energization;

d) Semiannually until the end of the warranty;

e) After the warranty period, conduct physical-chemical tests and dissolved gas analysis by chromatography annually.

Additionally, the standard specifies that other criteria should be followed in special conditions, such as overload or changes close to the limit values in significant oil properties, which may require more frequent analyses.

2.2.2. MOISTURE IN THE OIL

Moisture in oil and solid insulation has a significant impact on the operational condition and health of transformers, as changes in the dielectric can lead to failures. The main sources of moisture increase in transformer oil are atmospheric moisture ingress and degradation of solid insulation. Depending on the amount of water, insulation temperature, and oil aging degree, the moisture content of insulating oils influences: a) dielectric strength of the oil; b) cellulose insulation; c) oil and cellulose insulation aging rate [8].

2.2.2.1. PARAMETERS FOR TRANSFORMERS AND REACTORS

The oil from equipment in service can be classified as "meets limit values" or "does not meet limit values", based on the evaluation of its properties.

In accordance with ABNT NBR 10576:2017 [8], the moisture content limit varies according to the voltage class of transformers and reactors. The maximum values are shown in Table 2.1 [8], measured in mg/kg (or ppm). When these values are exceeded, corrective action should be taken to treat the oil.

Table 2.1 – Maximum moisture parameter for corrective action in the oil of transformers and reactors in use [8].

Voltage class range	Maximum moisture parameter
≤ 72.5 kV	40 mg/kg or ppm
> 72.5 kV and ≤ 145 kV	30 mg/kg or ppm
> 145 kV	20 mg/kg or ppm

2.3. WEIGHT OF EVIDENCE (WOE) AND INFORMATION VALUE (IV)

In this thesis, the use of WoE and IV metrics is proposed for the selection of categorical variables based on their predictive power in the context of assisting in the management of power transformers and reactors. The WoE metric indicates the difference between the proportion of good and bad samples and can be used to measure the performance of categories. The IV is used to indicate the degree of contribution of the variable to the prediction of a target variable [11]. The proposed method offers a quantitative way to evaluate and filter key information, generating more efficient planning scenarios [12].

Equation (1) [11] shows the calculation of the WoE metric, where *i* corresponds to each category and *T* to the total value of the categories.

$$WoE_i = \ln\left(\frac{\%\ bad\ sample_i}{\%\ good\ sample_i}\right)$$

or (1)

$$WoE_i = \ln\left(\frac{bad\ sample_i}{bad\ sample_T}\right) - \ln\left(\frac{good\ sample_i}{good\ sample_T}\right)$$

The ratio of bad and good samples over their respective total values is a number between 0 and 1, and the natural logarithm ($\ln$) of a number between 0 and 1 will always be negative. This occurs because the $\ln(x)$ function describes the exponent to which you need to raise the number e (approximately 2.71828) to get the value x. Since e^y for $y < 0$ results in a number between 0 and 1, the natural logarithm of any number in this range will be negative. The closer the ratio is to 1, meaning values of samples in the category are close to the total values, the greater the natural logarithm will be.

Still considering the Equation (1), when WoE is positive, the category is associated with undesired events, bad samples, and conversely, negative WoE is associated with desired events, good samples. A positive WoE value indicates that the predictor increases the likelihood of the outcome occurring, suggesting a positive association. On the other hand, a negative WoE value implies that the predictor decreases the likelihood of the outcome, indicating a negative association. In summary, the higher the WoE value, the greater the weight of evidence for undesired events. Table 2.2 provides a didactic example to illustrate the application of Equation (1).

Table 2.2 – Example application of the WoE metric.

Variable 1	**Number of Samples**		**Sample Ratio**		**ln**		**WoE**
	Bad	**Good**	**Bad**	**Good**	**Bad**	**Good**	
Category A1	1	7	0.125	0.350	-2.079	-1.050	-1.030
Category B1	4	8	0.500	0.400	-0.693	-0.916	0.223
Category C1	3	5	0.375	0.250	-0.981	-1.386	0.405
Total Samples	8	20					

Equation (2) [11] illustrates the calculation of the IV metric, where *i* corresponds to the categories, *T* to the total value of the categories, and *n* to the number of categories of a variable.

$$IV = \sum_{i=1}^{n} (\%\, bad\ sample_i - \%\, good\ sample_i) * WOE_i$$

or (2)

$$IV = \sum_{i=1}^{n} \left(\frac{bad\ sample_i}{bad\ sample_T} - \frac{good\ sample_i}{good\ sample_T} \right) * WOE_i$$

The IV will only present positive values, since if the difference between the ratio of bad and good samples is positive, it means that the ratio of bad samples is greater, hence the WoE is also positive. Conversely, if the ratio between bad and good samples is negative, it means that the ratio of good samples is greater, so the WoE will also be negative. The Equation involves summation because the IV evaluates the predictive power of the variable as a whole and not individual categories. Table 2.3 shows a didactic example to illustrate the application of Equation (2).

Table 2.3 – Example application of the IV metric.

Variable 1	**Bad Samples - Good Samples**	**WoE**	**IV**
Category A1	-0.225	-1.030	0.232
Category B1	0.100	0.223	0.022
Category C1	0.125	0.405	0.051
IV Variable 1			**0.305**
Variable 2	**Bad Samples - Good Samples**	**WoE**	**IV**
Category A2	-0.150	-0.470	0.071
Category B2	0.150	0.223	0.033
IV Variable 2			**0.104**

The highest IV will be from the variable where the sum of the product of WoE by the difference between event and non-event for each category is greater, regardless of desired or undesired events. In general, the higher the IV, the greater the degree of contribution of the variable. In the case of IV ≥ 0.50, the authenticity of the variable should be verified. After analysis, considering it authentic, the variable has strong predictive power. The description of the resulting IV values is presented in Table 2.4 [11].

Table 2.4 – Description of IV value [11].

IV	Description
$IV < 0.02$	Useless variable
$0.02 \le IV < 0.10$	Weak value variable
$0.10 \le IV < 0.30$	Medium value variable
$0.30 \le IV < 0.50$	Strong value variable
$IV \ge 0.50$	Value is too high, consider the authenticity of the variable

In the example shown in Table 2.3, variable 1 has a strong contribution level, while variable 2 has a medium contribution level.

3. SCOPING REVIEW

This chapter presents a scoping review of research on the methodologies and techniques used for the maintenance and condition assessment of power transformers, which are the main asset in the electrical power transmission sector. It addresses articles on asset management, monitoring and diagnostics, oil analysis, and insulation moisture, with these articles originating from twenty-five countries and being published in journals in the last fifteen years, with more than half of them published in the last five years.

The aim of this research is to map the literature linked to the topic in a broader and more exploratory manner and to identify any existing gaps in knowledge. The methodology used in the research, a synthesis of the articles researched, characteristics of the sources of evidence and conclusions and limitations of the study are presented.

3.1. SCOPING REVIEW METHODOLOGY

This scoping review was prepared following the guidelines provided by the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) protocol [13].

To identify potentially relevant articles, the bibliographic databases of the publishers Elsevier, IEEE, and Multidisciplinary Digital Publishing Institute (MDPI) were searched from October 2023 to June 2024. The complete electronic search strategies used in the Elsevier, IEEE and MDPI databases are presented below in such a way that it allows the reproduction of the searches carried out.

The electronic search strategy for the Elsevier database used keywords, subject areas, and article type as selection filters. The keywords used were power transformers, asset management, condition based maintenance, failure analysis, moisture, and oil samples; the selected subject areas were energy, engineering, computer science, and materials science; and the chosen article types were review articles and research articles.

The electronic search strategy for the IEEE database used keywords, article type, and publication topics as selection filters. The keywords used were power transformers and asset management and condition based maintenance and failure analysis and moisture or oil samples; were selected articles from journals and magazines classified as Qualis CAPES Engineering IV A1 and A2, and selected the following publication topics: moisture, condition monitoring, and fault diagnosis.

The electronic search strategy for the MDPI database used keywords, subjects, journals, and article type as selection filters. The keywords used were the same as the IEEE search, with the inclusion of the key word Condition monitoring, the subjects were engineering, computer science and mathematics, the journals selected were Energies and Sensors (considering the Qualis CAPES Engineering IV A1 and A2 classification) and the chosen article types were review articles and research articles.

3.1.1. Eligibility criteria

To be included in this review, papers needed to address the topics of asset management and condition assessment of power transformers. To select articles, the following keywords were used: power transformer, asset management, condition based maintenance or condition monitoring, and failure analysis.

Considering that dielectric failure is the main failure mode in transmission substation transformers in Brazil and worldwide [4], articles specifically addressing oil analysis and insulation moisture were also examined, as insulating oil has dielectric and cooling functions in power transformers. Physicochemical and gas chromatography tests are carried out on the equipment's insulating oil to confirm that operating conditions are adequate and to analyze dissolved gases, respectively. Water content is one of the properties evaluated in physical–chemical tests. The keywords oil analysis, oil samples, and moisture were included.

Peer-reviewed journal papers were included if they were published in the period of 2009–2024 and in journals classified as Qualis CAPES Engineering IV A1 and A2, with the aim of mapping recent publications subjected to rigorous reviews.

Another eligibility criterion was the article type: review articles and research articles were included, with the aim of including articles that provide a solid foundation on the topic and insights and guidance for further investigation.

3.1.2. Selection of sources of evidence

After identifying articles by conducting a search by keywords and subject, articles were selected using the eligibility criteria. In the next stage of selecting sources of evidence, the topics, titles, and abstracts were evaluated. Finally, an evaluation was carried out on the full text of all publications identified in searches for potentially relevant publications. Figure 3.1 shows a flow diagram that illustrates the steps taken to select sources of evidence and the number of articles selected per step.

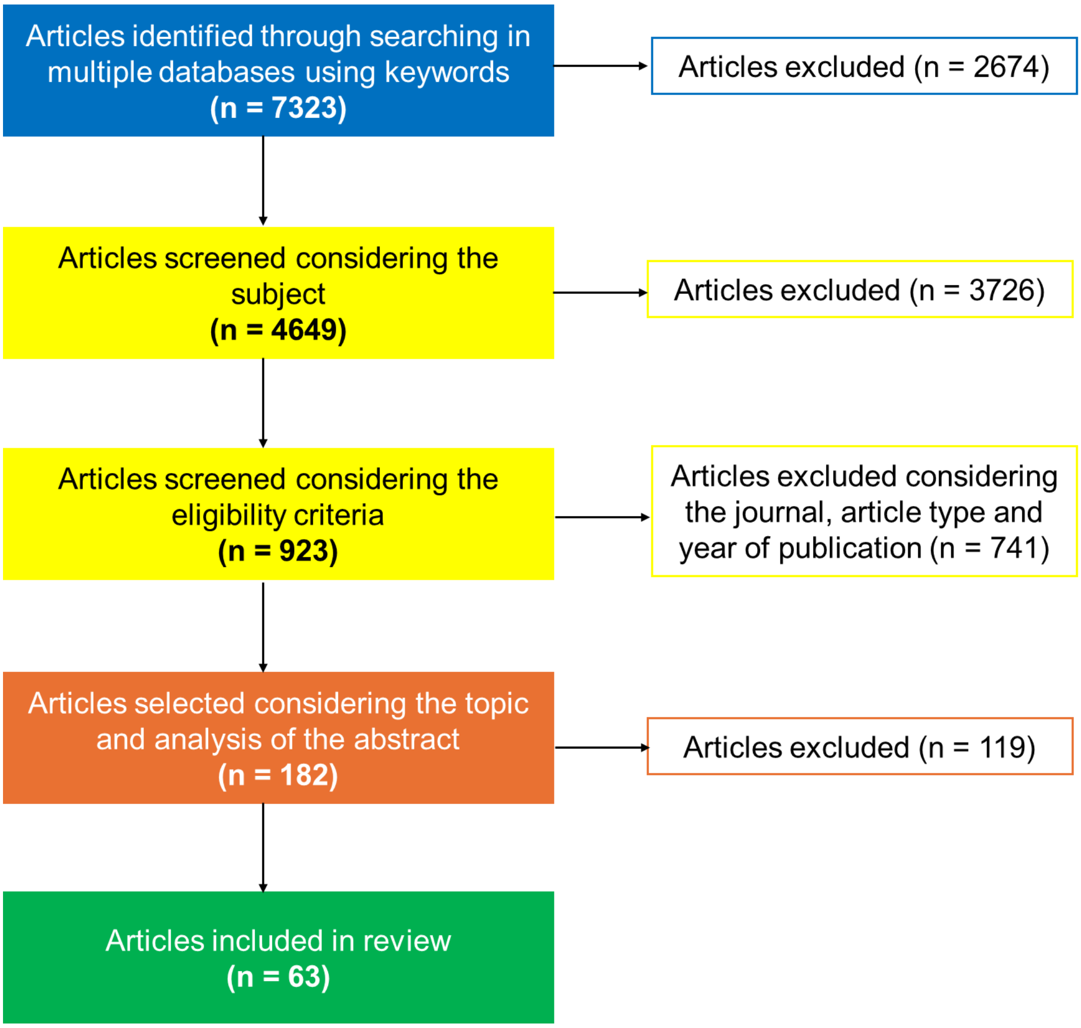


Figure 3.1 – Flow diagram illustrating the selection of sources of evidence [5].

3.1.3. Data Charting

The 63 articles included in this review are related to transformer asset management and condition assessment. After reading and analysis, it was deemed possible to classify them into subtopics, as some articles address the topic more broadly, while others specifically address diagnostic techniques and/or failure analyses. Table 3.1 shows the subtopic classification of the articles, considering the specificity of each work, and Figure 3.2 represents the hierarchy between the subtopics.

Table 3.1 – Classification of articles among the subtopics [5].

Subtopics	Number of articles
Asset management and condition assessment methodologies	10
Monitoring and diagnostic techniques and methods	22
Oil analysis	16
Moisture in insulation	15

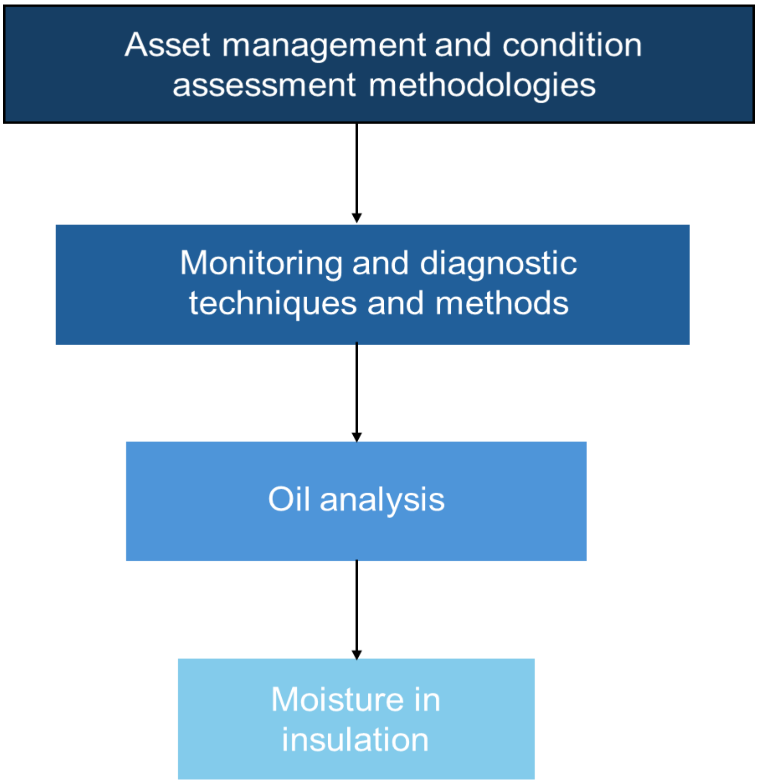


Figure 3.2 – Hierarchical relationship between the subtopics of the articles [5].

Articles related to transformer design, engineering optimization, and new technologies linked to manufacturing were not considered. Regarding articles related to sustainability it was considered only papers focusing on analysis that seeks better efficiency in the performance of power transformers articles focusing on specific environmental characteristics were not selected.

3.2. Synthesis

In the following sections, a synthesis of the articles is presented, focusing on their methodology, evaluated data, and contributions. At the end of each section, there is a table that summarizes the main subjects covered in each article, particularly the monitoring techniques, data analysis methods, and condition assessment methodologies and models.

3.2.1. Asset management and condition assessment methodologies

Abu-Elanien and Salama [14] and Velasquez-Contreras, Sanz-Bobi, and Arellano [15] addressed about the monitoring techniques Dissolved Gas Analysis (DGA) and Frequency Response Analysis (FRA).

Abu-Elanien and Salama also evaluate thermal analysis, vibration analysis, Partial Discharge (PD), and Recovery Voltage Measurement (RVM). Regarding the criteria for assessing the end-of-life physical condition of the equipment, the authors highlighted the degree of polymerization, maintained tensile strength, furanic compounds and addressed economic aging models considering linear depreciation and accelerated depreciation [14].

Velasquez-Contreras, Sanz-Bobi, and Arellano developed the General Asset Management Model for an Electric Utility, an integrated approach for managing power transformer assets within an electric utility company environment. The authors introduced a novel method to characterize the deterioration process of power transformers using the Multi-State Condition (MSC) Model and a methodology based on Hidden Markov Chains in order to estimate the failure rate using DGA tests. Anomaly detection modeling is carried out using oil temperature data and neural networks, and Decision Trees are employed as classifiers to evaluate FRA measurements of the transformers. An approach for maintenance scheduling was proposed using asset prioritization diagrams to support decision-making [15].

Soni and Mehta [16] presented a comprehensive review of the methods used for assessing the condition of power transformers, highlighting useful practices for evaluating the Health Index (HI), extending the lifecycle, and predicting failures. The following methods are addressed by the authors: DGA, FRA, RVM, PD testing, thermography testing, transformer turns ratio test, dielectric dissipation factor, winding resistance, core ground resistance, and insulation resistance calculation [16].

Murugan and Ramasamy [17], Koziel et al. [18], Gorginpour, Ghimatgar, and Toulab [19], Balanta et al. [20], and Biçen and Aras [21] proposed condition assessment methodologies of power transformers. Koziel et al. proposed a data quality management framework and emphasized the need to consider and quantify data quality for efficient asset management decisions [18] as Biçen and Aras introduced an intelligent asset management system for power transformers that enables a simultaneous and holistic evaluation of multiple input parameters for more accurate diagnostics, and it was based on a Failure-Sensitive Matrix (FSM) [21].

Murugan and Ramasamy developed statistical analysis of failures based on voltage level, geographic zone, and transformer components [17] while Gorginpour, Ghimatgar, and Toulab were based on the degree of polymerization and developed a Machine Learning (ML) algorithm with temperature, humidity, and daily scheduling data [19] and the method developed by Balanta et al. consider the insulation system degradation, risk index, Consequence Factor (CF), and economic impact [20]. The model proposed by Murugan and Ramasamy was based on a failure analysis of power transformers from two electric utility companies in Tamil Nadu, India, considering 196 failure cases from 2009 to 2013 [17] while the method developed by Gorginpour, Ghimatgar, and Toulab was validated using data from damaged transformers in Bushehr, Iran, with a predicted useful life accuracy error of less than 10% [19].

Jin et al. [22] and Jin, Kim, and Abu-Siada [23] addressed about DGA, PD, and vibration analysis. The authors also addressed the temperature measurement [22] and presented about the methods based on temperature and paper degradation value, and insulation system degradation [23]. The article underscores the importance of understanding the characteristics of each technique to design effective condition monitoring systems and anticipates the development of more comprehensive systems that not only report on the transformer's condition but also provide guidance for asset management and estimates of remaining useful life [22]. Regarding reliability assessments, the authors highlighted the use of the HI and the combination of inspection and monitoring data as an important tool [23].

Table 3.2 shows the main topics covered and the respective references.

Table 3.2 – Main subjects covered in each article on the subtopic of asset management and condition assessment methodologies [5].

Reference Index	**Monitoring techniques**	**Condition assessment methodologies/models**
[14]	Thermal analysis, vibration analysis, PD, DGA, RVM, and FRA.	Degree of polymerization, maintained tensile strength, and furanic compounds.
[15]	DGA, oil temperature measurement and FRA.	MSC model, Hidden Markov Chains, neural networks, and Decision Trees.

Reference Index	Monitoring techniques	Condition assessment methodologies/models
[16]	DGA, FRA, RVM, PD, thermography testing, transformer turns ratio test, dielectric dissipation factor, winding resistance, core ground resistance, and insulation resistance calculation.	
[17]		Statistical analysis of failures based on voltage level, geographic zone, and transformer components.
[18]		Data quality management framework.
[19]		Degree of polymerization and ML algorithm based on temperature, humidity, and daily scheduling.
[20]		Insulation system degradation, risk index, consequence factor, and economic impact.
[21]		FSM
[22]	DGA, PD, vibration analysis, and temperature measurement.	
[23]	DGA, PD, and vibration analysis.	Methods based on temperature and paper degradation value, and insulation system degradation.

3.2.2. Monitoring and diagnostic techniques and methods

Soni and Mehta [24] proposed an approach to identify the health indices of power transformers from the following tests: dielectric strength, acidity, breakdown voltage, DGA, furan compounds, the dielectric dissipation factor, the presence of moisture, interfacial tension, coil direct current resistance, and tangent delta. The authors utilize three mathematical analyses: Piecewise Linear Equations, the Analytic Hierarchy Process for weighting factors, and a residual analysis for curve fitting. Data from 100 transformers (20 healthy, 60 partially damaged, and 20 with high probability of failure) were analyzed [24].

Wong et al. [25] proposed the use of computational intelligence models through the analysis of techniques DGA, FRA, PD, Polarization and Depolarization Current (PDC) measurement, RVM, Infrared Thermography (IRT), and a Furan Analysis (FA) [25].

Ma, Saha, and Ekanayake [26], Guillen et al. [27], Didouche et al. [28], Sekatane and Bokoro [29], Beura, Wolters and Tenbohlen [30], and Liu et al. [31] addressed about PD measurement and analysis methods.

Ma, Saha, and Ekanayake addressed statistical learning techniques and highlighted the development of an algorithm based on a Support Vector Machine (SVM). To acquire knowledge about the statistical dependence between historical data and equipment conditions, data from PD in components collected from transformers were used [26]. Guillen et al. developed an algorithm to PD location in transformer windings. The authors used Discrete Wavelets Transform (DWT) and the Kullback–Leibler (KL) divergence [27]. Didouche et al. proposed a numerical localization method based on the Newton-Raphson iterative method which use mainly the finite elements method. This proposal reduces the number of sensors required for PD measurement [28].

Sekatane and Bokoro investigated the use of time reversal and the integration with Convolution Neural Networks (CNNs) for diagnosing PD in power transformers [29] and Beura, Wolters and Tenbohlen use Dijkstra's algorithm with additional line-of-sight propagation algorithms to determine the paths of the electromagnetic waves generated by PD sources. According to the authors, this algorithm can circumvent the time-consuming and computationally intensive process of simulating or collecting experimental data for an Artificial Intelligence (AI) based system of partial discharges in power transformers [30]. Liu et al. developed an ultrasonic detection system based on Fabry–Perot optical fiber sensor. The system employs a directional cross-localization method built on the Multiple Signal Classification (MUSIC) algorithm to precisely pinpoint dual PD sources into the transformer interior [31].

Senobari, Sadeha, and Borsi [32] and Seifi et al. [33] discussed about FRA as a method for assessing and detecting faults in transformers. Senobari, Sadeha, and Borsi realized a literature review regarding the topic and highlighted points such as the development of methods independent of the transformer structure and mathematical and statistical understanding of FRA comparison indices [32]. And Seifi et al. proposed the Sweep Reflection Coefficient (SRC) method, using FRA, for the early detection of inter-turn faults and mechanical faults in the winding. The proposal was validated in distribution transformers, with localization errors of less than 5% for ground resistance faults and less than 15% for other types of faults [33].

Velásquez and Lara [34], Medina et al. [35], and Costa, Silva, and Branco [36] analyzed condition and diagnostic data using neural networks, Fuzzy logic, and linear regressions, respectively. In the implementation of the neural network, three layers and fourteen neurons in the hidden layer were used, and a backpropagation algorithm was employed to predict the lifetime parameter of reactors. The proposed methodology was validated on power reactors from the Red de Energía del Perú through statistical tests, with a confidence level of 98.5% [34].

Medina et al. developed a Fuzzy Inference System (FIS) based approach to estimate the risk index of power transformers. Three FISs were used and tested on a fleet of 15 transformers: the HI, the CF, and the combination of both systems. The HI was calculated based on physical–chemical tests, considering the degree of polymerization. The failure CF was determined considering technical and operational parameters: load, oil volume, proximity to buildings, and penalties for asset unavailability. The third FIS combined the HI and CF to estimate the transformer's risk index [35].

Costa, Silva, and Branco proposed a model using linear regressions and temperature parameters based on environmental and historical data, along with gas concentration information [36].

Da Silva et al. [1] presented the development of a methodology to estimate the health of power transformers through a new diagnostic factor based on a historical average daily load curve and the identification of insulation conditions. And Zhou et al. [37] proposed a fault diagnosis model for power transformers based on a Probabilistic Neural Network (PNN) and optimized using an improved Gray Wolf Optimizer (GWO) algorithm. The methodology proposed by Da Silva et al. was applied and validated using data oil sample and electrical quantities from 204 power transformers in operation and its advantages include reducing the time needed to obtain results, allowing for the monitoring of solid transformer insulation degradation in 24-hour cycles [1]. The methodology proposed by Zhou et al. used 555 real fault samples from the Jiangxi Power Supply Company [37].

Huang et al. [38] proposed an unsupervised clustering method Entropy-Weighted (EW) K-means and the classical two-parameter Weibull model to assess the average failures of different groups of transformers. The methodology was validated in a distribution transformer park of the Chongqing Electric Power Company in China [38]. Islam et al. [39] presented the implementation of an intelligent framework based on a Multilayer Perceptron (MLP) generative model and a logistic regression classifier with the aim of assessing the condition of power transformers. This methodology was valid with data from 608 real transformers [39].

Wang et al. [40] proposed a new integrated multi-level and multi-parameter diagnostic method based on the integration of fault information. The method was validated using data from 628 fault samples from transformers across China [40]. Building upon the root cause analyses, in 2019 [41], the authors developed a systematic approach including traditional methods, Principal Component Analysis (PCA), data mining, and causal mapping. A total of 279 reactors in high-voltage systems in Latin America were analyzed, with the forecast of an additional 134 units in the coming years, and the methodology was validated in a case study on a 500 kV reactor [41].

Murugan and Ramasamy [42] proposed a practical approach based on statistical analysis to transformer maintenance using failure data and the HI, aiming to prevent failures and reduce the maintenance costs of this equipment. The article analyses 343 failures in distribution and power transformer components, with voltage classes ranging from 33 kV to 400 kV, over 11 years at the Tamil Nadu Electric Utility in India. A case study involving the calculation of the HI for seven power transformers was presented to validate the methodology [42].

Ariannik, Razi-Kazemi, and Lehtonen [43] presented the development of a lifespan estimation model for distribution transformers based on the degree of polymerization, considering the following variables: ambient temperature, load factor, and moisture content of the paper insulation. The calculation of the hot-spot temperature and a dynamic analysis of the polymerization degree profile were performed with variations in the operational conditions, and the optimal time to implement reductions in the paper insulation moisture content was suggested [43].

And Liu et al. [44] proposed an algorithm for monitoring abnormal conditions in distribution transformers using the Spearman correlation coefficient and the t-Statistics test. The authors developed a data acquisition system and utilized phase current data from transformers. The algorithm was validated based on nine cases of real data from the power distribution system of Zhejiang, China [44].

Table 3.3 shows the main topics covered and the respective references.

Table 3.3 – Main subjects covered in each article on the subtopic of monitoring and diagnostic techniques and methods [5].

Reference Index	**Monitoring techniques**	**Condition assessment methodologies/models**
[1]	An analysis data from oil samples and electrical quantities.	Models based on the historical average daily load curve.

Reference Index	Monitoring techniques	Condition assessment methodologies/models
[24]	The dielectric strength, acidity, breakdown voltage, DGA, furan compounds, dielectric dissipation factor, presence of moisture, interfacial tension, coil direct current resistance, and tangent delta.	Multi-Criteria Decision Making and Piecewise Linear Equations, Analytic Hierarchy Process for weighting factors, and residual analysis.
[25]	DGA, FRA, PD, PDC, RVM, IRT, and FA.	
[26]		SVM based ML algorithm using PD data.
[27]	PD.	DWT and KL.
[28]	PD.	The Newton-Raphson iterative method based on finite elements method.
[29]	PD.	The time reversal and integration with CNNs.
[30]	PD.	Dijkstra's algorithm.
[31]	PD.	Sensor: Fabry–Perot optical fiber. Algortimo: directional cross-localization based on the MUSIC.
[32]	FRA.	
[33]	FRA.	SRC.
[34]		Neural networks.
[35]	Physical–chemical tests.	Models: Fuzzy logic. Data: degree of polymerization, load, oil volume, proximity to buildings, and penalties for asset unavailability
[36]		Models: linear regressions. Data: temperature parameters and gas concentration information.
[37]		PNN and GWO.
[38]		Unsupervised clustering method of EW-Kmeans and the classical two-parameter Weibull model. Data: degree of polymerization.
[39]		MLP generative model and logistic regression classifier.
[40]	A comparative evaluation of the reliability of seven fault diagnostic methods in transformers, with emphasis on DGA.	Integrated multi-level and multi-parameter diagnostic method based on the integration of fault information.
[41]	An analysis based on data on mechanical, electrical, physical, and chemical dimensions.	PCA, data mining, and causal mapping.
[42]		A statistical analysis of failure data and identification of failure causes through maintenance records, direct observations in the repair yard, and consultations with experts.

Reference Index	Monitoring techniques	Condition assessment methodologies/models
[43]		The degree of polymerization, considering the variables ambient temperature, load factor, and moisture content of the paper insulation.
[44]		Spearman correlation coefficient and the t-Statistics test.

3.2.3. Oil analysis

The developments proposed by Chen et al. [45], Abu-Siada and Islam [46], Abu-Siada, Hmood, and Islam [47], Cui, Ma, and Saha [48], Tra, Duong, and Kim [49], Rajesh et al. [50], Velásquez and Lara [51], Aizpurua et al. [52], Bustamante et al. [53], Ward et al. [54], Menezes et al. [55], Soni and Mehta [56], [57], and Malik, Sharma, and Naayagi [58] were based on DGA data.

In addition to DGA, Velásquez and Lara used data from corrosive sulfur, and physical–chemical and electrical tests [51] while Aizpurua et al. evaluated oil quality, and solid insulation [52] and Ward et al. also addressed PD data [54]. Bustamante et al. discussed about DGA continuous online [53]. Soni and Mehta proposed methods using moisture content, furan compounds, and interfacial tension [56] and thermal stability, acidity, water content, breakdown voltage, and viscosity [57].

Vrsaljko, Haramija, and Hadzi-Skerlev [59] did not use DGA data, the authors proposed the use of High-Performance Liquid Chromatography (HPLC) to determine the contents of phenol, m-cresol, and o-cresol in transformer oil. They confirmed that phenol, m-cresol, and o-cresol were not present in new oils, indicating their presence as a result of insulation material degradation. Therefore, the detection of these compounds serves as an additional diagnostic tool to assess the normal or abnormal condition of transformer insulation, and elevated concentrations suggest equipment overheating [59].

Fuzzy logic was used by the authors in data analysis algorithms [47], [56], [57], [58]. Soni and Mehta proposed a combined approach using a Fuzzy Logic Controller (FLC) and Fuzzy C-Means (FCM) [56] and, in addition to Fuzzy Logic, clustering, and conditional probability [57].

The following methods have also been covered by the authors: Wavelet [44]; Gene Expression Programming (GEP) [46]; MLP Neural Networks [47]; Bayesian Particle Filtering [52]; ML classifiers: the J48 Decision Tree and random forest transformation with k-means, correlation based feature selection, and PCA [60], and Computational Intelligence based Decision Trees [52].

Cui, Ma, and Saha [48] and Tra, Duong, and Kim [49] proposed over-sampling methods with the aim of improving the diagnosis of the power transformer condition. The authors worked with oil analysis data, and the proposals were motivated by the fact that the nature of the data was not favorable for generalizing AI models. Both works used the Synthetic Minority Over-Sampling Technique (SMOTE), referred to by Cui, Ma, and Saha [48] as SMOTEBoost and by Tra, Duong, and Kim [49] as Adaptive SMOTE (ASMOTE), in their adaptive approach. To validate the effectiveness of the methods, SVM and k-Nearest Neighbor (kNN) classification algorithms were used by both authors.

Cui, Ma, and Saha [48] also used Decision Trees and radial basis function networks, while Tra, Duong, and Kim [49] also used the MLP. In 2023, Rajesh et al. [50] proposed the use of the Adaptive Synthetic Sampling Approach (ADASYN) technique, also a variant of the SMOTE. The authors used 4580 DGA samples from operational transformers to train the ML models.

The method proposed by Chen et al. [45] was validated in a total of 700 oil samples and the five proposed approaches out-perform the accuracy and efficiency of the conventional backpropagation neural network method [45]. The approach using GEP, proposed by Abu-Siada and Islam [46], combines Roger's ratio method, the IEEE method, and the CO/CO_2 ratio for interpreting the results and the critical classification of the transformer. Analyses are conducted on 338 oil samples collected from transformers with different classifications and lifespans [46], while the method proposed by Abu-Siada, Hmood, and Islam [47] was based on data from 2000 oil samples from different transformers, and the agreement of the method with real failures was tested on 70 samples with known failures [47].

Velásquez and Lara [51] analyzed failures in 61 transformers and a preliminary correlation study between sulfurinduced corrosion and PD activity was performed [51] and the proposal developed by Aizpurua et al. [52] was validated using data from a nuclear power plant [52].

The proposed approach by Soni and Mehta [56] was applied to data from 200 transformers. Different oil analysis techniques were analyzed: (a) Duval Triangle, (b) Gouda's Three-Ratio Method, (c) paper degradation based on the degree of polymerization and furans, and (d) insulation degradation based on moisture content and interfacial tension [56]. In the same year, the authors published [57] about the implementation of alternative fluids to replace mineral oil in transformers, citing environmental concerns, scarcity, the high costs of petroleum resources, and disposal issues.

Table 3.4 shows the main topics covered and the respective references.

Table 3.4 – Main subjects covered in each article on the subtopic of oil analysis [5].

Reference Index	Monitoring techniques	Condition assessment methodologies/models
[45]	DGA	Wavelet.
[46]	DGA	GEP.
[47]	DGA	Fuzzy logic.
[48]	DGA	SMOTE, SVM and kNN, and Decision Trees and radial basis function networks.
[49]	DGA	ASMOTE, SVM and kNN, and MLP.
[50]	DGA	ADASYN.
[51]	DGA, corrosive sulfur, and physical–chemical and electrical tests.	Data mining and MLP neural networks.
[52]	DGA, oil quality, and solid insulation.	Bayesian Particle Filtering.
[53]	DGA online contínuo.	
[54]	DGA and PD.	
[55]	DGA	Computational intelligence based Decision Trees.
[56]	DGA, moisture content, furan compounds, and interfacial tension	Degree of polymerization, FLC, and FCM.
[57]	Thermal stability, acidity, water content, DGA, dissolved gases, breakdown voltage, viscosity, and accelerated aging.	Fuzzy logic, clustering, and conditional probability.
[58]	DGA	Fuzzy logic, clustering, and conditional probability.
[59]	HPLC to determine phenol, m-cresol, and o-cresol contents in transformer oil.	
[60]		ML J48 Decision Tree and random forest, transformation with k-means, correlation based feature selection, and PCA.

3.2.4. Moisture in Insulation

Velásquez and Lara [61], Liu et al. [62], [63], Zukowski et al. [64], Hernandez and Ramirez [65], and Vatsa and Hati [66] addressed about the application of Frequency-Domain Spectroscopy (FDS) for the early detection of moisture in power transformer. The technique detects the degradation of the paper layer due to moisture.

Velásquez and Lara presented a study on the application of FDS in bushings.

The results of the technique and the conductivity associated with temperature changes in this layer is reflected in the characteristics of the dissipation factor and the capacitance of the bushing [61].

Liu et al. used the FDS with the algorithm-enhanced backpropagation neural network [62]. Liu et al. discussed conventional dielectric response measurement techniques in oil-immersed power transformers, such as RVM, FDS, and PDC. The authors proposed the use of Grey Relational Analysis (GRA) to assess insulation condition by analyzing various dielectric response data [63], while Zukowski et al. proposed an accurate way to determine the standard characteristics for the conductivity of paper impregnated with insulating oil and containing moisture. The authors used direct and alternating current conductivity frequencies obtained by FDS [64].

Hernandez and Ramirez proposed an approach based on Vector Fitting (VF), a rational approximation, to find positive real parameters for the Extended Debye Model (EDM) using FDS data [65] and Vatsa and Hati developed a deep learning based aging assessment technique combining FDS and squeeze-and-excitation-enabled CNN [66].

Singh, Sood, and Verma [67] and Arsad et al. [68] discussed the presence of moisture in the insulation by analyzing it with other physicochemical tests. Meanwhile, Przybylek [69] proposed an alternative measurement method to determine the moisture content in oil and Koch, Tenbohlen, and Stirl [70] developed an analysis relating moisture through water saturation while Medina et al. [71] discussed about the solid insulation degradation.

Singh, Sood, and Verma [67] presented a study on the influence of the age of transformers in operation on insulating oil, based on the following oil properties: moisture, dielectric strength, resistivity, the dissipation factor (tangent delta), interfacial tension, and flash point. They analyzed samples from 10 transformers in operation, with power ratings ranging from 16 to 20 MVA, manufactured between 1991 and 1997, installed in various substations in Punjab, India [67]. Medina et al. [71] used arithmetic Brownian Motion algorithms to estimate paper moisture. The method involved a holistic approach to assess the aging of solid insulation [71] while Przybylek [69] discussed methods for measuring moisture in liquid dielectrics, presenting the pros and cons of existing techniques and exploring the viability of Near-Infrared Spectroscopy (NIR) for this purpose [69].

Suleiman et al. [72], Liao et al. [73], Tu et al. [74], and Villarroela, Burgos, and García [75] discussed about alternative insulators and the relationship with moisture. Suleiman et al. [72] discussed the effect of moisture on the dielectric strength and physicochemical structure of biodegradable palm based insulating oils, used as alternatives to mineral oil. The Fourier Transform Infrared (FTIR) technique was employed to analyze the molecular structure changes in the oils at different moisture levels [72].

Liao et al. [73] investigated the effects of thermal aging on moisture equilibrium curves of a mineral oil–paper insulation system and a new paper–oil insulation system more resistant to aging [73] and Tu et al. [74] conducted an analysis of the moisture variation in oil and solid insulation during thermal aging for three types of papers [74].

Villarroela, Burgos, and García [75] presented an analysis of moisture dynamics in transformers insulated with natural esters and highlighted the growing use of these liquids as insulation in transformers, aligned with the search for greater sustainability and clean energy [75].

Table 3.5 shows the main topics covered and the respective references.

Table 3.5 – Main subjects covered in each article on the subtopic of moisture in insulation [5].

Reference Index	**Diagnosis of moisture level**	**Condition assessment**
[61]	The early detection of moisture in power transformer bushings.	Technique: FDS.
[62]	A prediction of moisture content.	Technique: FDS. Algorithms: Enhanced backpropagation neural network and ML.
[63]	A combination of techniques for evaluating dielectric response data.	Technique: RVM, FDS, and PDC. Analysis: GRA.
[64]	A way of determining the standard characteristics for the conductivity of paper impregnated with insulating oil and containing moisture.	Technique: FDS.
[65]	An approximation of moisture content.	Technique: FDS. Analysis: VF and EDM.
[66]	A technique to assess the aging status of paper.	Technique: FDS. Analysis: neural network.
[67]	Transformer age analysis and physicochemical tests on oil.	Oil properties analyzed: moisture, dielectric strength, resistivity, the dissipation factor (tangent delta), interfacial tension, and flash point.
[68]	Th methods for detecting and quantifying moisture in oil.	Oil properties analyzed: breakdown voltage and moisture. Analysis: accuracy, measurement time, and cost.

Reference Index	Diagnosis of moisture level	Condition assessment
[69]	An alternative method to the Karl Fischer Titration.	Technique: NIR.
[70]	The determination of moisture in oil and cellulose in power transformer.	Analysis: water saturation and representations of the isotherms
[71]	A method to estimate solid insulation degradation in power transformers.	Oil properties analyzed: temperature and moisture. Algorithm: Arithmetic Brownian Motion.
[72]	The effect of moisture on alternative insulating oils compared to mineral oil.	Technique: FTIR. Analysis: molecular structure changes in the oils at different moisture levels.
[73]	The effects of thermal aging on moisture equilibrium curves for mineral oil–paper insulation and proposal of a new paper–oil insulation system more resistant to aging.	Analysis: effects of moisture distribution based on the absorption capacity of oil and paper and the degree of polymerization.
[74]	The moisture variation in oil and solid insulation during thermal aging for three types of papers.	Oil properties analyzed: temperature and moisture. Analysis: saturated solubility and breakdown of molecular chains.
[75]	An analysis of moisture dynamics in transformers insulated with natural esters.	Oil properties analyzed: temperature and moisture

3.3. Characteristics and Summary of the Sources of Evidence

This section presents information about the characteristics and summary of the sources of evidence.

3.3.1. Characteristics of Each Source of Evidence

Table 3.6 shows the reference, subtopic, journal, year, and country of each source of evidence and the Figures 3.3 and 3.4 illustrate, respectively, the number of articles per year of publication and by country in which the work was developed and/or country of origin of the main author. More than half of the articles were published in the last 5 years, which confirms the topicality of the topic.

Table 3.6 – Information from each source of evidence [5].

Reference Index	Reference	Subtopic	Journal	Year	Country
[1]	Da Silva et al.	2.2.2	*Engineering Failure Analysis*	2021	Brazil
[14]	Abu-Elanien and Salama	2.2.1	*Electric Power Systems Research*	2010	Canada
[15]	Velasquez-Contreras, Sanz-Bobi, and Arellano	2.2.1	*Electric Power Systems Research*	2011	Spain

Reference Index	Reference	Subtopic	Journal	Year	Country
[16]	Soni and Mehta	2.2.1	*Engineering Failure Analysis*	2021	India
[17]	Murugan and Ramasamy	2.2.1	*Engineering Failure Analysis*	2015	India
[18]	Koziel et al.	2.2.1	*Applied Energy*	2021	Sweden
[19]	Gorginpour, Ghimatgar and Toulab	2.2.1	*IEEE Transactions on Dielectrics and Electrical Insulation*	2022	Iran
[20]	Balanta et al.	2.2.1	*Energies*	2023	Argentina
[21]	Biçen and Aras	2.2.1	*Engineering Failure Analysis*	2023	Australia
[22]	Jin et al.	2.2.1	*Energies*	2022	Australia
[23]	Jin, Kim, and Abu-Siada	2.2.1	*Engineering Failure Analysis*	2023	Turkey
[24]	Soni and Mehta	2.2.2	*Engineering Failure Analysis*	2022	India
[25]	Wong et al.	2.2.2	*Applied Soft Computing*	2022	Malaysia
[26]	Ma, Saha, and Ekanayake	2.2.2	*IEEE Transactions on Dielectrics and Electrical Insulation*	2012	Australia
[27]	Guillen et al.	2.2.2	*Electric Power Systems Research*	2016	Mexico
[28]	Didouche et al.	2.2.2	*Electric Power Systems Research*	2022	Algeria
[29]	Sekatane and Bokoro	2.2.2	*Energies*	2023	South Africa
[30]	Beura, Wolters and Tenbohlen	2.2.2	*Sensors*	2024	Germany
[31]	Liu et al.	2.2.2	*Energies*	2024	China
[32]	Senobari, Sadeha and Borsi	2.2.2	*Electric Power Systems Research*	2018	Iran
[33]	Seifi et al.	2.2.2	*International Journal of Electrical Power & Energy Systems*	2022	Germany
[34]	Velásquez and Lara	2.2.2	*Engineering Failure Analysis*	2018	Peru
[35]	Medina et al.	2.2.2	*Electric Power Systems Research*	2022	Argentina
[36]	Costa, Silva, and Branco	2.2.2	*Energies*	2022	Portugal
[37]	Zhou et al.	2.2.2	*Energies*	2021	China
[38]	Huang et al.	2.2.2	*Reliability Engineering & System Safety*	2023	China
[39]	Islam et al.	2.2.2	*Electric Power Systems Research*	2023	Bangladesh
[40]	Wang et al.	2.2.2	*IEEE Transactions on Dielectrics and Electrical Insulation*	2013	China
[41]	Velásquez and Lara	2.2.2	*Engineering Failure Analysis*	2019	Peru
[42]	Murugan and Ramasamy	2.2.2	*Engineering Failure Analysis*	2019	India

Reference Index	Reference	Subtopic	Journal	Year	Country
[43]	Ariannik, Razi-Kazemi, and Lehtonen	2.2.2	*Reliability Engineering & System Safety*	2020	Iran
[44]	Liu et al.	2.2.2	*IEEE Transactions on Power Delivery*	2019	China
[45]	Chen et al.	2.2.3	*IEEE Transactions on Power Delivery*	2009	China
[46]	Abu-Siada and Islam	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2012	Australia
[47]	Abu-Siada, Hmood, and Islam	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2013	Australia
[48]	Cui, Ma, and Saha	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2014	Australia
[49]	Tra, Duong, and Kim	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2019	South Korea
[50]	Rajesh et al.	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2023	India
[51]	Velásquez and Lara	2.2.3	*Engineering Failure Analysis*	2018	Peru
[52]	Aizpurua et al.	2.2.3	*Applied Soft Computing*	2019	Spain
[53]	Bustamante et al.	2.2.3	*Sensors*	2019	Spain
[54]	Ward et al.	2.2.3	*Sensors*	2021	Egypt
[55]	Menezes et al.	2.2.3	*IEEE Transactions on Power Delivery*	2022	Brazil
[56]	Soni and Mehta	2.2.3	*Electric Power Systems Research*	2023	India
[57]	Soni and Mehta	2.2.3	*Electric Power Systems Research*	2023	India
[58]	Malik, Sharma, and Naayagi	2.2.3	*IEEE Transactions on Dielectrics and Electrical Insulation*	2023	Singapore
[59]	Vrsaljko, Haramija, and Hadzi-Skerlev	2.2.3	*Electric Power Systems Research*	2012	Croatia
[60]	Senoussaoui, Brahami, and Fofana	2.2.3	*Energies*	2021	Algeria
[61]	Velásquez and Lara	2.2.4	*Engineering Failure Analysis*	2018	Peru
[62]	Liu et al.	2.2.4	*IEEE Transactions on Dielectrics and Electrical Insulation*	2022	China
[63]	Liu et al.	2.2.4	*Energies*	2017	China
[64]	Zukowski et al.	2.2.4	*Energies*	2021	Poland
[65]	Hernandez and Ramirez	2.2.4	*Energies*	2022	Mexico
[66]	Vatsa and Hati	2.2.4	*Engineering Applications of Artificial Intelligence*	2024	India

Reference Index	Reference	Subtopic	Journal	Year	Country
[67]	Singh, Sood, and Verma	2.2.4	*IEEE Transactions on Dielectrics and Electrical Insulation*	2012	India
[68]	Arsad et al.	2.2.4	*Energies*	2023	Malaysia
[69]	Przybylek	2.2.4	*Energies*	2022	Poland
[70]	Koch, Tenbohlen, and Stirl	2.2.4	*IEEE Transactions on Power Delivery*	2010	Germany
[71]	Medina et al.	2.2.4	*Electric Power Systems Research*	2017	Ecuador
[72]	Suleiman et al.	2.2.4	*IEEE Transactions on Dielectrics and Electrical Insulation*	2014	Malaysia
[73]	Liao et al.	2.2.4	*IEEE Transactions on Dielectrics and Electrical Insulation*	2015	China
[74]	Tu et al.	2.2.4	*IEEE Transactions on Dielectrics and Electrical Insulation*	2016	China
[75]	Villarroela, Burgos, and García	2.2.4	*International Journal of Electrical Power & Energy Systems*	2021	United Kingdom

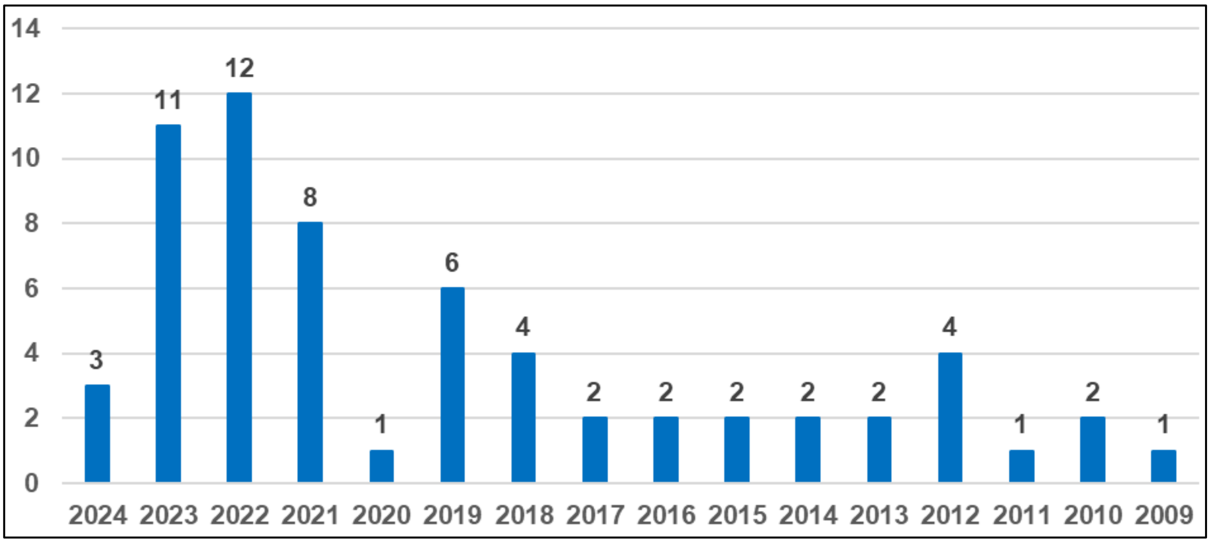


Figure 3.3 – Number of articles per year of publication.

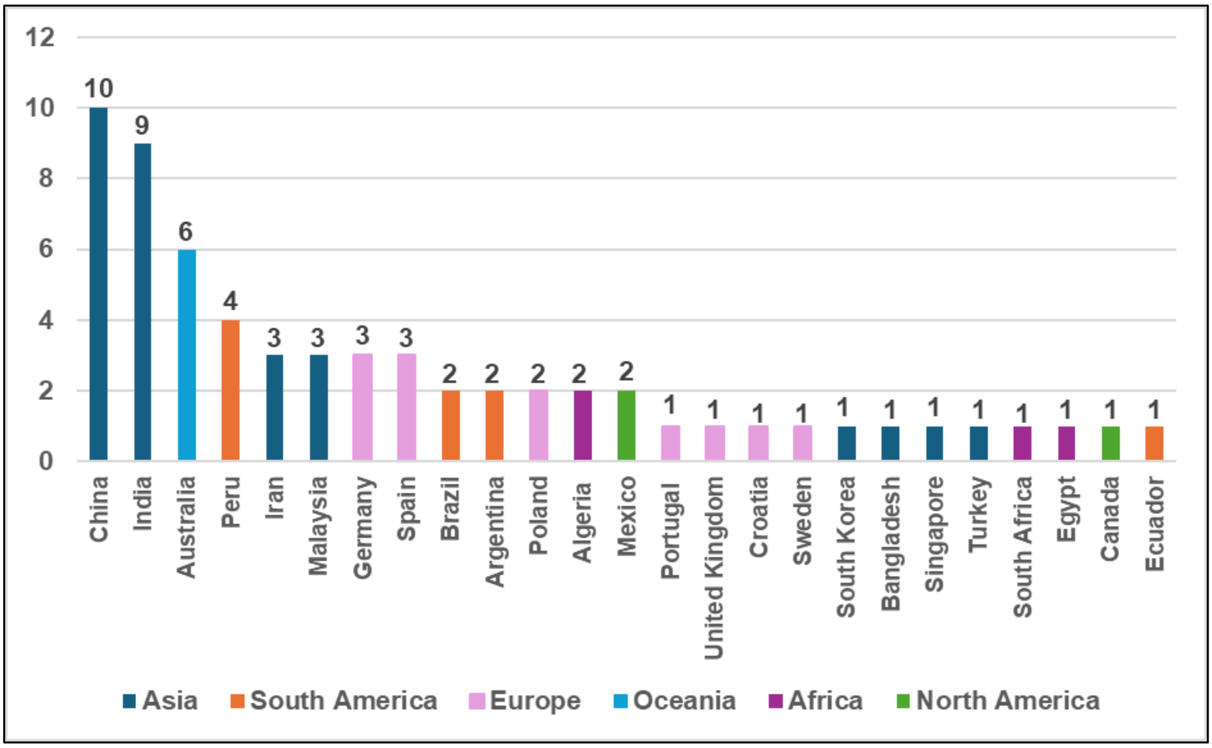


Figure 3.4 – Number of articles per country.

3.3.2. Summary of the sources of evidence

Articles [14], [15], [16], [17], [18], [19], [20], [21], [22], [23] addressed the importance of effective asset management for power transformers, focusing on monitoring techniques, condition assessments, and maintenance optimization. Models and methodologies were proposed to enhance reliability, reduce maintenance costs, and predict failures. Furthermore, there is a consensus on the importance of integrating diagnostic algorithms, monitoring techniques, and AI for fault prediction, with an emphasis on DGA and HI assessments.

Articles [1], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44] addressed a variety of methodologies and techniques for diagnosing faults and estimating the lifespan of power transformers. The authors highlighted the importance of DGA, PDs, statistical learning, online and offline diagnostic methods, lifespan models, and continuous monitoring techniques. Additionally, they emphasized the need for more accurate and efficient methods to ensure the reliability and operational safety of transformers, contributing to the optimization of maintenance strategies and the extension of the lifespan of this essential piece of equipment for power supply.

Articles [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60] discussed the techniques and methodologies used for diagnosing faults in power transformers, with an emphasis on insulating oil analysis and a particular focus on dissolved gases. The approaches included the use of statistical methods, AI techniques such as neural networks and Fuzzy logic, and physicochemical analyses. There is a search to find more effective and accurate methods, often combining multiple techniques to enhance result reliability.

Articles [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75] addressed the issue of moisture in power transformers, exploring different aspects related to its impact on the performance and integrity of insulation systems. The authors discussed methods for assessing moisture, the effects of thermal aging on insulating materials, the influence of moisture on the properties of insulating oils, and sustainable alternatives to traditional mineral oils. Additionally, there is an emphasis on the importance of the early detection and monitoring of moisture, highlighting the need for precise and efficient techniques to ensure the safe and reliable operation of transformers.

3.4. Discussions

This section presents the main limitations and conclusions of this scoping review.

3.4.1. Limitations

For this research, no specific systematic literature review tool was used. The search was carried out directly and only in the databases of Elsevier, IEEE, and MDPI. With the aim of mapping publications subjected to rigorous reviews, in this scoping review only journals classified as Qualis CAPES Engineering IV A1 and A2 were selected. Furthermore, a detailed description of each of the methods researched is not presented.

3.4.2. Conclusions

This scope review aimed to base the development of an original methodology to assist in asset management of power transformers. To this end, recent works published in twenty-five countries around the world were evaluated.

Asset management covers the entire life cycle of the equipment. The Operation and Maintenance stage, the largest part of the asset's useful life, includes Maintenance Management, which includes preventive, corrective and predictive maintenance. Figure 3.5 represents the hierarchical relationship between the asset management and the types of maintenance.

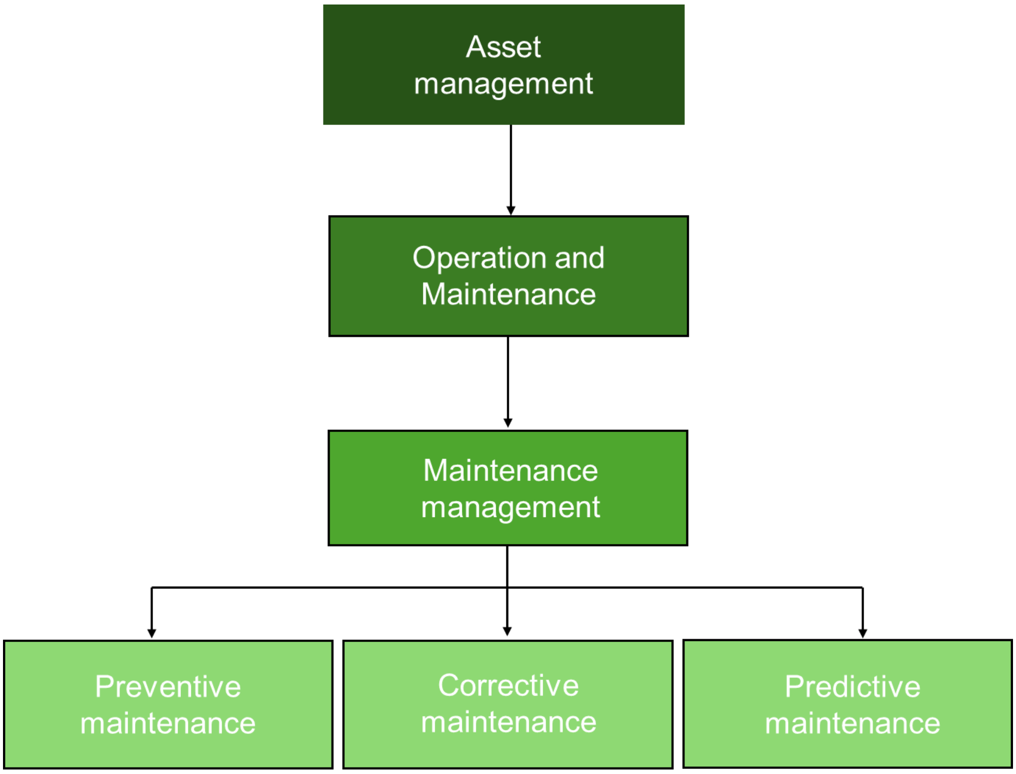


Figure 3.5 – Hierarchical relationship between the asset management and the types of maintenance [5].

One of the key research results is the finding that the data analysis methodologies related to identifying failures and aiding decision-making can add value to asset management, especially considering the aging of the transmission sector's infrastructure, the impracticality of replacing all depreciated assets from a regulatory perspective, the challenges posed by the system, including support for energy transition, and developments and advances in relation to smart grids.

Another important research result is the finding of the opportunity and importance of developing algorithms related to the evaluation of transformer insulating oil. This knowledge still largely depends on assessments by experts. The analysis of the physical–chemical results of the oil is an important line of research.

This research also found the importance of applying statistical tools prior to the application of AI algorithms. With the possibility of using diverse data from transformers, the prioritization of variables prior to applying predictive models can result in better performance.

As a future development trend, the application of data analysis tools and algorithms that contribute to a better evaluation of equipment stands out. Failure analysis, evaluation of maintenance data, equipment operating data and continuous online monitoring are important for the development of methodologies for diagnosis, prognosis of assets and estimation of their useful life. Figure 3.6 represents the hierarchy between these subjects.

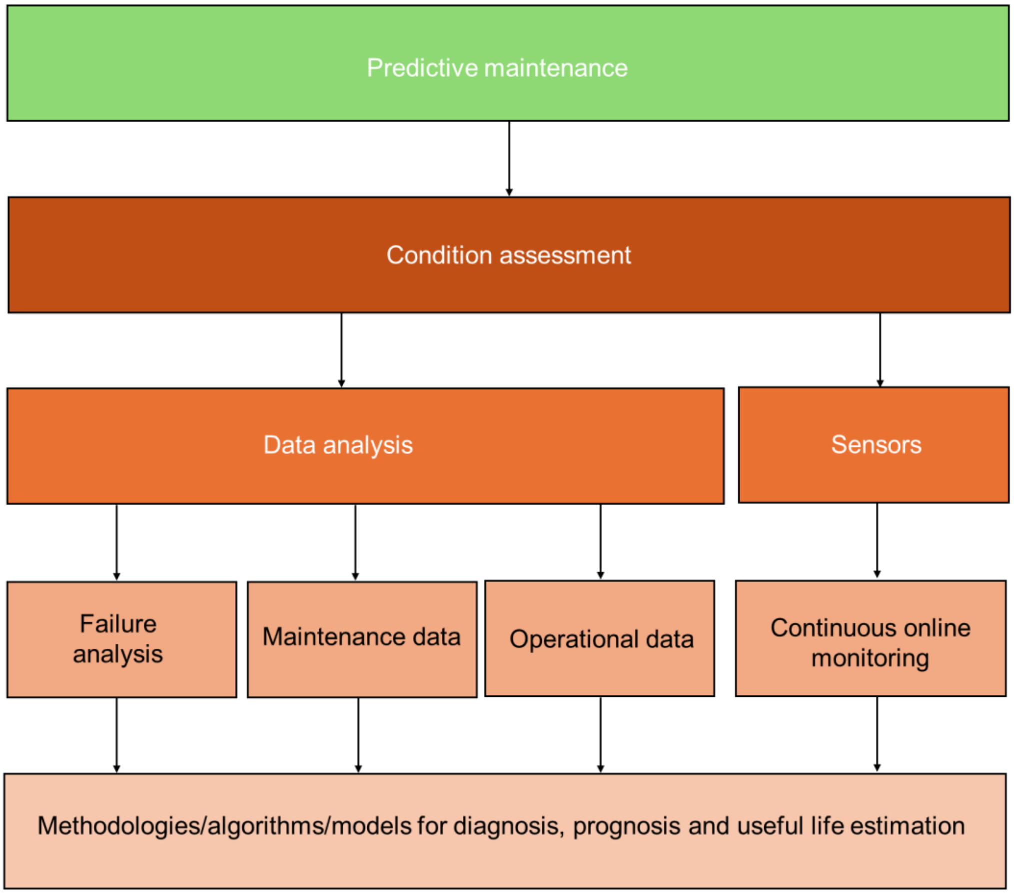


Figure 3.6 – Hierarchical relationship between the predictive maintenance and the data analysis methodologies [5].

4. METHODOLOGY

The proposed methodology aims to add value to the direction of maintenance investments in power transformers and reactors. It is a data analysis applied to mitigate the assumed risk given the scenario of aging infrastructure in the energy transmission sector and the impracticality of replacing all depreciated assets from a regulatory perspective.

The analyses presented in this thesis use the following technical information of the equipment as input data: voltage class, installation region (Regional), criticality, type, and age of transformers and reactors. The proposed methodology involves evaluating the predictive importance of each input categorical variable on the binary output – moisture below or above standardized values – in other words, how much each piece of information should be prioritized for an assessment of maintenance action direction.

Figure 4.1 illustrates the flowchart of the proposed analysis methodology, which includes the collection and evaluation of equipment data and dielectric related tests, descriptive analysis of the data, and application of statistical metrics to assess the predictive importance of the variables.

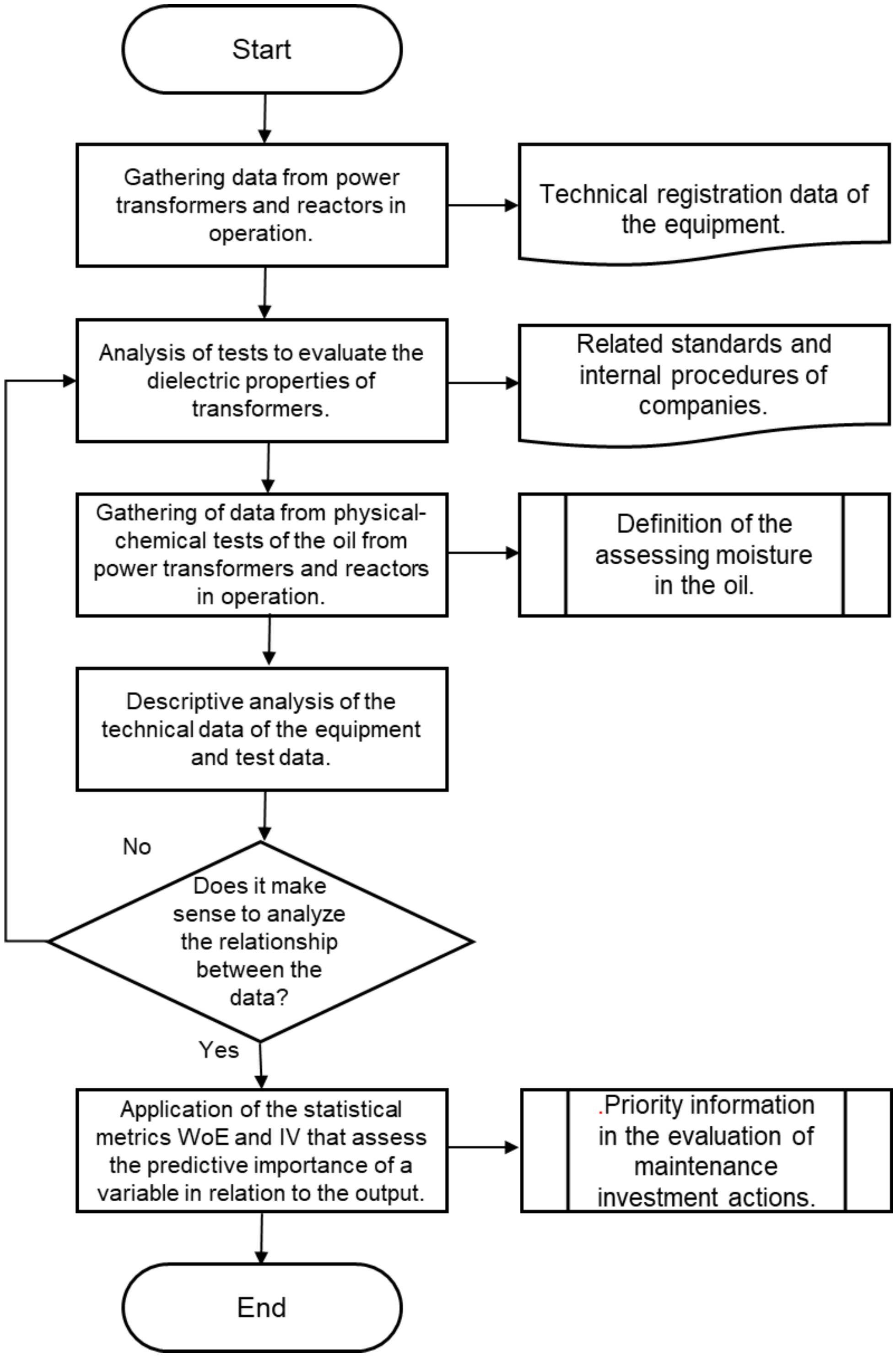


Figure 4.1 – Flowchart of the proposed methodology [6].

It is important to highlight that the information used as inputs for the proposed method consists of asset registration data, simple data that is available for all equipment, meaning no additional investment is required for the application of the methodology.

The following subsections describe the case study using such metrics in prioritizing maintenance actions in power transformers and the application of these metrics in different areas.

4.1. Case study of applying WoE and IV to assist in asset management of power transformers and reactors

This item presents the application of WoE and IV to assist in prioritizing maintenance actions in power transformers. Due to their similarity, reactors are also evaluated. As far as author’s knowledge, such metrics have not been applied for this purpose, which ensures the originality of the presented work.

The method was applied in the ISA CTEEP park, electrical energy transmission company in Brazil, responsible for approximately 95% of the energy transmitted in the state of São Paulo and about 30% of all energy in Brazil. Therefore, the variable values are according to the park in question. However, it should be noted that the characteristics are general registration for all power transformer and reactor.

4.1.1. Variables and their respective categories

The variables and their respective categories considered in the methodology will be presented below.

4.1.1.1. Voltage class

The Brazilian Electricity Regulatory Agency (*Agência Nacional de Energia Elétrica – ANEEL*) establishes criteria for the composition of the Basic Grid (*Rede Básica – RB*) and Other Transmission Installations (*Demais Instalações de Transmissão – DIT*), according to the voltage class. The *RB* includes transmission installations with voltage equal to or greater than 230 kV.

In this work, power transformers from the *RB* and *DITs* will be considered. As for the reactors in the ISA CTEEP park, they are installed in the *RB* due to the need for reactive compensation. Table 4.1 shows the voltage class ranges according to ABNT NBR 10576:2017 [8] and the respective voltage classes considered in each range in this work.

Table 4.1 – Ranges according to ABNT NBR 10576:2017 [8] and associated voltage classes [6].

Voltage class range (kV)	Voltage class (kV)
≤ 72.5	22 kV
	34.5 kV
	69 kV
> 72.5 and ≤ 145	88 kV
	138 kV
> 145	230 kV
	345 kV
	440 kV
	500 kV

4.1.1.2. INSTALLATION REGION (REGIONAL)

The installation region variable, referred to as Regional, is classified as follows: Bauru (TB), Cabreúva (TC), São Paulo (TS), Taubaté (TT), and National Expansion (TE). The Bauru, Cabreúva, São Paulo, and Taubaté Regionals belong to the state of São Paulo, while the installations of the National Expansion Regional represent ISA CTEEP assets outside the state of São Paulo. Each of the Regionals has specificities that influence maintenance actions, such as climate, served loads, and logistics for service. Figure 4.2 illustrates the areas covered by each of the Regionals.

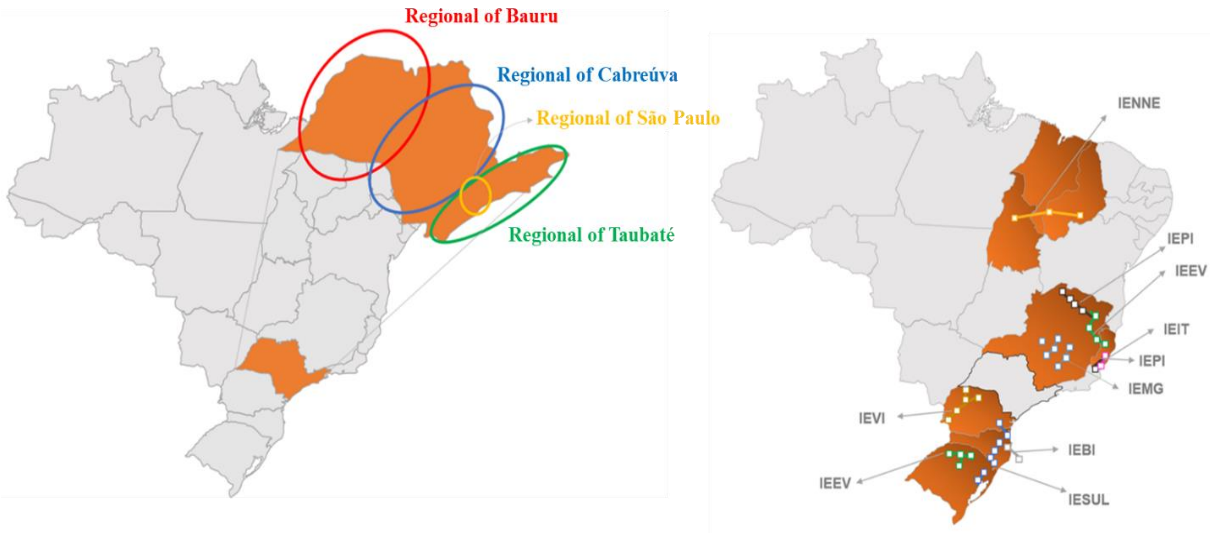


Figure 4.2 – Representation of the areas of the Regionals (a) Bauru, Cabreúva, São Paulo, Taubaté and (b) National Expansion [6].

4.1.1.3. CRITICALITY

According to the internal procedure of the concessionaire, the qualification of criticality takes into account five impacts: 1) Safety Impact; 2) Systemic Impact; 3) Environmental Impact; 4) Financial Impact; and 5) Image Impact. This involves an assessment of the potential consequences associated with the occurrence of a failure, considering these five aspects.

The Safety Impact measures the risk of exposure to individuals, taking into account the duration of people in the energized area and surroundings of the facilities, whether in preventive maintenance, corrective maintenance, outages, or failures. The level of exposure is measured through the frequency of attendance according to each locality. The risk of explosion is demonstrated on a scale according to the bay characterization.

The Systemic Impact considers: a) Busbars, bays, and circuit breakers of the General Module; b) Topology; c) Difficulty of release; d) Strategic installation; e) Annual planning of electrical operation. The Environmental Impact considers: a) Oil containment system; b) PCB content; c) SF6 gas volume.

The Financial Impact evaluates: a) Penalty cost; b) Replacement/repair cost; c) Attendance difficulty. And the Image Impact is valued through the perception factor by the distribution companies.

In accordance with the above considerations, a value from 0 to 5 is assigned to each of the five impacts.

Equation (3) represents the criticality model, with its respective weights, evaluated by bay, where *CQ* corresponds to the criticality qualification, I_{Sa} to the Safety Impact, I_S to the Systemic Impact, I_E to the Environmental Impact, I_F to the Financial Impact, and I_I to the Image Impact.

The model is specific to ISA CTEEP, however the criticality of assets is an assessment carried out by all energy transmission companies.

$$CQ = 0.31 * I_{Sa} + 0.21 * I_S + 0.20 * I_E + 0.14 * I_F + 0.14 * I_I \quad (3)$$

The final qualification is represented by A, B, and C, as shown in Table 4.2.

Table 4.2 – Conversion of the score into qualitative classification [6].

Criticality Qualification	Criticality	Description
≥ 3	A	High criticality

Criticality Qualification	Criticality	Description
< 3 e ≥ 2	B	Medium criticality
< 2	C	Low criticality

Table 4.3 shows, as an example, the calculation of *CQ* and their corresponding values of the parameters for three transformers.

Table 4.3 – Example of calculation of *CQ* and their corresponding values of the parameters [6].

Impacts	Weights	TR1 138 kV TB	TR2 88 kV TC	TR3 69 kV TE
I_{Sa}	0.31	5.00	4.25	3.50
I_S	0.21	3.12	3.61	0.39
I_E	0.20	2.67	1.00	1.00
I_F	0.14	1.40	3.50	2.00
I_I	0.14	2.00	0.00	2.00
CQ		**3.21**	**2.77**	**1.93**
Criticality		**A**	**B**	**C**

4.1.1.4. Type

The equipment will be classified by type as follows: power transformers, power autotransformers, and reactors, represented respectively by the abbreviations EQU-TRAFOP, EQU-AUTOTR and EQU-REACTO.

The windings of transformers with two or three windings are commonly referred to as primary, secondary, and tertiary windings. There are transformers that have only one winding, meaning the primary winding is connected to the secondary winding, so there is no insulation between them. These transformers are called autotransformers. In autotransformers, the primary and secondary windings are in contact. Each winding has at least three outputs, where electrical connections are made [76].

In view of the above, the Type variable takes into account structural differences regarding windings, considering the categories of transformers and autotransformers. In the case of reactors, the difference already involves the systemic function of the equipment. The methodology foresees the analysis of reactors based on their structural and testing similarity with transformers.

Throughout the text, except when referring to the Type variable, transformers and autotransformers are referred to as power transformers (or simply transformers).

4.1.1.5. AGE

According to the Asset Control Manual of the Electric Sector (*Manual de Controle Patrimonial do Setor Elétrico* – MCPSE) [77] developed by *ANEEL*, the regulatory service life, compatible with the depreciation period, of transformers is 35 years.

To consider this variable as categorical, four age ranges were considered: equipment up to 17 years old; between 17 and 35 years old; between 35 and 45 years old and from 45 years old.

The equipment's age at the time of the sample was taken into account, in other words, to calculate the equipment's age, the year the oil sample was analyzed is subtracted from the year the equipment was built.

4.1.1.6. MOISTURE IN THE OIL

As presented in item 3.2.2, moisture is an important variable for the evaluation of the oil in power transformers and reactors. The water content is one of the properties assessed in the physicochemical tests of the oil and is passable to corrective action.

The water samples presented in this study were measured in ppm and are considered as the output variable with binary valuation, that is, moisture below or above standardized values, according to the voltage class, as shown in Table 2.1.

4.1.1.7. SUMMARY OF THE VARIABLES AND CATEGORIES

The summary of the variables and categories considered in the analyses is shown in Table 4.4. Ranges were considered for the variables voltage class and age, and binary valuation was used for the output variable moisture in the oil.

Table 4.4 – Summary of the variables and categories considered in the analyses.

Variables	Categories
Voltage class range	≤ 72.5 kV
	> 72.5 kV and ≤ 145 kV
	> 145 kV
Installation region (Regional)	TT
	TS
	TE
	TC

Variables	Categories
	TB
Criticality	A
	B
	C
Type	EQU-TRAFOP
	EQU-AUTOTR
	EQU-REACTO
Age range	≤ 17 years
	> 17 years and < 35 years
	≥ 35 years and ≤ 45 years
	> 45 years
Moisture in the oil	≤ parameter
	> parameter

4.1.2. PERFORMANCE OF CATEGORIES AND PREDICTIVE IMPORTANCE OF VARIABLES

The proposed methodology was applied to a dataset consisting of 795 assets and nearly 10 thousand oil samples. Data processing and analysis were performed using Python, utilizing the Pandas and Matplotlib libraries. The Jupyter Lab interface was used for development. A data frame, named df, was created with the data. Figure 4.3 shows the information from the df, followed by the listing of the registration data of the assets and the data related to the oil tests.

```
df = pd.read_excel('C:/doctoral files/moisture_in_the_oil.xlsx')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9733 entries, 0 to 9732
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   id_equ             9733 non-null   int64
 1   description_equ    9733 non-null   object
 2   criticality_equ    9733 non-null   object
 3   voltage_class      9733 non-null   float64
 4   type_equ           9733 non-null   object
 5   regional           9733 non-null   object
 6   year_construction  9733 non-null   int64
 7   date_analysis      9733 non-null   datetime64[ns]
 8   water_content      9733 non-null   int64
 9   maximum_parameter  9733 non-null   int64
dtypes: datetime64[ns](1), float64(1), int64(4), object(4)
memory usage: 760.5+ KB
```

Figure 4.3 – Information from the df.

Registration data of the asset:

- id_equ: asset identification number;
- description_equ: asset description;
- criticality_equ: asset criticality;
- voltage_class: asset voltage class;
- type_equ: asset type;
- regional: asset installation region;
- year_construction: asset construction year.

Data related to the oil tests:

- date_analysis: date of oil analysis;
- water_content: water content in oil;
- maximum_parameter: maximum water content parameter, in accordance with ABNT NBR 10576:2017 [8].

Figure 4.4 shows the first and last lines of the df.

	id_equ	description_equ	criticality_equ	voltage_class	type_equ	regional	year_construction	date_analysis	water_content	maximum_parameter
0	142551	TR 440/345 kV - ATR-RES (1)	B	440.0	EQU-AUTOTR	TT	1983	2003-04-14	4	20
1	142551	TR 440/345 kV - ATR-RES (1)	B	440.0	EQU-AUTOTR	TT	1983	2004-04-19	4	20
2	142551	TR 440/345 kV - ATR-RES (1)	B	440.0	EQU-AUTOTR	TT	1983	2007-08-03	7	20
3	142551	TR 440/345 kV - ATR-RES (1)	B	440.0	EQU-AUTOTR	TT	1983	2009-04-25	7	20
4	142551	TR 440/345 kV - ATR-RES (1)	B	440.0	EQU-AUTOTR	TT	1983	2011-10-23	6	20
...	...	...	...	...	...	...	...	...	...	...
9728	501374	TR 440/88 kV - TR-3 - Phase A	C	440.0	EQU-TRAFOP	TC	2011	2021-06-12	7	20
9729	501374	TR 440/88 kV - TR-3 - Phase A	C	440.0	EQU-TRAFOP	TC	2011	2022-06-21	11	20
9730	501376	TR 440/88 kV - TR-3 - Phase V	C	440.0	EQU-TRAFOP	TC	2011	2015-08-18	5	20
9731	501376	TR 440/88 kV - TR-3 - Phase V	C	440.0	EQU-TRAFOP	TC	2011	2021-06-25	8	20
9732	501376	TR 440/88 kV - TR-3 - Phase V	C	440.0	EQU-TRAFOP	TC	2011	2022-06-21	10	20

9733 rows × 10 columns

Figure 4.4 – df with asset registration data and oil moisture test data.

Based on moisture values (water_content) and parameters (maximum_parameter), the status_moisture was created as a binary valuation for the moisture output. Additionally, the data frame df_above was created containing the samples above the standardized values.

Table 4.5 presents the sample and equipment values. Out of the total samples (9733), 8.26% (804 samples) are above the parameters and 91.74% (8929 samples) are below the parameters; 23.90% of the assets (190 assets) have at least one sample with moisture above the threshold, while 76.10% (605 assets) do not have any sample above the parameters.

Table 4.5 – Quantity of samples and equipment [6].

Data frame	Quantity of samples	Quantity of equipments
df	9733	795
df_above	804	190

No specific methodology was used to identify inaccuracies. The missing of data for certain equipment was noted, and these samples were excluded from the sample universe. With the exception of these samples discarded due to missing data, the entire database provided by the power transmission company was considered, based on the hypothesis of data quality, since these are accepted by the company, which has its processes standardized by standards.

From a statistical point of view, the samples should represent the dataset in a meaningful way.

The good and bad samples shown in Equations (1) and (2) are related to the oil moisture status in Table 4.6. The good sample is the one in which the moisture content value is below the established limit (Table 2.1) and the bad sample is the sample in which the moisture content value is above the established limit (Table 2.1). Event represents a bad sample, that is, event is a sample which the moisture content value is above the established limit, and non-event represents a good sample, that is, non-event is a sample which the moisture content value is below the established limit. The df_above is a data frame consisting of event data, in other words, data from bad samples.

Table 4.6 – Status of oil moisture samples [6].

Oil moisture status			
Above parameter	1	Bad sample	Event
Below parameter	0	Good sample	Non-event

Using Equations (1) and (2), the WoE of each category and the IV of the variables were calculated. The results are presented in Table 4.7 and will be discussed in Chapter 5.

Table 4.7 – WoE of the categories and IV of the variables [6].

Variables	Categories	Total	Event	Non-event	% Event	% Non-event	WoE	% Event - %Non-event	IV
Voltage class range	≤ 72.5 kV	309	9	300	1.12%	3.36%	**-1.099**	-2.24%	0.025
	> 72.5 kV and ≤ 145 kV	1536	34	1502	4.23%	16.82%	**-1.381**	-12.59%	0.174
	> 145 kV	7888	761	7127	94.65%	79.82%	**0.170**	14.83%	0.025
Voltage class range									**0.224**
Installation region (Regional)	TT	1781	93	1688	11.57%	18.90%	**-0.491**	-7.34%	0.036
	TS	2865	440	2425	54.73%	27.16%	**0.701**	27.57%	0.193
	TE	144	3	141	0.37%	1.58%	**-1.443**	-1.21%	0.017
	TC	2484	128	2356	15.92%	26.39%	**-0.505**	-10.47%	0.053
	TB	2459	140	2319	17.41%	25.97%	**-0.400**	-8.56%	0.034
Installation region (Regional)									**0.334**
Criticality	A	2130	255	1875	31.72%	21.00%	**0.412**	10.72%	0.044
	B	5578	514	5064	63.93%	56.71%	**0.120**	7.22%	0.009
	C	2025	35	1990	4.35%	22.29%	**-1.633**	-17.93%	0.293
Criticality									**0.346**
Type	EQU-TRAFOP	7011	587	6424	73.01%	71.95%	**0.015**	1.06%	0.000
	EQU-AUTOTR	1065	43	1022	5.35%	11.45%	**-0.761**	-6.10%	0.046
	EQU-REACTO	1657	174	1483	21.64%	16.61%	**0.265**	5.03%	0.013
Type									**0.060**
Age range	≤ 17 years	3078	40	3038	4.98%	34.02%	**-1.923**	-29.05%	0.558
	> 17 years and < 35 years	2957	138	2819	17.16%	31.57%	**-0.609**	-14.41%	0.088
	≥ 35 years and ≤ 45 years	2515	416	2099	51.74%	23.51%	**0.789**	28.23%	0.223
	> 45 years	1183	210	973	26.12%	10.90%	**0.874**	15.22%	0.133
Age range									**1.002**
		9733	804	8929					

4.2. COMPARATIVE ANALYSIS OF THE APPLICATION OF THE WOE AND IV METRICS IN PREDICTIVE MODELS OF DIFFERENT AREAS

Below are presented applications of the WoE and IV metrics in predictive models from different areas.

Zhou et al [78], Alsabhan et al [79], Miao et al [80], Niu et al [81], Yang et al [82], Zhang et al [83], Wang, Kang and Wang [84], Apu et al [85], Bhandari, Dhakal and Tsou [86], and Li et al [87] present methods to predict geological landslide risks, a problem that involves several variables and mainly affects mountainous areas, using IV. The metric is used to determine the predictive power of landslide-causing factors, and the prediction models categorize the levels of zones susceptible to landslides.

Zhou et al [78], Miao et al [80], Niu et al [81], Zhang et al [83], Wang, Kang and Wang [84], Li et al [87], and Yang et al [82] present cases from regions in China: the Three Gorges Reservoir region, in Hubei Province [78], [80], Shaanxi province, in the South [81], Sichuan province in the Southwest [83], Jiaxian County in Shaanxi province on the Loess Plateau [84], Taihe in the northwest of Anhui province [87], and the Western Tibetan Plateau [82].

Alsabhan et al [79] and Bhandari, Dhakal and Tsou [86] present cases from the Himalayan region – Himachal Pradesh, India [79] and Nepal [86] – and Apu et al [85] present a case from Khagrachari, Bangladesh.

Variables such as slope [79], [81], [83], [85], [87], elevation [79], [83], vegetation [79], [83], [85], [87], soil type [79], [81], aspect [79], [83], curvature [79], [83], [87], precipitation [83], soil moisture [85], and relief [81], [85] are used in the studies.

In addition to IV, the following methods are also used to develop the models: Frequency Ratio (FR) [79], [85], [86], Neural Networks [78], [80], [87], SVM [78], [80], Logistic Regression [78], [83], Random Forest [80], Decision Tree [83], and Kernel Extreme Learning Machine [84].

In the cases presented above [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], the application of the WoE and IV metrics aided in the formulation of strategies for preventing and controlling landslide-related disasters.

Ghasemzadeh et al [88] present a geochemical model aimed at mineral exploration and use IV to quantify the importance of geochemical anomaly classes relative to mineral deposits. The proposed model makes geochemical data relevant information for mineral exploration strategies, standing out compared to models generated by Fuzzy operators. The authors state that the use of this statistical metric reveals spatial patterns of geochemical signals around mineral deposits, facilitating the exploration of undiscovered locations. The methodology was applied to a dataset from the Kerman province in Iran.

Zhao, Yuan, and Chen [11] propose a model for managing air traffic flow and use IV to select and weigh relevant variables for prediction, such as air traffic delay at the departure airport and estimated flow at the first passing point. The authors implemented the method using SVM and Particle Swarm Optimization (PSO). The prediction accuracy of the IV-PSO-SVM model reached 96.4%, surpassing SVM and PSO-SVM by 13.5% and 9.1%, respectively, demonstrating that IV contributed to selecting more relevant features for the desired prediction, which reflected in the model's forecast. The methodology was validated with data from airports and air passing points in China.

Addis [89] proposes a study to map flood susceptibility in the Abay River basin in Ethiopia using the FR and IV metrics to identify flood-prone areas. The model initially predicted twelve flood conditioning factors, including slope, elevation, aspect, land use, curvature, distance from roads, distance from rivers, precipitation, lithology, and soil texture. The conditioning factors were integrated with training data to determine weights using both models, and susceptibility maps were reclassified into classes ranging from very low to very high. The FR and IV models proved effective in mapping flood susceptibility, and the resulting maps can be valuable for flood planning and mitigation decisions by local government.

5. RESULTS AND DISCUSSIONS

The IV values of the variables are graphically represented in Figure 5.1.

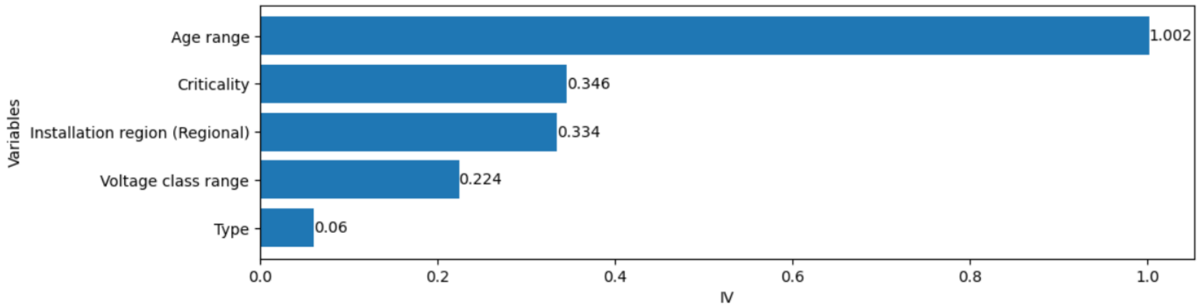


Figure 5.1 – IV of the variables.

The age range variable presented IV > 0.50. The IV was calculated for different category ranges of this variable and in all cases the result was greater than 0.5. In this case, as shown in Table 2.4 [11], one must consider the authenticity of the variable. According to the ABNT NBR 10576:2017 [8], the main sources of moisture increase in transformer oil are atmospheric moisture ingress and solid insulation degradation. The age of assets influences both. Therefore, it is considered authentic and has strong predictive power.

Table 5.1 describes the predictive values of the variables according to the IV value.

Table 5.1 – Predictive value of the variables according to the IV.

Variables	Predictive value according to the IV
Age range	Strong
Criticality	Strong
Installation region (Regional)	Strong
Voltage class range	Medium
Type	Weak

With the aim of prioritizing variables with the greatest impact on investment decisions for maintenance actions and determining the most significant variables for a predictive model, variables with weak and medium predictive value were disregarded.

Table 5.2 presents the main information resulting from the analyses presented in this thesis: variables to be prioritized, and order of prioritization, in the evaluation of investments in maintenance, considering the moisture in the oil of power transformers and reactors, and which must be considered in a predictive model.

Table 5.2 – Variables to be prioritized in maintenance investment evaluation and which must be considered in a predictive model.

N	Variables
1	Age range
2	Criticality
3	Installation region (Regional)

The proposed methodology indicated the variables of age, criticality, and Regional as having the greatest predictive power. The degradation of solid insulation and the deterioration of the sealing are directly related to the age of transformers. Regional refers to the equipment's installation location, thus influenced by climate factors such as weather, temperature, and humidity. Criticality encompasses systemic impact as one of the qualifying factors. Systemic position is directly linked to operational conditions, such as load and operating cycles. In addition, the Regional variable also takes into account human aspects, management, processes, and people. The way maintenance is performed and the individuals carrying it out can also influence the moisture content in the oil.

The degradation of solid insulation and the deterioration of the sealing in transformers and reactors significantly impact the water content in the oil. As the insulation of paper degrades, it releases moisture into the oil. Additionally, the deteriorated sealing loses efficiency, allowing moisture from the air to enter.

Regarding climate conditions, the ambient temperature influences the oil's ability to retain water. Heat causes the paper's ability to hold water to decrease, and the moisture present in the transformer's solid insulation migrates to the oil, increasing the oil's capacity to retain water in solution. Thus, during periods of high temperatures or in warm climates, the oil's moisture content may increase. On the other hand, low temperatures reduce the oil's ability to hold dissolved moisture, which can result in the formation of free water (water droplets) in the oil. Water is, therefore, highly detrimental to the dielectric.

The relative humidity of the air is another climate factor that influences moisture in transformer oil. High air humidity can increase moisture in the oil. Transformers exposed to humid air, especially if there are sealing system failures, may allow ambient moisture to infiltrate the oil over time.

In relation to operating conditions, high loads increase the transformer's operating temperature, which also brings about the aforementioned thermal effect. Load and unload cycles (turning the transformer on and off or fluctuations in the load), in turn, cause periodic temperature changes in the transformer. This cyclical process of moisture absorption and release over time can increase the oil's moisture content.

Therefore, it is concluded that the results obtained through IV are consistent.

Regarding the WoE values, as described in item 4.1, a positive value indicates that the category increases the probability of the event occurring, while a negative value implies that the category decreases the probability of the event. In this case, the event refers to the sample having a moisture content above the parameters. In summary, the higher the WoE value, the greater the weight of evidence for samples exceeding the parameters.

Figures 5.2, 5.3, and 5.4 graphically show the WoE values for each category of the variables with strong predictive power (highlighted in Table 5.2): age range, criticality, and Regional, respectively.

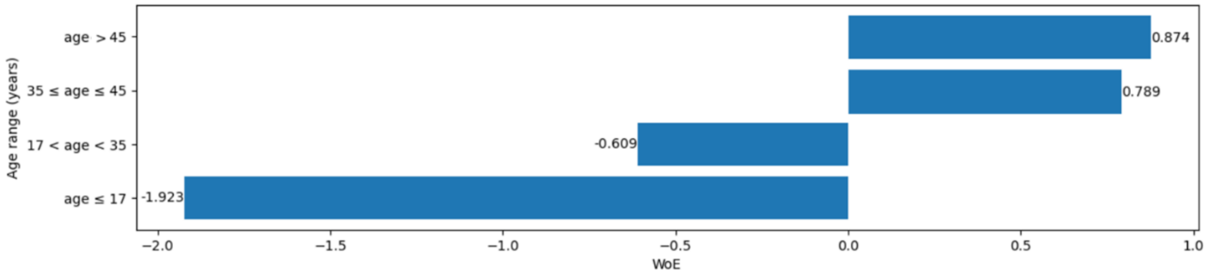


Figure 5.2 – WoE values of the age range variable.

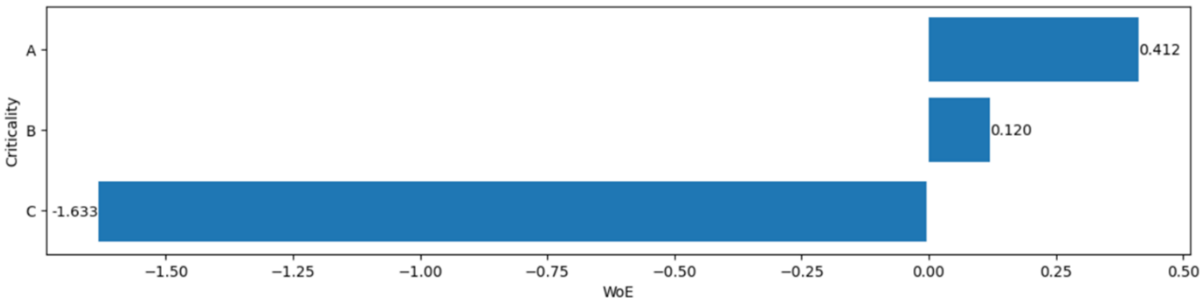


Figure 5.3 – WoE values of the criticality variable.

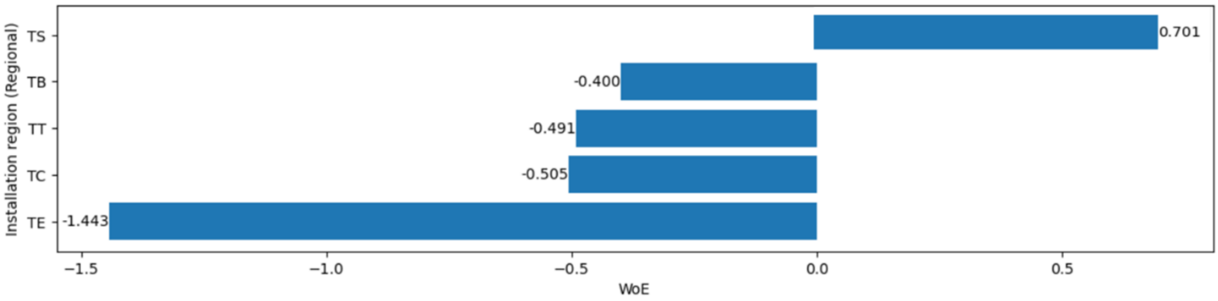


Figure 5.4 – WoE values of the Regional variable.

The categories over 35 years old show positive WoE values, meaning that, based on this dataset, assets with expired regulatory useful life tend to have more samples with moisture content in the oil above the parameters.

Regarding the criticality variable, categories A and B have positive WoE values, with category A holding the greatest weight of evidence for events. For the Regional variable, only the TS Regional shows positive WoE values.

The categories with the highest weight of the variables with the strong predictive power are related, that is, the older equipment is located in the TS Regional and represents the most critical assets responsible for transmitting electrical energy to the city of São Paulo and its metropolitan area. On the other hand, the categories with the lowest weight of these variables are from the newer assets (in the age range younger than 17 years), with lower criticality, and from the TE Regional. The concessions outside the state of São Paulo, from the TE Regional, are new installations acquired through transmission auctions, representing the newest assets of the Company.

Therefore, it is concluded that the results obtained through WoE are consistent.

To obtain an initial understanding of the data and identify patterns, trends and possible correlations, a preliminary descriptive analysis of the data was carried out. The main results are presented in the APPENDIX. It is worth highlighting that, through this preliminary analysis, it was possible to identify the initial patterns, however it was not possible to identify the variables with greater predictive power, an objective achieved with the use of WoE and IV metrics.

6. CONCLUSIONS

It is concluded that the methodology using the WoE and IV metrics arrived at consistent results, considering factors that significantly influence the water content in the oil.

The online treatment of oil, or in other words, treatment while the equipment is in operation, or the replacement of the oil load are the possible corrective actions, considering the water content in the oil, applied according to the parameters shown in Table 2.1. This prioritization is currently carried out at ISA CTEEP considering the tacit knowledge of experts. The acquisition cost of machines that perform online treatment is high, and for the installation of the machine or replacement of the oil, it is necessary to schedule the shutdown of the transformer or reactor, which incurs penalties for unavailability. Moreover, corrective activities, whether online treatment or oil replacement, involve aspects of safety and the environment.

Information on which data should be prioritized also helps guide investments in preventive actions. Acting preventively, in turn, aims to minimize occurrences and failures, and consequently, the need for corrective actions that cause significant impacts. Through data analysis, the proposed methodology reveals trends and, in this way, provides information for decision-making, which minimizes corrective actions on the overall system.

Therefore, from the perspective of applying the methodology to issues in the electrical sector, the analysis of the predictive importance of the variables to be evaluated constitutes an important contribution, especially considering large equipment parks, for directing preventive actions, in order to assess which categories should be prioritized in investment analysis.

The existing maintenance practices in the Brazilian power transmission sector are, for the most part, periodic preventive maintenance, that is, time based preventive maintenance, and corrective maintenance. The proposed methodology helps determine preventive maintenance based on the condition of transformers, and it can also guide adjustments to the frequency of periodic preventive maintenance.

From the perspective of added value through the innovative application of data analysis tools, assessing the predictive importance of a variable relative to the output before developing predictors can lead to better performing models.

An example of a possible application of the proposed methodology is in guiding investments in continuous online monitoring of transformers. Installing monitoring systems throughout the entire operating park is not feasible; therefore, an analysis should be conducted to prioritize this investment. The proposed approach directs which variables should be prioritized in this analysis and the category of the selected variables with the highest weight of evidence.

Figure 6.1 illustrates a real case in the ISA CTEEP park of installation a continuous online oil monitoring system on a 39-year-old transformer, with criticality A, located in the TS Regional, the variables and categories indicated by the proposed methodology using IV and WoE metrics.

(a) (b)

Figure 6.1 – Real case of (a) continuous online oil monitoring system installed on a (b) power transformer in operation, installed in an ISA CTEEP substation.

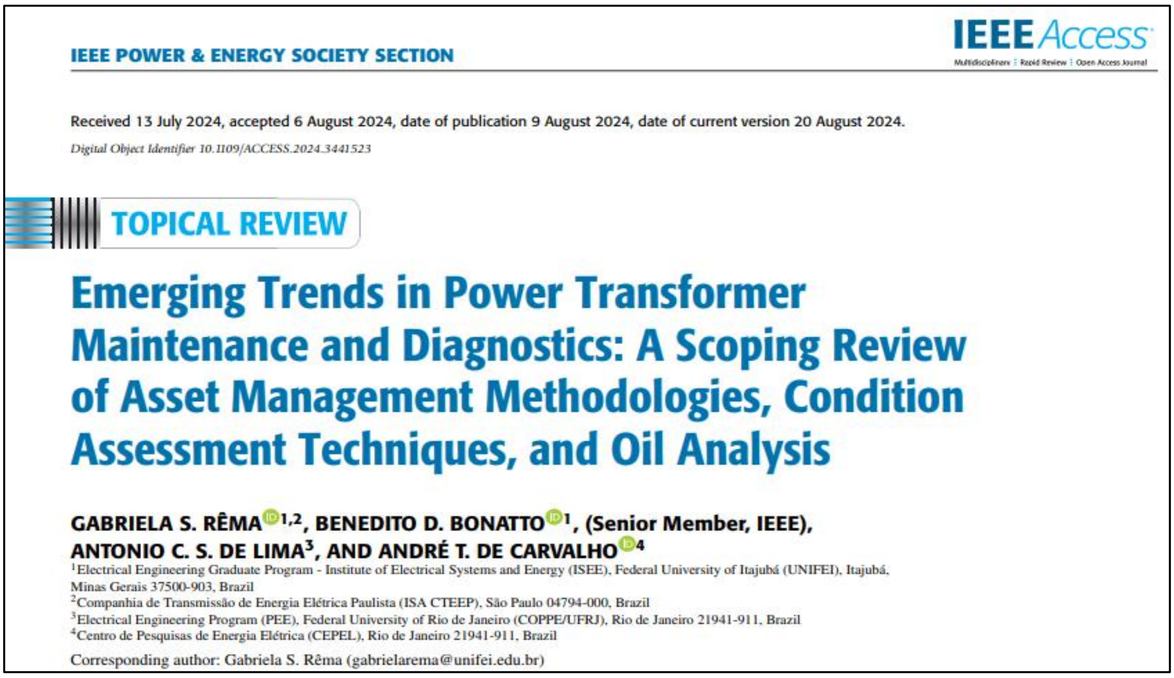
Therefore, the company's actions are aligning with the results of the methodology, indicating that there is potential for implementation.

The following presents the publications and possibilities for future work.

6.1. Publications

The scoping review carried out which was the basis for the original contribution of this thesis was published in the article "Emerging Trends in Power Transformer

Maintenance and Diagnostics: A Scoping Review of Asset Management Methodologies, Condition Assessment Techniques, and Oil Analysis" [5] in IEEE Access, as illustrated in Figure 6.2.


IEEE POWER & ENERGY SOCIETY SECTION

IEEE Access

Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 July 2024, accepted 6 August 2024, date of publication 9 August 2024, date of current version 20 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3441523

TOPICAL REVIEW

Emerging Trends in Power Transformer Maintenance and Diagnostics: A Scoping Review of Asset Management Methodologies, Condition Assessment Techniques, and Oil Analysis

GABRIELA S. RÊMA[1,2], BENEDITO D. BONATTO[1], (Senior Member, IEEE), ANTONIO C. S. DE LIMA[3], AND ANDRÉ T. DE CARVALHO[4]

[1]Electrical Engineering Graduate Program - Institute of Electrical Systems and Energy (ISEE), Federal University of Itajubá (UNIFEI), Itajubá, Minas Gerais 37500-903, Brazil

[2]Companhia de Transmissão de Energia Elétrica Paulista (ISA CTEEP), São Paulo 04794-000, Brazil

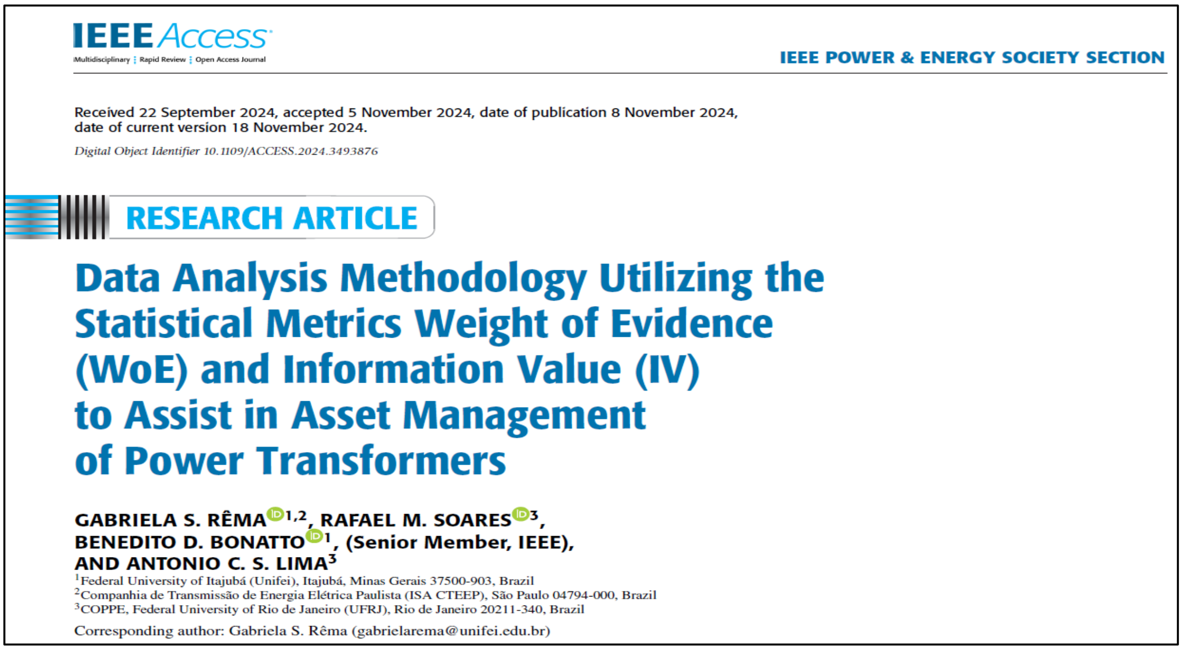
[3]Electrical Engineering Program (PEE), Federal University of Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro 21941-911, Brazil

[4]Centro de Pesquisas de Energia Elétrica (CEPEL), Rio de Janeiro 21941-911, Brazil

Corresponding author: Gabriela S. Rêma (gabrielarema@unifei.edu.br)

Figure 6.2 – Article with the scope review of the thesis published in IEEE Access [5].

The methodology and results of the thesis are presented in the article “Data analysis methodology utilizing the statistical metrics Weight of Evidence (WoE) and Information Value (IV) to assist in asset management of power transformers” published in IEEE Access, as illustrated in Figure 6.3.


IEEE Access

Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

IEEE POWER & ENERGY SOCIETY SECTION

Received 22 September 2024, accepted 5 November 2024, date of publication 8 November 2024, date of current version 18 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3493876

RESEARCH ARTICLE

Data Analysis Methodology Utilizing the Statistical Metrics Weight of Evidence (WoE) and Information Value (IV) to Assist in Asset Management of Power Transformers

GABRIELA S. RÊMA[1,2], RAFAEL M. SOARES[3], BENEDITO D. BONATTO[1], (Senior Member, IEEE), AND ANTONIO C. S. LIMA[3]

[1]Federal University of Itajubá (Unifei), Itajubá, Minas Gerais 37500-903, Brazil

[2]Companhia de Transmissão de Energia Elétrica Paulista (ISA CTEEP), São Paulo 04794-000, Brazil

[3]COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro 20211-340, Brazil

Corresponding author: Gabriela S. Rêma (gabrielarema@unifei.edu.br)

Figure 6.3 – Article with the methodology and results of the thesis published in IEEE Access [6].

Table 6.1 shows the articles published, accepted and submitted at congresses, during the doctorate, linked to failure analysis techniques and assessment of the condition of power transformers.

Table 6.1 – Articles published, accepted and submitted at congresses.

N	Article title	Congress	Article status	Congress location	Month and year	Authorship
1	Black-Box Modeling of Power Transformers at High Frequencies	International Conference on Power Systems Transients (IPST)	Published	Perpignan, França	June, 2019	Main author
2	*Modelagem do Transformador de Potência em Altas Frequências através de Medições em Campo*	*XXV Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Published	Belo Horizonte, MG	November, 2019	Main author
3	*Análise da Interação entre Equipamentos de Alta Tensão e Eventos Transitórios Eletromagnéticos de Altas Frequências do Sistema Elétrico*	*XXV Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Published	Belo Horizonte, MG	November, 2020	Main author
4	*Aplicação de Tecnologias da Indústria 4.0 no Monitoramento Online de Ativos do Setor Elétrico*	*36º CBMGA - Congresso Brasileiro de Manutenção e Gestão de Ativos*	Published	Online	October, 2021	Main author
5	*A técnica de resposta em frequência aplicada ao diagnóstico de falhas e modelagem de transformadores auxiliando na gestão de ativos*	*XXVI Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Published	Rio de Janeiro, RJ	May, 2022	Main author
6	*Análise do Estado de Transformadores de Potência a partir do Monitoramento Online da Condição Operativa*	*X WORKSPOT- 10° Workshop Internacional sobre Transformadores de Potência, Equipamentos, Subestações e Materiais*	Published	Foz do Iguaçu, PR	November, 2022	Main author
7	*Análise sobre a priorização da manutenção de transformadores de potência pautada na condição e criticidade do ativo*	*XIX Encontro Regional Ibero-Americano do CIGRE (ERIAC)*	Published	Foz do Iguaçu, PR	May, 2023	Co-author
8	*Análise dos dados de monitoramento da condição como prevenção de ocorrências*	*XXVII Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Published	Brasília, DF	November, 2023	Co-author

N	Article title	Congress	Article status	Congress location	Month and year	Authorship
9	*Avaliação da condição de transformadores de potência baseado nas manutenções preventivas e avanços em relação ao monitoramento online contínuo*	*XI WORKSPOT-11° Workshop Internacional sobre Transformadores de Potência, Equipamentos, Subestações e Materiais*	Accepted	Rio de Janeiro, RJ	November, 2024	Main author
10	*Estruturação de arquitetura de rede e processos envolvendo o monitoramento online contínuo e análise da condição de ativos da ISA CTEEP*	*XI WORKSPOT-11° Workshop Internacional sobre Transformadores de Potência, Equipamentos, Subestações e Materiais*	Accepted	Rio de Janeiro, RJ	November, 2024	Main author
11	*Metodologia para qualificação da condição de transformadores de potência à óleo baseado em normas técnicas e análise estatística de dados*	*XXVIII Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Submitted	Recife, PE	October, 2025	Co-author
12	*Importância da gestão de ativos no setor de transmissão de energia elétrica: implementação da ISO 55.001 na ISA CTEEP e os desafios da revisão 2024*	*XXVIII Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE)*	Submitted	Recife, PE	October, 2025	Co-author

6.2. POSSIBILITIES FOR FUTURE WORK

As possibilities for future work highlight the comparison with other variable selection methods, which consider predictive power.

It is also suggested to consider other physical-chemical characteristics of the oil, besides moisture content, as, in this way, the maintenance actions to be evaluated and prioritized will take into account other properties for evaluating the transformers' dielectric.

In addition to the physical-chemical properties of the oil, operational variables of the assets, such as oil and winding temperatures, voltage, and current, can provide important information for diagnostics. It is proposed to use data from these variables to correlate with the presented variables, as well as failure data of the assets.

Also noteworthy as a possibility for future work is the development of a prediction model aiming to predict the oil sample results regarding moisture content.

The predictive model should be designed based on the variables identified as having the highest predictive power through the analysis using IV. These variables should be used as input variables in the model to predict the occurrence of increased moisture content in transformers and reactors oil.

REFERENCES

[1] da Silva, D. G. T., Braga Da Silva, H. J., Marafão, F. P., Paredes, H. K. M., & Gonçalves, F. A. S. (2021). Enhanced health index for power transformers diagnosis. Engineering Failure Analysis, 126. https://doi.org/10.1016/j.engfailanal.2021.105427

[2] de Castro, N., Martini, S. & Gianelloni, F. (2020). *Gestão de Ativos de Transmissão e a Importância de Processos Padronizados.* GESEL – *Grupo de Estudos do Setor Elétrico*, UFRJ.

[3] Plano Decenal de Expansão de Energia 2031. Ministério de Minas e Energia e Empresa de Pesquisa Energética. Brasil - Brasília: MME/EPE, 2022.

[4] Tenbohlen, S. et al. "Transformer Reliability Survey". Working Group A2.37 – Cigrè. December, 2015.

[5] G. S. Rêma, B. D. Bonatto, A. C. S. de Lima and A. T. de Carvalho, "Emerging Trends in Power Transformer Maintenance and Diagnostics: A Scoping Review of Asset Management Methodologies, Condition Assessment Techniques, and Oil Analysis," in IEEE Access, vol. 12, pp. 111451-111467, 2024, doi: 10.1109/ACCESS.2024.3441523.

[6] G. S. Rêma, R. M. Soares, B. D. Bonatto and A. C. S. de Lima. "Data analysis methodology utilizing the statistical metrics Weight of Evidence (WoE) and Information Value (IV) to assist in asset management of power transformers" in IEEE Access, vol. 12, pp. 165322-165334, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3493876.

[7] Tenbohlen, S. et al. "Results of a Standardized Survey about the Reliability of Power Transformers". The 20th International Symposium on High Voltage Engineering. Buenos Aires, Argentina, August 27 – September 01, 2017.

[8] ABNT NBR 10576:2017 – *Óleo mineral isolante de equipamentos elétricos — Diretrizes para supervisão e manutenção.*

[9] Soni, R., & Mehta, B. (2023). A review on transformer condition monitoring with critical investigation of mineral oil and alternate dielectric fluids. In Electric Power Systems Research (Vol. 214). Elsevier Ltd. https://doi.org/10.1016/j.epsr.2022.108954

[10] Ariannik, M., Razi-Kazemi, A. A., & Lehtonen, M. (2020). An approach on lifetime estimation of distribution transformers based on degree of polymerization. Reliability Engineering and System Safety, 198. https://doi.org/10.1016/j.ress.2020.106881

[11] Zhao, Z., Yuan, J., & Chen, L. (2023). Research on Air Traffic Flow Management Delay Distribution Prediction Based on IV Value and PSO-SVM. IEEE Access, 11, 84035–84047. https://doi.org/10.1109/ACCESS.2023.3300373

[12] Niu, J., Li, X., Tian, Z., & Yang, H. (2023). A framework for quantifying the value of information to mitigate risk in the optimal design of distributed energy systems under uncertainty. Applied Energy, 350. https://doi.org/10.1016/j.apenergy.2023.121717

[13] Tricco et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine, 169(7). https://doi.org/10.7326/M18-0850

[14] Abu-Elanien, A. E. B., & Salama, M. M. A. (2010). Asset management techniques for transformers. Electric Power Systems Research, 80(4), 456–464. https://doi.org/10.1016/j.epsr.2009.10.008

[15] Velasquez-Contreras, J. L., Sanz-Bobi, M. A., & Galceran Arellano, S. (2011). General asset management model in the context of an electric utility: Application to power transformers. Electric Power Systems Research, 81(11), 2015–2037. https://doi.org/10.1016/j.epsr.2011.06.007

[16] Soni, R., & Mehta, B. (2021). Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies. In Engineering Failure Analysis (Vol. 128). Elsevier Ltd. https://doi.org/10.1016/j.engfailanal.2021.105634

[17] Murugan, R., & Ramasamy, R. (2015). Failure analysis of power transformer for effective maintenance planning in electric utilities. Engineering Failure Analysis, 55, 182–192. https://doi.org/10.1016/j.engfailanal.2015.06.002

[18] Koziel, S., Hilber, P., Westerlund, P., & Shayesteh, E. (2021). Investments in data quality: Evaluating impacts of faulty data on asset management in power systems. Applied Energy, 281. https://doi.org/10.1016/j.apenergy.2020.116057

[19] Gorginpour, H., Ghimatgar, H., & Toulabi, M. S. (2022). Lifetime Estimation and Optimal Maintenance Scheduling of Urban Oil-Immersed Distribution-Transformers Considering Weather-Dependent Intelligent Load Model and Unbalanced Loading. IEEE Transactions on Power Delivery, 37(5), 4154–4165. https://doi.org/10.1109/TPWRD.2022.3146154

[20] Balanta, J. Z., Rivera, S., Romero, A. A., & Coria, G. (2023). Planning and Optimizing the Replacement Strategies of Power Transformers: Literature Review. In Energies (Vol. 16, Issue 11). MDPI. https://doi.org/10.3390/en16114448

[21] Biçen, Y., & Aras, F. (2023). Smart asset management system for power transformers coupled with online and offline monitoring technologies. Engineering Failure Analysis, 154. https://doi.org/10.1016/j.engfailanal.2023.107674

[22] Jin, L., Kim, D., Abu-Siada, A., & Kumar, S. (2022). Oil-Immersed Power Transformer Condition Monitoring Methodologies: A Review. In Energies (Vol. 15, Issue 9). MDPI. https://doi.org/10.3390/en15093379

[23] Jin, L., Kim, D., & Abu-Siada, A. (2023). State-of-the-art review on asset management methodologies for oil-immersed power transformers. In Electric Power Systems Research (Vol. 218). Elsevier Ltd. https://doi.org/10.1016/j.epsr.2023.109194

[24] Soni, R., & Mehta, B. (2022). Evaluation of power transformer health analysis by internal fault criticalities to prevent premature failure using statistical data analytics approach. Engineering Failure Analysis, 136. https://doi.org/10.1016/j.engfailanal.2022.106213

[25] Wong, S. Y., Ye, X., Guo, F., & Goh, H. H. (2022). Computational intelligence for preventive maintenance of power transformers. In Applied Soft Computing (Vol. 114). Elsevier Ltd. https://doi.org/10.1016/j.asoc.2021.108129

[26] Ma, H., Saha, T. K., & Ekanayake, C. (2012). Statistical Learning Techniques and Their Applications for Condition Assessment of Power Transformer. In IEEE Transactions on Dielectrics and Electrical Insulation (Vol. 19, Issue 2).

[27] Guillen, D., Idarraga-Ospina, G., Mombello, E., & Cabral, S. (2016). Partial discharges location in transformer winding using wavelets and Kullback-Leibler divergence. *Electric Power Systems Research*, *136*, 398–405. https://doi.org/10.1016/j.epsr.2016.03.020

[28] Didouche, S., Nacer, A., Ziani, A., Moulai, H., & Mazighi, K. (2022). Iterative method for partial discharges location in power transformers. Electric Power Systems Research, 211. https://doi.org/10.1016/j.epsr.2022.108225

[29] Sekatane, P. M., & Bokoro, P. (2023). Time Reversal vs. Integration of Time Reversal with Convolution Neural Network in Diagnosing Partial Discharge in Power Transformer. *Energies*, *16*(23). https://doi.org/10.3390/en16237872

[30] Beura, C. P., Wolters, J., & Tenbohlen, S. (2024). Application of Pathfinding Algorithms in Partial Discharge Localization in Power Transformers. *Sensors*, *24*(2). https://doi.org/10.3390/s24020685

[31] Liu, F., Du, J., Shi, Y., Zhang, S., & Wang, W. (2024). Localization of Dual Partial Discharge in Transformer Windings Using Fabry–Pérot Optical Fiber Sensor Array. *Energies*, *17*(11). https://doi.org/10.3390/en17112537

[32] Khalili Senobari, R., Sadeh, J., & Borsi, H. (2018). Frequency response analysis (FRA) of transformers as a tool for fault detection and location: A review. In Electric Power Systems Research (Vol. 155, pp. 172–183). Elsevier Ltd. https://doi.org/10.1016/j.epsr.2017.10.014

[33] Seifi, S., Werle, P., Shayegani Akmal, A. A., & Mohseni, H. (2022). Sweep reflection coefficient method for localization of internal faults in transformer windings. International Journal of Electrical Power and Energy Systems, 137. https://doi.org/10.1016/j.ijepes.2021.107781

[34] Velásquez, R. M. A., & Lara, J. V. M. (2018). Life estimation of shunt power reactors considering a failure core heating by floating potentials. Engineering Failure Analysis, 86, 142–157. https://doi.org/10.1016/j.engfailanal.2018.01.004

[35] Medina, R. D., Zaldivar, D. A., Romero, A. A., Zuñiga, J., & Mombello, E. E. (2022). A fuzzy inference-based approach for estimating power transformers risk index. Electric Power Systems Research, 209. https://doi.org/10.1016/j.epsr.2022.108004

[36] Costa, J. V., da Silva, D. F. F., & Branco, P. J. C. (2022). Large-Power Transformers: Time Now for Addressing Their Monitoring and Failure Investigation Techniques. Energies, 15(13). https://doi.org/10.3390/en15134697

[37] Zhou, Y., Yang, X., Tao, L., & Yang, L. (2021). Transformer fault diagnosis model based on improved gray wolf optimizer and probabilistic neural network. Energies, 14(11). https://doi.org/10.3390/en14113029

[38] Huang, W., Shao, C., Hu, B., Li, W., Sun, Y., Xie, K., Zio, E., & Li, W. (2023). A restoration-clustering-decomposition learning framework for aging-related failure rate estimation of distribution transformers. Reliability Engineering and System Safety, 232. https://doi.org/10.1016/j.ress.2022.109043

[39] Islam, N., Khan, R., Das, S. K., Sarker, S. K., Islam, M. M., Akter, M., & Muyeen, S. M. (2023). Power transformer health condition evaluation: A deep generative model aided intelligent framework. Electric Power Systems Research, 218. https://doi.org/10.1016/j.epsr.2023.109201

[40] Wang, X., Li, Q., Li, C., Yang, R., Su, Q. (2013). Reliability Assessment of the Fault Diagnosis Methodologies for Transformers and a New Diagnostic Scheme Based on Fault Info Integration. In IEEE Transactions on Dielectrics and Electrical Insulation (Vol. 20, Issue 2).

[41] Arias Velásquez, R. M., & Mejía Lara, J. V. (2019). Root cause analysis for shunt reactor failure in 500 kV power system. Engineering Failure Analysis, 104, 1157–1173. https://doi.org/10.1016/j.engfailanal.2019.06.076

[42] Murugan, R., & Ramasamy, R. (2019). Understanding the power transformer component failures for health index-based maintenance planning in electric utilities. Engineering Failure Analysis, 96, 274–288. https://doi.org/10.1016/j.engfailanal.2018.10.011

[43] Ariannik, M., Razi-Kazemi, A. A., & Lehtonen, M. (2020). An approach on lifetime estimation of distribution transformers based on degree of polymerization. Reliability Engineering and System Safety, 198. https://doi.org/10.1016/j.ress.2020.106881

[44] Liu, S., Zhao, Y., Lin, Z., Ding, Y., Yan, Y., Yang, L., Wang, Q., Zhou, H., & Wu, H. (2019). Data-Driven Condition Monitoring of Data Acquisition for Consumers' Transformers in Actual Distribution Systems Using t-Statistics. IEEE Transactions on Power Delivery, 34(4), 1578–1587. https://doi.org/10.1109/TPWRD.2019.2912267

[45] Chen, W., Pan, C., Yun, Y., & Liu, Y. (2009). Wavelet networks in power transformers diagnosis using dissolved gas analysis. IEEE Transactions on Power Delivery, 24(1), 187–194. https://doi.org/10.1109/TPWRD.2008.2002974

[46] Abu-Siada, A., & Islam, S. (2012). A New Approach to Identify Power Transformer Criticality and Asset Management Decision Based on Dissolved Gas-in-Oil Analysis. In IEEE Transactions on Dielectrics and Electrical Insulation (Vol. 19, Issue 3).

[47] Abu-Siada, A., Hmood, S., & Islam, S. (2013). A New Fuzzy Logic Approach for Consistent Interpretation of Dissolved Gas-in-Oil Analysis. In IEEE Transactions on Dielectrics and Electrical Insulation (Vol. 20, Issue 6).

[48] Cui, Y., Ma, H., & Saha, T. (2014). Improvement of power transformer insulation diagnosis using oil characteristics data preprocessed by SMOTEBoost technique. IEEE Transactions on Dielectrics and Electrical Insulation, 21(5), 2363–2373. https://doi.org/10.1109/TDEI.2014.004547

[49] Tra, V., Duong, B. P., & Kim, J. M. (2019). Improving diagnostic performance of a power transformer using an adaptive over-sampling method for imbalanced data. IEEE Transactions on Dielectrics and Electrical Insulation, 26(4), 1325–1333. https://doi.org/10.1109/TDEI.2019.008034

[50] Rajesh, K. N. V. P. S., Mohan Rao, U., Fofana, I., Rozga, P., & Paramane, A. (2023). Influence of Data Balancing on Transformer DGA Fault Classification With Machine Learning Algorithms. IEEE Transactions on Dielectrics and Electrical Insulation, 30(1), 385–392. https://doi.org/10.1109/TDEI.2022.3230377

[51] Arias Velásquez, R. M., & Mejía Lara, J. V. (2018). Corrosive Sulphur effect in power and distribution transformers failures and treatments. Engineering Failure Analysis, 92, 240–267. https://doi.org/10.1016/j.engfailanal.2018.05.018

[52] Aizpurua, J. I., Stewart, B. G., McArthur, S. D. J., Lambert, B., Cross, J. G., & Catterson, V. M. (2019). Improved power transformer condition monitoring under uncertainty through soft computing and probabilistic health index. Applied Soft Computing, 85. https://doi.org/10.1016/j.asoc.2019.105530

[53] Bustamante, S., Manana, M., Arroyo, A., Castro, P., Laso, A., & Martinez, R. (2019). Dissolved gas analysis equipment for online monitoring of transformer oil: A review. In Sensors (Switzerland) (Vol. 19, Issue 19). MDPI AG. https://doi.org/10.3390/s19194057

[54] Ward, S. A., El-faraskoury, A., Badawi, M., Ibrahim, S. A., Mahmoud, K., Lehtonen, M., & Darwish, M. M. F. (2021). Towards precise interpretation of oil transformers via novel combined techniques based on DGA and partial discharge sensors. Sensors, 21(6), 1–21. https://doi.org/10.3390/s21062223

[55] Menezes, A. G. C., Araujo, M. M., Almeida, O. M., Barbosa, F. R., & Braga, A. P. S. (2022). Induction of Decision Trees to Diagnose Incipient Faults in Power Transformers. IEEE Transactions on Dielectrics and Electrical Insulation, 29(1), 279–286. https://doi.org/10.1109/TDEI.2022.3148453

[56] Soni, R., & Mehta, B. (2023). Diagnosis and prognosis of incipient faults and insulation status for asset management of power transformer using fuzzy logic controller & fuzzy clustering means. Electric Power Systems Research, 220. https://doi.org/10.1016/j.epsr.2023.109256

[57] Soni, R., & Mehta, B. (2023). A review on transformer condition monitoring with critical investigation of mineral oil and alternate dielectric fluids. In Electric Power Systems Research (Vol. 214). Elsevier Ltd. https://doi.org/10.1016/j.epsr.2022.108954

[58] Malik, I. M., Sharma, A., & Naayagi, R. T. (2023). A Comprehensive and Practical Method for Transformer Fault Analysis With Historical Data Trend Using Fuzzy Logic. IEEE Transactions on Dielectrics and Electrical Insulation, 30(5), 2277–2284. https://doi.org/10.1109/TDEI.2023.3286795

[59] Vrsaljko, D., Haramija, V., & Hadži-Skerlev, A. (2012). Determination of phenol, m-cresol and o-cresol in transformer oil by HPLC method. Electric Power Systems Research, 93, 24–31. https://doi.org/10.1016/j.epsr.2012.07.006

[60] Senoussaoui, M. E. A., Brahami, M., & Fofana, I. (2021). Transformer oil quality assessment using random forest with feature engineering. Energies, 14(7). https://doi.org/10.3390/en14071809

[61] Arias Velásquez, R. M., & Mejía Lara, J. V. (2018). Bushing failure in power transformers and the influence of moisture with the spectroscopy test. Engineering Failure Analysis, 94, 300–312. https://doi.org/10.1016/j.engfailanal.2018.08.006

[62] Liu, J., Ding, Z., Fan, X., Geng, C., Song, B., Wang, Q., & Zhang, Y. (2022). A BPNN Model-Based AdaBoost Algorithm for Estimating Inside Moisture of Oil-Paper Insulation of Power Transformer. IEEE Transactions on Dielectrics and Electrical Insulation, 29(2), 614–622. https://doi.org/10.1109/TDEI.2022.3157909

[63] Liu, J., Zheng, H., Zhang, Y., Wei, H., & Liao, R. (2017). Grey relational analysis for insulation condition assessment of power transformers based upon conventional dielectric response measurement. *Energies*, *10*(10). https://doi.org/10.3390/en10101526

[64] Zukowski, P., Rogalski, P., Koltunowicz, T. N., Kierczynski, K., & Bondariev, V. (2021). Precise measurements of the temperature-frequency dependence of the conductivity of cellulose—insulating oil—water nanoparticles composite. Energies, 14(1). https://doi.org/10.3390/en14010032

[65] Hernandez, G., & Ramirez, A. (2022). Dielectric Response Model for Transformer Insulation Using Frequency Domain Spectroscopy and Vector Fitting. *Energies*, *15*(7). https://doi.org/10.3390/en15072655

[66] Vatsa, A., & Hati, A. S. (2024). Insulation aging condition assessment of transformer in the visual domain based on SE-CNN. *Engineering Applications of Artificial Intelligence*, *128*. https://doi.org/10.1016/j.engappai.2023.107409

[67] Singh, J., Sood, Y. R., & Verma, P. (2012). The Influence of Service Aging on Transformer Insulating Oil Parameters. In IEEE Transactions on Dielectrics and Electrical Insulation (Vol. 19, Issue 2).

[68] Arsad, S. R., Ker, P. J., Jamaludin, M. Z., Choong, P. Y., Lee, H. J., Thiviyanathan, V. A., & Yang Ghazali, Y. Z. (2023). Water Content in Transformer Insulation System: A Review on the Detection and Quantification Methods. In Energies (Vol. 16, Issue 4). MDPI. https://doi.org/10.3390/en16041920

[69] Przybylek, P. (2022). Application of Near-Infrared Spectroscopy to Measure the Water Content in Liquid Dielectrics. Energies, 15(16). https://doi.org/10.3390/en15165907

[70] Koch, M., Tenbohlen, S., & Stirl, T. (2010). Diagnostic application of moisture equilibrium for power transformers. IEEE Transactions on Power Delivery, 25(4), 2574–2581. https://doi.org/10.1109/TPWRD.2010.2048343

[71] Medina, R. D., Romero, A. A., Mombello, E. E., & Rattá, G. (2017). Assessing degradation of power transformer solid insulation considering thermal stress and moisture variation. Electric Power Systems Research, 151, 1–11. https://doi.org/10.1016/j.epsr.2017.04.006

[72] Suleiman, A. A., Muhamad, N. A., Bashir, N., Murad, N. S., Arief, Y. Z., & Phung, B. T. (2014). Effect of moisture on breakdown voltage and structure of palm based insulation oils. IEEE Transactions on Dielectrics and Electrical Insulation, 21(5), 2119–2126. https://doi.org/10.1109/TDEI.2014.004431

[73] Liao, R., Lin Y., Guo, P. & Liu H. (2015). Thermal Aging Effects on the Moisture Equilibrium Curves of Mineral and Mixed Oil-paper Insulation Systems. IEEE Transactions on Dielectrics and Electrical Insulation, 22 (2), 842–850. https://doi.org/10.1109/TDEI.2014.004371

[74] Tu, Y., Chen, J., Wang, S., & Wang, W. (2016). Moisture migration in oil-impregnated film insulation under thermal ageing. IEEE Transactions on Dielectrics and Electrical Insulation, 23(2), 1135–1141. https://doi.org/10.1109/TDEI.2015.005406

[75] Villarroel, R., García de Burgos, B., & García, D. F. (2021). Moisture dynamics in natural-ester filled transformers. International Journal of Electrical Power and Energy Systems, 124. https://doi.org/10.1016/j.ijepes.2020.106172

[76] FRONTIN, S. O, et al. "*Equipamentos de Alta Tensão – Prospecção e Hierarquização de Inovações Tecnológicas*". 1ª ed. Brasília, 2013.

[77] *Manual de Controle Patrimonial do Setor Elétrico (MCPSE). Agência Nacional de Energia Elétrica.* Rev. 2. Brasil: ANEEL, 2014.

[78] Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., & Pourghasemi, H. R. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. Computers and Geosciences, 112, 23–37. https://doi.org/10.1016/j.cageo.2017.11.019

[79] Alsabhan, A. H., Singh, K., Sharma, A., Alam, S., Pandey, D. D., Rahman, S. A. S., Khursheed, A., & Munshi, F. M. (2022). Landslide susceptibility assessment in the Himalayan range based along Kasauli – Parwanoo road corridor using weight of evidence, information value, and frequency ratio. Journal of King Saud University - Science, 34(2). https://doi.org/10.1016/j.jksus.2021.101759

[80] Miao, F., Ruan, Q., Wu, Y., Qian, Z., Kong, Z., & Qin, Z. (2023). Landslide Dynamic Susceptibility Mapping Base on Machine Learning and the PS-InSAR Coupling Model. Remote Sensing, 15(22). https://doi.org/10.3390/rs15225427

[81] Niu, H., Shao, S., Gao, J., & Jing, H. (2024). Research on GIS-based information value model for landslide geological hazards prediction in soil-rock contact zone in southern Shaanxi. Physics and Chemistry of the Earth, 133. https://doi.org/10.1016/j.pce.2023.103515

[82] Yang, Y., Guo, Y., Chen, H., Tang, H., Li, M., Sun, A., & Bian, Y. (2024). Application of a Hybrid Model in Landslide Susceptibility Evaluation of the Western Tibet Plateau. Applied Sciences (Switzerland), 14(2). https://doi.org/10.3390/app14020485

[83] Zhang, J., Qian, J., Lu, Y., Li, X., & Song, Z. (2024). Study on Landslide Susceptibility Based on Multi-Model Coupling: A Case Study of Sichuan Province, China. Sustainability (Switzerland), 16(16). https://doi.org/10.3390/su16166803

[84] Wang, Y., Kang, L., & Wang, J. (2024). Landslide risk assessment combining kernel extreme learning machine and information value modeling-A case study of Jiaxian Country of loess plateau, China. Heliyon, 10(17). https://doi.org/10.1016/j.heliyon.2024.e37352

[85] Apu, S.I., Sharmili, N., Gazi, M.Y. et al. Remote Sensing and GIS-Based Landslide Susceptibility Mapping in a Hilly District of Bangladesh: A Comparison of Different Geospatial Models. J Indian Soc Remote Sens (2024). https://doi.org/10.1007/s12524-024-01988-x

[86] Bhandari, B. P., Dhakal, S., & Tsou, C. Y. (2024). Assessing the Prediction Accuracy of Frequency Ratio, Weight of Evidence, Shannon Entropy, and Information Value Methods for Landslide Susceptibility in the Siwalik Hills of Nepal. Sustainability (Switzerland), 16(5). https://doi.org/10.3390/su16052092

[87] Li, H., Hu, B. X., Lin, B., Zhu, S., Meng, F., & Li, Y. (2024). Collapse Susceptibility Assessment in Taihe Town Based on Convolutional Neural Network and Information Value Method. Water (Switzerland), 16(5). https://doi.org/10.3390/w16050709

[88] Ghasemzadeh, S., Maghsoudi, A., Yousefi, M., & Mihalasky, M. J. (2022). Information value-based geochemical anomaly modeling: A statistical index to generate enhanced geochemical signatures for mineral exploration targeting. Applied Geochemistry, 136. https://doi.org/10.1016/j.apgeochem.2021.105177

[89] Addis, A. (2023). GIS – based flood susceptibility mapping using frequency ratio and information value models in upper Abay river basin, Ethiopia. Natural Hazards Research, 3(2), 247–256. https://doi.org/10.1016/j.nhres.2023.02.003

APPENDIX – PRELIMINARY DESCRIPTIVE ANALYSIS

As an initial phase of data exploration, a preliminary descriptive analysis was conducted to identify patterns, trends, and possible correlations among variables. Scatter plots and contingency tables were used to visualize and understand data relationships and distributions. To achieve this, the *plot* and *crosstab* commands from the Pandas and Matplotlib libraries were used, respectively. Both tools help evaluate the relationship between two variables, providing insights into how different categories relate to each other.

Scatter plots

The scatter plots are shown in Figures A.1 to A.5. Samples exceeding the parameter thresholds were plotted for each variable, considering the oil analysis date.

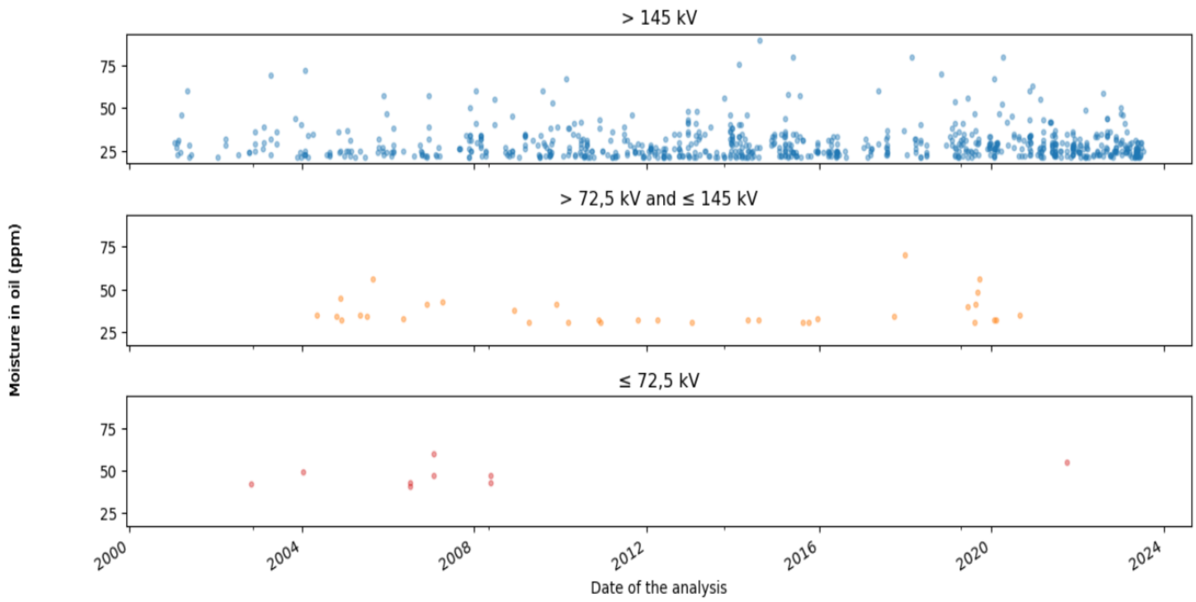


Figure A.1 – Scatter plots classified by voltage class range and considering oil moisture samples above the parameters (in ppm).

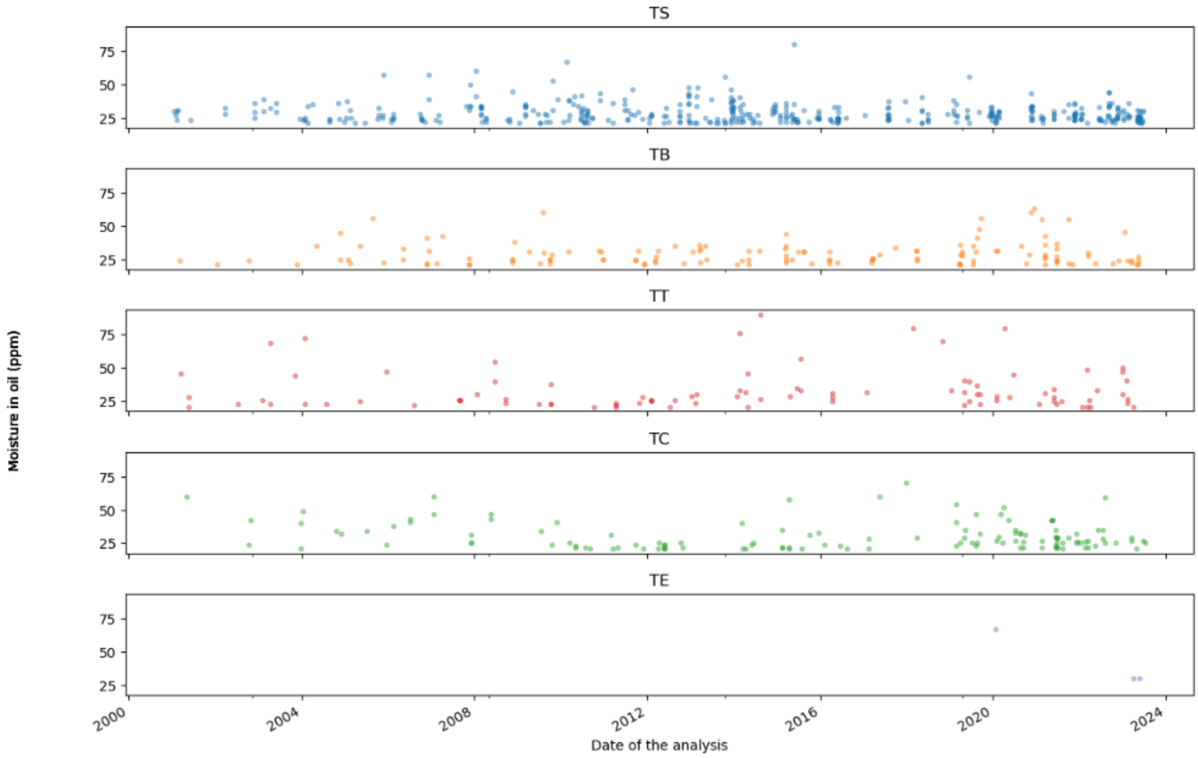


Figure A.2 – Scatter plots classified by Regional and considering oil moisture samples above the parameters (in ppm) [5].

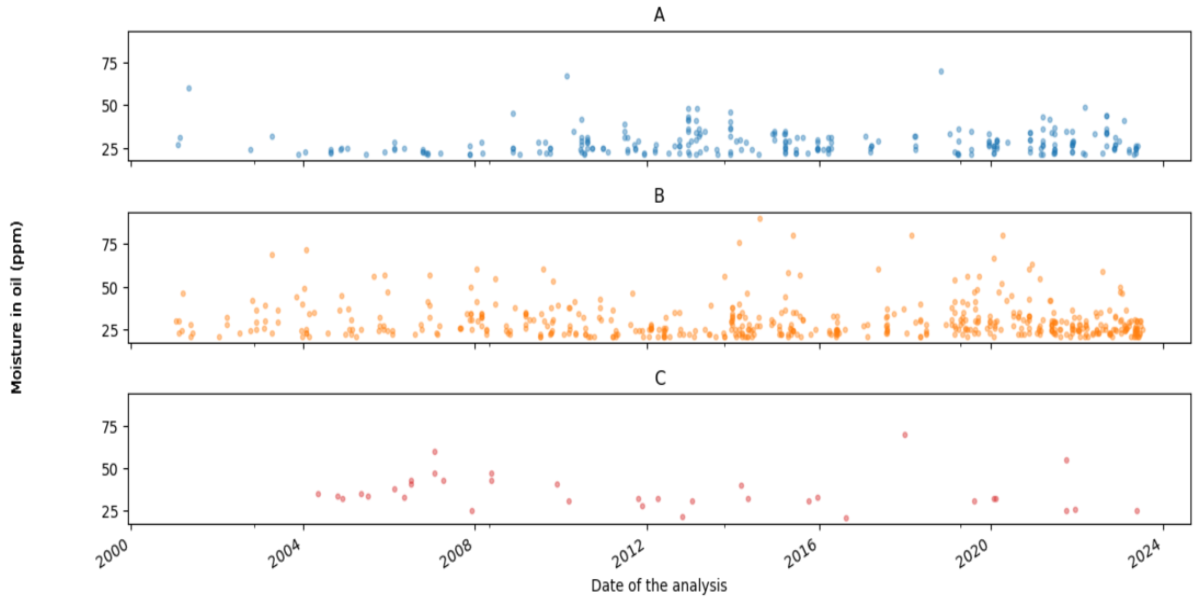


Figure A.3 – Scatter plots classified by criticality and considering oil moisture samples above the parameters (in ppm).

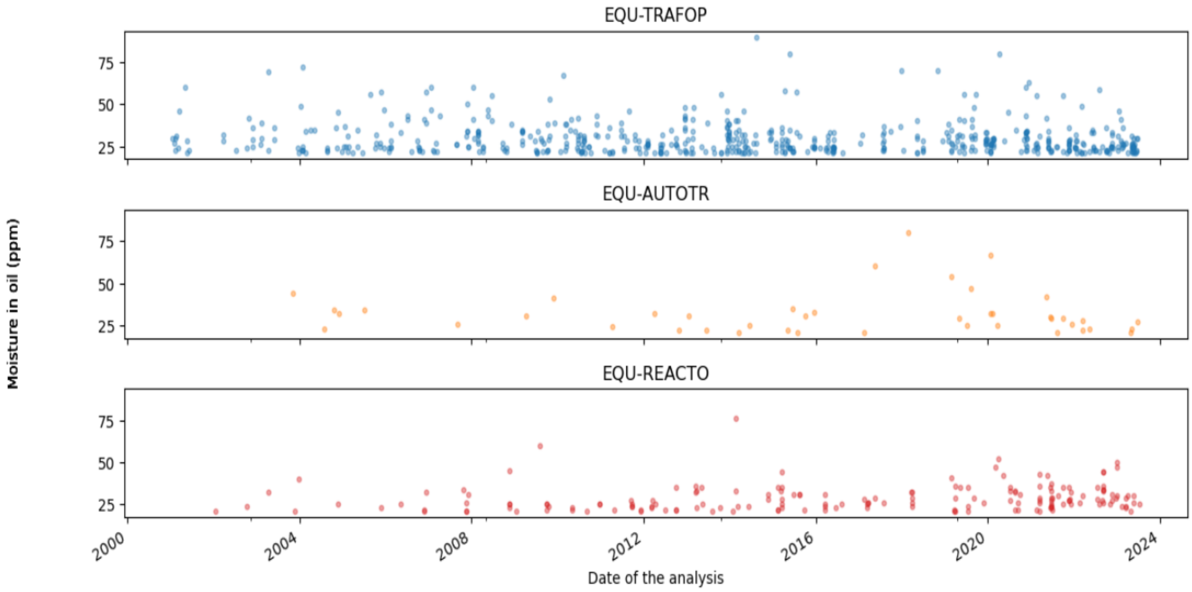


Figure A.4 – Scatter plots classified by type and considering oil moisture samples above the parameters (in ppm).

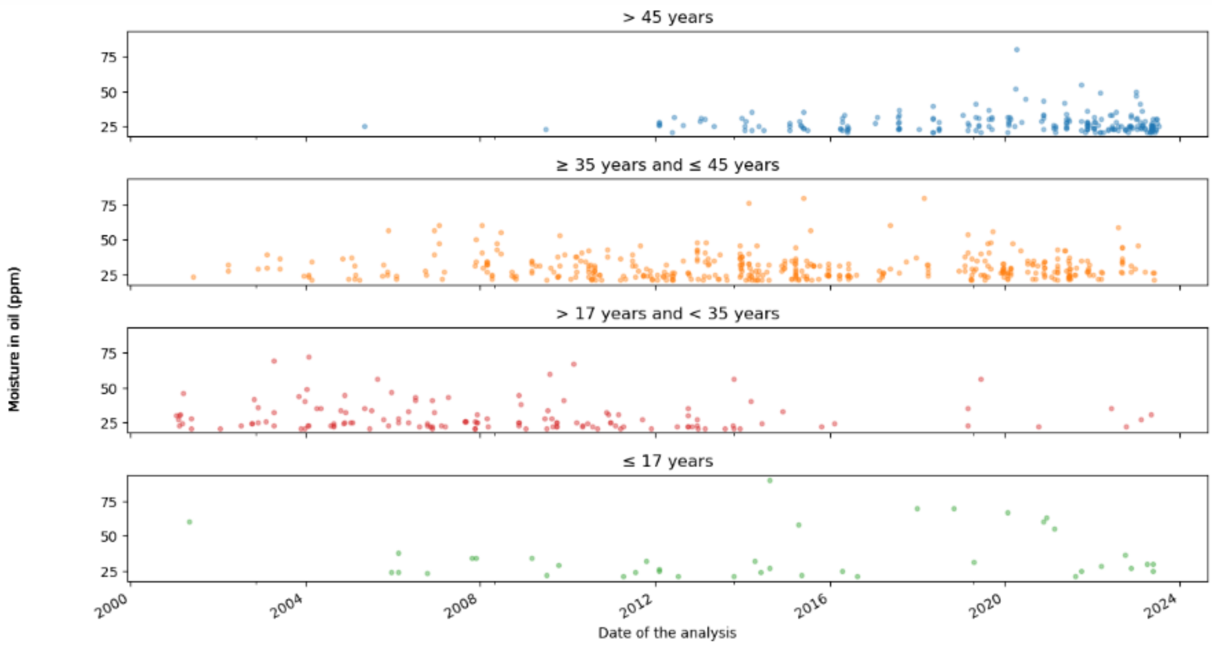


Figure A.5 – Scatter plots classified by age range and considering oil moisture samples above the parameters (in ppm).

These analyses illustrate the distribution of samples exceeding the parameter thresholds for each category of the variables, considering the analysis date.

For the variables voltage class range and Regional, the categories with the highest number of samples above the parameters are, respectively, >145 kV and TS, the same categories indicated by the WoE calculation as having the greatest weight of evidence. However, for the variables criticality, type, and age range, the categories with the highest number of samples above the parameters are, respectively, B, EQU-TRAFOP, and 35 ≤ age ≤ 45, which are different from the categories indicated by the WoE calculation as having the greatest weight of evidence.

This is because, according to Equation (1), WoE also considers the total number of samples above and below the parameters, not just scalar values, thus considering the complete data distribution and providing a more informative analysis.

Contingency tables

The contingency table is widely used for categorical data, presenting the frequency of occurrences of category combinations for two variables. Given that there are five categorical input variables, there are ten possible combinations. Tables A.1 to A.10 show the occurrence frequencies of the category combinations in percentage values, considering the entire dataset (df) and the dataset of samples above the thresholds (df_above).

Table A.1 – Contingency table relating, in percentage values (%), the voltage class range and Regional variables of (a) df and (b) df_above.

Regional / **Voltage class range**	**TB**	**TC**	**TE**	**TS**	**TT**
≤ 72.5 kV	7.39	6.25	0.00	0.00	2.15
> 72.5 kV and ≤ 145 kV	0.74	1.03	0.00	0.00	1.41
> 145 kV	17.14	18.25	1.48	29.44	14.74

(a)

Regional / **Voltage class range**	**TB**	**TC**	**TE**	**TS**	**TT**
≤ 72.5 kV	3.23	0.87	0.00	0.00	0.12
> 72.5 kV and ≤ 145 kV	0.12	1.00	0.00	0.00	0.00
> 145 kV	14.05	14.05	0.37	54.73	11.44

(b)

Table A.2 – Contingency table relating, in percentage values (%), the voltage class range and criticality variables of (a) df and (b) df_above.

Criticality / **Voltage class range**	**A**	**B**	**C**
≤ 72.5 kV	0.10	6.48	9.20
> 72.5 kV and ≤ 145 kV	0.00	0.70	2.48
> 145 kV	21.78	50.13	9.13

(a)

Criticality / **Voltage class range**	**A**	**B**	**C**
≤ 72.5 kV	0.00	1.87	2.36

Criticality / Voltage class range	A	B	C
> 72.5 kV and ≤ 145 kV	0.00	0.25	0.87
> 145 kV	31.72	61.82	1.12

(b)

Table A.3 – Contingency table relating, in percentage values (%), the voltage class range and type variables of (a) df and (b) df_above.

Type / Voltage class range	EQU-AUTOTR	EQU-REACTO	EQU-TRAFOP
≤ 72.5 kV	2.17	0.00	13.61
> 72.5 kV and ≤ 145 kV	0.00	0.00	3.17
> 145 kV	8.77	17.02	55.25

(a)

Type / Voltage class range	EQU-AUTOTR	EQU-REACTO	EQU-TRAFOP
≤ 72.5 kV	1.37	0.00	2.86
> 72.5 kV and ≤ 145 kV	0.00	0.00	1.12
> 145 kV	3.98	21.64	69.03

(b)

Table A.4 – Contingency table relating, in percentage values (%), the voltage class range and age range variables of (a) df and (b) df_above.

Age range / Voltage class range	≤ 17 years	> 17 years and < 35 years	≥ 35 years and ≤ 45 years	> 45 years
≤ 72.5 kV	0.58	1.00	0.80	0.80
> 72.5 kV and ≤ 145 kV	2.83	5.30	5.02	2.63
> 145 kV	28.22	24.08	20.01	8.72

(a)

Age range / Voltage class range	≤ 17 years	> 17 years and < 35 years	≥ 35 years and ≤ 45 years	> 45 years
≤ 72.5 kV	0.00	0.50	0.50	0.12
> 72.5 kV and ≤ 145 kV	0.25	1.74	1.87	0.37
> 145 kV	4.73	14.93	49.38	25.62

(b)

Table A.5 – Contingency table relating, in percentage values (%), the Regional and criticality variables of (a) df and (b) df_above.

Criticality / Regional	A	B	C
TB	3.80	16.24	5.22
TC	1.23	14.54	9.75
TE	0.00	1.12	0.36
TS	13.37	14.17	1.90
TT	3.48	11.24	3.58

(a)

Criticality / Regional	A	B	C
TB	9.20	6.59	1.62
TC	0.87	12.69	2.36
TE	0.00	0.37	0.00
TS	19.40	35.07	0.25
TT	2.24	9.20	0.12

(b)

Table A.6 – Contingency table relating, in percentage values (%), the Regional and type variables of (a) df and (b) df_above.

Type / Regional	EQU-AUTOTR	EQU-REACTO	EQU-TRAFOP
TB	3.44	7.84	13.98
TC	3.17	3.04	19.31
TE	0.23	0.83	0.42
TS	1.04	3.80	24.60
TT	3.06	1.51	13.73

(a)

Criticality / Regional	EQU-AUTOTR	EQU-REACTO	EQU-TRAFOP
TB	0.75	10.07	6.59
TC	3.36	5.72	6.84
TE	0.12	0.25	0.00
TS	0.37	4.98	49.38
TT	0.75	0.62	10.20

(b)

Table A.7 – Contingency table relating, in percentage values (%), the Regional and age range variables of (a) df and (b) df_above.

Age range / Regional	≤ 17 years	> 17 years and < 35 years	≥ 35 years and ≤ 45 years	> 45 years
TB	9.78	6.84	5.95	2.69
TC	7.83	8.52	6.70	2.48
TE	1.24	0.03	0.21	0.00
TS	6.20	10.64	8.92	3.68
TT	6.58	4.35	4.07	3.31

(a)

Age range / Regional	≤ 17 years	> 17 years and < 35 years	≥ 35 years and ≤ 45 years	> 45 years
TB	1.00	4.60	9.45	2.36
TC	1.00	3.36	8.96	2.61
TE	0.37	0.00	0.00	0.00
TS	1.49	6.97	29.73	16.54
TT	1.12	2.24	3.61	4.60

(b)

Table A.8 – Contingency table relating, in percentage values (%), the criticality and type variables of (a) df and (b) df_above.

Type / **Criticality**	**EQU-AUTOTR**	**EQU-REACTO**	**EQU-TRAFOP**
A	0.10	5.59	16.19
B	8.11	9.71	39.49
C	2.73	1.73	16.35

(a)

Type / **Criticality**	**EQU-AUTOTR**	**EQU-REACTO**	**EQU-TRAFOP**
A	0.00	12.19	19.53
B	3.86	9.33	50.75
C	1.49	0.12	2.74

(b)

Table A.9 – Contingency table relating, in percentage values (%), the criticality and age range variables of (a) df and (b) df_above.

Age range / **Criticality**	**≤ 17 years**	**> 17 years and < 35 years**	**≥ 35 years and ≤ 45 years**	**> 45 years**
A	3.86	10.43	6.02	1.57
B	19.47	14.18	15.71	7.95
C	8.29	5.77	4.11	2.63

(a)

Age range / **Criticality**	**≤ 17 years**	**> 17 years and < 35 years**	**≥ 35 years and ≤ 45 years**	**> 45 years**
A	0.87	7.09	19.53	4.23
B	3.36	8.58	30.72	21.27
C	0.75	1.49	1.49	0.62

(b)

Table A.10 – Contingency table relating, in percentage values (%), the criticality and age range variables of (a) df and (b) df_above.

Age range / **Type**	**≤ 17 years**	**> 17 years and < 35 years**	**≥ 35 years and ≤ 45 years**	**> 45 years**
EQU-AUTOTR	4.39	1.99	2.85	1.72
EQU-REACTO	6.31	5.53	4.41	0.78
EQU-TRAFOP	20.93	22.86	18.59	9.66

(a)

Age range / **Type**	**≤ 17 years**	**> 17 years and < 35 years**	**≥ 35 years and ≤ 45 years**	**> 45 years**
EQU-AUTOTR	0.50	0.87	2.36	1.62
EQU-REACTO	0.50	4.60	14.55	1.99
EQU-TRAFOP	3.98	11.69	34.83	22.51

(b)

Table A.11 presents the categories with the highest occurrence frequency for each combination of variables.

Table A.11 – Categories with the highest frequency of occurrence for each combination of variables.

Variables	**Combinations of categories with the highest frequency of occurrence**	
	df	**df_above**
Voltage class range and Regional	> 145 kV and TS	> 145 kV and TS
Voltage class range and criticality	> 145 kV and B	> 145 kV and B
Voltage class range and type	> 145 kV and EQU-TRAFOP	> 145 kV and EQU-TRAFOP
Voltage class range and age range	> 145 kV and ≤ 17 years	> 145 kV and ≥ 35 years and ≤ 45 years
Regional and criticality	TB and B	TS and B
Regional and type	TS and EQU-TRAFOP	TS and EQU-TRAFOP
Regional and age range	TS and > 17 years and < 35 years	TS and ≥ 35 years and ≤ 45 years
Criticality and type	B and EQU-TRAFOP	B and EQU-TRAFOP
Criticality and age range	B and ≤ 17 years	B and ≥ 35 years and ≤ 45 years
Type and age range	EQU-TRAFOP and > 17 years and < 35 years	EQU-TRAFOP and ≥ 35 years and ≤ 45 years

These categories represent the largest number of samples, considering the entire dataset (df) and the dataset of samples above the parameters (df_above).

Based on these analyses, an opportunity was identified to seek contribution through metrics that enable a more informative and assertive analysis of the predictive power of categorical variables.