



UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO - SALAMANCA
DIVISIÓN DE INGENIERÍAS

*“Short circuit detection in electrical transformers through
statistical analysis of vibration signals in the time and
frequency domains”*

TESIS

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN INGENIERÍA ELÉCTRICA

(Opción: Instrumentación y Sistemas Digitales)

PRESENTA:

Ing. Daniel Alejandro Zambrano Román

DIRECTORES:

Dr. David Camarena Martínez

Dr. Arturo García Pérez

Salamanca Gto., a 26 de septiembre del 2023.

M. en I. HERIBERTO GUTIÉRREZ MARTIN
COORDINADOR DE ASUNTOS ESCOLARES
PRESENTE.-

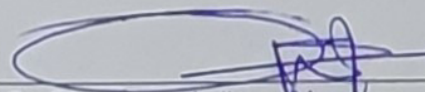
Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno(a) Daniel Alejandro Zambrano Román del *Programa de Maestría en Ingeniería Eléctrica (Instrumentación y Sistemas Digitales)* y cuyo número de NUA es: 826853 del cual soy director. El título de la tesis es: Short circuit detection in electrical transformers through statistical analysis of vibration signals in the time and frequency domains.

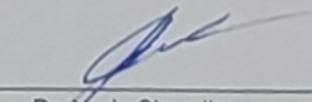
Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

ATENTAMENTE


Dr. David Camporena Martínez
DIRECTOR DE TESIS
SECRETARIO


Dr. Arturo García Pérez
DIRECTOR DE TESIS


Dr. Oscar Gerardo Ibarra Manzano
PRESIDENTE


Dr. Yuriy Shmally
VOCAL

Dedictory

To my parents, Mr. Marcos Herminio Zambrano Duque and Mrs. Maria Teresa Roman Pernia.

To my sister Marianela Zambrano Roman.

To my beloved Gledys Nathaly.

Acknowledgements

To Dr. David Camarena Martinez, for allowing me to return to a classroom and be part of this master's degree program

To Dr. Arturo Garcia Perez, for all the support during the investigation of this thesis, his advice, and his way of sharing knowledge

To Dr. Jesus Yañez, for his advice and lessons during the development of the chapters.

To Dr. Roberto Huerta, for the bits of advice he taught me related to the importance of this research.

To Fernando, Nacho, Oscar, Alberto, and Franklin, my colleagues in the telematics research laboratory, for their support and advice during the important moments of this research.

To Mexico and its people for allowing me to live and have great experiences in this beautiful country.

To those who supported me during the development of this project of thesis and the master's degree...thank you

Acknowledgements

I want to express my deepest gratitude to the University of Guanajuato, especially to the Irapuato - Salamanca Engineering campus (DICIS) for the academic development and the financial support received during my staying at this institution.



To the National Experimental University of Táchira (Venezuela) where I learned an enormous part of my academic and personal knowledge.



This work of thesis was developed thanks to the invaluable support received through the Consejo Nacional de Humanidades Ciencia y Tecnología, CONACYT from Mexico, Under scholarship number 826853 and CVU 1161100.



Contents

Dedicatory	ii
Acknowledgements	iii
Institutional Acknowledgements	iv
Contents	x
List of Figures	xi
List of Tables	xiv
Abstract	xiv
1 Introduction	1
1.1 Introduction	1

CONTENTS

1.2	Background	3
1.3	Justification	6
1.4	Objectives	7
1.4.1	General objective	7
1.4.2	Specific objectives	7
1.5	Organization of the thesis	7
2	Theoretical background	9
2.1	Electrical Transformer	9
2.1.1	Ideal transformer	10
2.1.2	Practical transformer	12
2.2	Faults in Electrical transformers	13
2.3	Techniques for conditions assessment in electrical transformers	14
2.3.1	Dissolved gas analysis	14
2.3.2	Partial discharge monitoring (PDM) systems	15
2.4	Signal analysis for conditions assessment in electrical transformers	16
2.4.1	Frequency response analysis	16
2.4.2	Vibration signals as a method for evaluating conditions in different applications	16
2.4.2.1	What is a vibration signal?	16
2.4.2.2	Vibration signals in electrical transformers	17
2.5	Feature extraction	18
2.6	Classifiers	19

CONTENTS

2.6.1	K Nearest Neighbors	20
2.6.2	Naive Bayes	21
2.6.3	Support vector machine	22
2.7	Evaluation of classifiers	24
2.7.1	Confusion matrix	24
2.7.2	Accuracy	25
3	Methodology	26
3.1	Signal acquisition	27
3.2	Frequency domain analysis	28
3.2.1	Decimation	29
3.2.2	Fast Fourier Transform	30
3.2.2.1	FFT base 2 algorithm	31
3.2.3	Statistical frequency features	32
3.2.4	Features extraction	33
3.3	Time domain analysis	35
3.3.1	Statistical time features	35
3.3.2	Features extraction	37
3.4	Data normalization	37
3.5	Feature selection	37
3.5.1	Kruskall - Wallis method	38
3.5.2	Principal Component Analysis	42

3.6	Classification	43
3.6.1	K Nearest Neighbors (K-NN)	43
3.6.2	Naive Bayes	44
3.6.3	Support Vector Machines	45
4	Results	48
4.1	Vibration signals dataset	49
4.2	Frequency domain analysis	50
4.2.1	Decimation of the signal	51
4.2.2	Spectral representation	51
4.2.3	Statistical frequency features	54
4.2.4	Features selection	58
4.2.4.1	Kruskall - Wallis method	58
4.2.5	Classification and fault identification	59
4.2.5.1	K-NN classifier results	60
4.2.5.2	Naive Bayes classifier results	60
4.2.5.3	SVM classifier results	61
4.2.6	Dimensionality reduction with PCA	62
4.2.7	Classification and fault identification	62
4.2.7.1	K-NN classifier results	63
4.2.7.2	Naive Bayes classifier results	63
4.2.7.3	SVM classifier results	64

CONTENTS

4.3	Time domain analysis	65
4.3.1	Statistical features extraction	65
4.3.2	Features selection	72
4.3.2.1	Kruskall Wallis test	72
4.3.3	Classification and fault identification	73
4.3.3.1	K-NN classifier results	74
4.3.3.2	Naive Bayes classifier results	75
4.3.3.3	SVM classifier results	75
4.3.4	Dimensionality reduction with PCA	76
4.3.5	Classification and fault identification	76
4.3.5.1	K-NN classifier results	77
4.3.5.2	Naive Bayes classifier results	77
4.3.5.3	SVM classifier results	78
4.4	Results discussion	79
5	Conclusion	83
5.1	Conclusion	83
	References	85

List of Figures

2.1	Representation of a transformer [Swafford, 2016]	10
2.2	Ideal transformer circuit diagram [Singuor, 2012]	12
2.3	Diagram of a practical transformer equivalent circuit [Kulkarni, 2013]	13
2.4	K-NN algorithm	21
2.5	Optimal hyperplane, i.e., the one with the maximum distance from the nearest training samples [Duda et al., 2001]	24
3.1	General scheme of the proposal	27
3.2	General scheme of the proposal	28
3.3	General scheme of the proposal	29
3.4	Complex plane representation of the roots of unity [Heckbert, 1995]	30
3.5	Butterfly diagram	32
3.6	General scheme of the proposal	35

LIST OF FIGURES

3.7	Features arranged for applying the method iteratively	40
3.8	Example of the same feature across all the conditions taken	41
3.9	Data matrix used as a parameter for the Kruskal - Wallis test	41
4.1	Vibration signal measured in the three axes for healthy condition	49
4.2	Vibration signal measured in the three axes for 30 SCTs condition	50
4.3	Segmentation of the signal in five portions	51
4.4	Triaxial vibration signal	51
4.5	FFT spectrum triaxial vibration signal	52
4.6	FFT vibration signal	53
4.7	FFT vibration signal	53
4.8	FFT vibration signal	54
4.9	FFT vibration signal 30 SCTs	54
4.10	Boxplot representation of all conditions features extracted axis X	55
4.11	Boxplot representation of all conditions features extracted axis Y	56
4.12	Boxplot representation of all conditions features extracted axis Z	57
4.13	Features selected with Kruskal - Wallis test	59
4.14	Confusion matrix K-NN classifier results	60
4.15	Confusion matrix Naive Bayes classifier results	61
4.16	Confusion matrix SVM classifier results	61
4.17	Subspace created after applying PCA	62
4.18	Confusion matrix K-NN classifier results	63

LIST OF FIGURES

4.19	Confusion matrix Naive Bayes classifier results	64
4.20	Confusion matrix SVM classifier results	64
4.21	Boxplot representation of all conditions features extracted axis X, Part A	66
4.22	Boxplot representation of all conditions features extracted axis X, Part B	67
4.23	Boxplot representation of all conditions features extracted axis Y, Part A	68
4.24	Boxplot representation of all conditions features extracted axis Y, Part B	69
4.25	Boxplot representation of all conditions features extracted axis Z, Part A	70
4.26	Boxplot representation of all conditions features extracted axis Z, Part B	71
4.27	Features selected with Kruskal - Wallis test	74
4.28	Confusion matrix K-NN classifier results	74
4.29	Confusion matrix Naive Bayes classifier results	75
4.30	Confusion matrix SVM classifier results	76
4.31	Subspace created after applying PCA	77
4.32	Confusion matrix K-NN classifier results	78
4.33	Confusion matrix Naive Bayes classifier results	78
4.34	Confusion matrix SVM classifier results	79
4.35	Classification results with machine learning models and Kruskal - Wallis	80
4.36	Classification results with machine learning models and Kruskal - Wallis	82

List of Tables

2.1	Confusion Matrix multiclass	25
3.1	Frequency domain features	33
3.2	Time domain features	36
4.1	Probabilities computed with Kruskal Wallis test	58
4.2	Index of the features sorted	59
4.3	Probabilities computed with Kruskal Wallis test	73
4.4	Index of the features sorted	73
4.5	Classification results with features selection in the frequency domain	80
4.6	Classification results with features selection in the time domain	81

Abstract

Electrical transformers are essential elements in industrial processes where the reliability and life span of these machines can be significantly improved by modeling a diagnosis strategy to monitor the working conditions in these devices. The short circuit turns (SCTs) in the windings significantly contribute to damage in electrical transformers; therefore, early detection during the initial stages is vital to prevent more extensive deterioration and schedule the appropriate maintenance. This work proposes a methodology based on the vibration signals analysis in the time and frequency domain to detect and classify a healthy state and several levels of SCTs. For both studies, the proposal consists of several steps. Starting with signal processing of the steady state, decimation, and computing of the Fast Fourier Transform (FFT) in the frequency domain, and the use of the raw steady state signal in the time domain. Followed by the extraction of statistical parameters. Subsequently, two techniques are used to select and reduce the features and keep those with the highest quality to distinguish the damage severity. Finally, three machine learning algorithms are trained and validated to classify the conditions under test. A comparative examination is made with the classification results achieved in each domain after applying the methodology to ensure a feasible and satisfactory assessment of the operating conditions of the transformer. Each step of the procedure was developed using mathematical software, and the results achieved show the effectiveness of this proposal in precisely identifying and classifying the severity of short circuit damage in the windings with the addition of low computational cost and fast processing time.

CHAPTER 1

Introduction

1.1 Introduction

Electrical transformers are essential in power systems distribution, enabling reliable and continued transmission of electrical power for different processes by increasing or decreasing the electrical voltage as it flows through the power grid and playing a crucial role in ensuring the availability of electricity in homes, businesses, industrial facilities, and other profitable organizations [AJ et al., 2018]. The main components of a transformer are an iron core, two or more coils called windings, and an insulating oil to reduce the heat generated during its operation time. These devices work on the principle that electromagnetic induction can transfer electrical energy from one circuit to the other; this indicates that when an alternating current flows through a coil, it generates a changing magnetic field within the core of the transformer. This changing magnetic field generates a fluctuating electromotive force (EMF) across any additional coils wound around the same core. This phenomenon allows the transfer of electrical energy between these distinct coils without needing a direct metallic (conductive) connection between the two circuits. With the adjustment of the number of turns of each coil, the voltage in the unit can be increased or decreased [Reclamation et al., 2011].

Transformers are robust machines that must be highly reliable during their operation life. However, despite their importance, they hold limitations, experiencing various types of failures in the internal components that can induce significant damage to the power grid and the surrounding equipment and even endanger human lives. Preventing such failures and maintaining transformers in optimal operating conditions is a critical concern for organizations, and a prevailing industry trend to move from conventional time-based maintenance programs to condition-based maintenance approaches requires methodologies and procedures capable of delivering quick and precise diagnoses [Wang et al., 2002].

Windings are one of the most vulnerable components in the transformer; therefore, they must rely on a high-quality insulation film to prevent electrical current from flowing between them (the coils) and the core; one of the most common failures in this component is the insulation failure, the gradually weakened dielectric strength of the insulation is caused by electrical stresses such as switching surges, lightning impulses, and frequent overloading. Another factor contributing to that degradation is the increased operating temperatures, which are related to heightened contact resistance, partial discharge, and issues with the cooling system. In addition, mechanical deformation in windings can occur due to short-circuit currents caused by high electrodynamic forces and transportation. The combination of all these stresses and problems, as well as moisture and contamination, accelerates the aging process of the insulation, leading to electrical failure at an expedited rate [Hu et al., 2019].

Numerous methods have been studied for fault detection and condition assessment in transformers, with good results but some drawbacks as a reduction in its operation and, in other cases, a complete shutdown of the devices. The analysis of vibration signals has gained attention as an emerging and effective tool for detecting internal faults in transformers in a non-invasive manner since the vibrations are generated primarily on the iron core and the windings. Changes in the mechanical properties of the windings will lead to changes in the vibration response, e.g., when an anomalous condition emerges [Huerta-Rosales et al., 2020a]. The processing of these signals uses distinct domains (e.g., time, frequency, and time-frequency domains) for evaluation in the operation conditions in transformers. A key element in employing vibration signals is extracting fault-sensitive features capable of discovering different levels of issues. The pattern recognition approach involves using different feature extraction, selection, and reduction combinations to classify and identify abnormalities from the acquired data. In literature, Statistical parameters (SP) are a popular set of features capable of extracting relevant signal properties or trends about the behavior of systems due to their sensitivity [Yanez-Borjas et al., 2020]. The extraction of SP in different domains is a powerful tool for pattern recognition in the conditions assessment of electrical transformers.

1.2 Background

Considerable studies have been executed in the past to predict failures in electrical transformers, e.g., [Dai et al., 2017] applied DGA to study the faults due to gases generated because of changes in the insulating oil, machine learning techniques were then used, in which a database was created for chromatographic analysis of the gases present in the oil; after all measurements were collected the effect of discharge current and superheat were also analyzed, finally found good classification results, applying a neural network. [Kim et al., 2005] applied frequency response analysis (FRA) in a three-phase transformer to identify a faulted phase (according to the author, the most common failure is the windings displacement) as an alternative to the dissolved gas analysis (DGA) technique, FRA generates responses of magnitude and phase of current and voltage signals which were then compared with other frequency responses used as patterns.

The phenomenon of inrush current during the start-up process in transformers is explained in [Gopika R, 2017]; the authors concluded that the phase-by-phase ignition allows attenuation of the current intensity since the current transient does not produce significant intensity. This is why they developed three methodologies as a solution for this controlled ignition. However, the disadvantage of this procedure is the high economic cost of implementation.

Signal processing is a practice used for assessing conditions in a system. This approach aims to extract and select features capable of determining even a small change in the operating conditions of that system. Among the signal processing methods, the time, frequency, and time-frequency domains are employed to observe pattern changes, and statistical parameters are a set of features qualified for interpreting the behavior of a system since a signal with different operating conditions will have different statistical criteria. One of the first vibration signals processing research in electrical transformers is found in [Garcia et al., 2006]; they developed a model for a monitoring system to estimate transformer tank vibration. The model calculates vibration on the transformer tank starting from input variables that can be measured in several parts of the transformer, including the tank vibrations. If the transformer were in a healthy state, it would show a good concordance between estimated and measured values. A big difference between estimated and measured vibration would appear in the case of vibrations due to a deforming winding vibration. The authors tested their experimental model successfully with the generation of deformations in the windings to validate their assumption.

[[Helmi and Forouzantabar, 2019](#)] employed vibration signals from an induction motor to analyze time and frequency domains. For the time domain, 15 statistical features were computed (STF) from the raw signal, and for the frequency domain, they employed one of the first techniques for signal processing in the frequency domain (the Fast Fourier Transform (FFT)). With the magnitude and frequency vectors from this calculation, the procedure was followed by extracting 13 statistical frequency features (SFFs). Then, 28 parameters with healthy and several faulty conditions were employed in a neural network, reaching an accuracy of 98.2 %. They emphasized that their method did not recognize any faults as healthy since the planned maintenance was essential to them. Another example of studying two domains for body signals is in [[Narayan, 2021](#)], where they compared a set of statistical features in the time domain (TD) and frequency domain (FD) by using LDA and ANN to classify and identify six different hand movements. Discrete Wavelet Transform was employed mainly for de-noising the sEMG signal before the feature extraction. Finally, a feature vector is formed, which consists of all TD and FD features for classification purposes. ANN exhibited 96.4 % accuracy and was found to be better than the LDA classifier, whereas the classification accuracy of the LDA classifier was found to be 94.5 %. The result exhibits that ANN had a more remarkable ability for sEMG signal classification than the LDA classifier. In their work, [[Yanez-Borjas et al., 2019](#)] presented a methodology for detecting structural damage in a bridge based on vibration analysis. This approach extracted statistical features from the time (10 parameters) and frequency domains (4 parameters). The method employed linear discriminant analysis and a neural network classifier to assess the health condition of the structure and several damage levels caused by corrosion. Overall, they proposed a methodology that integrated various techniques to effectively evaluate the structural integrity of civil infrastructure, reaching an accuracy of 93.8 %.

In the same way, [[Huerta-Rosales et al., 2021](#)] used a database of vibration signals from a single-phase transformer and simulated different levels of fault scenarios; statistical features in the time domain were extracted from the raw signals, and after applying optimization techniques, the best features were determined. PCA (Principal Component Analysis) and LDA (Linal Discrimator Analysis) were applied to reduce dimensionality; once the features were determined, SVM and ANN (Automatic Neural Network) were used as classification tools, finding that SVM delivered the best results, and finally implemented the classifier in FPGA.

Recently [[Rodríguez-López, 2022](#)] analyzed and used vibration signals and artificial intelligence to detect faults in transformers. The author used a database with different levels of short-circuits in the windings and employed more complex signal processing techniques such

as Wavelet, EMD (Empirical mode decomposition), EEMD (Ensembled EMD), and VMD (Variational Mode Decomposition) after the IMF was delivered, extracted features such as Entropy, Energy, and Kurtosis, building combinations between them. These combinations were input into a neural network. The results showed that VMD and the extraction of kurtosis and energy gave the best fault classification. However, the author pointed out the high computational cost of using these techniques. [Mejia-Barron et al., 2018] applied as well techniques such as EMD, EEMD, and CEEMD to analyze inrush current in transformers. The authors concluded that the IMF obtained by CEEMD gave the best decomposition results since the first method performed poorly and the second required a long processing time.

Another example of vibration signal processing can be found in [Roa-Terán, 2017]; in this case, for vibration signals captured from an induction motor. The study included processing techniques such as EMD, EEMD, and CEEMD (Complete EEMD) to extract the IMF (intrinsic mode function); with these IMF, computed statistical features to classify faults in the motor, among classifiers tested were Adaboost, SVM (Support Vector Machine) and Decision Tree, and the classification reached 100 % accuracy with the SVM classifier. [Contreras-Hernandez et al., 2019] is another example of this procedure, where quaternion signal analysis (QSA) was conducted. The authors segmented the signal in the time domain and, once they obtained the QSA, calculated statistics features of the segmented sample and a set with 10 types of faults for classification, 6 simple and 4 combined, using different algorithms such as LDA (Lineal discriminator analysis), Decision Tree, KNN (K nearest neighbor) and Convolutional Neural Networks LSTM (Long Short Term Memory) achieving 96 % accuracy with the KNN classifier and 4000 samples per data segment. Moreover, they emphasized accuracy and metrics such as precision and recall in discussing their results.

The NMD (No Lineal Mode Decomposition) technique utilized by [Huerta-Rosales et al., 2020b] is another example of signal processing methodology. The authors selected a group of features to detect faults in the windings, even in the early stages. Also, the RMS value was computed using the Hilbert transform (HT) proposed as a fault indicator, demonstrating sensitivity to fault severity. Finally, a fuzzy logic system was developed for automatic fault diagnosis. The authors highlighted that the technique is suitable for the employment of the transient and steady state of the signal for fault detection, and the method is considered to them as a low complex solution since three steps were required: the NMD technique, RMS computed through HT, and the automated diagnosis.

1.3 Justification

Electrical transformers are important power distribution devices for industrial processes and household electricity. Although these devices are robust machines, the possibility of damage or abnormality is quite significant and, consequently, interrupts the power supply. In addition, the detection of the issue is not a simple task. A transformer fault can occur in the iron core, the insulating oil, and the windings due to mechanical or electrical problems [AJ et al., 2018]. Therefore, locating and identifying these issues is a complex assignment, given the wide range of methods for evaluating conditions in the machine, and adding the extra inconvenience of an unscheduled shutdown, this could bring economic losses to the industrial process where the transformer is installed [Mishra et al., 2020].

Fault detection in electrical transformers in early or incipient stages has considerable importance. If the malfunction is not corrected or detected in a timely manner, the damage caused can be significant and affect the performance of the power distribution network or, in the worst case, affect human lives in the surrounding areas. Thus, implementing a system that detects faults in a non-invasive way, such as vibration signal analysis, with techniques that require low computational effort and extracting features with relevant information from those signals for the identification of the problem is a significant advantage, as it will allow the generation of the corrective maintenance, the reduction in the repairing cycles, a decrement in the probability of fault recurrence, and evade an expensive replacement of the complete machine.

In transformers, as in other applications, a healthy vibration will have different components in energy and frequency than a faulty state; depending on the environmental conditions of the experimental setup and the methodology for measuring the signals, the proposal for studying the vibrations may deliver that relevant information for assessing damage levels [Helmi and Forouzantabar, 2019].

Therefore, this research proposes a methodology to detect several levels of short circuit turns in the windings of a transformer as well as a healthy state for comparison purposes, with the analysis in the time and frequency domains applying signal processing techniques. For each domain, a set of statistical attributes is extracted, and the relevant ones are selected to classify and identify with machine-learning techniques the operating conditions of the transformer at the lowest possible computational cost.

1.4 Objectives

1.4.1 General objective

Detect short-circuit faults in the windings of electrical transformers by applying vibration signals processing techniques and classifying the damage severity through machine learning models.

1.4.2 Specific objectives

1. Review the state-of-the-art of recent approaches that apply vibration signals analysis for damage detection in electrical machines.
2. Develop a methodology employing a dataset of vibration signals that detect and classify the short-circuit turns in electrical transformers.
3. Study the vibration signals in the time and frequency domains and extract relevant parameters that allow a good separation between the operating conditions.
4. Implement machine learning models to classify the short-circuit turns in a timely manner and compare the achieved results of each domain.

1.5 Organization of the thesis

This thesis is divided into 4 chapters, which are briefly described below:

Chapter 1: Introduction. This chapter contains a brief overview of some background work on the matter under study, followed by the objectives and the justification for the motive of this investigation.

Chapter 2: Theoretical background. The concepts of electrical transformers and the faults in the device are exposed, followed by background techniques to detect faults and their results. As well as the vibration signal analysis approach as an alternative to detect and diagnose faults in the windings. Finally, the feature extraction and selection procedure and the machine learning techniques are described.

Chapter 3: Methodology. During the development of this chapter, it will be explained the method proposed for the study of the signals in the time and frequency domains, the features extraction and the method for selection, the machine learning models to perform classification tasks, and the accuracy to compare the classification results.

Chapter 4: Results. In this chapter, the results are deeply exposed and explained after applying the methodology.

Chapter 5: Conclusions. It presents the conclusions on the obtained results after applying the method.

Theoretical background

In this chapter, it is presented an overview of the basic theoretical concepts related to electrical transformers, including the ideal transformers and the equations that rule the model, followed by the voltage and current transformation relation. After that, the practical transformer is explained with the losses in the core and the flux. Followed by the faults in the device associated with the operation. In addition, a brief presentation of other approaches and techniques studied for conditions assessment in the machine. Last but not least, an explanation of vibration signals analysis as a low-cost and non-invasive method to evaluate mechanical changes in the windings and how those changes allow fault detection in the transformer.

2.1 Electrical Transformer

The electrical transformer is a simple, static electromagnetic device that works on the principle of Faraday's law of induction to convert electrical energy from one value to another [Calvert, 2006]. It consists of two or more coils of conducting wire, called windings, which are wrapped around a core of iron sheets. The windings are not directly connected but are linked by a magnetic flux inside the core. When one winding, the primary, is connected to a power

source, the second winding, the secondary, delivers power to the loads. Figure 2.1 shows a diagram of that transformer, showing the core with magnetic flux ϕ , the primary winding of N_p turns, and the secondary winding of N_s turns.

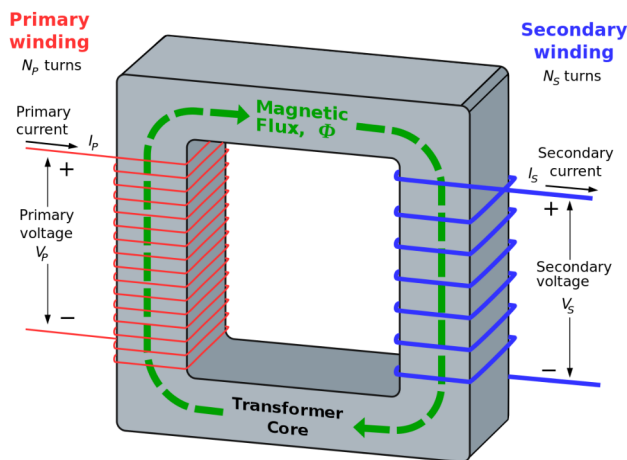


Figure 2.1. Representation of a transformer [Swafford, 2016]

2.1.1 Ideal transformer

The ideal transformer is linear, has no losses, and is perfectly coupled, so the energy transfer from the primary winding to the secondary winding is perfect. To be considered an ideal transformer, all the flux produced in the primary must link to the secondary winding without leaking out of the magnetic core. In an actual transformer, the windings are wound on top of each other, not on separate legs, to reduce leakage flux. In the usual shell-type transformer, both primary and secondary are on one leg and are surrounded by the core. A core-type transformer has windings covering the core legs. [Calvert, 2006].

When an alternating voltage is applied to the primary winding, a current will flow, setting a magnetomotive force, thus, an alternating flux in the core. This alternating flux linking both windings induces an electromagnetomotive force in each of them. In the primary winding, this is the back electromagnetomotive force, so if the transformer is ideal, it will oppose the primary applied voltage to the extent that no current would flow. But what happens is that the current that flows is the transformer magnetizing current [Heathcore, 1998].

The induced electromagnetomotive force is the secondary open-circuit voltage in the secondary winding. If a load is connected to the secondary winding, the flow of the secondary

current creates a demagnetizing magnetomotive force. Thus destroying the balance between the primary applied voltage and back electromagnetomotive force. An increased primary current must be drawn from the supply to restore the balance to provide a precisely equivalent magnetomotive force. So, that equilibrium is once more established when this additional primary current creates an ampere-turns balance with those of the secondary. Since there is no difference between the voltage induced in a single turn and whether it is part of the primary or the secondary winding, the total voltage induced in each winding by the common flux must be proportional to the number of turns [Heathcore, 1998].

The turns ratio is expressed with two numbers, like 2:1 or 2 to 1. The first number represents the relative number of turns of the primary, while the second number represents the relative number of turns of the secondary. The turns ratio of a transformer is calculated by applying the following mathematical expression:

$$\frac{N_p}{N_s} \tag{2.1}$$

where N_p is the number of turns in the primary winding, and N_s is the number of turns in the secondary winding.

Thus, the relation of transformation in terms of voltage and current is established by:

$$\frac{N_p}{N_s} = \frac{E_1}{E_2} \tag{2.2}$$

$$\frac{N_p}{N_s} = \frac{I_2}{I_1} \tag{2.3}$$

E, I, and N are the induced voltages, the currents, and the number of turns in the windings identified respectively by the appropriate subscripts. Hence, the voltage is proportional to the number of turns in the respective windings. The currents are in inverse proportion (and the relationship holds for both instantaneous and R.M.S. values) [Heathcore, 1998], figure 2.2 shows an ideal transformer circuit diagram, the primary winding of N_p turns, the secondary winding of N_s turns, the primary Voltage V_p , the secondary voltage V_s , the current in the primary I_p , and the current in the secondary I_s . .

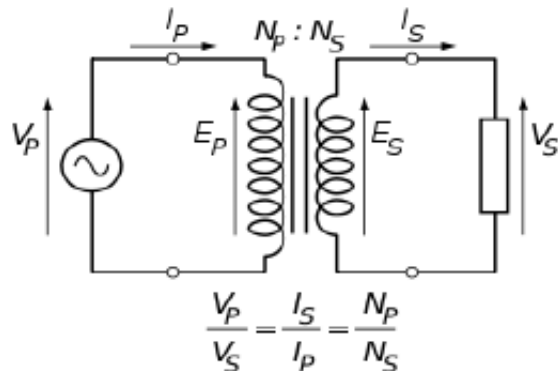


Figure 2.2. Ideal transformer circuit diagram [Singuor, 2012]

2.1.2 Practical transformer

The transformation relation in the device is not perfect. For this reason, the concept of a real transformer is the most accurate model for the losses inside the device that occur when part of the flux leaks through the core and causes hysteresis losses. All transformers have winding resistance, a core with finite permeability, leakage flux, hysteresis, and eddy current losses. These can be represented by an equivalent circuit, allowing an appropriate device analysis.

In the initial stages of transformer development, engineers considered leakage reactance a disadvantage that needed to be reduced to the greatest extent possible within reasonable economic limitations. However, as power stations, transmission networks, and distribution systems became larger and more intricate, the concept of leakage reactance, or more practically, impedance (including the resistance of transformer windings), began to be appreciated for its significant role in controlling fault currents [Heathcore, 1998].

The transformer is supposed to be a purely inductive device. But, as the primary and secondary coils are made of conducting materials with an insulating layer between them, it can be likened to a case of two conductors separated by a dielectric medium. This gives rise to capacitive effects. As these capacitances are unintended by design, they are called parasitic capacitances. For low-frequency transformers, parasitic capacitances do not play a significant role. However, capacitive effects become significant as the frequency increases, and with a high turns ratio, those effects start to dominate [Bath, 2022].

Figure 2.3 illustrates the schematic diagram of a practical transformer, which is

represented by an equivalent circuit. Within this circuit model, the ideal transformer is denoted by E_1 and E_2 . The circuit incorporates various elements, namely the primary winding Joule losses, R_1 and X_{L1} , as well as the secondary winding losses, R_2 and X_{L2} . Additionally, the model accounts for core losses and reactances through the shunt leg impedances: core or iron losses denoted as R_c , and magnetizing reactance represented by X_m . These losses and reactances primarily arise from hysteresis and eddy current effects within the core and are proportionate to the square of the core flux at a given frequency [Chapman, 2000]

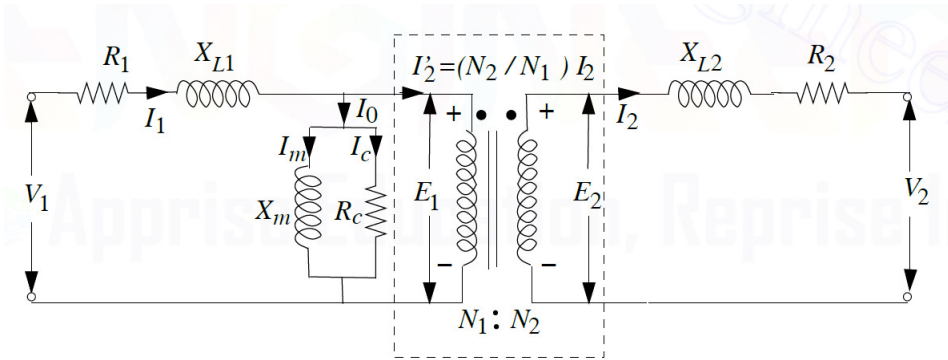


Figure 2.3. Diagram of a practical transformer equivalent circuit [Kulkarni, 2013]

2.2 Faults in Electrical transformers

Electrical transformers are a complex network of components prone to faults, which can be divided into three categories: electrical, mechanical, and thermal. In addition, each one can be categorized as an internal or external component type. Internal component failures include insulation deterioration, loss of winding clamping, overheating, moisture, solid contamination in the insulating oil, partial discharge, winding resonance, and design and manufacturing defects. External component failures commonly are lightning strikes, system switching operations, system overload, and system failures such as a short circuit [AJ et al., 2018].

These devices could be subjected to mechanical forces during transportation, short circuit events caused by electromagnetic forces, and inrush current during the ignition, which facilitates the degradation of the insulation, decreasing the life of the transformer -expected around 25 to 30 years according to the design- due to the dependence of the insulation on the expected performance of the device.

When a transformer is new in function, it has sufficient electrical and mechanical

strength to overcome unusual system conditions. As transformers age, their insulation strength can degrade, so they cannot resist events such as short-circuit faults or transient over-voltages [Wang et al., 2002]. The end of the life of a transformer may be determined by one factor or a combination of several factors. The principal cause of this event is related mainly to the thermal factor, where overheating is a challenge engineers face while proceeding with the maintenance of the device [Zeng et al., 2021].

Apart from normal aging, some events damage the transformer and lead to a situation where the aging process of the device is faster than expected, increasing the probability of an even worse abnormal situation that could reduce the expected life. Thermal aging of insulating materials for transformers is related to the chemical reactions in the materials. These chemical reactions are caused by pyrolysis, oxidation, and hydrolysis and are accelerated by increased temperature, oxygen, and humidity levels. Associated with the chemical reaction of the cellulose paper, there is a reduction in the mechanical properties. The insulation paper becomes brittle until it almost falls apart but still retains an acceptable level of dielectric strength [Wang et al., 2002].

Furthermore, regular overloading and short-circuit incidents on aged transformers may lead to unexpected premature faults and damage due to power supply interruption. Moreover, the faults of transformers can damage the environment through oil leakages. It could be dangerous to the personnel by creating fires and explosions, resulting in costly repairs, significant revenue losses, and damage to the human lives in the surrounding area [Islam, 2018].

2.3 Techniques for conditions assessment in electrical transformers

2.3.1 Dissolved gas analysis

The Dissolved Gas Analysis (DGA) method is widely employed for online incipient fault diagnosis; the effective use of the method requires routine oil sampling and modern online gas monitoring technologies. Abnormal electrical or thermal behavior in insulation oil can release small quantities of gases, and the gas composition depends on the specific fault type. Detecting certain gas levels in oil-filled transformers during service often serves as the first indication of a malfunction, which, if left unaddressed, may result in faults in the transformer.

Gasses can be generated due to various mechanisms, including arcing, partial discharge, low-energy sparking, and insulation overheating due to severe overloading. Fault identification in oil-filled transformers is based on the types of gases produced, which usually include combustible gases such as hydrogen (H_2), methane (CH_4), ethylene (C_2H_4), ethane (C_2H_6), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2).

DGA interpretation schemes rely on empirical assumptions and practical knowledge gathered by experts globally. These schemes are typically grounded in specific principles, such as gas concentrations, key gases and ratios, and graphical representations. Commonly used schemes include Key Gas Analysis, Dornenberg and Rogers Ratio Methods, Nomograph, IEC Ratio, Duval Triangle, and CIGRE Method. These approaches assist in fault diagnosis, ensuring the continued function and longevity of transformer systems [Sun et al., 2012].

2.3.2 Partial discharge monitoring (PDM) systems

Partial discharge (PD) occurs within a transformer when the electric field surpasses the local dielectric strength of the insulation. Several potential reasons for PD occurrence include insulation damage resulting from overvoltages and lightning strikes, incipient weaknesses due to manufacturing defects, or gradual deterioration due to natural aging processes. PD may appear insignificant initially, but its destructive nature begins with chemical decomposition and material erosion. If left unattended, the affected area can expand, ultimately posing a significant risk of electrical breakdown [Judd et al., 2005].

Monitoring PD in electrical systems is crucial for the early detection of potential issues and for preventing catastrophic failures. Several methods are employed to monitor PD, each with advantages and limitations. Some standard monitoring techniques include acoustic emission, electromagnetic interference monitoring, and electrical, chemical, and optical detection. Most modern PD monitoring systems can provide real-time data and can be integrated into overall condition monitoring platforms. By continuously monitoring partial discharges, operators can take proactive measures to address potential issues, schedule maintenance, and prevent unplanned downtime and costly equipment failures. Regular monitoring can also extend the operational life of electrical assets with condition-based maintenance instead of time-based maintenance and improve overall system reliability and safety [Hussain et al., 2021].

2.4 Signal analysis for conditions assessment in electrical transformers

2.4.1 Frequency response analysis

Frequency response analysis is an off-line comparative-based diagnostic test that, due to its high sensitivity, has the potential to detect faults that other condition assessment techniques may not be able to find. A signal is injected into one of the terminals of the transformer, and the response gathered is measured at the other terminal to perform an FRA measurement [Abu-Siada et al., 2013]. This signal, which can be a sweep frequency, an impulse signal, or a step voltage, is measured to the ground (the tank of the transformer) and used as the reference signal for FRA calculation; this reference signal could be a low voltage broadband impulse signal (IFRA) or a sinusoidal voltage in a wide frequency range (SFRA).

The frequency response magnitude is the ratio of the reference and response signals and is often represented in dB. In the same way, the frequency response phase angle is the phase angle shift of the response relative to the input signal. Any abnormalities in these diagrams with respect to former times indicate the presence of changes or failures in the transformer [Kim et al., 2005].

2.4.2 Vibration signals as a method for evaluating conditions in different applications

2.4.2.1 What is a vibration signal?

Vibration is a mechanical event in which oscillations occur about an equilibrium point, and the time series that carries the information of those oscillations is called a vibration signal. These oscillations from the equilibrium point must be acquired at a high sampling rate. These signals are categorized as steady or stationary and non-steady (mostly).

A stationary signal can be denoted by a sine wave with a constant period if its frequency or spectral contents are not changing to time. In contrast, a non-stationary signal would have a sine wave with varying periods because the frequency in the non-stationary signal varies with time [khushnandan, 2021].

Although good results have been obtained with above mentioned traditional methods, vibration analysis has attracted particular attention since it is an effective way to study the performance or deterioration of various applications, such as induction motors, civil structures, and electrical transformers, because the vibration signals are directly correlated with the mechanical changes in the windings.

The diagnosis of electrical transformers using vibration signal techniques can be divided into three groups: signal-based, model-based, and knowledge-based. Signal-based algorithms usually extract information from the vibration signal, e.g., [Mejia-Barron et al., 2018] applied signal processing techniques, such as CEEMD, then, after the decomposition was performed, extracted functional indices from inrush current signals and deduced the fault severity of a transformer. Model-based algorithms establish a conjunction between vibrations and input parameters, such as current, voltage, and temperature, e.g., [Garcia et al., 2006] proposed a tank vibration model, which used the fundamental component. Knowledge-based algorithms focus on pattern recognition and decision making, e.g., [Wu et al., 2018] compared the vibration of normal and DC-biased conditions of transformers, using least square SVM classifier to recognize DC-biased conditions; their suggested methodology helped target DC.

2.4.2.2 Vibration signals in electrical transformers

The primary source of vibrations in a transformer are the core and windings during the operation. In the core, the magnetostriction phenomenon and magnetic forces produce vibrations; these forces change the shape of a ferromagnetic material under a magnetic field. After a voltage is applied, the forces are induced into the core; hence, the fundamental frequency of core vibration is double the transformer excitation voltage frequency (60 Hz in Mexico), giving 120 Hz of frequency voltage harmonics and inducing others at high-frequency and random magnitudes. The magnitude of the transformer core vibration F_c is proportional to the square of the excitation voltage (V) as seen in (2.4) [Garcia et al., 2006, Huerta-Rosales et al., 2021].

$$F_c \propto V^2 \tag{2.4}$$

Vibrations in the windings are produced by electromagnetic forces caused by the interaction of the current flowing in the windings and the magnetic leakage flux. Axial forces act by compressing and creating vert deformations in the windings, making these

forces proportional to the square of the current. Therefore, the vibration acceleration F_w is proportional to the square of the current (I), as shown in (2.5). Likewise, the main acceleration component in the core is twice the fundamental excitation voltage frequency; the current also creates harmonics at high frequencies with random magnitudes.

$$F_w \propto I^2 \tag{2.5}$$

Therefore, the current increases considerably when short-circuited windings are in the device. Consequently, the acting forces cause higher deformations and higher magnitude vibrations on these windings, notoriously different from a transformer in a healthy state [Rodríguez-López, 2022].

2.5 Feature extraction

Analysis of the vibration signals generated by the transformer allows identification and diagnosis of abnormal operating conditions. Voltage, current, and vibration signals are the most effective indicators for assessing the condition of the machine.

By examining these signals, deviations from normal behavior can be detected, providing valuable insight into transformer performance and condition. Vibration signals provide essential indications of mechanical stresses and potential faults in the transformer. The problems and behavior of the device can be modeled by studying and using certain features that contain crucial information from the acquired signal.

Feature extraction converts the raw data of a signal into practical numerical attributes that, once calculated and processed, preserve the necessary information of the data and provide better results for machine learning than applying that raw data. It is commonly used in various fields, including image and video processing and natural language processing, to enhance the performance and efficiency of data analysis and pattern recognition algorithms. Having these features arranged in a feature set, which usually is a matrix where rows are every observation, and the columns are the features calculated, is possible to reduce the computational cost and resources necessary for pattern recognition by applying well-studied methodologies that can help with the diagnose of issues in applications, however, to avoid the risk of losing potentially necessary information from the signal, it is not always possible to improve the computational

processes when there is the high dimensionality in the features [Kang et al., 2016].

Vibration signals can be studied with several approaches; recently, many works have focused on time and frequency domain analysis to explain and extract relevant information about the signals; in the frequency domain, the Fast Fourier transform (FFT) can disclose information that cannot be discovered in the time domain [Helmi and Forouzantabar, 2019], the initial analysis of the frequency spectrum after applying the FFT shows that as the size of the defect increases, the magnitude of the components increases. New components will appear at high frequencies. In the frequency and time domains, statistical features (SFs) are used to find relevant properties of the phenomenon under study since vibration signals of distinct operating conditions will have different statistical behavior; this allows for differentiating a healthy from a faulty condition. Some advantages of SFs are that they require less mathematical operations than other methods and can deliver results almost in real-time, at the same time, can be used for non-stationary signals, such as vibration, electroencephalogram, rotatory machines, and civil structures [Yanez-Borjas et al., 2019].

2.6 Classifiers

Machine learning is a subarea of the artificial intelligence field that allows computers to learn without being explicitly programmed. This exceptional capability enables computers to distinguish complex patterns within specific data by leveraging algorithms and data. Machine learning finds diverse applications across various disciplines, including image and speech recognition, natural language processing, autonomous vehicles, and more.

The techniques are divided into three subcategories: supervised, unsupervised, and reinforcement. Classification algorithms are supervised machine learning models trained with labeled data sets so that the models learn and become more accurate over time. For example, an algorithm would be trained with images of dogs and other things labeled by humans, and the machine would learn to identify images of dogs on its own. Supervised machine learning is the most common type of model used today.

In this work, supervised machine learning techniques are used with the SFs to detect faults in electrical transformers, with three main classifiers: KNN, Naive Bayes, and Support vector machine (SVM).

2.6.1 K Nearest Neighbors

K nearest neighbors (KNN) is a non-parametric algorithm, which means it does not make any assumption on the underlying data. This type of algorithm is used for binary or multiclass classification, as well as for regression [Mirbozorgi, 2020], with the object of this work being employed as an instance-based classifier, where a sample is classified into a category according to its nearest neighbors. It is also known as a lazy learner algorithm since it does not immediately learn from the training set. Instead, it stores the data set and acts on it at classification time, i.e., when it receives a new data sample, it classifies that sample into a category that is very similar to the new data [Shah, 2022].

The number of neighbors (K) is chosen in the first step. Then, the nearest neighbors are identified based on the distance of the new point to those neighbors (the distance metric can be Euclidean, Mahalanobis, cosine similarity, etc.). Equation (2.6) shows the Euclidean distance, a prevalent measure implemented for this technique. Among these K neighbors, it is counted the number of data points in each category and assigned the new data points to that category for which the number of the neighbors is maximum.

$$d = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (2.6)$$

A good selection of the value for the K is necessary to avoid problems when the model is under test [Moldagulova and Sulaiman, 2017]. A low K value leads to noise problems and adverse effects on the outcome. While a large K value is good, it sometimes leads to difficulties, as the computational calculation is complex and time-consuming [Imandoust and Bolandraftar, 2013]. There are no pre-defined statistical methods to find the most favorable K value, and two methods are a good starting point to find the optimal K value; the square root of the number of training samples per class 2.7 [Zhang et al., 2018] and base two logarithm of the number of training samples per class 2.8, and if the classes are unbalanced, the value of K is taken as the one obtained with the class with the lowest number of samples.

$$K = \sqrt{N} \quad (2.7)$$

$$K = \log_2(N) \quad (2.8)$$

Figure 2.4 shows an example of the K-NN algorithm with five neighbors, starting with a new observation placed close to the training data points; the distance from that sample is

measured with respect to the training points, the K (5) smallest distances are sorted, and since 3 of the 5 smallest distances correspond to the triangles class, the majority decision place the new observation in that class.

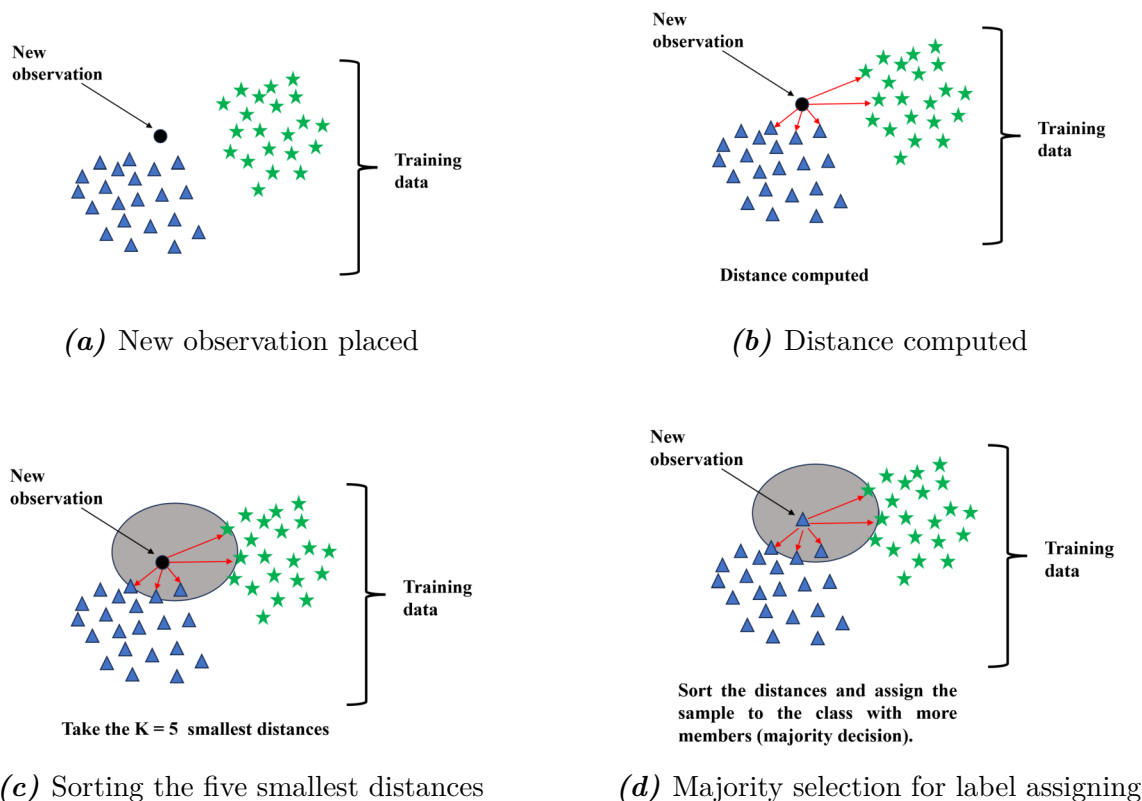


Figure 2.4. K-NN algorithm

2.6.2 Naive Bayes

The naive Bayes classifier is a collection of algorithms for machine-learning models based on the probabilistic theorem of Bayes. The term naive implies the assumption of independence of the features between them and an equal contribution from them to the outcome [Murphy et al., 2006]. This theorem shows the probability of an event occurring given the knowledge before the event; equation 2.9 is a general representation of that theorem, where $P(A)$ is the probability that A occurs; $P(B)$ is the probability that B occurs; $P(A|B)$ is the probability that A occurs given B; likewise $P(B|A)$ is the probability that B occurs given A.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2.9)$$

Multinomial Naive Bayes is one of the variants of the Naive Bayes algorithms, which is quite helpful in large datasets with multiple labels [Kharwal, 2021]. On the one hand, the advantage of this classifier is the ease, fast implementation, and good performance. Still, on the other hand, the most significant disadvantage is that it requires the predictors to be independent. In real-life problems, this is not a very common scenario and hinders the performance of the model.

2.6.3 Support vector machine

Support Vector Machines (SVM) are a group of supervised learning algorithms developed by Vladimir Vapnik at the AT &T Bell laboratories [Vapnik, 1995] to solve regression and classification problems. The main goal of the model is to find a hyperplane of separation with the highest margin representing a boundary between data groups or classes. The technique can solve linear problems efficiently, and non-linear and non-separable problems are solved by transforming them into separable problems where no linear characteristic remains. Kernel functions are used to smooth this process. Some kernel functions are linear, polynomial, sigmoid, and Gaussian, the last one the most popular for other authors [Kari et al., 2018].

It finds many applications in binary classification but can also be employed in multiclass classification. The mechanism by which this classification is achieved is the one-against-other, with two approaches. Each classifier is placed in series in the first case, while in the second case, the one-against-other method is put in parallel [Huerta-Rosales et al., 2021].

Given a set of n point of the form $[(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)]$, where y_n represents the label of which every x_n belongs, the algorithm will find the maximum hyperplane that divides the points with its corresponding class, the equation 2.10 is used for SVM to find that hyperplane of separation, where \mathbf{x} is the data points, \mathbf{w} is the vector of wights and \mathbf{b} is the bias

$$w^T x + b = 0 \tag{2.10}$$

And if y_i have two possible outcomes (-1,1), the equation for the support vectors of each class is given by 2.11:

$$\begin{aligned}w^T x + b &\geq 1, \forall x \in w_1 \\w^T x + b &\leq -1, \forall x \in w_2\end{aligned}\tag{2.11}$$

where w_1 and w_2 correspond to each class, to find the optimal hyperplane, is necessary to solve the equation 2.12 of the quadratic problem minimization and subject to 2.13:

$$J(w) = \frac{1}{2} \| w \|^2\tag{2.12}$$

$$y_i(w^T x_i + b) \geq 1\tag{2.13}$$

and the final decision is obtained as follows:

$$w = \sum_{i=1}^N y_i \alpha_i x_i\tag{2.14}$$

the terms of equation 2.14 α_i (Lagrange multipliers), $y_i = \pm 1$ (class indicator for each training x_i) are know as a support vector machine [Vapnik, 1995].

Figure 2.5 is a representation of training data that are linearly separable. The support vectors that are (uniformly) close to the optimal hyperplane separate the two classes of data; the area bounded by these two hyperplanes is called the "margin," which is the most significant separation between classes.

Having normalized data, these planes can be described by equations 2.15

$$\begin{aligned}w^T x + b &= 1 \\w^T x + b &= -1\end{aligned}\tag{2.15}$$

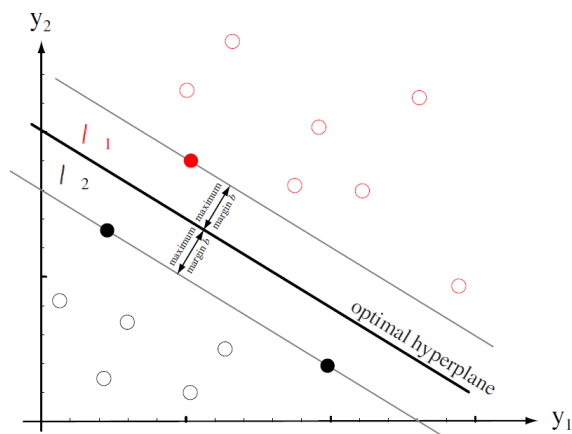


Figure 2.5. Optimal hyperplane, i.e., the one with the maximum distance from the nearest training samples [Duda et al., 2001]

2.7 Evaluation of classifiers

Once the features are extracted and selected, the classification models are trained and tested, delivering results; it is time to interpret them. Classifiers are usually evaluated using a variety of approaches, including numerical metrics such as accuracy or a graphical representation of performance such as a Receiver Operating Characteristic (ROC) curve. In the case of multiclass classification, F1 score, Precision, Recall, and Sensitivity are very useful for better interpretation of performance as well as for comparison of models [Ali et al., 2022].

2.7.1 Confusion matrix

It is a table that summarizes how the predictions of the classification model are performed. It is a correlation matrix between the actual labels and the classification of the model (predicted labels). The confusion Matrix results fall into one of four categories for binary classification problems. The main diagonal, true positive (TP) and true negative (TN), represents the correct classification of the model. False positives (FP) and false negatives (FN) denote the number of incorrectly classified instances [Ali et al., 2022]. In the case of a multiclass problem (which is the case in this thesis work), the positive class is the label for which the computation is performed, the negative class is the remaining label, and the table form corresponds to the number of classes $N \times N$. Table 2.1 is a representation of a multiclass confusion matrix where the main diagonal represents the correct classification.

Table 2.1. Confusion Matrix multiclass

		Predicted		
		A	B	C
Actual	A	T		
	B		T	
	C			T

2.7.2 Accuracy

The percentage of correctly predicted classes (the main diagonal) out of the total data samples. This is the simplest and most straightforward performance measure [Ali et al., 2022]. However, the accuracy is only valid if the class distribution is symmetric, i.e., if the number of data samples in each class is the same. Equation 2.16 computes the accuracy of the model:

$$Accuracy = \frac{TP + TN}{FP + TP + FN + TN} \quad (2.16)$$

CHAPTER 3

Methodology

Fault detection in electrical transformers is continuously studied for management procedures and control in industrial facilities. An adequate methodology for the assessment of the conditions could avoid severe damage and the increment in repair costs; another advantage is the reduction of scheduled stops for preventive maintenance with an extension of the device service life. To develop a fault detection procedure, it is required to have a detailed study of the configuration of the transformer (power, number of turns, etc.) and the operating conditions of the device, where non-invasive and non-stop method like vibration signal analysis is suitable for evaluating those conditions and detecting device damage.

A methodology is proposed for this investigation to identify short-circuit turns using vibration signals and several conditions simulated in an electrical transformer. For detecting the short-circuit faults, two approaches will be studied with the vibration signals; the first is an analysis in the frequency domain with the Fast Fourier Transform (FFT), and the second is in the time domain of the raw steady state signal. In both approaches, statistical features (SFs) will be computed to observe changes in the patterns of the signals since an abnormal condition would lead to changes in the SFs. Once the calculation is completed, feature selection and dimensionality reduction techniques will be tested, and machine learning models will be employed to detect and classify short-circuit turns (SCTs) conditions in the windings. At the

same time, the performance of the classification models will be tested through the computation of the accuracy. Finally, the results of both proposals will be compared to comprehend the advantages and drawbacks each domain could deliver. Figure 3.1 shows a general scheme of the proposal, where each domain analysis side will be further explained.

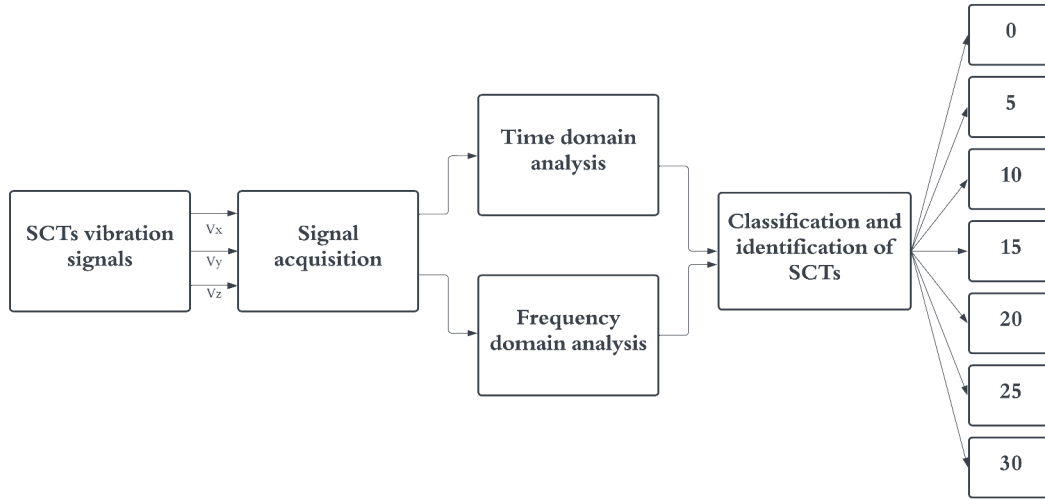


Figure 3.1. General scheme of the proposal

3.1 Signal acquisition

The vibration signals were acquired in the systems and electrical equipment laboratory of the Autonomous University of Queretaro (UAQ); the transformer used is a single-phase of 1.5 kVa with 135 turns in its primary winding that operates at 120 V and 60 Hz (operation frequency of Mexico) without load during its energization; the device was adjusted to emulate the SCTs faults. The experimental setup includes an autotransformer, two solid-state relays SAP4050D, a data acquisition system (DAC), the transformer under test, an accelerometer, and a PC. Figure 3.2 shows the scheme for signal acquisition. The autotransformer is used with one of the relays model SAP4050D to energize the testing transformer; this is done to demagnetize the core at the end of every test; the SCTs tests are conducted with the other relay. For the acquisition, a triaxial accelerometer model 8395A from KISTLER with a frequency sampling of 6000 Hz and a bandwidth from 0 to 1000 Hz is used to capture signals in three axes A_x, A_y, A_z . The selected location is on the clamping frame of the transformer since it can receive the vibrations of the core and windings symmetrically. The data acquisition system (DAS) is the National Instruments NI-USB6211 board, which has a 16-bit analog-to-digital

converter configured with a sampling frequency of 6000 samples/s.

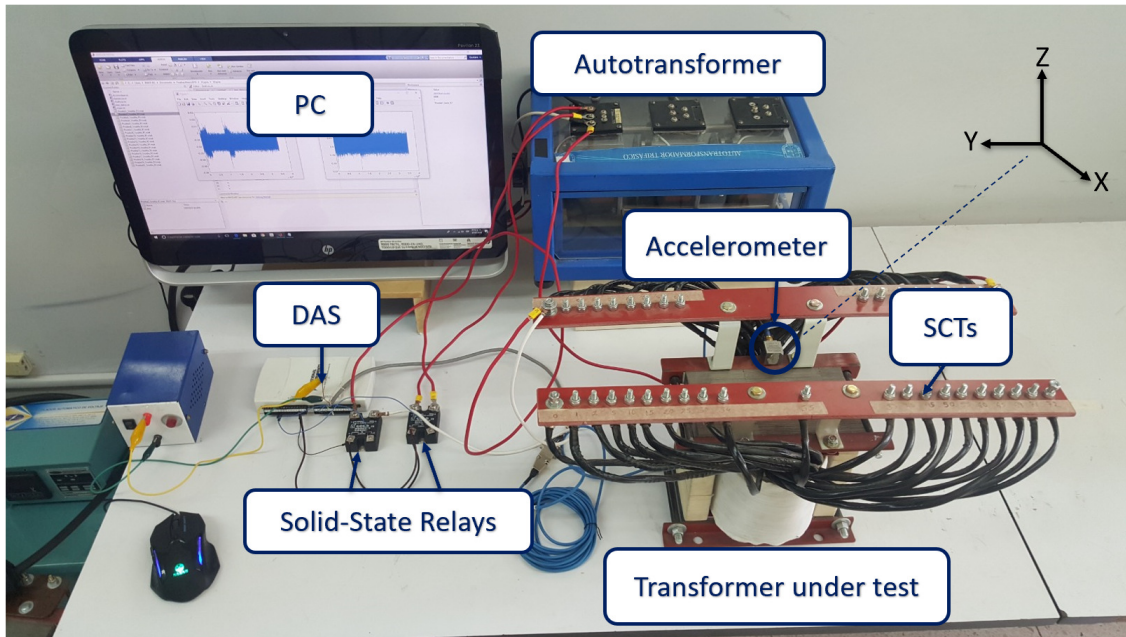


Figure 3.2. General scheme of the proposal

For the simulation of the short-circuit turns (SCTs) in the transformer, the method included a healthy state or complete normal operation and six levels of short-circuit in the windings; the SCTs were taken starting in 5 and finishing in 30 SCTs, the increment is of 5 windings per step. Hence, the conditions applied for this work are 0, 5, 10, 15, 20, 25, and 30 (SCTs) windings. All the signals acquired have a transient and steady state. The conditions were taken in that manner, as it was the most convenient layout to make the necessary connections to take the short circuits; in figure 3.2, the strip at the top of the transformer under test shows the terminals where the connections were made. If the transformer allows it, the short circuits can be simulated in any increasing step.

3.2 Frequency domain analysis

The vibration signals were studied with several techniques in the frequency domain, including Fast Fourier transform (FFT), MUSIC, and Spectrogram, selecting the FFT as the main technique for the study. The main goal of a domain change is to observe an explicit difference in the vibration patterns in the windings and to disclose data that cannot be discovered in the time domain. With the extraction of statistical features in the frequency

domain (SFF) of those vibration patterns, the assessment of the operating conditions in the transformer could be achieved with pattern recognition. Figure 3.3 shows a scheme of the method to study the signals in the frequency domain, starting from a required decimation to every signal in each axis, followed by computing of the frequency domain technique and the computing of the statistical features, next the features selection or dimensionality reduction, followed by the classification with machine learning techniques, and finally, the calculation of the accuracy to compare the results of the models.

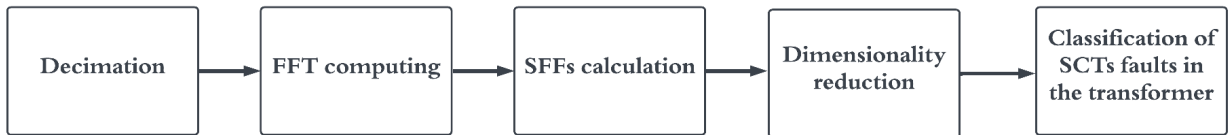


Figure 3.3. General scheme of the proposal

3.2.1 Decimation

Sampling is a process of converting a continuous signal into a sequence of discrete values. The sampling theorem explains that if the highest frequency of a signal X_a is F_{max} and the signal is sampled at a frequency $F_s > 2F_{max}$, where F_s is also called Nyquist frequency. Then X_a is perfectly reconstructed. [Proakis and Manolakis, 2006].

Decimation implicates the reduction of the sampling rate of a signal. While "downsampling" typically refers to a singular step in this process, the terms are occasionally used interchangeably. When decimation is applied to a sequence of samples derived from a signal or another continuous function, it generates an approximation of the sequence that would have been obtained had the signal been sampled at a lower rate. The decimation factor, generally an integer or a rational fraction greater than one, determines how the sampling interval is multiplied or how the sampling rate is divided equivalently. It is worth noting that only the sampling rate frequency and corresponding Nyquist frequency have changed. As long as the new Nyquist frequency is larger than the signal frequency, no aliasing will occur [Parker, 2017].

The signals were sampled at 6000 Hz, and according to the Nyquist theorem, the highest frequency perceptible in a frequency spectrum is 3000 Hz, but the accelerometer sensor has a bandwidth restriction of 1000 Hz. Therefore, a decimation is applied to the vibration signals since there is no guarantee that any frequency beyond 1000 Hz could be a real measurement

captured by the sensor. To preserve the first 1000 Hz of bandwidth, the downsampling is carried out by a factor of 3, making the new sampling frequency to be of 2000 Hz and the highest frequency perceptible 1000 Hz. Other techniques could be tested, for example, a lowpass filter, but depending on the filter configuration, this could incur an attenuation in the magnitude components.

3.2.2 Fast Fourier Transform

The Fast Fourier Transform (FFT) is a mathematical algorithm based on the Discrete Fourier Transform (DFT) to convert a signal from the time domain to the frequency domain; the domain change has individual spectral components that deliver frequency information about the signal [Proakis and Manolakis, 2006]. Equation (3.1) defines the DFT, where the main goal is computing a sequence $X[K]$ of N complex values given another data sequence $x[n]$ of N length.

$$X[K] = \sum_{n=1}^{N-1} x[n] e^{-\frac{j2\pi kn}{N}}, \quad \text{for } k = 0, \dots, N - 1 \quad (3.1)$$

From the equation, $e^{-\frac{j2\pi k}{N}}$ for $k = 0, \dots, N - 1$, N^{th} is the root of unity, receiving this name since in complex arithmetic $(e^{-j2\pi k}) = 1$ for all k values. These roots are vertices of a regular polygon inscribed in the unit circle of the complex plane. Figure 3.4 shows three circles and the roots of unity graphed for $N = 2$, $N = 4$, and $N = 8$ in the complex plane [Heckbert, 1995].

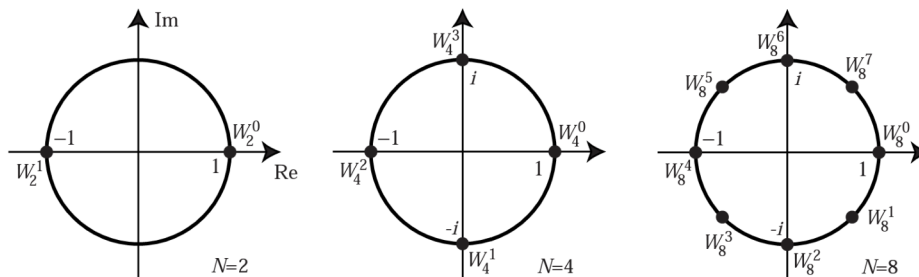


Figure 3.4. Complex plane representation of the roots of unity [Heckbert, 1995]

It is noteworthy that for every k value, computing $X[K]$ directly requires $O(N^2)$ operations: N complex multiplications ($4N$ real multiplications) and $N - 1$ complex additions

($4N - 2$ real additions). The FFT algorithm determines the DFT of an input quite faster than calculating it directly; this means that it reduces the number of operations needed for a problem of size N from $O(N^2)$ to $O(N \log_2 N)$ stages, and the difference is noticeable as N is larger [Proakis and Manolakis, 2006].

3.2.2.1 FFT base 2 algorithm

The FFT can be separated into two subsequences (even and odd) using the divide and conquer method; this method is based on the decomposition of a DFT of N points into smaller and successive DFT transforms, and the sequence will follow the next rule.

$$\begin{cases} n = 2r & \text{if } \text{even}, \\ n = 2r + 1 & \text{if } \text{odd}. \end{cases}$$

where $r = 1, 2, \dots, \frac{N}{2} - 1$.

The general mathematical expression of the summation of both terms is resumed in 3.2, and computing concurrently the indexed sub-sequences represents an advantage of this approach [Maklin, 2019].

$$X[K] = x_{\text{even}}[k] + e^{\frac{-j2\pi k}{N}} x_{\text{odd}}[k] \quad (3.2)$$

The diagram process for computing and creating two complex numbers from two complex numbers is named the butterfly diagram; generally, the diagram implies a complex multiplication and two complex additions. For $N = 2^v$, there are $N/2$ butterfly and $\log_2 N$ stages. Each butterfly takes two complex numbers, such as x and y , and computes two other numbers from them, $x + \alpha y$ and $x - \alpha y$, where α is a complex number; figure 3.5 illustrates a butterfly operation; this diagram is used for complex multiplication and addition and can be expanded to any N number of points [Heckbert, 1995].

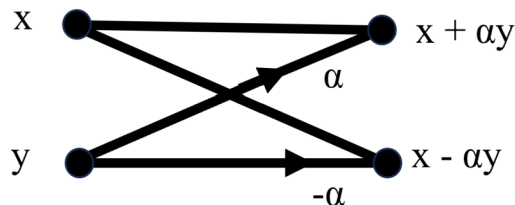


Figure 3.5. Butterfly diagram

Once the butterfly computing has been executed over a couple of complex numbers to create two new complex numbers, the algorithm requires that the new numbers be stored in the same position in memory of the original numbers. Therefore, a storing fixed space is necessary, i.e., $2N$ storing registers for the results [Proakis and Manolakis, 2006].

Summarising, the base 2 algorithm instantly takes two memory data points, executes the butterfly operations, and returns the values to the memory. This process is repeated several times $((N \log_2 N)/2)$ and represents the key to the efficiency of the DFT computation for reducing the number of actual arithmetic operations required [Proakis and Manolakis, 2006]. The final result of the algorithm is the frequency domain representation of the input sequence $x[n]$, containing complex numbers that represent the magnitude and phase of different frequency components present in the original input.

3.2.3 Statistical frequency features

SFFs can be computed in the frequency domain after applying a domain change with the FFT to describe the shape of the vibration signals from different standpoints [Wang et al., 2015]. Table 3.1 presents the mathematical expressions of 9 SFFs examined in this work [Helmi and Forouzantabar, 2019, Motahari-Nezhad and Jafari, 2020], where i is the number of spectrum lines for $i = 1, 2, \dots, N$, \mathbf{p}_i is a spectrum magnitude, and \mathbf{f}_i is the frequency value, for the i th spectrum line.

From table 3.1, MF tells the vibration energy in the frequency domain, FC and RMS_f describe the position change of the main frequency, RVF is a measurement of the centralized or decentralized degree of the spectrum power energy [Wang et al., 2015], Kurtosis is defined as a ratio of a fourth-order central moment to the square of a second-order central moment, SK is a measure of the asymmetry of the analyzed spectrum, while KU is a measure of the peakedness, i.e., higher kurtosis means more of the variance is due to infrequent extreme

deviations, as opposed to frequent modestly-sized deviations [Benedetto and Benedetto, 2011], the 1st spectral moment is an alternative statistical analysis way to extract features from the vibrations power spectrum, VMF and RV and variance measurement in the power spectrum and frequency respectively.

Table 3.1. Frequency domain features

Parameter name	Mathematical expression
<i>Mean Frequency</i>	$MF = \frac{1}{N} \sum_{i=1}^N p_i$
<i>Frequency Center</i>	$FC = \frac{\sum_{i=1}^N f_i p_i}{\sum_{i=1}^N p_i}$
<i>1st Spectral Moment</i>	$SM1 = \sum_{i=1}^N p_i f_i$
<i>Variance of Mean Frequency</i>	$VMF = \frac{\sum_{i=1}^N (p_i - MF)^2}{N-1}$
<i>Root Variance</i>	$RV = \sqrt{\frac{\sum_{i=1}^N (f_i - FC)^2 p_i}{N}}$
<i>Root Mean Square</i>	$RMS_f = \sqrt{\frac{\sum_{i=1}^N f_i^2 p_i}{\sum_{i=1}^N p_i}}$
<i>Skewness</i>	$SK = \frac{\sum_{i=1}^N f_i^2 (f_i - FC)^3 p_i}{(RV)^3 \cdot N}$
<i>Kurtosis</i>	$KU = \frac{\sum_{i=1}^N f_i^2 (f_i - FC)^4 p_i}{(RV)^4 \cdot N}$
<i>Root Variance Frequency</i>	$RVF = \sqrt{\frac{\sum_{i=1}^N (f_i - MF)^2 p_i}{\sum_{i=1}^N p_i}}$

3.2.4 Features extraction

The vibration signals with all the conditions are stored as Microsoft Access Tables (.mat) files; the extraction of the 9 features from table 3.1 starts with a segmentation of the signal in several portions. Then, a decimation is performed due to the aforementioned reasons; the FFT is computed in every data frame, the magnitude \mathbf{p}_i and frequency vector \mathbf{f}_i resultant are used for the feature computing, creating a feature vector of 1×9 for one segment sample. Since there are 20 triaxial signals for every condition and the 5 segments segmentation for oversampling, the total of samples is 100 for the X-axis, 100 for the Y-axis, and 100 for the Z-axis, replicating the procedure for every condition. The FFT calculation and the subsequent feature extraction will be performed in each of those 100 samples. The resulting matrix will be 100×9 for each one of the axes, delivering 100 triaxial samples per condition. Finally, each dataset is stored in mat files for selection and posterior classification. Algorithm 1 shows the

process explained for the feature extraction in the frequency domain.

Algorithm 1 Features extraction in the frequency domain

Require: Signal: File with vibration measurements

Require: n: conditions

Require: k: Number of tests

Require: j: Number of segments

Require: Fs: Sampling frequency

```
Vx ← Signal                                ▷ Column 4 steady state from sample 9001 to 36000
Vy ← Signal                                ▷ Column 5 steady state from sample 9001 to 36000
Vz ← Signal                                ▷ Column 6 steady state from sample 9001 to 36000
conditions ← [0 ... 30]
for n = 1 to 7 do
  n ← conditions(i)
  for j = 1 to 5 do
    X ← remove - offset(Vx)
    Y ← remove - offset(Vy)
    Z ← remove - offset(Vz)
    for j = 1 to j do
      Vx ← X((j - 1) * 5400 + 1 : j * 5400)
      Vxdec ← decimate(Vx, 3)
      [pxj fxj] ← FFT(Vxdec, Fs/3)
      featx ← SFF(pxj, fxj)
      Vy ← Y((j - 1) * 5400 + 1 : j * 5400)
      Vydec ← decimate(Vy, 3)
      [pyj fyj] ← FFT(Vydec, Fs/3)
      featy ← SFF(pyj, fyj)
      Vz ← Z((j - 1) * 5400 + 1 : j * 5400)
      Vzdec ← decimate(Vz, 3)
      [pzj fzj] ← FFT(Vzdec, Fs/3)
      featz ← SFF(pzj, fzj)
    end for
  end for
end for
return featx                                ▷ SFF signal axis X
return featy                                ▷ SFF signal axis Y
return featz                                ▷ SFF signal axis Z
```

3.3 Time domain analysis

The raw vibration signals will be analyzed in the time domain in order to assess the SCTs conditions in the transformers accurately. This analysis aims to extract statistical patterns that effectively represent the working conditions. Once the statistical information from the vibrations is obtained, pattern recognition techniques can be employed to evaluate the level of damage. The method, as illustrated in Figure 3.6, involves several steps. Statistical time features (STFs) are first computed from the raw signal. Secondly, feature selection and dimensionality reduction techniques are applied. Finally, machine learning algorithms are utilized for classification purposes, and the accuracy is computed for each model.

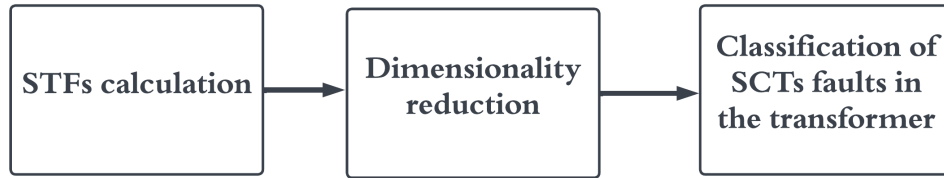


Figure 3.6. General scheme of the proposal

3.3.1 Statistical time features

In the signal processing field, estimating patterns capable of being associated with diverse states of a system is essential to describe and predict its behavior. The computation of statistical time features (STFs) accomplishes this task since different operation conditions will have different statistical parameters even when those pattern changes are very small [Chen and Lee, 2021] with the advantage of requiring fewer mathematical operations and without needing prior signal transformation for this analysis. Table 3.2 presents the mathematical formulations of 20 STFs examined in this approach, where \mathbf{x}_i represents the signal for $\mathbf{i} = \{1, 2, 3, \dots, \mathbf{N}\}$, where \mathbf{N} is the number of data points captured. The steady state of the raw signal will be used to calculate the features from the table in the three axes V_x, V_y, Z_z s measured by the accelerometer.

Considering the numerous STFs available, each measuring different signal characteristics, it is clear that not all can provide relevant information for accurate diagnosis.

Table 3.2. Time domain features

Parameter name	Mathematical expression
Mean	$\mu = \frac{1}{N} \sum_{i=1}^N x_i$
Standard deviation	$\sigma = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i)^2}$
Square mean root	$SMR = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{x_i} \right)^2$
Range	$Med = \max(x_i) - \min(x_i)$
Mode	$Mo = \operatorname{argmax}_{x_i} P(X = x_i)$
Median	$Magnitude \left(\frac{N+1}{2} \right)$
Root mean square	$RMS = \sqrt{\left(\frac{1}{N} \sum_{i=1}^N \{x_i\}^2 \right)}$
Skewness	$Skw = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N \cdot \sigma^3}$
Kurtosis	$Kf = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N \cdot \sigma^4}$
Normalized 5th moment	$5thM = \frac{\sum_{i=1}^N (x_i - \mu)^5}{N \cdot \sigma^5}$
Normalized 6th moment	$6thM = \frac{\sum_{i=1}^N (x_i - \mu)^6}{N \cdot \sigma^6}$
Shape factor with RMS	$SF_{rms} = \frac{RMS}{\frac{1}{N} \sum_{i=1}^N x_i }$
Shape factor with SMR	$SF_{smr} = \frac{SMR}{\frac{1}{N} \sum_{i=1}^N x_i }$
Impulse factor	$IF = \frac{\max(x_i)}{\frac{1}{N} \sum_{i=1}^N x_i }$
Crest factor	$CF = \frac{\max(x_i)}{RMS}$
Latitude factor	$LF = \frac{\max(x_i)}{SMR}$
Shannon entropy	$SE = - \sum_{i=1}^N x_i^2 \log(x_i^2)$
Median absolute deviation	$MAD = \frac{1}{N} \sum_{i=1}^N x_i - \mu $
Skewness factor	$F_{skw} = \frac{Skw}{RMS^3}$
Kurtosis factor	$F_{kf} = \frac{kf}{RMS^4}$

Therefore, evaluating the discriminative power of different variables in distinguishing between two or more classes is important [Yanez-Borjas et al., 2019]. At the same time, dimensionality reduction is beneficial as it helps eliminate noise and improve the performance of the machine learning models.

3.3.2 Features extraction

The extraction of the 20 features starts with the segmentation of all steady-state signals, but for this domain, fewer steps are taken for the extraction. After the segmentation is completed, the features from table 3.2 are computed, creating a feature vector of 1×20 for one portion sample. Since there are 20 triaxial signals for every condition and the 5 segments segmentation for oversampling as in the frequency domain, there will be a total of 100 samples for every axis and condition. The feature extraction will be performed in each of those 100 samples. The resulting dataset is 100×20 for each one of the axes and conditions. Finally, each dataset is stored in mat files for selection and posterior classification; 21 mat files are created (3 files per axis and 7 conditions). Algorithm 2 shows the process explained for the feature extraction in the time domain.

3.4 Data normalization

The feature matrix has to be normalized to be more suitable for machine learning models, and the zscore technique is selected to carry out the normalization; this process starts by taking the mean and standard deviation on every features column, then every value (\hat{x}) is scaled by subtracting the mean and dividing by the standard deviation of the column as in equation 3.3. The normalization was performed on each one of the columns features.

$$\hat{x} = \frac{x - \mu}{\sigma} \quad (3.3)$$

Although the features were normalized, a high dimensionality problem still has to be managed to avoid redundant or noisy data that could decrease the performance of the classification model.

3.5 Feature selection

The gathered features in different files with conditions and axes must be arranged in a dataset where all the conditions are included to select the best attributes for pattern recognition. The classification of the SCTs faults will be carried out by machine learning tools.

Algorithm 2 Features extraction in the time domain

Require: Signal: File with vibration measurements**Require:** n: conditions**Require:** k: Number of tests**Require:** j: Number of segments

```
Vx ← Signal                                ▷ column 4 steady state from sample 9001 to 36000
Vy ← Signal                                ▷ column 5 steady state from sample 9001 to 36000
Vz ← Signal                                ▷ column 6 steady state from sample 9001 to 36000
conditions ← [0 ... 30]
for n = 1 to 7 do
  N1 ← conditions(i)
  for k = 1 to 20 do
    X ← remove - offset(Vx)
    Y ← remove - offset(Vy)
    Z ← remove - offset(Vz)
    for j = 1 to 5 do
      Wx ← X((j - 1) * 5400 + 1 : j * 5400)
      featx ← STF(Wx)
      Wy ← Y((j - 1) * 5400 + 1 : j * 5400)
      featy ← STF(Wy)
      Wz ← Z((j - 1) * 5400 + 1 : j * 5400)
      featz ← STF(Wz)
    end for
  end for
end for
return featx                                ▷ STF signal axis X
return featy                                ▷ STF signal axis Y
return featz                                ▷ STF signal axis Z
```

3.5.1 Kruskal - Wallis method

The Kruskal-Wallis test is a nonparametric method used for comparing and identifying differences between multiple datasets, being particularly useful when dealing with data with a non-normal or unknown distribution. It is an alternative to traditional analysis of variance (ANOVA) that has a stronger assumption that the populations have normal distributions. The primary goal of the Kruskal-Wallis method is to evaluate the probability of making an

incorrect conclusion when stating that there are no differences among three or more treatment groups [Theodorsson-Norheim, 1986].

To accomplish this, the Kruskal-Wallis method considers the variations between the data groups and the variations within each group. It calculates a probability value, commonly known as the p-value, which indicates the likelihood of rejecting the null hypothesis. The null hypothesis is that the samples come from the same population [Kruskal and Wallis, 1952]. In this case, the null hypothesis affirms that all the average values of the datasets are equal. If the computed p-value falls below a predetermined significance level, the null hypothesis is rejected, concluding that at least one of the group means is significantly different from the others.

As a result, lower p-values indicate a greater ability of the selected feature to differentiate between the chosen dataset and the remaining datasets. In essence, the Kruskal-Wallis method provides a statistical framework for quantifying the significance of differences between multiple datasets, enabling the inference of meaningful conclusions regarding the comparability of these datasets [Bashar and Bhuiyan, 2016].

Algorithm 3 shows the process of the test application; the method starts by arranging all features and conditions in a database; for example, having 100 samples and 20 features computed per condition, the database will have the shape of 100×140 , where the first 20 columns are the features computed for the healthy condition, the columns 21 to 40 are the 5 SCTs statistical features and so on to the 30 SCTs conditions statistical features in the columns 121 to 140, the process is executed in the three axes separately. Figures 3.7, 3.8, and 3.9 illustrate how the function from the MATLAB toolbox works iteratively; a feature across all the conditions will be selected (7 columns will be taken), then that same feature is going to be arranged in a matrix, the Kruskal Wallis function will compute the probability for rejecting the hypothesis of similarities of that statistical parameter among all the operating conditions. The computed probability for each feature is stored in an array of the size 1×20 , and it will be replicated for the other axes; this is performed for feature selection in further steps.

Algorithm 3 Kruskal - Wallis test for each feature

Require: Vx : Vibration axis X features

Require: Vy : Vibration axis Y features

Require: Vz : Vibration axis Z features

Require: $Labels$: Set with labels

Require: N : Number of features

Require: $features$: Number of features

for $i = 1$ to $features$ **do**

$Px \leftarrow kruskallwallis(Vx(i) \ Vx(i + 20) \ \dots \ Vx(i + 120), labels)$ \triangleright X-axis probabilities

end for

return Px \triangleright The probability of similarity for each feature and condition

for $i = 1$ to $features$ **do**

$Py \leftarrow kruskallwallis(Vy(i) \ Vy(i + 20) \ \dots \ Vy(i + 120), labels)$ \triangleright Y-axis probabilities

end for

return Py \triangleright The probability of similarity for each feature and condition

for $i = 1$ to $features$ **do**

$Pz \leftarrow kruskallwallis(Vz(i) \ Vz(i + 20) \ \dots \ Vz(i + 120), labels)$ \triangleright Z-axis probabilities

end for

return Pz \triangleright The probability of similarity for each feature and condition

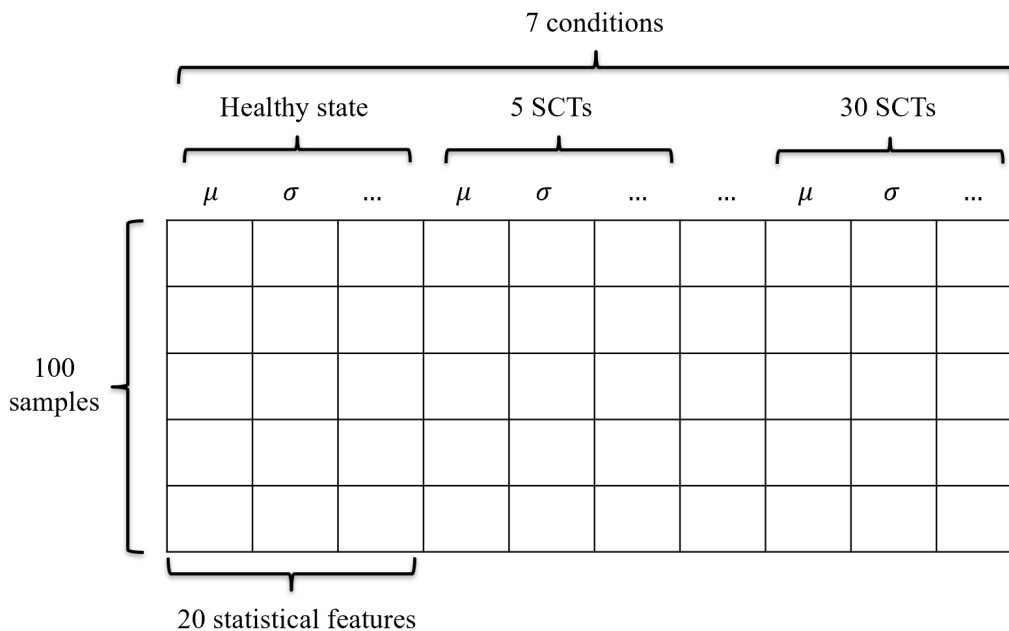


Figure 3.7. Features arranged for applying the method iteratively

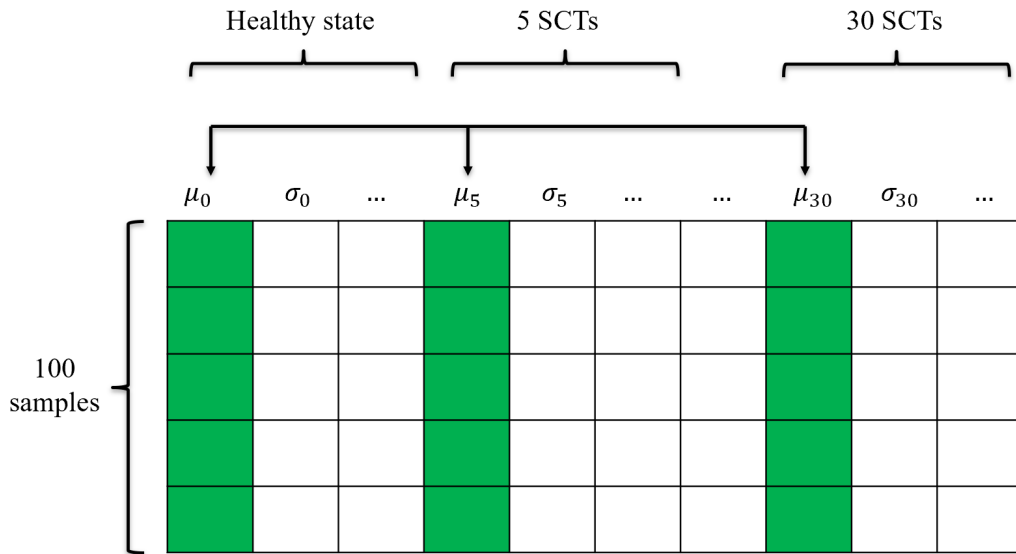


Figure 3.8. Example of the same feature across all the conditions taken

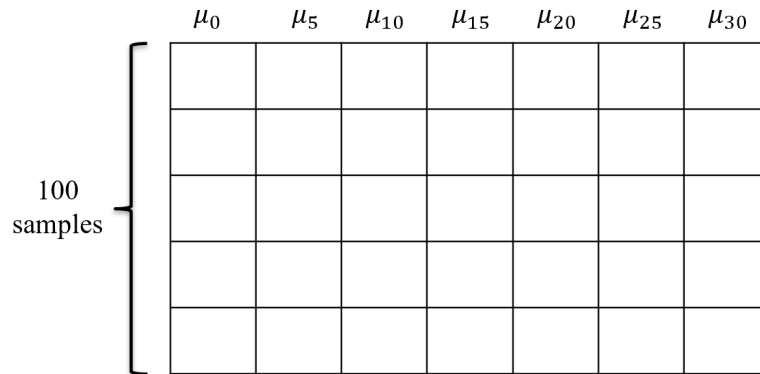


Figure 3.9. Data matrix used as a parameter for the Kruskal - Wallis test

Algorithm 4 shows a quick algorithm to select the features; once the probabilities have been computed and stored, the algorithm needs them arranged in a single array; this array will be sorted so that the lowest probabilities are in the first positions. As explained before, a low probability indicates that the populations are not similar to each other, meaning that the most capable statistical parameters for separating the SCTs conditions will be employed in the machine learning models, and several selections will be made to test those models.

Algorithm 4 Selection applying Kruskal Wallis probabilities test

Require: Px : Probabilities array X-axis features**Require:** Py : Probabilities array Y-axis features**Require:** Pz : Probabilities array Z-axis features**Require:** PKW : Vector where all probabilities will be arranged $PKW \leftarrow [Px \ Py \ Pz]$ $feats, index \leftarrow sort(PKW)$ \triangleright Sorted values where lower probabilities will be on the top**return** $index$ \triangleright The position of each column feature

3.5.2 Principal Component Analysis

Principal Component Analysis (PCA) is a machine learning technique utilized in unsupervised learning. It involves statistical analysis and an orthogonal transformation to convert correlated data samples into linearly uncorrelated variables. PCA serves as a valuable tool for dimensionality reduction while preserving the essential features of the data [S. Karamizadeh and Hooman, 2013].

Let $\mathbf{X} \in \mathbb{R}^{a \times b}$ represent the dataset, where a denotes the number of observations and b represents the number of features. PCA applies an algorithm to identify a lower-dimensional subspace (with k dimensions and j lower-dimension, where $j \ll k$) whose basis vectors correspond to the directions of maximum variance in the original feature space [Pankaj and Wilscy, 2013]. The process begins by calculating the mean vector μ for each feature. The covariance matrix Σ is then computed using the mean vector. Then, the covariance matrix Σ is decomposed to obtain the eigenvalues $[\lambda_1, \lambda_2, \dots, \lambda_n]$ (representing the distribution of energy in the source data) and the corresponding eigenvectors (denoted as V) [S. Karamizadeh and Hooman, 2013]. This can be expressed as:

$$\Sigma = V \cdot D \cdot V^{-1} \quad (3.4)$$

Once the vector j complies with a certain threshold of cumulative energy represented by $1 < j < \lambda_n$, the eigenvectors matrix $V = [1 : j]$ will be employed to project the data points in a new matrix S , as in equation 3.5.

$$S = M \cdot V[1 : j] \quad (3.5)$$

PCA offers several advantages. It exhibits low sensitivity to noise, requires less memory capacity, and enhances efficiency by eliminating data redundancy through the use of orthogonal components. It finds applications in face recognition, data denoising, and more [Pankaj and Wilscy, 2013].

3.6 Classification

The feature vector with the statistical parameters extracted and selected begins the classification and identification process of the operating conditions in the windings. In order to obtain the best results that validate the methodology proposed, the following three machine-learning models will be used in the time and frequency domains:

- K nearest neighbors (K-NN)
- Naive Bayes (NB)
- Support vector machines (SVM)

The configuration of the models is done with the machine learning toolbox from MATLAB®. In addition, each model is deployed with stratified k-fold cross-validation to classify SCTs conditions and the healthy state. K-fold divides the database into k subsets, where one subset is used for testing, and the k-1 folds are put together in a training set [Xiong et al., 2020]. The stratification will guarantee that all the conditions have the same percentage of samples for each target class.

3.6.1 K Nearest Neighbors (K-NN)

The K-NN classifier was used with the Euclidean distance from equation 2.6. The optimal value for K neighbors was selected by applying the square root of the number of training samples per class (N) of equation 2.7 [Zhang et al., 2018]; having a balance of 100 samples in each condition, the approximate value for the K parameter is around 10.

The process for classifying the corresponding fold set of test samples with K - NN is shown in algorithm 5; in this sequence, the normalized dataset that follows the procedure of

Algorithm 5 Classification with K-NN algorithm

Require: DB : Database with all features normalized**Require:** T : Labels for each sample and condition**Require:** F : Number of folds for cross-validation**Require:** N : Number of neighbors**Require:** TC : Targets classified**Require:** AC : Accumulates the number of correct classifications

```
 $F \leftarrow 10$                                 ▷ Setting in 10 the folds for cross-validation
 $N \leftarrow 10$                                 ▷ Setting the number of neighbors
 $AC \leftarrow 0$                                 ▷ Setting the classification accumulator
 $Model \leftarrow KNN(DB, T, N, F)$             ▷ The function will return the model trained
 $TC \leftarrow Kfoldpredict(Model)$           ▷ Assigns labels to the K test fold
if  $TC$  is =  $T$  then
     $AC \leftarrow AC + 1$ 
end if
 $Accuracy \leftarrow AC / length(T)$           ▷ Computes accuracy
 $Lost \leftarrow 1 - accuracy$                 ▷ Computes the lost in the classification
return  $Accuracy$                             ▷ The classification accuracy of the model
```

equation 3.3 and with a feature selection to reduce dimensionality is prepared to be used. Then, the variables with the labels vector, the number of folds for the cross-validation, the number of neighbors, and a variable that will store the accuracy of the model. The classification model is applied with *fitcknn* function from MATLAB. The parameters included in the function are the dataset, the labels vector, the type of distance, the number of neighbors, and the option for k-fold that has to be enabled with the inclusion of the number of folds. The model will be trained and stored in the object "Model", the "Kfoldpredict" object will use the trained model for assigning labels to the test set randomly created, and the results will be stored in the variable TC . The accuracy of the model is computed by logically comparing the actual labels with the predicted ones; the loss in the model is also computed for analysis purposes. Finally, the algorithm will return the accuracy of the model.

3.6.2 Naive Bayes

Multinomial Naive Bayes was used as one of the alternative forms of Naive Bayes algorithms for multiclass classifications; the algorithm assumes that the predictors are conditionally independent, given the class, the function *fitcnb* is applied to test the dataset

and classify the SCTs conditions of the transformer.

Algorithm 6 Classification with Naive Bayes algorithm

Require: DB : Database with all features normalized

Require: T : Labels for each sample and condition

Require: F : Number of folds for cross-validation

Require: TC : Targets classified

Require: AC : Accumulates the number of correct classifications

```
 $F \leftarrow 10$  ▷ Setting in 10 the folds for cross-validation  
 $AC \leftarrow 0$  ▷ Setting the classification accumulator  
 $Model \leftarrow NB(DB, T, F)$  ▷ The function will return the model trained and ready for  
testing the k fold set  
 $TC \leftarrow Kfoldpredict(Model)$  ▷ Assigns labels to the test fold  
if  $TC$  is =  $T$  then  
     $AC \leftarrow AC + 1$   
end if  
 $Accuracy \leftarrow AC / length(targets)$  ▷ Computes accuracy  
 $Lost \leftarrow 1 - accuracy$  ▷ Computes the lost in the classification  
return  $Accuracy$  ▷ The classification accuracy of the model
```

Algorithm 6 has a similar process to Algorithm 5 but with fewer steps; as explained before, the dataset is managed similarly. Then, the variables with the labels vector, the number of folds for the cross-validation, and a variable that stores the accuracy (AC) of the model are created. Followed by assigning an initial value of 10 in the number of folds and 0 to AC . The classification model is applied with the aforementioned function from MATLAB represented with NB in the algorithm. The parameters included in the function are the normalized dataset, the labels vector, and the option for k-fold that has to be enabled, including the number of folds. As well as in the K - NN classifier, the model will be trained and stored in the object "Model" the "Kfoldpredict" object will use the trained model for assigning labels to the test set randomly created, and the results will be stored in the variable TC . The accuracy is computed following the same steps as in the previous classifier. Finally, the algorithm will return the accuracy of the model executed.

3.6.3 Support Vector Machines

SVM classifier was employed under the same cross-validation, with a second-degree

polynomial kernel configuration to achieve the classification; the function employed is error-correcting output codes, "*fitcecoc*" that returns a multiclass model.

Algorithm 7 Classification with SVM algorithm

Require: *DB*: Database with all features normalized

Require: *T*: Labels for each sample and condition

Require: *F*: Number of folds for cross-validation

Require: *PL*: Polynomial order

Require: *TC*: Targets classified

Require: *AC*: Accumulates the number of correct classifications

$F \leftarrow 10$ ▷ Setting in 10 the folds for cross-validation

$PL \leftarrow 2$ ▷ Setting the polynormal order

$AC \leftarrow 0$ ▷ Setting the classification accumulator

$Template \leftarrow kernel, polynomial$ ▷ Setting the classification accumulator

$Model \leftarrow SVM(DB, T, Template, F)$ ▷ The function will return the model trained and ready for testing the k fold set

$TC \leftarrow Kfoldpredict(Model)$ ▷ Assigns labels to the test fold

if TP is = T **then**

$AC \leftarrow AC + 1$

end if

$Accuracy \leftarrow AC/length(targets)$ ▷ Computes accuracy

$Lost \leftarrow 1 - accuracy$ ▷ Computes the lost in the classification

return *Accuracy* ▷ The classification accuracy of the model

Algorithm 7 follows an equivalent process as the other algorithms, repeating the normalized dataset. Then, the variables with the labels vector, the number of folds for the cross-validation, and a variable that stores the accuracy (*AC*) of the model are created. Followed by assigning an initial value of 10 in the number of folds and 0 to *AC*. Before the application of the model, the learning template has to be defined, and the kernel trick is employed; the template includes a *kernel* function that shows a transformation of the data, *polynomial* is related to the type of kernel function to use (in this case is a polynomial grade 2). After the template is set, the classification model is applied and represented in the algorithm with *SVM*. The parameters included in the function are the normalized dataset, the labels vector, the template, and the option for K-fold that has to be enabled, including the number of folds. As well as in the K - NN classifier, the model will be trained and stored in the object "Model" the "Kfoldpredict" object will use the trained model for assigning labels to the test set randomly created, and the results will be stored in the variable *TC*. The accuracy is computed

following the same steps as in the previous classifier. Finally, the algorithm will return the accuracy of the model executed.

CHAPTER 4

Results

The development of this chapter is focused on the exposition of the results after the application of the proposal. The vibration signals acquired from a single-phase transformer include a healthy state and six levels of SCTs in the windings; 20 signals were captured for every condition in total. The reason for acquiring only 20 signals is related to the safety of the transformer since repeating the short circuits simulated could yield damage to the device. Given this drawback, several segmentations were done for each signal to oversample since, according to statistics, the larger the number of samples, the better the inferences made from the statistical calculations.

The chapter structure is divided into two stages using MATLAB for all the methodology steps. First is the frequency domain analysis that includes the FFT application to the steady state of the signal. Second is the analysis in the time domain of the raw signal. For both steps, statistical features were calculated. Then, once the attributes are gathered, feature selection and dimensionality reduction are executed to remove noisy or redundant data. Next, the classification of the faults through machine learning tools is accomplished. In this step, the classification accuracy achieved for each model is analyzed with confusion matrices to measure the performance of the selected models. Lastly, a contrast of the results is discussed.

4.1 Vibration signals dataset

For this investigation, 20 signals were measured for each SCTs condition; each signal has an acquisition time of 5 s (30000 samples), where the transient state is around 0.5 s (3000 samples), and the steady state is around 4.5 s (27000 samples). In total, the dataset has 140 signals with all the SCTs conditions; as explained in Chapter 3, the conditions are the healthy state and 6 damage severities. The accelerometer measured each signal in three axes A_x , A_y , A_z . Figures 4.1 and 4.2 show an example of the three axes vibration signal in steady-state for the healthy condition (0 SCTs) and the most severe damage (30 SCTs).

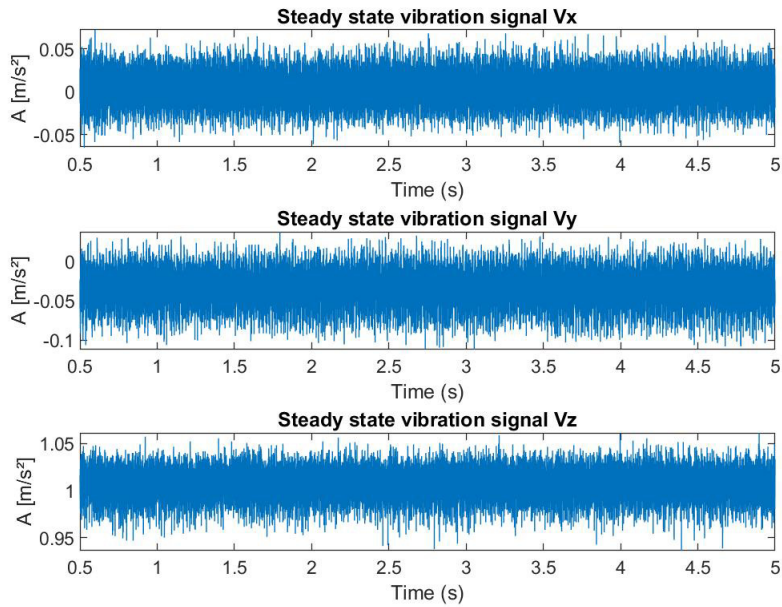


Figure 4.1. Vibration signal measured in the three axes for healthy condition

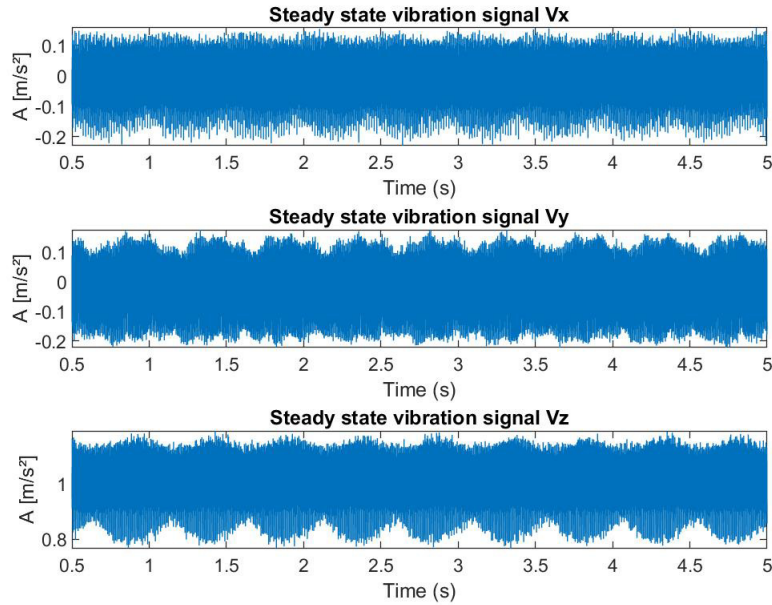


Figure 4.2. Vibration signal measured in the three axes for 30 SCTs condition

4.2 Frequency domain analysis

For this analysis, the Fast Fourier Transform (FFT) with a set of statistical features was employed. Before computing it, it is important to process the signals in order to avoid erroneous interpretations of the results given the acquisition conditions. Firstly, the offset of each axis signal was removed; the reason to do this is to delete any level added to the signal during the measurement. Secondly, the frequency sampling was reduced by three times, given the reasons explained in the previous chapter. Once these steps were executed, every signal was segmented into five segments, where each frame is of 0.9 seconds (5400 samples), allowing the possibility of acquiring 100 triaxial signals per condition and increasing the database from 140 to 700 raw signals for all conditions. Figure 4.3 shows an example of segmentation performed on the signal; the total length of the signal is 27000 samples, and each frame has 5400 samples.

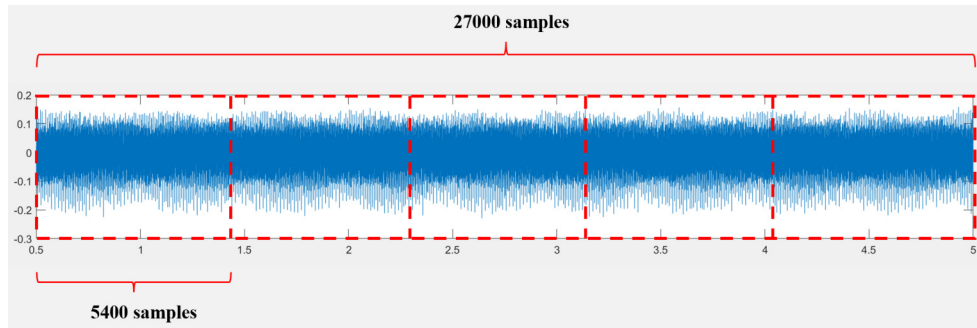
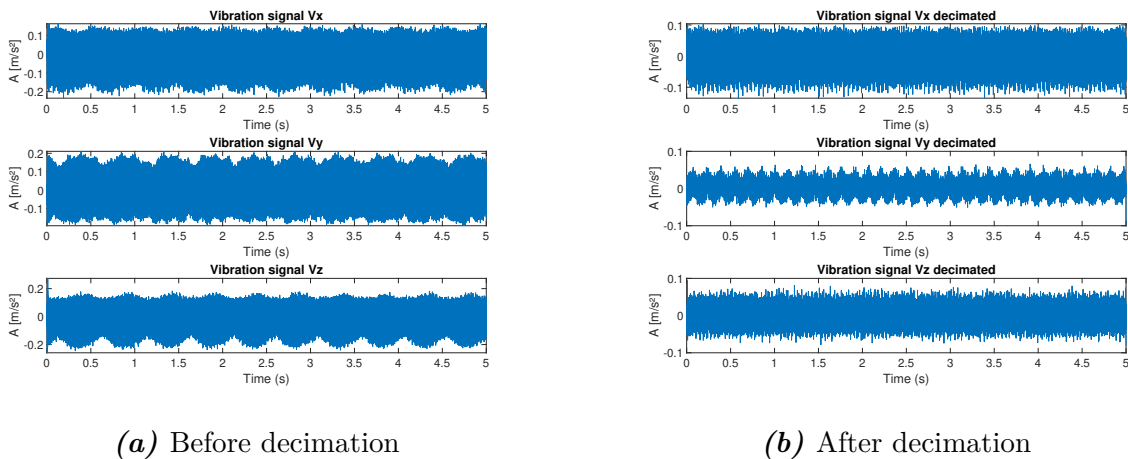


Figure 4.3. Segmentation of the signal in five portions

4.2.1 Decimation of the signal

The offset is removed from each one of the signals, and then decimation is performed to reduce the frequency sampling. Figure 4.4 shows an example of a triaxial signal with transient and steady-state before and after decimation. It is worth noting that this is a general explanation of how it works the decimation process, but for the particular case of this work, the procedure was done to every segment of the signal before computing the FFT.



(a) Before decimation

(b) After decimation

Figure 4.4. Triaxial vibration signal

4.2.2 Spectral representation

The FFT was computed after the signals were decimated; the procedure of this domain change delivered two arrays, one with the magnitudes spectrum and the other with the frequencies. Since the FFT performs a bilateral spectrum calculation, a correction was done to obtain a single-sided amplitude spectrum. Figure 4.5 shows an example of a 30 SCTs

vibration signal in the X-axis, where differences in the frequency spectrum before and after executing the decimation of the signal are observed (Vibration signals in the Y-axis and Z-axis have similar spectrums with low noise), where the first 1000 Hz of spectrum remains without attenuations or information lost. However, spectral components are visible beyond 1000 Hz. Given the restrictions of frequencies in the accelerometer, there is no guarantee that those energy components at high frequencies were real information.

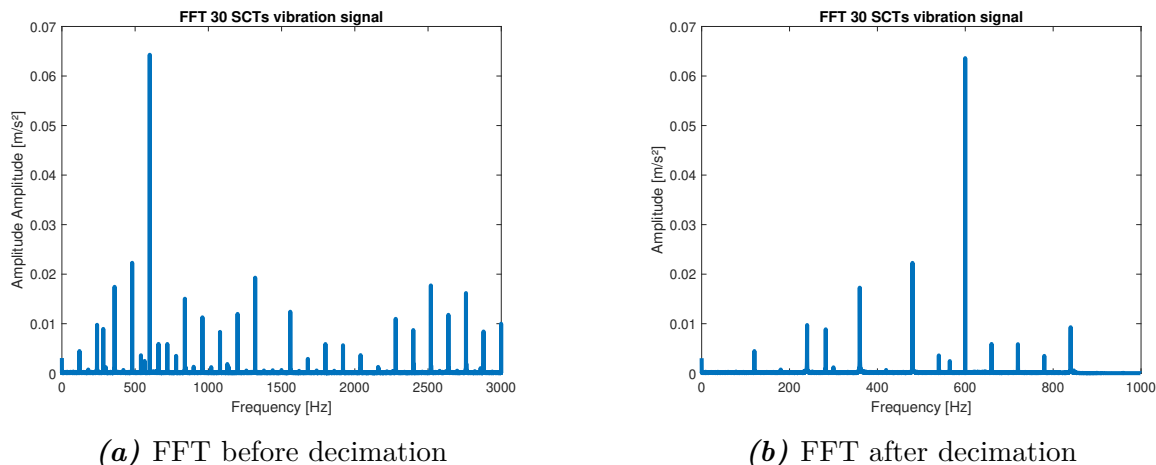


Figure 4.5. FFT spectrum triaxial vibration signal

It is important to acknowledge that the FFT does have certain limitations. One such limitation is spectral leakage, which occurs when the frequency components of a signal extend beyond the resolution capabilities of the FFT due to an inadequate sampling frequency. In such cases, the frequency components can overlap, leading to inaccuracies and loss of information in the resulting spectrum. For vibration signal analysis, other techniques such as MUSIC (Multiple Signal Classification) and Spectrogram were employed to observe in time-frequency spectrums if some frequency components only showed up during the transient of the signal. The results obtained were not concluding, and since FFT delivered a quite clean spectrum of frequencies, with possibilities to evaluate the operating conditions in the transformer with statistical analysis, the other techniques were discarded.

After the two processing steps were completed, the FFT was computed in each segmented signal and every one of the axes; each computation has one side magnitude spectrum array \mathbf{p} magnitude and frequency array \mathbf{f} . This means that with the aforementioned framing procedure and the total signal captured, one operation condition, e.g., the healthy state, has 100 FFT computed in the X-axis, 100 in the Y-axis, and 100 in the Z-axis.

Analyzing the acceleration vibration components in the frequency spectrum of the

signals, a noticeable change in the energy was observed, but not on a consistent basis. The main acceleration component of the vibrations is 120 Hz (two times the frequency of voltage operation), and harmonic content related to that main component is also visible. The energy did not increase linearly in one frequency component as the short circuit fault increased. Still, there is a remarkable energy increase that, with statistical analysis, would have enough capacity to assess conditions in the operation of the device. Figures 4.6, 4.7 4.8, and 4.9 show an example of a vibration signal in the X-axis with an energy increment. The magnitude axis was fixed in all frequency spectrums so it can be noticeable the change content. The frequency spectrum graphics were split into different figures to appreciate the energy component better as the faults increase.

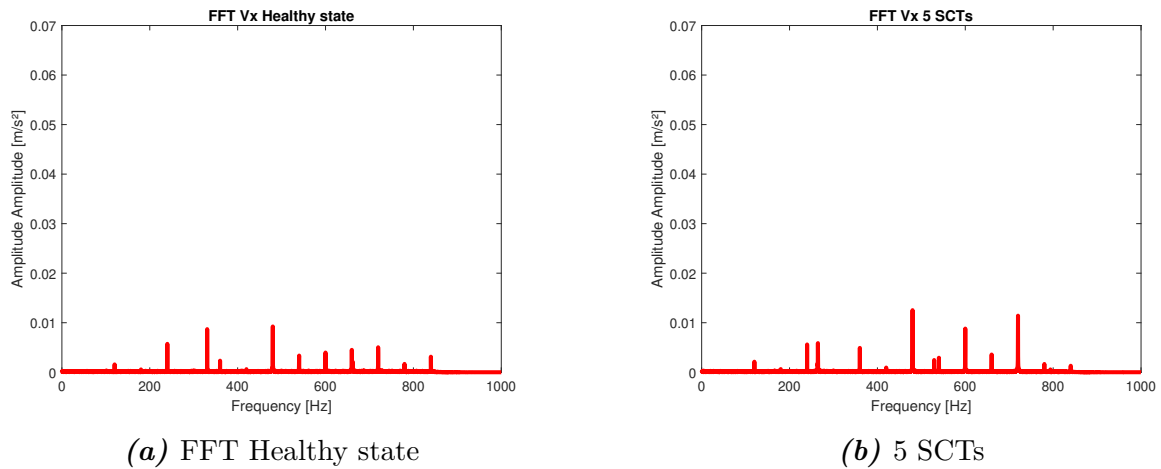


Figure 4.6. FFT vibration signal

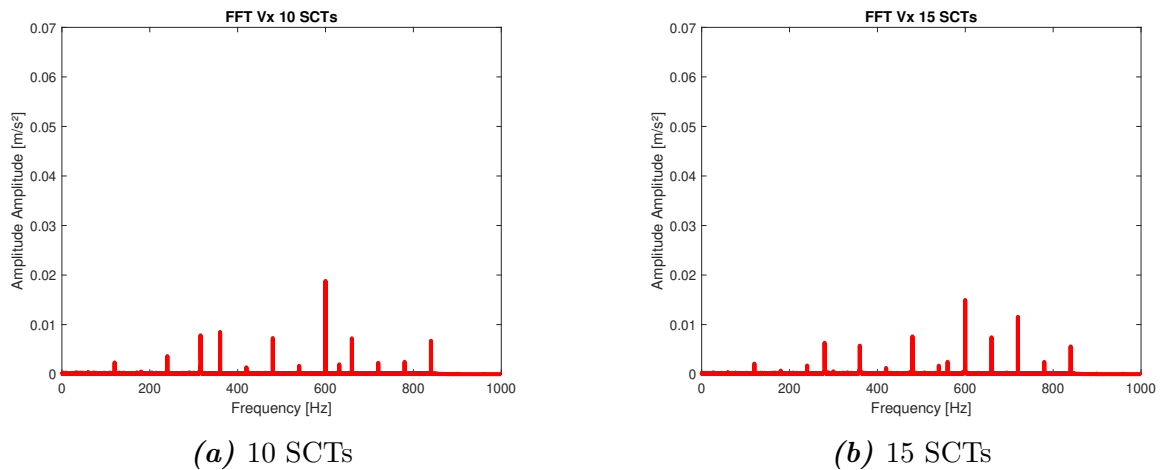


Figure 4.7. FFT vibration signal

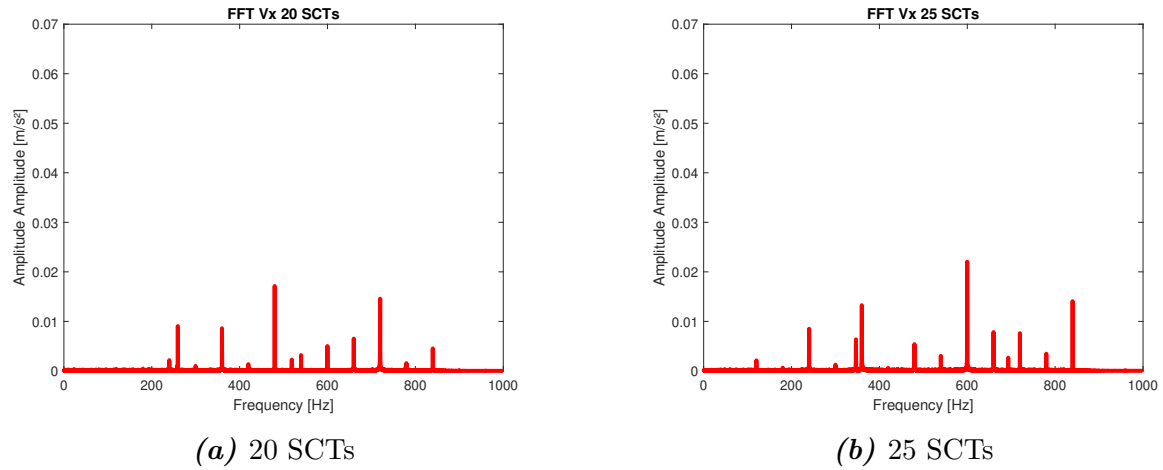


Figure 4.8. FFT vibration signal

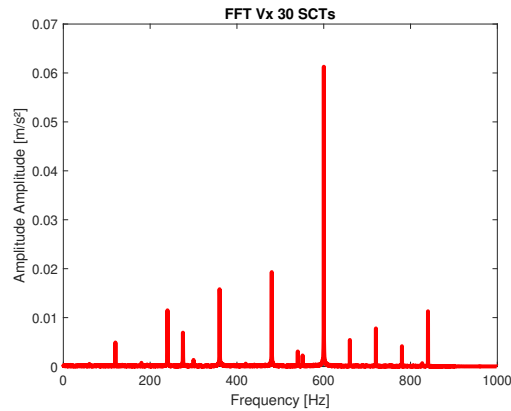


Figure 4.9. FFT vibration signal 30 SCTs

4.2.3 Statistical frequency features

The 9 SFFs from table 3.1 were computed once it was completed the calculation of the FFT using the \mathbf{p} magnitude spectrum array and \mathbf{f} frequency array for each segmented signal and condition as described in algorithm 1. Every feature computed from one portion sample was arranged in an array of the shape 1×9 . After iterating by condition, test, and segment, all features were arranged in a matrix, where one condition had 100×9 for A_x , 100×9 for A_y , and 100×9 for A_z and saved as Microsoft access tables (mat files) for selection and reduction in further steps. Figures 4.10, 4.11, and 4.12 show boxplot representations of A_x , A_y , and A_z features, each figure shows one feature and its behavior across all the SCTs conditions in the windings. Visually, this tool is quite helpful in the discovery of those attributes capable of separating and classifying an abnormality.

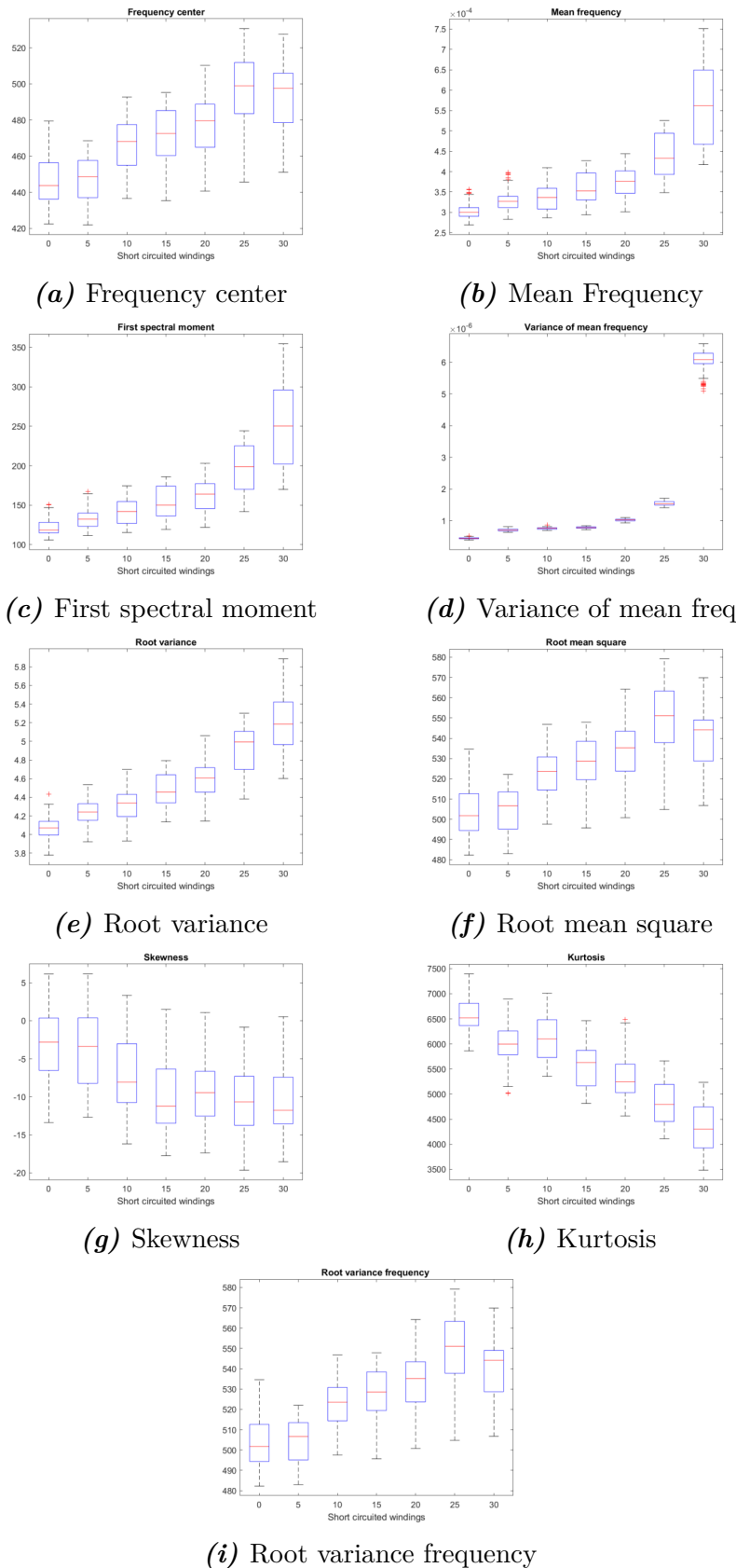


Figure 4.10. Boxplot representation of all conditions features extracted axis X

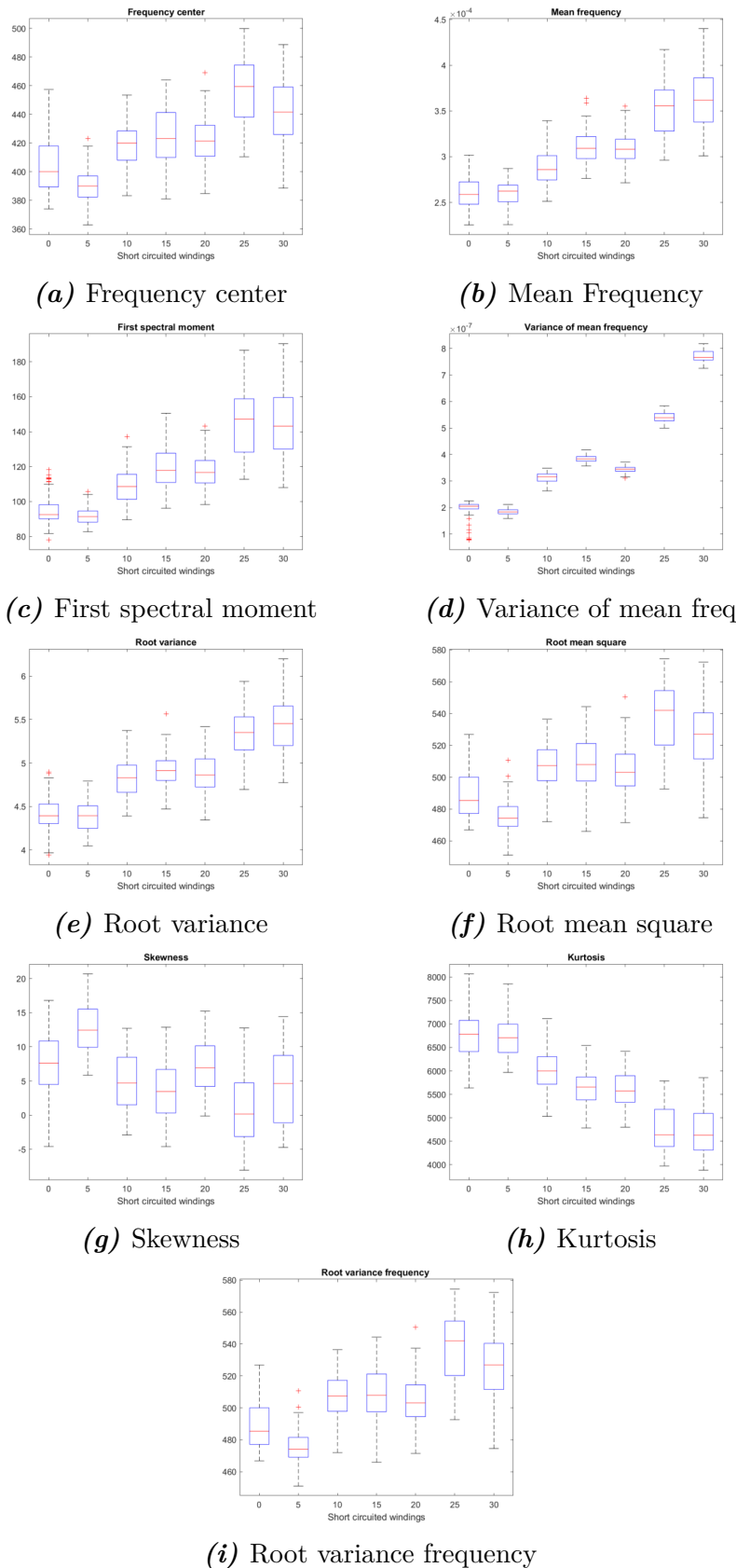


Figure 4.11. Boxplot representation of all conditions features extracted axis Y

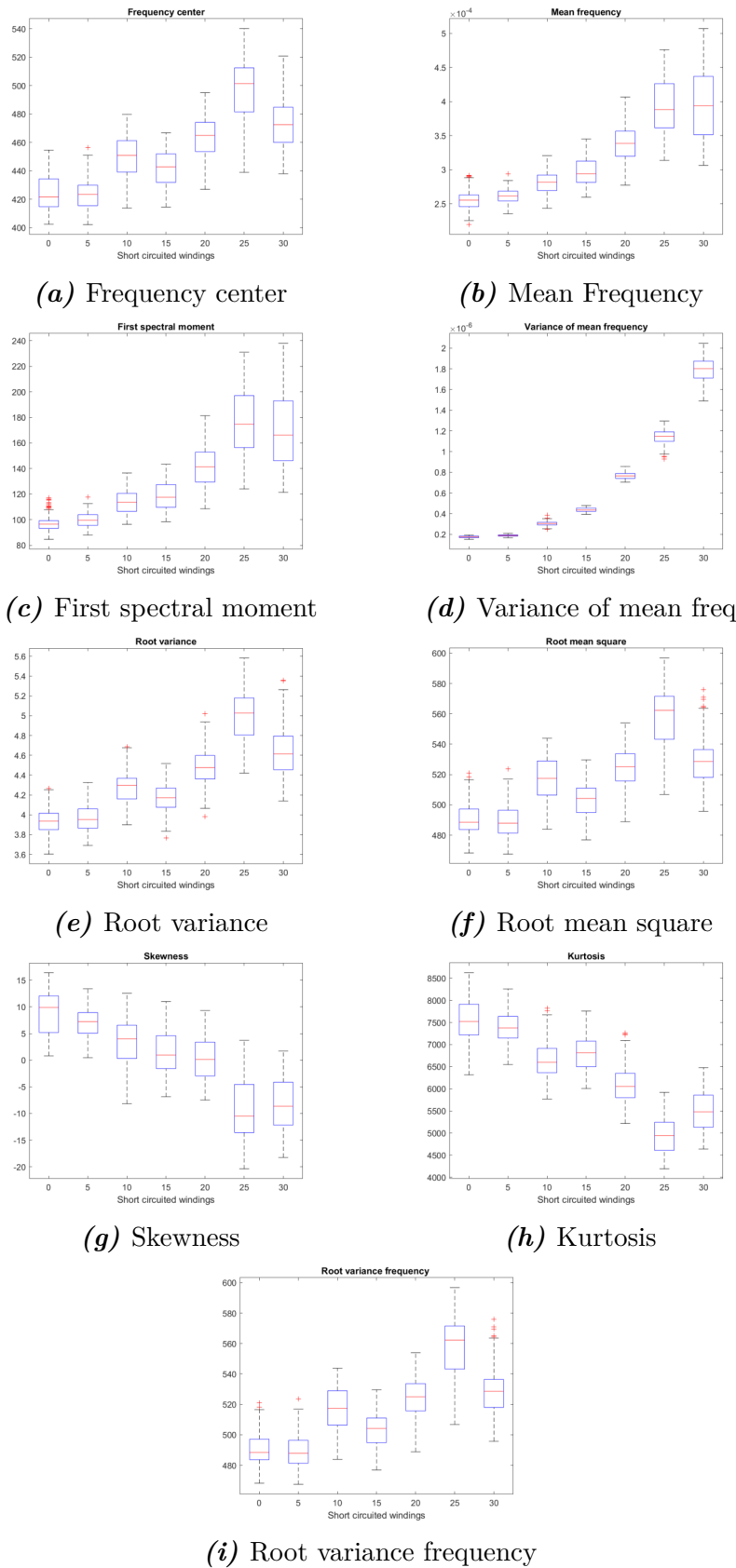


Figure 4.12. Boxplot representation of all conditions features extracted axis Z

4.2.4 Features selection

4.2.4.1 Kruskal - Wallis method

This study utilized the Kruskal-Wallis test to identify the most effective attributes for detecting SCTs faults in the windings. The method involved computing probabilities by comparing the same feature through all conditions for each axis separately. The procedure started with the iterative importation of mat files of 100×9 shape and stored in a variable, resulting in a total matrix size of 100×63 . The first 9 columns represent features in the healthy state, columns 10 to 18 contain 5 SCTs features, and so on. With 9 features in total, a loop was used to compute the probability for each feature. For instance, in the first iteration, the function considered the frequency center of the healthy state (column 1), the 5 SCTs (column 10) continuing with the same attribute to the final condition to compute the probability. This process was repeated for all features, and the resulting probabilities were stored in an array.

The algorithm was applied to the features of the three axes. The computed probability results were arranged in arrays for easier interpretation; Table 4.1 presents the probabilities computed. Notably, the fourth row (VMF) exhibited remarkable results in terms of separation between conditions, as observed from the graphical representation, aligned with the theory of the method. This feature not only displayed distinct medians but also exhibited significant differences throughout the complete boxplot. The Kruskal-Wallis test confirmed these observations, yielding very small probabilities and validating the graphical findings. Consequently, this feature was considered suitable for condition identification and classification.

Table 4.1. Probabilities computed with Kruskal Wallis test

Kruskall Wallis test			
Legend	Probability Vx	Probability Vy	Probability Vz
'MF'	4.0108e-85	2.0773e-78	8.6113e-107
'MP'	8.6797e-104	3.6742e-117	5.494e-122
'1 st SP'	1.1249e-102	3.6659e-112	2.9611e-121
'VMF'	4.2875e-139	2.5715e-142	3.1434e-143
'RV'	1.9536e-113	1.7442e-106	4.6528e-116
'RMS'	3.7108e-87	3.8446e-83	1.45795e-103
'SKF'	3.5617e-39	4.7136e-50	2.5605e-101
'KUF'	8.6584e-110	4.3814e-113	2.6695e-117
'RVF '	3.7108e-87	3.8447e-83	1.4579e-103

The features presented in the above table were arranged in an array where the first

9 probabilities are from X-axis features, then Y-axis and Z-axis. The array was sorted in a descending manner, the lower feature probabilities were placed at the top of the array, and the indexes of those column features were taken for selection purposes. Several selections were made for testing, starting from 1 attribute and increasing progressively to a final selection of 12 features. These features selected were normalized and used in the machine learning models with the respective accuracy computing. Table 4.2 shows the index of the twelve most discriminative features after the sort; the features are VMF, RV, KUF for the X-axis; VMF, MP, KUF and 1st SP for the Y-axis; and VMF, MP, 1st SP, KUF, and RV for the Z-axis. Figure 4.13 shows a 2D and 3D representation of the best features sorted, where a good separation among conditions is observed.

Table 4.2. Index of the features sorted

Axis	Z	Y	X	Z	Z	Z	Y	Z	X	Y	Y	X
Feature index	22	13	4	20	21	26	11	23	5	17	12	8

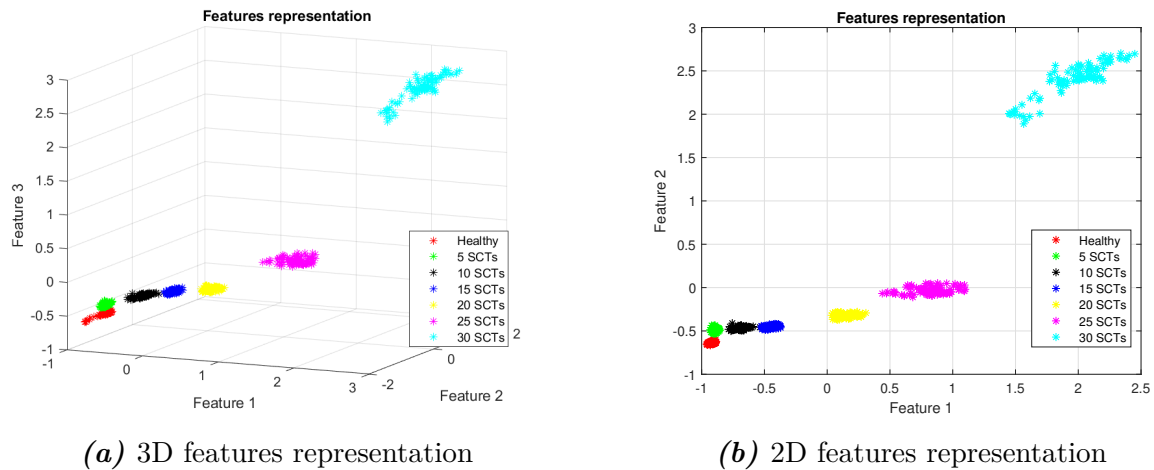


Figure 4.13. Features selected with Kruskal - Wallis test

4.2.5 Classification and fault identification

In this step of the methodology, three algorithms were trained and tested to classify the SCTs faults in the windings. The selected statistical features that allow a good separation among classes are used with the K-NN, Naive Bayes, and SVM models. Each model has its internal parameters set to achieve the best accuracy possible, and the validation was done with K-fold cross-validation, where the number of folds was selected in 10 to achieve this task; each fold consisted of 90 samples for training and 10 samples for testing per class, randomly shuffled

as the K-fold was selected. In order not to show confusion matrices for the classification result for all the feature combinations, only the results for the four features selected will be shown since the best results were achieved with this feature combination.

4.2.5.1 K-NN classifier results

The K-NN classifier was used with the Euclidean distance from the equation 2.6; according to the expression of equation 2.9, the optimal K value is 10 neighbors given the number of samples. Figure 4.14 shows the confusion matrix of the classification results. The model reached an accuracy of 99.86%, delivering an appropriate classification result with two samples misclassified between the healthy state and the 5 SCTs conditions.

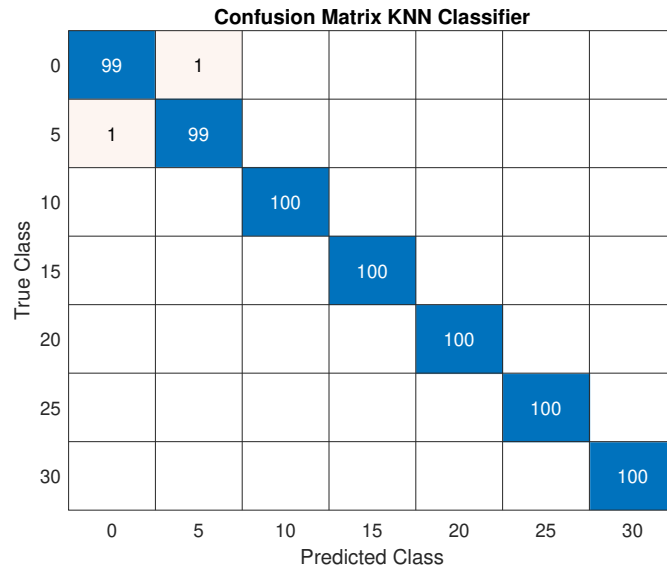


Figure 4.14. Confusion matrix K-NN classifier results

4.2.5.2 Naive Bayes classifier results

The Naive Bayes classifier was applied using the function from the MATLAB toolbox; its function parameters were the normalized dataset with the features selected by the technique, the labels vector, and the option for Kfold cross-validation. The model was trained and validated. Figure 4.15 shows the confusion matrix of classification results. The model achieved a perfect classification in all the conditions.

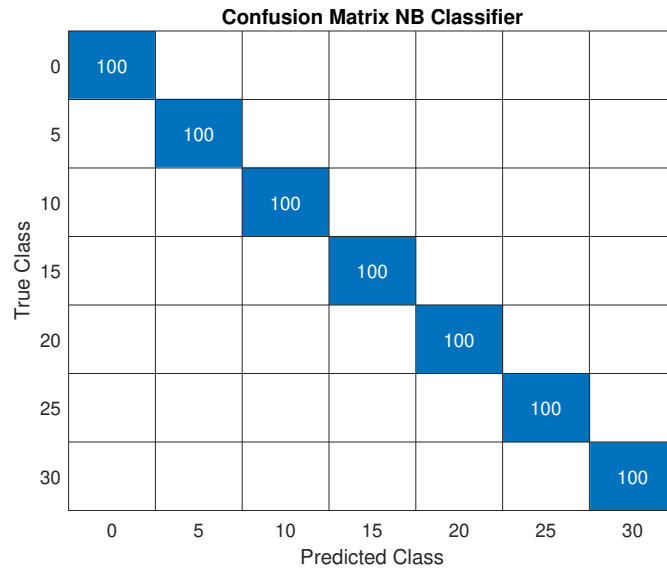


Figure 4.15. Confusion matrix Naive Bayes classifier results

4.2.5.3 SVM classifier results

The SVM was trained using the kernel trick template of polynomial grade 2; this template was added to the function and the normalized dataset, the labels vector, and the K-fold cross-validation. Figure 4.16 shows the confusion matrix of the SVM model; the model achieved a perfect classification in all the conditions.

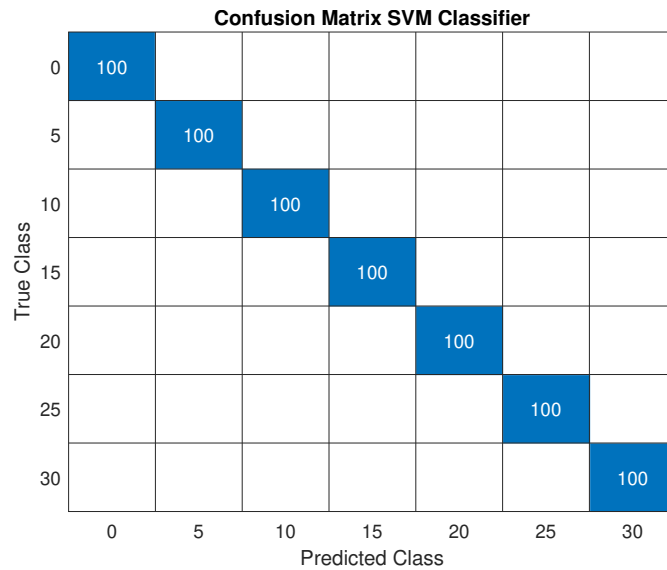


Figure 4.16. Confusion matrix SVM classifier results

4.2.6 Dimensionality reduction with PCA

An alternative dimension reduction method was tested to compare the results with the previous technique. The PCA managed and reduced the dimensions of the features to avoid redundant or noisy data that could lower the performance of the classification models. For this, the feature matrix with all the conditions was normalized to start the subspace transformation, and the aforementioned normalization was performed on each of the 27 feature columns. The PCA function provided by MATLAB was used and, after execution, returned the eigenvalues calculated from the covariance matrix and the coefficient matrix for transformation. When the eigenvalues were analyzed, it was noted that 15 of the 27 eigenvalues maintained 99.9% of the source energy of the data. Following the procedure for the PCA algorithm, every eigenvalue has an eigenvector, so those 15 principal components were selected in the eigenvector features matrix to transform the original standardized features. Once completed, a new subspace of features was created, as shown in figure 4.17, where a good separation is observed among the seven conditions. It is worth mentioning that the three most significant features out of the fifteen were plotted, as a 3-D graphic represents the maximum achievable dimensionality.

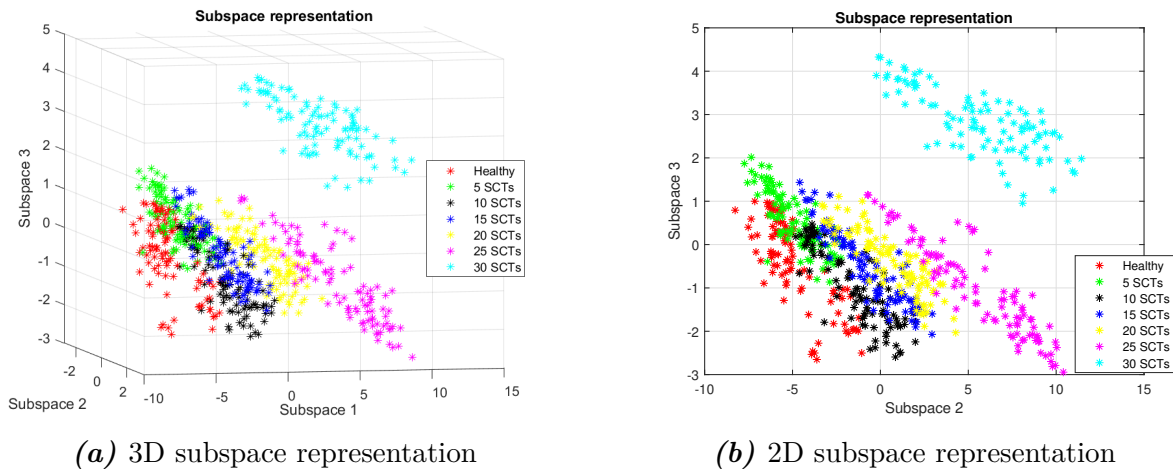


Figure 4.17. Subspace created after applying PCA

4.2.7 Classification and fault identification

As well as in the Kruskal - Wallis method, the same three algorithms were trained, and the validation of each model was also done with the same K-fold for cross-validation to classify the SCTS faults.

4.2.7.1 K-NN classifier results

The K-NN classifier was used with the Euclidean distance. Figure 4.18 shows the confusion matrix of the K - NN model with that optimal K value ($K = 10$). The model reached an accuracy of 94.29%, delivering a good classification result even when some samples were misclassified.

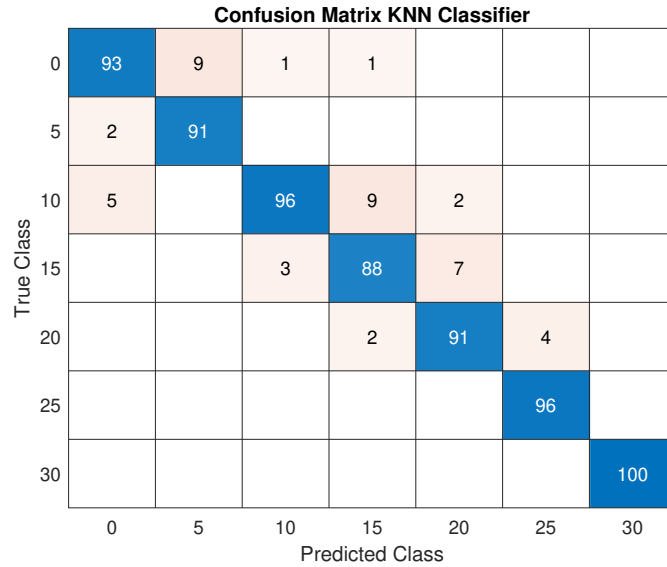


Figure 4.18. Confusion matrix K-NN classifier results

4.2.7.2 Naive Bayes classifier results

The Naive Bayes classifier was applied using the function from the MATLAB toolbox; the model was trained for assigning a label to a new sample and validated with K-fold cross-validation. Figure 4.19 shows the confusion matrix of the Naive Bayes model. The model obtained a regular classification among the healthy state and 5, 10, and 15 SCTs conditions; the accuracy of the model was 96.29%.

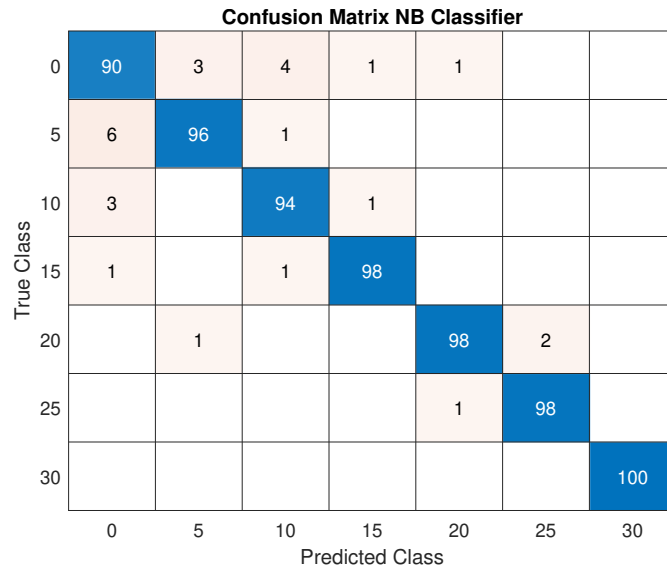


Figure 4.19. Confusion matrix Naive Bayes classifier results

4.2.7.3 SVM classifier results

The SVM algorithm classifier was trained using the kernel trick template of polynomial grade 2; this template was added to the function. Figure 4.20 shows the confusion matrix of the SVM model. The model achieved a good classification in almost all the conditions, and the accuracy of this model was 98.71%. There were misclassifications between the healthy state and 5 SCTs condition and the 15 with 20 SCTs conditions.

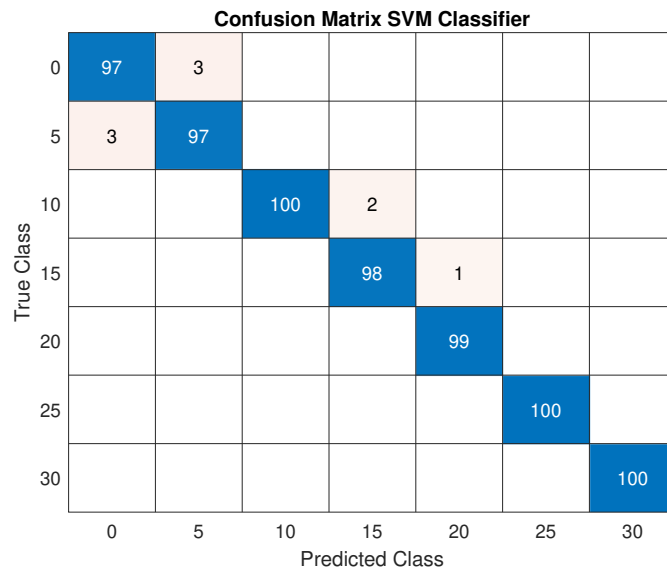


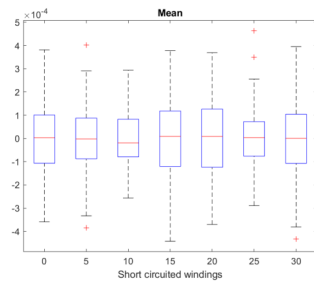
Figure 4.20. Confusion matrix SVM classifier results

4.3 Time domain analysis

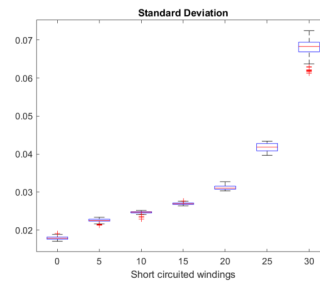
An alternative study for the vibration signals was done with a time domain analysis of the raw signals for assessing conditions in the windings of electrical transformers. Several steps were followed to achieve the classification. Firstly, the offset of each axis signal was removed; then, every signal was divided into five segments as aforementioned in previous steps for oversampling and increasing the database from 140 signals to 700 raw signals for all the conditions. Finally, statistical features were computed and arranged in a dataset.

4.3.1 Statistical features extraction

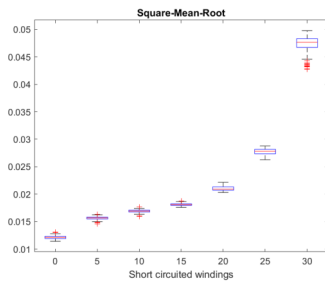
The 20 STFs from table 3.2 were computed from one portion sample and arranged in an array of the shape 1×20 . After iterating by condition, test, and segments, all features were arranged in a matrix, where one condition had 100×20 for A_x , 100×20 for A_y , and 100×20 for A_z and stored as Microsoft access tables (mat files) for selection and reduction in further steps. Figures 4.21, 4.22 show a boxplot representation of A_x vibration, 4.23, 4.24 of A_y , and 4.25, 4.26 of A_z . Each boxplot represents one statistical feature and its behavior across all the operating conditions in the windings. Visually, this tool is quite helpful in the discovery of those attributes capable of separating and classifying an abnormality. At the same time, it is possible to corroborate it with mathematical analysis.



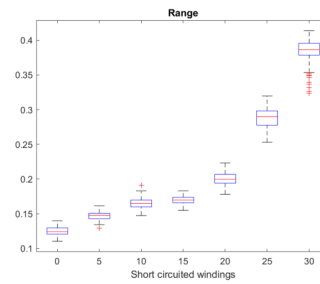
(a) Mean



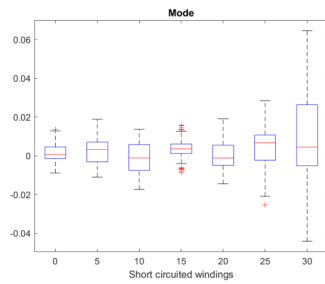
(b) Standard Deviation



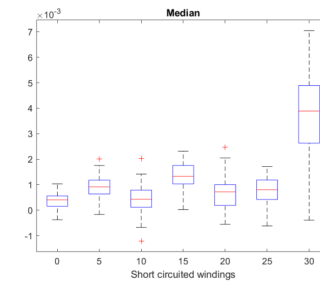
(c) Square Mean Root



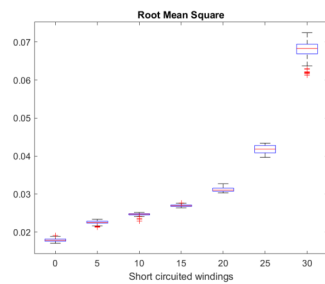
(d) Range



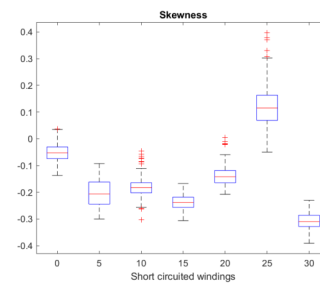
(e) Mode



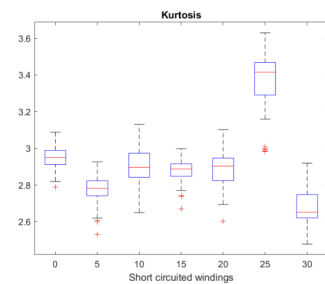
(f) Median



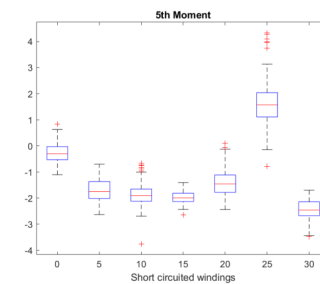
(g) Root Mean Square



(h) Skewness



(i) Kurtosis



(j) 5th Moment

Figure 4.21. Boxplot representation of all conditions features extracted axis X, Part A

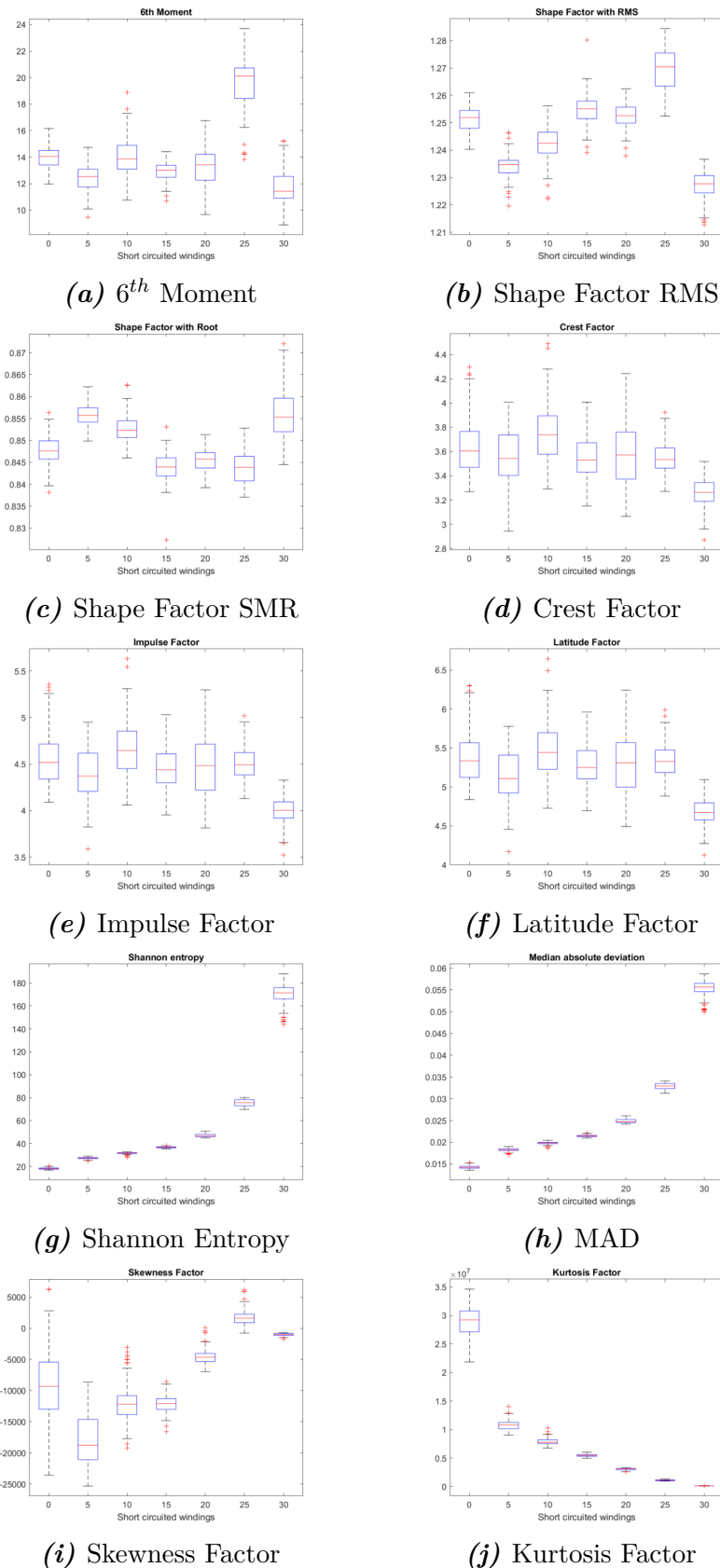


Figure 4.22. Boxplot representation of all conditions features extracted axis X, Part B

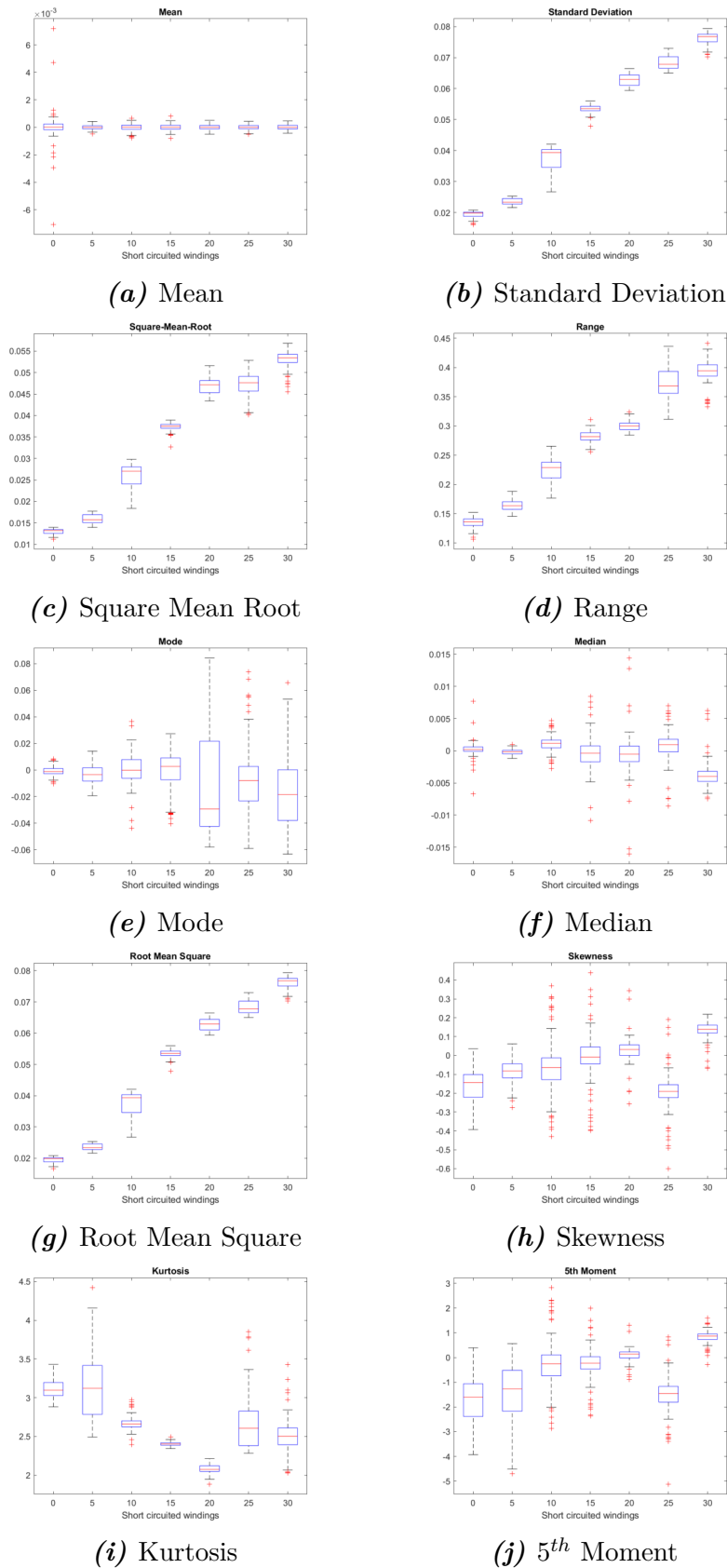


Figure 4.23. Boxplot representation of all conditions features extracted axis Y, Part A

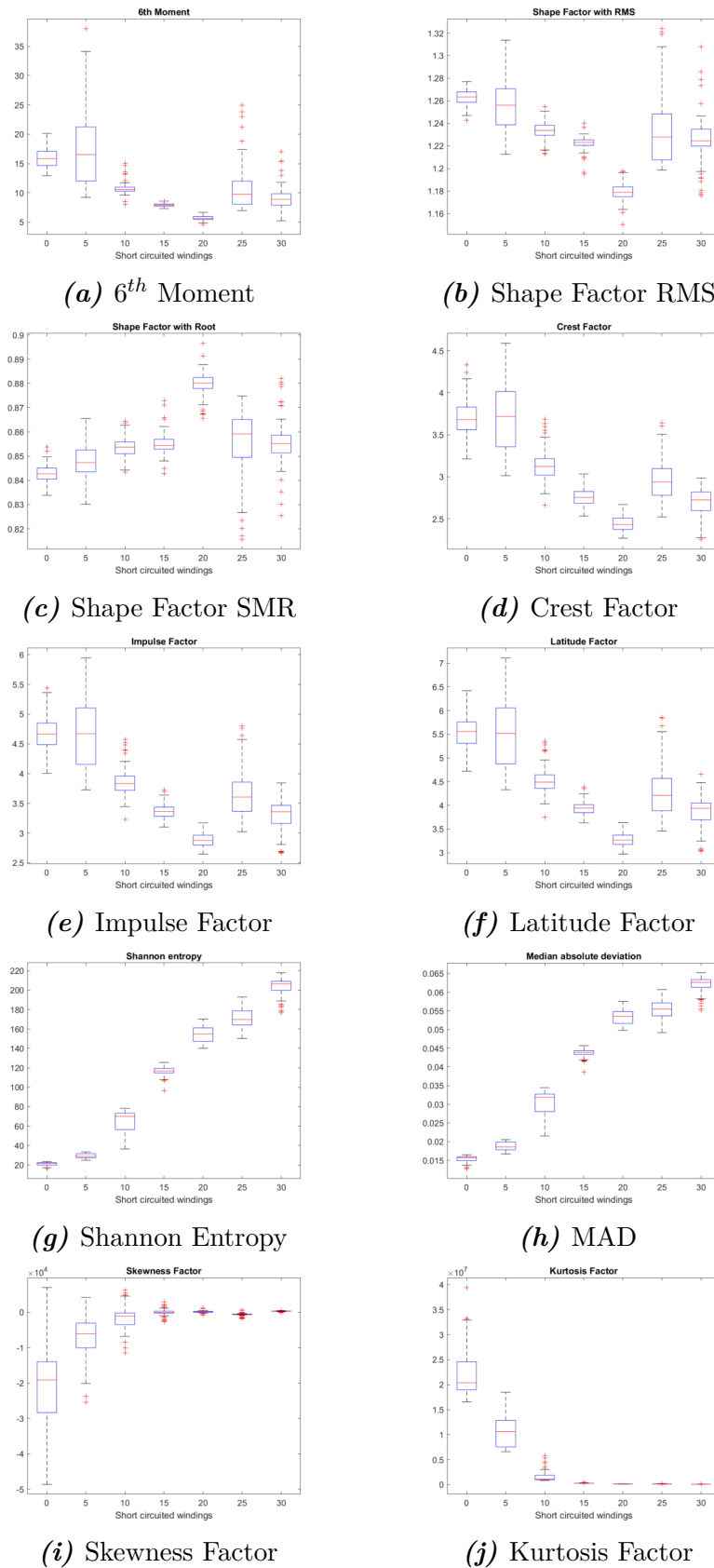


Figure 4.24. Boxplot representation of all conditions features extracted axis Y, Part B

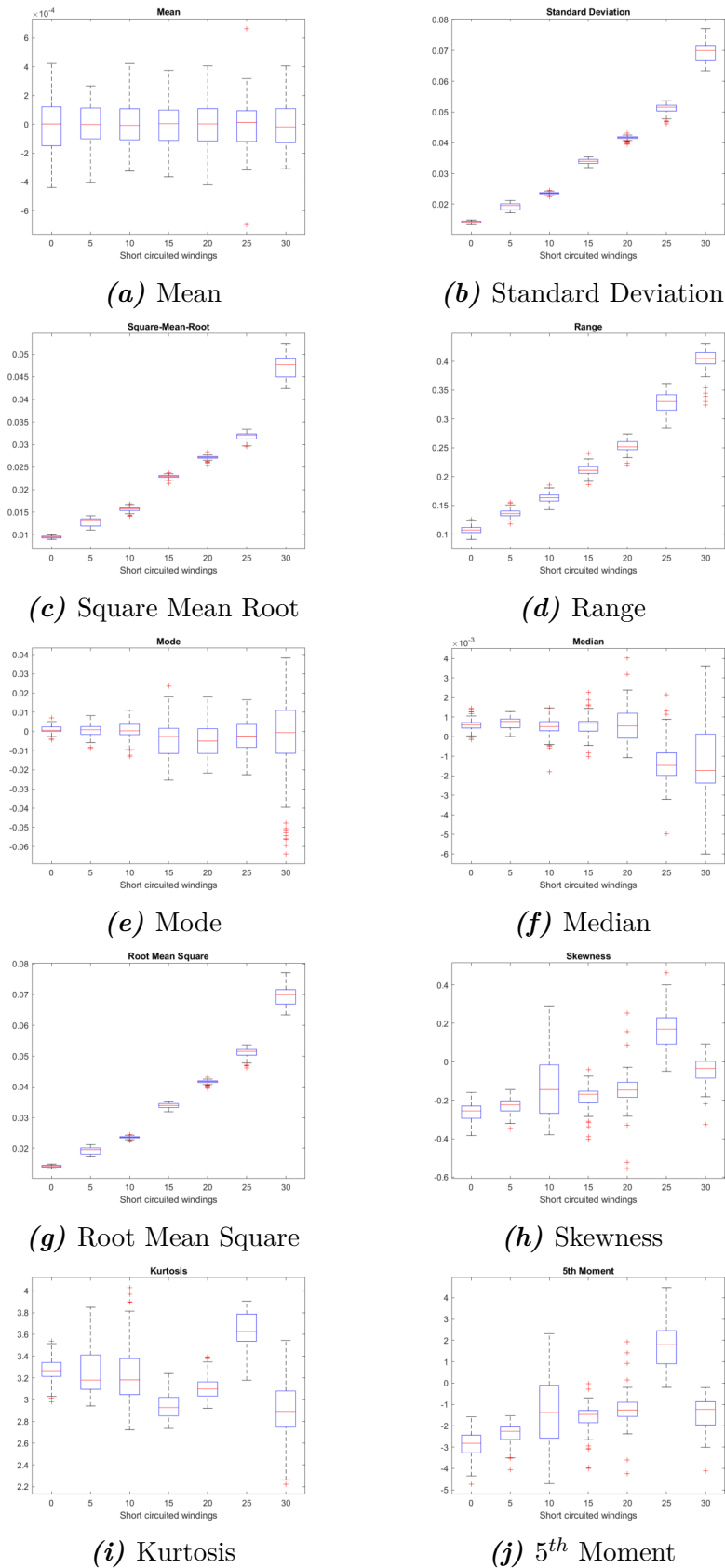


Figure 4.25. Boxplot representation of all conditions features extracted axis Z, Part A

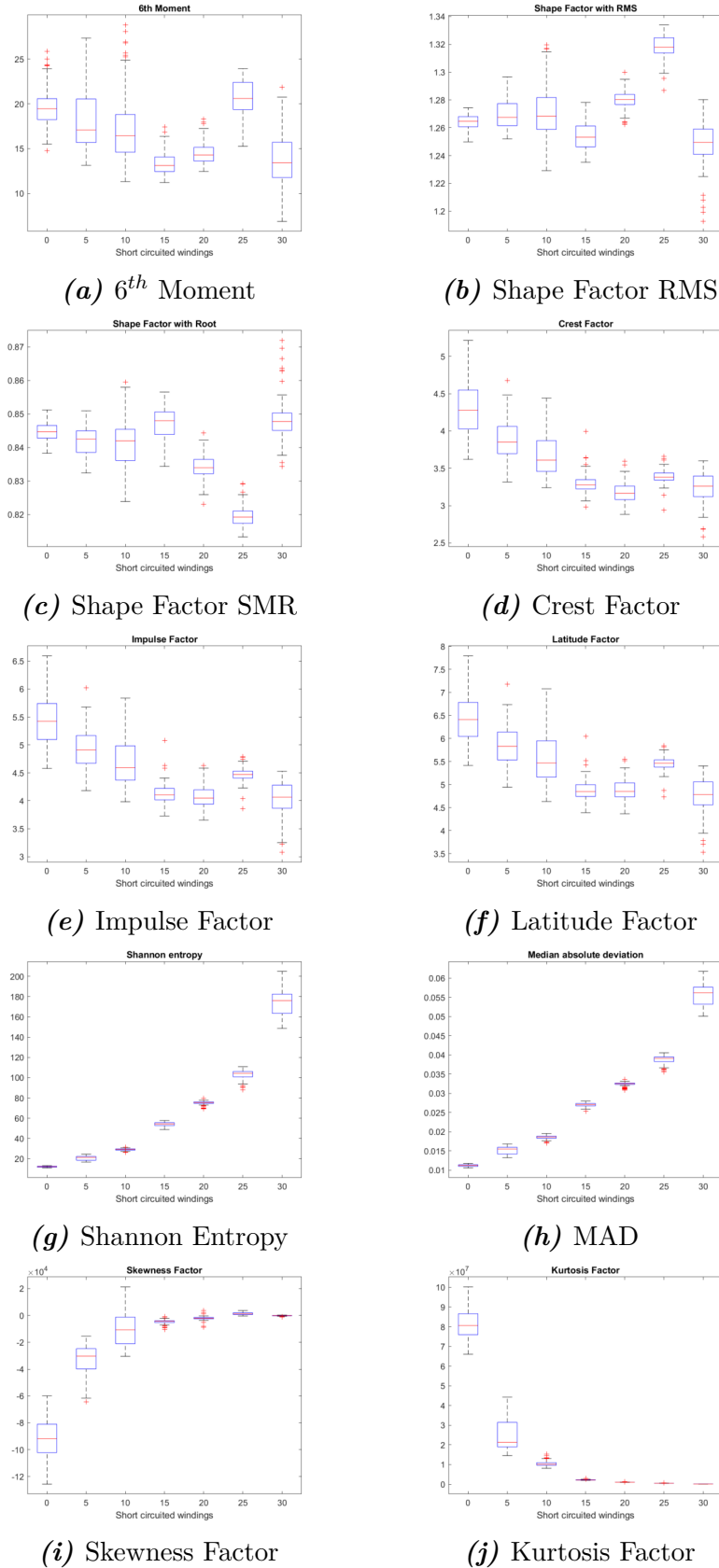


Figure 4.26. Boxplot representation of all conditions features extracted axis Z, Part B

4.3.2 Features selection

4.3.2.1 Kruskal Wallis test

The Kruskal-Wallis test was used in the time domain to identify the most effective attributes for detecting SCTs faults in the windings. The method computed probabilities by separately comparing the same feature across all conditions and axes, as in the frequency domain. The procedure applied iterative importation of 100×20 shape mat files, with the feature matrix stored in a variable, resulting in a total matrix size of 100×140 . The first 20 columns represent features in the healthy state, columns 21 to 40 contain 5 SCTs features, and so on. Following a similar algorithm procedure as in the frequency domain, a loop was used to compute the probability for each feature. For instance, in the first iteration, the function considered the mean of the signal of the healthy state (column 1), the 5 SCTs (column 21) until 30 SCTs (column 61) to compute the probabilities. This process was repeated for all features, and the resulting probabilities were stored in an array. The same algorithm was applied for the features from the Y-axis and the Z-axis, and these results were arranged in arrays for easier interpretation.

The arrays with the results for each axis were arranged; table 4.3 presents the probabilities computed. For the selection, the features were rearranged in a single array, where the first twenty positions are X-axis features, the next twenty are Y-axis, and the last twenty are Z-axis features. The selection was made by sorting in descending order the features with lower probabilities at the top of the list, and the column index of those features was used to create a final features dataset to be used in the machine learning models. Several selections were made, and the models were tested, starting from one feature and finishing with fifteen features. The probabilities computed validated appropriately the separation among conditions of some features visually achieved by the boxplots of the above figures, e.g., features such as standard deviation (SD), square mean root (SMR), root mean square (RMS), and median absolute deviation (MAD) are statistical parameters that allow a good separation among classes.

Table 4.4 shows the index of the fifteen most discriminative features after the sort; the features are SD, RMS, SE, MAD, KUF, and RG for the Z-axis; SMR, SD, and RMS for the Y-axis; and SD, RMS, SE, MAD, SMR, and KUF for X-axis. Figure 4.27 shows a 2D and 3D representation of the first three features sorted, where a good separation among conditions is observed. Finally, the features were normalized to be suitable for machine learning models.

Table 4.3. Probabilities computed with Kruskal Wallis test

<i>Kruskall Wallis test</i>			
Legend	Probability Vx	Probability Vy	Probability Vz
'ME'	0.9996	0.9998	0.9999
'SD'	1.2616e-144	2.2834e-144	1.2076e-144
'SMR'	1.2832e-144	6.1441e-141	1.2106e-144
'RG'	7.0930e-142	6.6995e-142	2.2002e-144
'MO'	5.4251e-07	5.0096e-13	1.5642e-08
'MED'	1.8572e-79	1.412e-52	1.4223e-53
'RMS'	1.2616e-144	2.2834e-144	1.2076e-144
'SKW'	3.8409e-128	2.5899e-87	2.738e-96
'KUR'	9.995e-103	1.7382e-110	8.3107e-87
'5 th M'	6.867e-113	9.9500e-94	3.4212e-90
'6 th M'	1.6561e-84	1.0897e-115	8.144e-88
'SFrms'	6.4771e-124	5.0409e-99	4.1933e-98
'SFsmr'	9.7042e-106	1.4540e-86	2.6103e-98
'CF'	3.3555e-47	6.0420e-124	1.1967e-108
'IF'	8.5219e-51	1.2108e-122	4.8458e-109
'LF'	7.4925e-53	3.2676e-121	1.0518e-107
'SE'	1.2646e-144	1.6294e-143	1.2076e-144
'MAD'	1.2739e-144	3.3098e-142	1.2076e-144
'SKF'	6.6971e-119	5.1048e-99	3.5892e-125
'KUF'	1.3434e-144	2.1379e-141	1.2106e-144

Table 4.4. Index of the features sorted

Axis	Z	Z	Z	Z	Y	Z	X	X	X	X	X	X	Z	Y	Y
Feature index	42	47	57	58	43	60	2	7	17	18	3	20	44	22	27

4.3.3 Classification and fault identification

The three algorithms to classify SCTs faults in the windings used in the frequency domain were replicated in the time domain. The chosen statistical features effectively distinguish between different classes and were utilized in combination with K-NN, Naive Bayes, and SVM models. To evaluate the performance of each model, K-fold cross-validation is employed. The samples for testing are randomly shuffled for each selected fold. Subsequently, the algorithms undergo comprehensive testing for their usefulness and accuracy. To avoid displaying many classification results for each set of features selected, only the classification with ten features is displayed in all the models.

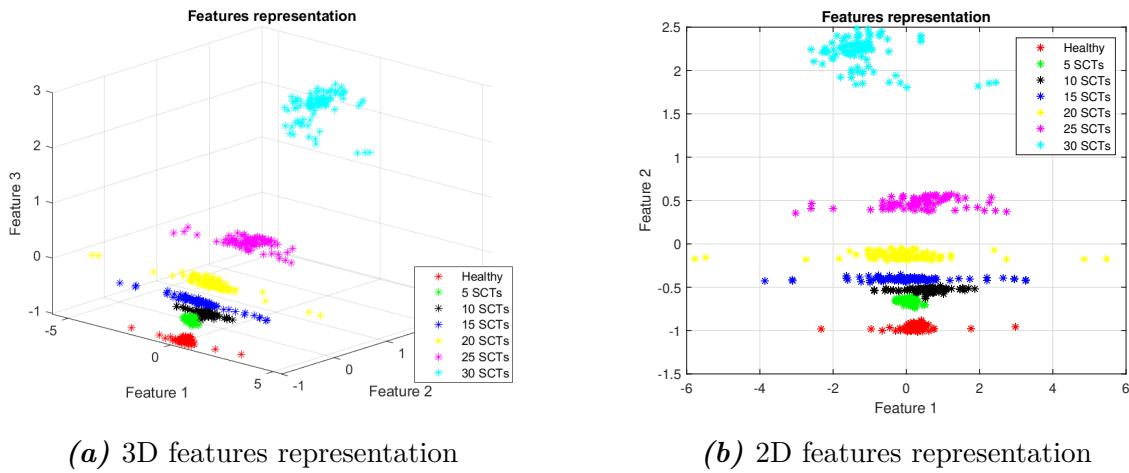


Figure 4.27. Features selected with Kruskal - Wallis test

4.3.3.1 K-NN classifier results

The K-NN classifier was employed, utilizing the Euclidean distance as per Equation 2.6. The optimal neighbor value computed was $K = 10$. Figure 4.28 shows the confusion matrix of the K-NN model, which reached an accuracy of 100%, delivering an excellent classification result. All the conditions were classified correctly, and the healthy state was distinguished from the SCTs conditions.

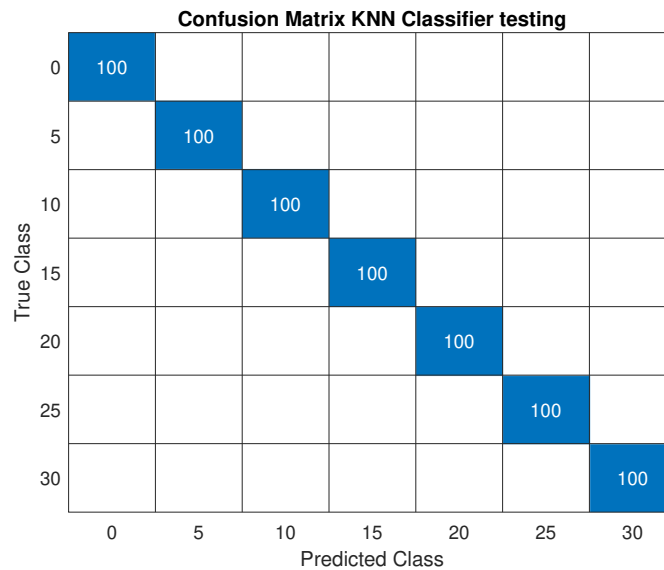


Figure 4.28. Confusion matrix K-NN classifier results

4.3.3.2 Naive Bayes classifier results

The Naive Bayes classifier was applied using the function from the MATLAB toolbox. The model was trained for assigning a label to a new sample and validated with K-fold cross-validation. Figure 4.29 shows the confusion matrix of the Naive Bayes model. The model achieved a significant performance with an accuracy of 99.86%; only one sample was misclassified between the 5 and 10 SCTs.

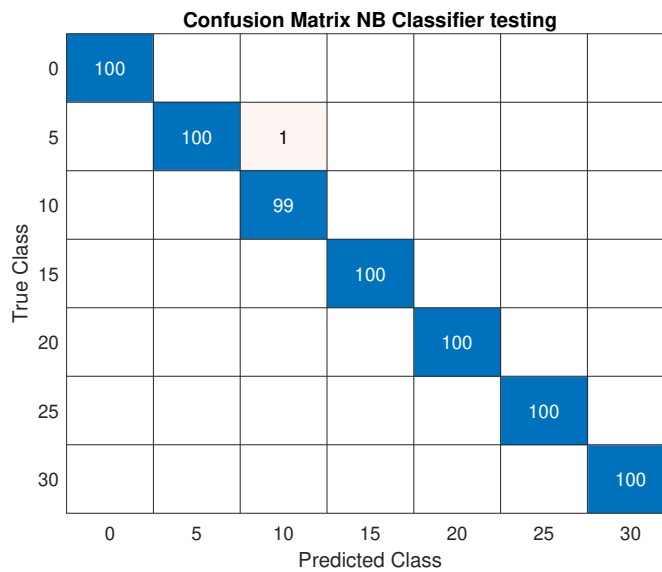


Figure 4.29. Confusion matrix Naive Bayes classifier results

4.3.3.3 SVM classifier results

The SVM algorithm classifier was trained using the kernel trick template of polynomial grade 2; this template was added to the function, as well as the normalized dataset, the labels vector, and the option for cross-validation testing. Figure 4.30 shows the confusion matrix of the SVM model. The model achieved a remarkable accuracy of 100%, indicating an outstanding classification performance. It correctly classified all the conditions, successfully distinguishing the healthy state from the SCTs conditions.

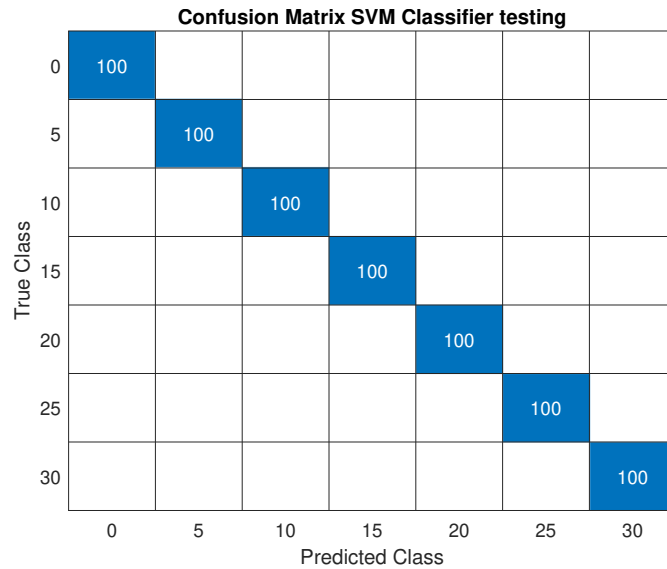


Figure 4.30. Confusion matrix SVM classifier results

4.3.4 Dimensionality reduction with PCA

The PCA technique was employed in the time domain as an alternative for dimensionality reduction more suitable for machine learning models; normalization was applied to all feature columns. The MATLAB function specifically designed for PCA was utilized, and the feature matrix was processed with this function. The outcome of the function included eigenvalues, which were calculated based on the covariance matrix and the coefficient matrix required for transformation.

During the analysis of eigenvalues, it became evident that 30 out of the 60 eigenvalues retained approximately 99.9% of the original data energy. Consequently, following the standard procedure for the PCA algorithm, each eigenvalue was associated with an eigenvector. In this case, the 30 principal components corresponding to these eigenvectors were selected to create the eigenvector features matrix, which was used to transform the original standardized features. Once completed, a new subspace of features is generated, as shown in figure 4.31, where a good separation is observed among the seven conditions. The three most significant features out of the thirty were plotted, as a 3-D graphic represents the maximum achievable dimensionality.

4.3.5 Classification and fault identification

As well as in the frequency domain, the focus was on training and testing three distinct

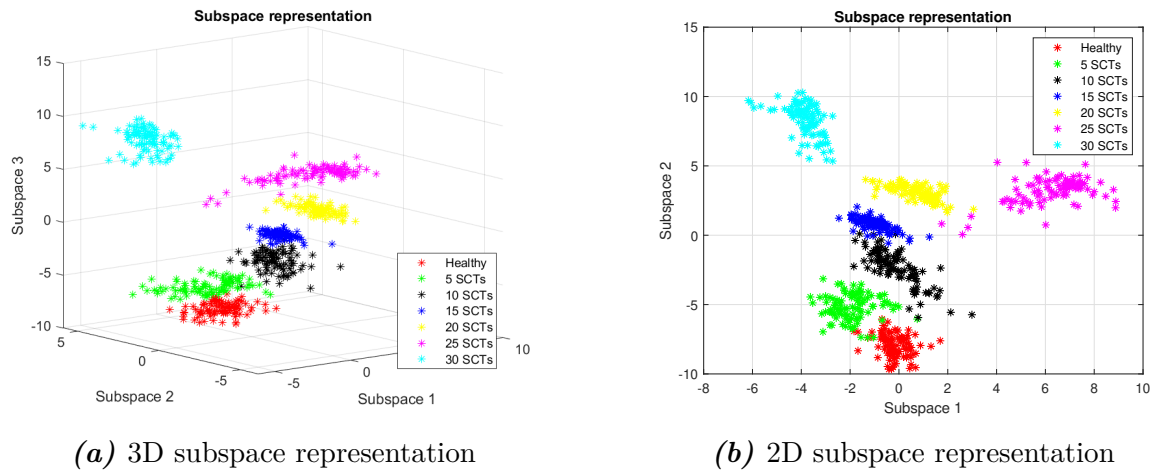


Figure 4.31. Subspace created after applying PCA

algorithms to classify SCTS faults effectively. The algorithms employed for this task were the K-NN, Naive Bayes, and SVM models. A K-fold cross-validation approach was adopted to evaluate the performance of each model comprehensively.

4.3.5.1 K-NN classifier results

The K-NN classifier was used with the Euclidean distance. Figure 4.32 shows the confusion matrix of the K - NN model with that optimal K value ($K = 10$). The model reached an accuracy of 99.57%, delivering an appropriate classification result even when some samples were misclassified along the 5 SCTs and 10 SCTs conditions.

4.3.5.2 Naive Bayes classifier results

The Naive Bayes classifier was applied using the function from the MATLAB toolbox.

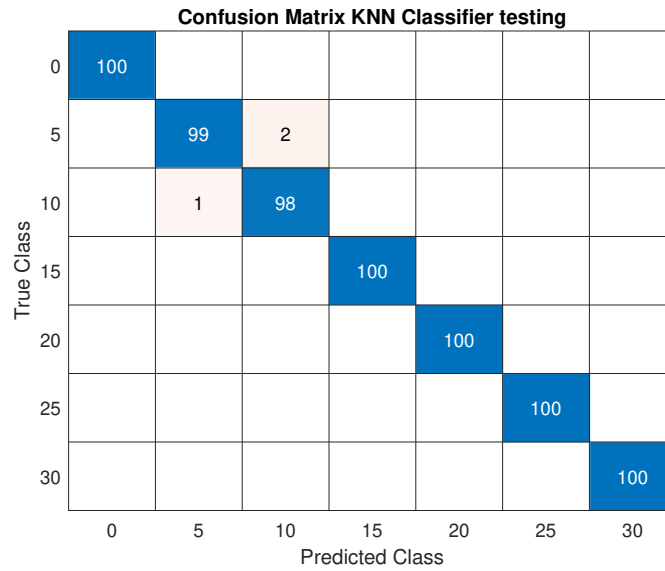


Figure 4.32. Confusion matrix K-NN classifier results

The model was trained for assigning a label to a new sample and validated with K-fold cross-validation. Figure 4.33 shows the confusion matrix of the Naive Bayes model. The model achieved 98.71% accuracy, delivering an appropriate classification result; some samples between 10 and 15 SCTs conditions were misclassified.

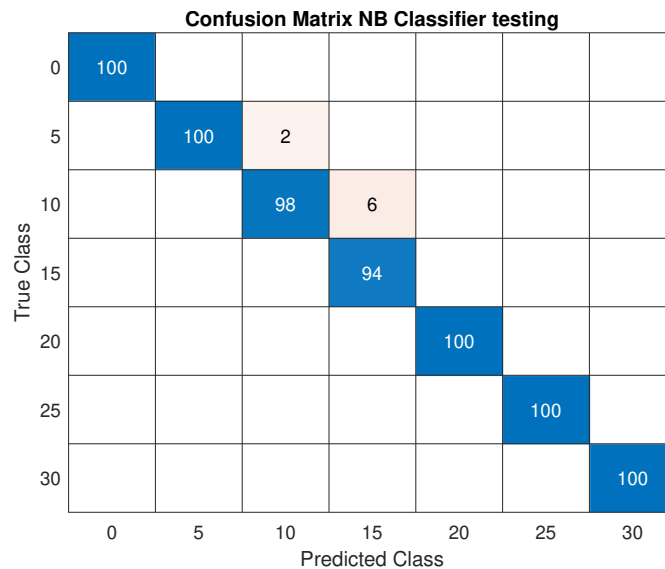


Figure 4.33. Confusion matrix Naive Bayes classifier results

4.3.5.3 SVM classifier results

The SVM algorithm classifier was trained using the kernel trick template of polynomial

grade 2; this template was added to the function, as well as the normalized dataset, the labels vector, and the option for cross-validation testing. Figure 4.34 shows the confusion matrix of the SVM model. The model only misclassified a sample of 25 SCTs conditions and achieved 99.86% classification accuracy.

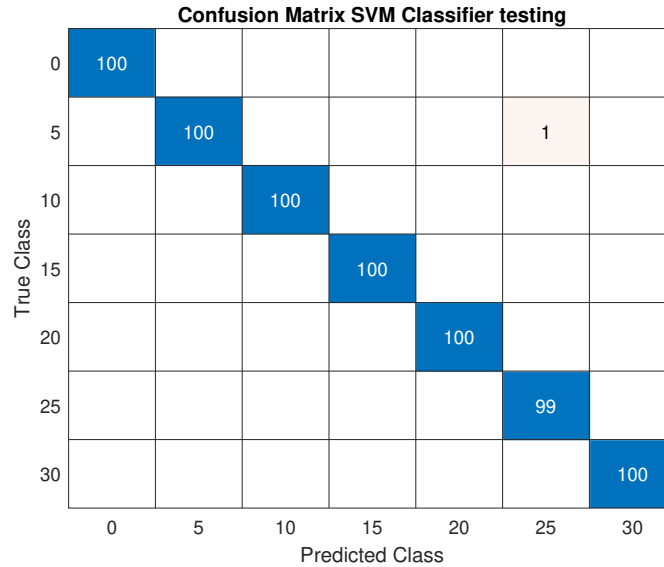


Figure 4.34. Confusion matrix SVM classifier results

4.4 Results discussion

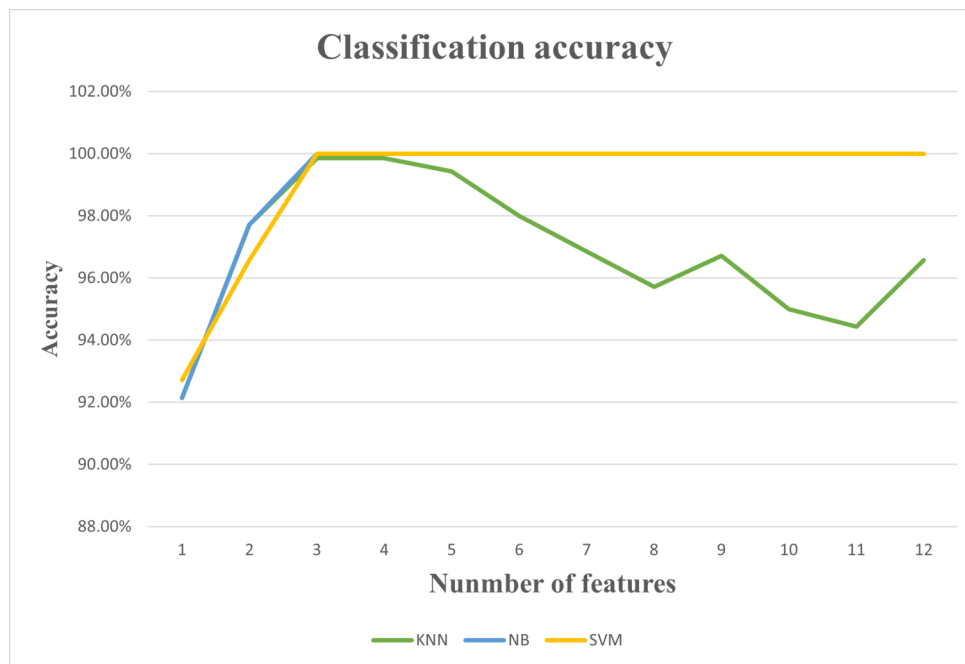
After displaying all the classification results achieved with each selection technique and machine learning model in the time and frequency domains, the discussion and comparison deliver the final thoughts of the study with these vibration signals.

In the frequency domain, there is a marked difference between the classification results achieved, where the features selected using the Kruskal - Wallis test performed better than space transformation with PCA. However, it is worth noting that with features selected, the K-NN model reached a maximum classification, and that performance decreased as the number of features increased (all tests were done with the same neighbor optimal value). The other models remained with constant classification results as the feature increased. Table 4.5 shows the classification result accuracies with different feature selections; the peak performance of K-NN was achieved with four features, and figure 4.35 shows the graphical representation of these classification results.

The transformation made with PCA performed in the frequency domain regularly;

Table 4.5. Classification results with features selection in the frequency domain

Feats	KNN	NB	SVM
1	92.14%	92.14%	92.71%
2	97.71%	97.71%	96.57%
3	99.86%	100%	100%
4	99.86%	100%	100%
5	99.43%	100%	100%
6	98%	100%	100%
7	97%	100%	100%
8	95.71%	100%	100%
9	96.71%	100%	100%
10	95%	100%	100%
11	94.43%	100%	100%
12	96.57%	100%	100%

**Figure 4.35.** Classification results with machine learning models and Kruskal - Wallis

while it was able to separate the conditions, this separation between the first 4 conditions was not very effective. Studying the features computed, a slight statistical variation was noted in most of them, and a small group of features delivered a significant difference between the conditions. The decrement in the classification results achieved with the models evidences the small separation between some groups of conditions.

In the time domain study, steady results are found in the classifications achieved with

the PCA and Kruskal - Wallis techniques, where both had a suitable performance, capable of separating the classes satisfactorily. Small variations were located in the K-NN and SVM techniques. For example, K-NN delivered a perfect classification with Kruskal - Wallis and the same classifier misclassified three samples between 5 and 10 SCTs after applying PCA, SVM performed accurately with Kruskal - Wallis and only misclassified one sample of 25 SCTs condition. Furthermore, Naive Bayes had slight variation comparing both techniques, whereas Kruskal - Wallis only misclassified one sample, and the same technique after using PCA misclassified eight samples between 10 and 15 SCTs. Finally, the results explain that both techniques are suitable for feature selection or dimensionality reduction in the time domain with this dataset of signals.

Unlike what was observed achieved in the frequency domain with feature selection after applying Kruskal - Wallis, in the time domain, after several selections starting from the first three and finishing in the first fifteen features, the classification results with the three techniques remain constant or steady. An explanation for this is that a big group of features in the time domain rightly separate the conditions, or in other words, those statistical parameters reflected the significant variations of the vibration signals as the SCTs conditions changed. Table 4.6 shows the classification result accuracies with different feature selections in the time domain, and figure 4.36 shows the graphical representation of these classification results.

Table 4.6. Classification results with features selection in the time domain

Feats	KNN	NB	SVM
3	100%	99.86%	100%
4	100%	99.86%	100%
5	100%	99.86%	100%
6	99.86%	99.71%	100%
7	99.86%	99.71%	100%
8	99.86%	99.71%	100%
9	99.71%	99.71%	100%
10	99.86%	99.86%	100%
11	100%	99.71%	100%
12	99.71%	99.71%	100%
13	100%	99.71%	100%
14	100%	99.71%	100%
15	100%	99.71%	100%

The execution time of the machine learning models was another exceptional outcome of the methodology. The elapsed time for classifying a new sample after the model is trained and validated was measured since the good separation between the categories inferred a fast

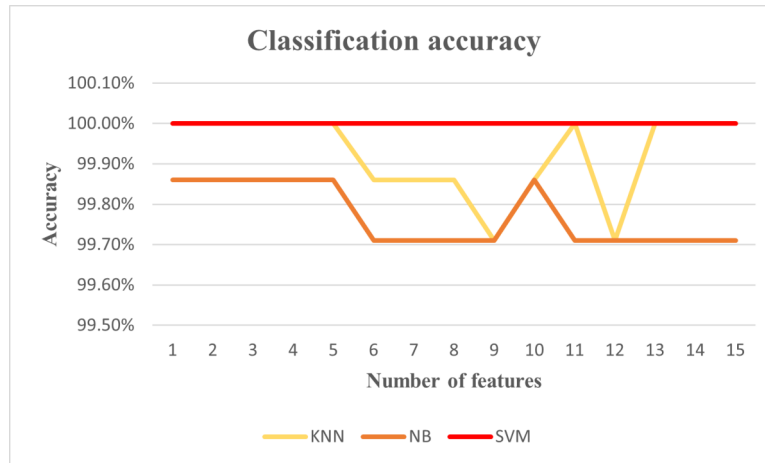


Figure 4.36. Classification results with machine learning models and Kruskal - Wallis

classification with these models proposed; from this, the average classification time obtained to classify a new testing sample was 1.66 *ms* with K - NN, 881.25 μs with Naive Bayes, and 800 μs with SVM.

A final remark of the results is that the methodology was capable of assessing the SCTs in the windings without an intricate computational process or complex signal-processing technique and in a fast manner. At the same time, conducting the features extraction, selection, and classification of the conditions was successfully achieved with this dataset.

5.1 Conclusion

This thesis work proposed and conducted a methodology based on integrating vibration signals processed in the time and frequency domains, extracting statistical features in both domains and pattern recognition tools to classify SCTs in electrical transformers. The conditions for classification included a healthy state (0 SCTs) and six levels of SCTs severities. The vibration signals were acquired in three axes with an accelerometer from a single-phase transformer with no load adjusted to emulate the different SCTs conditions.

The methodology was executed in both domains independently for a comparative analysis of the results. The procedure included processing the steady state of the signals, where each one was segmented for oversampling, and the offset level was removed; in the frequency domain, other steps were applied, including a decimation and a computation of the FFT. Then, Statistical features were computed on each axis and condition; once the attributes were gathered, they were arranged in a features matrix for pattern recognition purposes. Three algorithms were executed for each technique separately after selecting the best features, primarily with the Kruskal - Wallis test and secondly with PCA as an alternative technique

for dimensionality reduction. The training and validation of each model were done with K-fold cross-validation, and the accuracy of each model was computed to assess the classification performance.

The results show that the statistical features analysis in the time domain delivered the best classification result with lower computing time and in a rapid manner, where SVM achieved an accuracy of 100 % with features selected with the Kruskal - Wallis test and 99.86% after the reduction with PCA, while in the frequency domain, the statistical parameters with Naive Bayes and SVM achieved an accuracy of 100 % after selecting the features with Kruskal - Wallis test. However, the classification performance slightly decreased after applying PCA in this domain. It is worth noting that the misclassifications in some conditions could be improved by employing other statistical parameters capable of delivering a substantial separation between the healthy state, 5 and 10 SCTs conditions. A more minor setback of the frequency domain study is a slightly higher computation time since executing two more processing steps is necessary. However, there is no significant gap in computing the machine learning models in each domain. At the same time, as in previous research studies, the use of statistical features in both domains allowed a good identification of abnormalities in these electrical machines.

On the one hand, the main advantage of the frequency domain approach and these vibration signals is that with the FFT, which is one of the most basic techniques for spectrum computing and domain transformation, it was possible to precisely assess the operation conditions in the windings without the need of using signal processing techniques with higher computational cost and therefore longer processing times. On the other hand, in the time domain approach, the method achieved excellent classification results with little computation time using the raw steady state of the signal. Yet, it should be considered that this is a restrictive model since vibration signals acquired from a transformer with various types of loads (lineal and no lineal) and configurations should be studied to confirm that with a domain transformation and Fourier analysis is possible to assess the operating conditions in the device.

A significant contribution of this work is the use of straightforward signal processing techniques and, consequently, the improvement of the processing times than the predecessors achieved with more complex processing tools and machine learning models. From this, it follows that these particular signals can be studied and classified accurately in less time, and the methodology can be extrapolated for a real-time processing environment.

References

- [Abu-Siada et al., 2013] Abu-Siada, A., Hashemnia, N., Islam, S., and Masoum, M. A. (2013). Understanding power transformer frequency response analysis signatures. *IEEE Electrical Insulation Magazine*, 29(3):48–56.
- [AJ et al., 2018] AJ, C., Salam, M., Rahman, Q., Wen, F., Ang, S., and Voon, W. (2018). Causes of transformer failures and diagnostic methods – a review. *Renewable and Sustainable Energy Reviews*, 82:1442–1456.
- [Ali et al., 2022] Ali, K., Shaikh, Z. A., Khan, A. A., and Laghari, A. A. (2022). Multiclass skin cancer classification using efficientnets – a first step towards preventing skin cancer. *Neuroscience Informatics*, 2(4):100034.
- [Bashar and Bhuiyan, 2016] Bashar, S. K. and Bhuiyan, M. I. H. (2016). Classification of motor imagery movements using multivariate empirical mode decomposition and short time fourier transform based hybrid method. *Engineering Science and Technology, an International Journal*, 19(3):1457–1464.
- [Bath, 2022] Bath, R. (2022). Inside transformers. Accessed on April 11, 2023.
- [Benedetto and Benedetto, 2011] Benedetto, A. and Benedetto, F. (2011). Remote sensing of soil moisture content by gpr signal processing in the frequency domain. *IEEE Sensors Journal*, 11(10):2432–2441.

REFERENCES

- [Calvert, 2006] Calvert, J. (2006). Inside transformers. Accessed on April 11, 2023.
- [Chapman, 2000] Chapman, S. (2000). *Máquinas eléctricas*. McGraw Hill, London.
- [Chen and Lee, 2021] Chen, H.-Y. and Lee, C.-H. (2021). Deep learning approach for vibration signals applications. *Sensors*, 21(11).
- [Contreras-Hernandez et al., 2019] Contreras-Hernandez, J. L., Almanza-Ojeda, D. L., Ledesma-Orozco, S., Garcia-Perez, A., Romero-Troncoso, R. J., and Ibarra-Manzano, M. A. (2019). Quaternion signal analysis algorithm for induction motor fault detection. *IEEE Transactions on Industrial Electronics*, 66(11):8843–8850.
- [Dai et al., 2017] Dai, J., Song, H., Sheng, G., and Jiang, X. (2017). Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. *IEEE Transactions on Dielectrics and Electrical Insulation*, 24(5):2828–2835.
- [Duda et al., 2001] Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley, New York, 2 edition.
- [Garcia et al., 2006] Garcia, B., Burgos, J., and Alonso, A. (2006). Transformer tank vibration modeling as a method of detecting winding deformations-part i: theoretical foundation. *IEEE Transactions on Power Delivery*, 21(1):157–163.
- [Gopika R, 2017] Gopika R, S. D. (2017). Study on power transformer inrush current. *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, 24(5):59–63.
- [Harlow, 2004] Harlow, J. (2004). *ELECTRIC POWER TRANSFORMER ENGINEERING*. CRC Press LLC, London.
- [Heathcore, 1998] Heathcore, M. (1998). *The J P, Transformer book*. Planta Tree, Boston.
- [Heckbert, 1995] Heckbert, P. (1995). Fourier transforms and the fast fourier transform (fft) algorithm. *Computer Graphics*, 2(1995):15–463.
- [Helmi and Forouzantabar, 2019] Helmi, H. and Forouzantabar, A. (2019). Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and anfis. *IET Electric Power Applications*, 13(5):662–669.
- [Hergli et al., 2018] Hergli, K., Marouani, H., and Zidi, M. (2018). Hysteresis identification models: A review. pages 189–198.

REFERENCES

- [Hu et al., 2019] Hu, Y., Zheng, J., and Huang, H. (2019). Experimental research on power transformer vibration distribution under different winding defect conditions. *Electronics*, 8(8).
- [Huerta-Rosales et al., 2020a] Huerta-Rosales, J. R., Granados-Lieberman, D., Amezcua-Sanchez, J. P., Camarena-Martinez, D., and Valtierra-Rodriguez, M. (2020a). Vibration signal processing-based detection of short-circuited turns in transformers: A nonlinear mode decomposition approach. *Mathematics*, 8(4).
- [Huerta-Rosales et al., 2020b] Huerta-Rosales, J. R., Granados-Lieberman, D., Amezcua-Sanchez, J. P., Camarena-Martinez, D., and Valtierra-Rodriguez, M. (2020b). Vibration signal processing-based detection of short-circuited turns in transformers: A nonlinear mode decomposition approach. *Mathematics*, 8(4).
- [Huerta-Rosales et al., 2021] Huerta-Rosales, J. R., Granados-Lieberman, D., Garcia-Perez, A., Camarena-Martinez, D., Amezcua-Sanchez, J. P., and Valtierra-Rodriguez, M. (2021). Short-circuited turn fault diagnosis in transformers by using vibration signals, statistical time features, and support vector machines on fpga. *Sensors*, 21(11).
- [Hussain et al., 2021] Hussain, M. R., Refaat, S. S., and Abu-Rub, H. (2021). Overview and partial discharge analysis of power transformers: A literature review. *IEEE Access*, 9:64587–64605.
- [Imandoust and Bolandraftar, 2013] Imandoust, S. B. and Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background.
- [Islam, 2018] Islam, M.M.; Lee, G. H. S. (2018). A review of condition monitoring techniques and diagnostic tests for lifetime estimation of power transformers. *Electrical Engineering*, 100(2):581–605.
- [Judd et al., 2005] Judd, M., Yang, L., and Hunter, I. (2005). Partial discharge monitoring of power transformers using uhf sensors. part i: sensors and signal interpretation. *IEEE Electrical Insulation Magazine*, 21(2):5–14.
- [Kang et al., 2016] Kang, M., Islam, M. R., Kim, J., Kim, J.-M., and Pecht, M. (2016). A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics. *IEEE Transactions on Industrial Electronics*, 63(5):3299–3310.

REFERENCES

- [Kari et al., 2018] Kari, T., Gao, W., Zhao, D., Abiderexiti, K., Mo, W., Wang, Y., and Luan, L. (2018). Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm. *IET Generation, Transmission, Distribution*, 12(21).
- [Kharwal, 2021] Kharwal, A. (2021). Multinomial naive bayes in machine learning. Accessed on May 22, 2023.
- [khushnandan, 2021] khushnandan, R. (2021). Vibration data and a few techniques to analyze it. Accessed on May 19, 2023.
- [Kim et al., 2005] Kim, J.-W., Park, B., Jeong, S. C., Kim, S. W., and Park, P. (2005). Fault diagnosis of a power transformer using an improved frequency-response analysis. *IEEE Transactions on Power Delivery*, 20(1):169–178.
- [Kruskal and Wallis, 1952] Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.
- [Kulkarni, 2013] Kulkarni, S., K. S. (2013). *Transformer Engineering Design, Technology, and Diagnosticss*. CRC Press, Taylor Francis group. (visited 2023-07-15).
- [Leung, 2022] Leung, K. (2022). Micro, macro weighted averages of f1 score, clearly explained understanding the concepts behind the micro average, macro average, and weighted average of f1 score in multi-class classification with simple illustrations. Accessed on July 09, 2023.
- [Maklin, 2019] Maklin, C. (2019). Fast fourier transform. Accessed on August 26, 2023.
- [Mejia-Barron et al., 2018] Mejia-Barron, A., Valtierra-Rodriguez, M., Granados-Lieberman, D., Olivares-Galvan, J. C., and Escarela-Perez, R. (2018). The application of emd-based methods for diagnosis of winding faults in a transformer using transient and steady state currents. *Measurement*, 117:371–379.
- [Mirbozorgi, 2020] Mirbozorgi, B. (2020). Multinomial naive bayes in machine learning. Accessed on May 22, 2023.
- [Mishra et al., 2020] Mishra, D., Baral, A., Haque, N., and Chakravorti, S. (2020). Condition assessment of power transformer insulation using short-duration time-domain dielectric spectroscopy measurement data. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4404–4411.

REFERENCES

- [Moldagulova and Sulaiman, 2017] Moldagulova, A. and Sulaiman, R. B. (2017). Using knn algorithm for classification of textual documents. In *2017 8th International Conference on Information Technology (ICIT)*, pages 665–671.
- [Motahari-Nezhad and Jafari, 2020] Motahari-Nezhad, M. and Jafari, S. M. (2020). Anfis system for prognosis of dynamometer high-speed ball bearing based on frequency domain acoustic emission signals. *Measurement*, 166:108154.
- [Murphy et al., 2006] Murphy, K. P. et al. (2006). Naive bayes classifiers. *University of British Columbia*, 18(60):1–8.
- [Narayan, 2021] Narayan, Y. (2021). Hb vsemg signal classification with time domain and frequency domain features using lda and ann classifier. *Materials Today: Proceedings*, 37:3226–3230. International Conference on Newer Trends and Innovation in Mechanical Engineering: Materials Science.
- [Pankaj and Wilscy, 2013] Pankaj, D. S. and Wilscy, M. (2013). Comparison of pca, lda and gabor features for face recognition using fuzzy neural network. pages 413–422.
- [Parker, 2017] Parker, M. (2017). Chapter 3 - sampling, aliasing, and quantization. In Parker, M., editor, *Digital Signal Processing 101 (Second Edition)*, pages 21–30. Newnes, second edition edition.
- [Proakis and Manolakis, 2006] Proakis, J. G. and Manolakis, D. K. (2006). *Digital Signal Processing (4th Edition)*. Prentice Hall, 4 edition.
- [Reclamation et al., 2011] Reclamation, B., of Reclamation, B., Interior, U., and Research, T. (2011). *Transformers: Basics, Maintenance and Diagnostics*. Books Express Publishing.
- [Roa-Terán, 2017] Roa-Terán, Y. L. (2017). Análisis de funciones modales usando entropía para la detección de fallas en motores de inducción. Master’s thesis, Universidad de Guanajuato.
- [Rodríguez-Lopéz, 2022] Rodríguez-Lopéz, D. A. (2022). Desarrollo de una metodología a través de análisis de señales de vibración para la detección de la falla de corto circuito en transformadores eléctricos. Master’s thesis, Universidad de Guanajuato.
- [S. Karamizadeh and Hooman, 2013] S. Karamizadeh, S. Abdullah, A. M. M. Z. and Hooman, A. (2013). An overview of principal component analysis. *Journal of Signal and Information Processing*, 4(3B):173–175.

REFERENCES

- [Shah, 2022] Shah, A. (2022). Classifying customer churn using k-nearest neighbour supervised machine learning algorithm. Accessed on May 22, 2023.
- [Singur, 2012] Singur, R., S. P. P. N. S. D. (2012). Simulation of single phase transformer with different supplies.
- [Still, 1919] Still, A. (1919). *Principles of Transformer Design*. Jhon Wiley Sons, London.
- [Sun et al., 2012] Sun, H.-C., Huang, Y.-C., and Huang, C.-M. (2012). A review of dissolved gas analysis in power transformers. *Energy Procedia*, 14:1220–1225.
- [Swafford, 2016] Swafford, Z. (2016). Transformer losses. Accessed on July 15, 2023.
- [Theodorsson-Norheim, 1986] Theodorsson-Norheim, E. (1986). Kruskal-wallis test: Basic computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. *Computer Methods and Programs in Biomedicine*, 23(1):57–62.
- [Udegbonam, 2020] Udegbonam, S. (2020). Inside transformers. Accessed on April 13, 2023.
- [Vapnik, 1995] Vapnik, V.; Cortes, C. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- [Wang et al., 2002] Wang, M., Vandermaar, A., and Srivastava, K. (2002). Review of condition assessment of power transformers in service. *IEEE Electrical Insulation Magazine*, 18(6):12–25.
- [Wang et al., 2015] Wang, X., Zheng, Y., Zhao, Z., and Wang, J. (2015). Bearing fault diagnosis based on statistical locally linear embedding. *Sensors*, 15(7):16225–16247.
- [Wu et al., 2018] Wu, X., Li, L., Zhou, N., Lu, L., Hu, S., Cao, H., and He, Z. (2018). Diagnosis of dc bias in power transformers using vibration feature extraction and a pattern recognition method. *Energies*, 11(7).
- [Xiong et al., 2020] Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., and Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Computational Materials Science*, 171:109203.
- [Yanez-Borjas et al., 2019] Yanez-Borjas, J. J., Camarena-Martinez, D., Valtierra-Rodriguez, M., SaucedoDorantes, J. J., and Amezquita-Sanchez, J. P. (2019). Methodology based on statistical features and linear discriminant analysis for damage detection in a truss-type bridge. pages 1–6.

REFERENCES

- [Yanez-Borjas et al., 2020] Yanez-Borjas, J. J., Valtierra-Rodriguez, M., Camarena-Martinez, D., and Amezquita-Sanchez, J. P. (2020). Statistical time features for global corrosion assessment in a truss bridge from vibration signals. *Measurement*, 160:107858.
- [Zeng et al., 2021] Zeng, R., Lu, Y., Long, S., Wang, C., and Bai, J. (2021). Cardiotocography signal abnormality classification using time-frequency features and ensemble cost-sensitive svm classifier. *Computers in Biology and Medicine*, 130:104218.
- [Zhang et al., 2018] Zhang, S., Li, X., Zong, M., Zhu, X., and Wang, R. (2018). Efficient knn classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 29(5):1774–1785.
- [Zheng et al., 2019] Zheng, J., Huang, H., and Pan, J. (2019). Detection of winding faults based on a characterization of the nonlinear dynamics of transformers. *IEEE Transactions on Instrumentation and Measurement*, 68(1):206–214.