Universidad Autónoma de Querétaro

Campus San Juan del Río

Facultad de Ingeniería

Compression and data mining for the diagnosis of incipient short-circuit faults in single-phase transformers with electric load

Compresión y minería de datos para el diagnóstico de fallas incipientes de cortocircuito en transformadores monofásicos con carga eléctrica

T E S I S

Que como parte de los requisitos para obtener el grado de Maestra en Ciencias Mecatrónica

PRESENTA:

Ing. Arantxa Contreras Valdes

Director de tesis:

Dr. Martín Valtierra Rodríguez

San Juan del Río, Querétaro

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias Mecatrónica

Compresión y minería de datos para el diagnóstico de fallas incipientes de cortocircuito en transformadores monofásicos con carga eléctrica

TESIS

Que como parte de los requisitos para obtener el grado de
Maestra en Ciencias Mecatrónica

Presenta:
Ing. Arantxa Contreras Valdes

Dirigido por:
Dr. Martín Valtierra Rodríguez

Dr. Martín Valtierra Rodríguez
Presidente

Dr. Luis Morales Velázquez
Secretario

Dr. René de Jesús Romero Troncoso
Vocal

Dr. Arturo Yosimar Jaen Cuellar
Suplente

Dr. Jesús Rooney Rivera Guillén
Suplente

Centro Universitario, Querétaro, Qro.
Agosto, 2020
México

Table of contents

Summary

Timely detection of short-circuit faults in transformers is an important challenge to ensure reliability and safety in the generation and transmission of electrical energy. With this detection, the reduction of costs and time required for device repair can be achieved and, at the same time, severe damages to the equipment and large economic losses for both electricity providers and end-users can be also avoided. Investigations related to the detection of electrical faults in transformers have shown that the most common fault that occurs is the short-circuit fault. The methodology developed in this research work allows the diagnosis of the incipient fault condition of a single-phase transformer with an electric load. In general, the methodology is based on the acquisition of multiple physical quantities such as temperature, current, voltage, and vibrations. In this sense, it should be taken into account that the methodology presented proposes the implementation of a data compression algorithm based on the discrete wavelet transform to increase performance during data characterization and eliminate the noise contained in the signals. Then, the data is decompressed for analysis. The variance is extracted and, subsequently, the results are standardized to be analyzed by a data mining algorithm. A support vector machine is used as a classifier in order to detect the severity of the fault under different load conditions. The results obtained reflect the effectiveness and usefulness of the proposed methodology since it is possible to identify and classify the short-circuit faults induced in the transformer.

Keywords: Incipient fault detection, transformer, data compression, data mining, support vector machine.

Dedication

A mis papas, porque éste y cada uno de mis logros son siempre para ellos.

Los amo con mi vida entera.

Acknowledgments

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo brindado bajo la beca número 892305.

Al Dr. Martín Valtierra Rodríguez porque con usted la palabra gracias se queda pequeña. Porque en el primer recuerdo de cada cosa que he aprendido aquí, está usted enseñándome siempre dispuesto a ayudarme en lo que sea con una sonrisa. Porque fue quien me empujó a transformar todos los "no creo que pueda" en "adelante, inténtalo". Gracias por enseñarme a no conformarme con los caminos obvios y no tener miedo a los retos. Porque es un placer trabajar con usted sabiendo que juntos sacamos adelante cualquier protocolo, artículo, review o lo que nos propongamos. Todo mi cariño y agradecimiento. Gracias por confiar en mí.

Al Dr. Luis Morales Velázquez por su infinita paciencia y por ayudarme con cualquier cuestión o pregunta que le haya hecho con toda la amabilidad del mundo.

Al Dr. Arturo Yosimar Jaen Cuellar por todas nuestras conversaciones y por haber sido siempre tan generoso con su tiempo las veces que lo he necesitado.

Al Dr. Jesús Rooney Rivera Guillén por su sensatez, su criterio y sus buenas preguntas.

Al Dr. René de Jesús Romero Troncoso por ser un ejemplo para todos y formar al mejor y único director de tesis que he tenido.

A todos ustedes por ser los testigos de mis primeros pasos en este mundo de la investigación y acogerme con los brazos abiertos.

Y principalmente a Dios y mis papás, mis pilares y guardianes y la principal razón por la que cada uno de los días que he pasado aquí hayan merecido la pena.

Gracias por todo y tanto.

List of figures

List of tables

Chapter 1

1. Introduction

Today, the industry of energy and electricity demands to increase the availability and reliability of electrical equipment. With the motivations of detecting the existence of abnormal changes in the internal conditions of the network elements, determining if the changes could lead to an incipient failure, and improving the efficiency of the use of the assets through a comprehensive monitoring approach, it has emerged the concept of Smart Grids (REI, *Redes Eléctricas Inteligentes*), which refers to a system that integrates the generation, transmission, distribution and commercialization of electricity using information and communication technology through the employment of smart meters and control systems.

According to the Electric Industry Law (LIE, *Ley de la Industria Eléctrica*), a REI integrates advanced measurement, monitoring, communication and operation technologies, in order to improve the reliability, safety, quality or efficiency of the national electrical system.

Within the various elements that constitute the electrical system, such as switches, disconnectors, transmission lines, among others, the power transformer is an extremely important device since it provides a vital link in the chain of other elements to supply electricity to consumers. The reliable and continuous performance of the power transformer is the key to the transmission of electricity.

Failures in electrical transformers are always a matter of great concern because they can cause power interruption and possible significant economic and safety impacts. Therefore, it is desirable to detect the existence of abnormal changes in the internal condition of the electrically charged transformer and determine whether such changes could be the cause of an incipient fault. The identification of incipient faults, as well as automatic detection, can be feasible through the monitoring of abnormal changes in some parameters (e.g., temperature, vibrations, currents, voltages, among others) and the use of diagnostic methods (e.g., artificial intelligence, machine learning, statistics and database systems) that provide an evaluation of the condition of the equipment.

The existing electricity grid infrastructure represents a challenge, but also new opportunities for system modernization. A great research effort is being generated to introduce new information and data processing systems that meet the needs and demands associated with the concept of smart grid, one of the most important is the generation and transmission of huge volumes of data that result of the constant and online monitoring of the equipment that is part of the REI. For this reason, it is crucial to direct efforts to the development of processing techniques that involve compression data and data mining techniques to meet the need of handling large data sets.

1.1 Background

In May 2016, the Ministry of Energy (SENER, *Secretaría de Energía*) launched the REI program derived from the energy reform (SENER, 2016). This program proposes the implementation and operation of the REI to increase the efficiency, reliability, quality, and safety of the National Electric System of Mexico in an economically viable way, in addition to making significant changes in the issues related to the operation of computer systems.

The current need, not only in Mexico but in the world, in terms of advanced infrastructure for the different equipment that make up the electrical system, particularly the transformer, is of special interest since it is considered one of the most important elements (AJ *et al.,* 2018). The constant monitoring of a transformer for fault detection represents one of the main objectives of the REI program, which is the increase in the reliability of the electric distribution networks.

Therefore, the monitoring of the condition of a transformer needs to be carried out through new technologies that allow the timely diagnosis of incipient faults. An incipient fault is a fault that is just beginning, i.e., a fault of low severity; therefore, its detection is a very challenging task. As these types of failures develop slowly, they can become a major fault over time if the cause is not detected and corrected on time. The analysis of multiple variables such as current (Zheng *et al.,* 2018), voltage (Shuin *et al.,* 2016), temperature (Xuewei & Hanshan, 2018) and vibrations (Bagheri & Phung, 2016) can provide collaborative information for the detection of such failures. However, the problem of collecting such data arises with the market penetration of the smart grid and smart measuring devices due to the excessive generation of data produced by a large number of devices operating permanently. In fact, modern technologies related to electrical systems require online monitoring, which also increases the volumes of data. In this regard, the development and application of techniques that allow, on one hand, handling a large amount of data and, on the other hand, offering an alternative for the identification of relevant information extracted from multiple variables are of paramount importance.

There are many works dedicated to studying the main faults in transformers as can be reviewed in the next section.

1.1.1 Main transformers faults

There are many failures that a transformer can present; however, they can be classified into mechanical failures and electrical failures (Bagheri *et al.,* 2018).

For the detection of mechanical failures, different types of analysis have been used. For instance, Yang *et al.* (2019) diagnose the on-load tap changer through a strategy based on dynamic time warping. Zhan *et al.* (2018) detect the winding condition and diagnose winding fault through time-frequency analysis of the transformer transient vibration response. Karimifard *et al.* (2012) use the vector fitting method, while Rahimpour *et al.* (2010) use statistical features, for the detection and location of the most important mechanical defects in transformers. These defects are: disk-space variation, radial deformation, and axial displacement.

Regarding the study of techniques for diagnosing electrical failures, Li et al. (2019) propose a fractional-order model that can accurately calculate the distribution of currents into the transformer windings and the dynamic characteristics under a single-phase ground fault condition. Rahman & Nirgude (2019) study the partial discharge inception due to particles in the transformer oil by using electric field analysis. Rajmani & Chakravorti (2012) consider the multi-dimensional wavelet network method to classify the electrical faults in transformers. These electrical failures include insulation failure, short-circuit, partial discharge, and electric arc. But not only the above-mentioned methods have been used for this purpose, there are several useful techniques that have proven to be of great help in detecting electrical failures and prolonging the life of power and distribution transformers; for example, in-line chromatography, oil testing by Ultraviolet-Visible Spectroscopy (UV-VIS), Sweep Frequency Response Analyzer (SFRA), Furano analyzer, Dielectric Dissipation Factor (DDF) or Tan Delta, water-in-oil analyzer, on-line winding temperature, routine inspection monitor and indicator device using microprocessor (μP), thin-film condenser sensor and the use of High Temperature Superconducting transformers (HTS) have been reported (Malik *et al.*, 2011).

As previously reviewed, there are several failures that transformers can have throughout their useful life and various diagnostic techniques that can be used. According to the statistics of modern transformer failures, a range of 70% to 80% of the total transformer failures are due to short-circuit failures (Mohseni *et al.*, 2016); for this reason, it is important to consider the study of these failures by means of the use of various methodologies and the measurement of different physical variables for timely diagnosis.

1.1.1.1 Short-circuit faults

Short-circuit faults are the most frequent and dangerous cause of power transformer failure (Mohseni *et al.*, 2016). This fact highlights the importance in its diagnosis since if the fault is not detected in its incipient stage, it will become more severe (Figure 1.1), which could result in an irreversible damage to the transformer, unexpected interruptions, and, consequently, economic losses.

In order to avoid the possibility of sudden failures and improve the reliability of energy systems, authors such as Ballal *et al.* (2016) describe the development and implementation of a transformer inter-turn fault detection system at an early stage. This system is based on the calculation and application of correction coefficients. When making the diagnosis in the initial stage, the effects caused to the imbalance of supply voltage, asymmetry constructions, instrumental errors, and non-uniform load distribution are compensated.

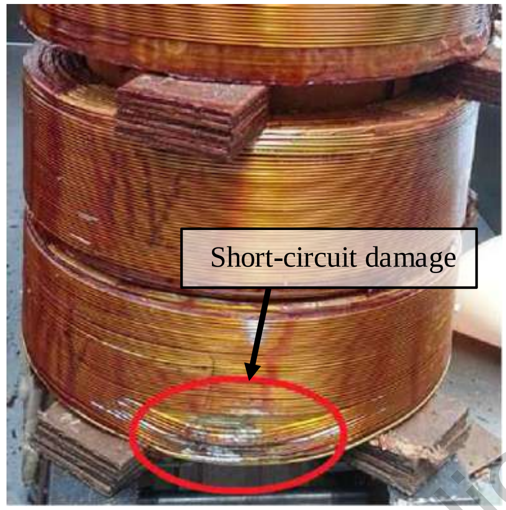


Figure 1.1 Picture of a transformer damaged by short-circuit fault (Gutten et al., 2016).

Another contribution is presented by Mohseni *et al.* (2016). They develop a methodology for inter-turn short-circuit faults which can be detected in a transformer with an electric load. In the methodology, a pulse signal is injected into the transformer winding terminal through a coupling capacitor. The results show that the short-circuit between turns can be visibly detected using the in-line pulse technique in the low-frequency range. Higher levels of severity of failure cause similar changes in resonance and anti-resonance, as well as the magnitude of the transfer function. The results provided in this contribution demonstrate that early diagnosis can be made with the transformer in service using the impulse technique. On the other hand, Mohammadpour & Dashti (2011) present a new method to locate short-circuit faults by considering turn-turn, phase and core, phase and ground, primary winding and secondary winding. In that study, a dynamic transformer model is simulated. The current and voltage signals of the transformer's input and output terminals are measured for different types of faults. Fourier transform of all measured signals is calculated. Then, harmonics of all simulated faults are extracted and saved as special indexes in a database; finally, the fault location process is completed with pattern recognition. This last step determines the location of the fault. Subsequently, Aljohani & Abu-Siada (2015) investigate the impact of short-circuit failure on the signature of Frequency Response Analysis (FRA) using a three-dimensional finite element analysis to emulate the physical conditions of the operation of power transformers with an electric load. The results show that the impact of short-circuit faults of less than 5% on the FRA signature is difficult to detect visually through conventional FRA interpretation techniques and advanced techniques based on digital image processing should be adopted to allow the identification of incipient failures. Short-circuit failure levels greater than 5% will have a visible impact on the amplitude of the FRA signature throughout the entire frequency range and will influence the resonance frequencies in the high-frequency range. Alternatively, at the local level, research work has also been carried out to detect short-circuit faults in transformers. Mejía *et al.* (2018) report the application of methods based on Empirical Mode Decomposition (EMD), such as classical EMD, Ensemble EMD (EEMD, *Ensemble Empirical Mode Decomposition*) and Complete EEMD (CEEMD, *Complete Empirical Mode Decomposition*) for inrush current signal analysis. This analysis leads to the detection of short-circuited turns in transformers. Also, Álvarez *et al.* (2018) propose a methodology to obtain characteristics that help to quantify short-circuit faults using a methodology based on the Wavelet Packet Transform (WPT), whose function is to decompose a signal into its different frequency components. The analysis is based on energizing current signals obtained experimentally in a single-phase transformer, where the current signal is divided into two parts: the first one corresponds to the transient part of the signal and the second part corresponds to the analysis of the stable state of the signal. Shannon's entropy is proposed as an index for the detection and quantification of the fault. Practically, various short-circuit fault conditions between winding windings are analyzed, which are 10, 20, 30 and 40 short-circuited turns, in addition to the transformer's healthy state. As can be seen in the aforementioned works, there are very few investigations that address the analysis of transformers with electric load; further, the diagnosis of the short-circuit failure at an incipient level is not presented, even at the simulation level. Incipient faults will not cause severe damage temporarily; yet, in a long time, incipient faults will gradually develop, leading to more serious failures. Therefore, it is necessary to monitor the transformer with electric load and diagnose the fault in a timely manner. It is well known that the detection of transformer failures is crucial for the safety and quality of the electricity supply derived from the reduction of operating and maintenance costs; however, the scope in the detection of an incipient failure is of vital importance since it could lead to a sudden failure and, eventually, to the failure of the entire transformer.

1.1.1.2 Incipient faults

Timely determination of any anomaly or an incipient fault in the transformers improves decision making, reliability, and continuity of the power supply.

In the literature there are multiple works at international level where the implementation of algorithms for this need is proposed; as an example, Rigatos & Piccolo (2012) use a diffuse neural network to model the thermal condition of the power transformer under operating conditions without failures. The output of the neural-fuzzy network is compared with measurements from the power transformer and the obtained residuals undergo statistical processing according to the fault detection and isolation algorithm. If a failure threshold is exceeded, the deviation from the normal operation can be detected in its initial stages and an alarm can be initiated. The document presented by Iorgulescu *et al.* (2010) describe the diagnosis of electrical failures based on vibration monitoring. These failures may appear in any electrical equipment, but their document focuses on the transformer, and three-phase and single-phase induction motors. To perform the diagnosis, they analyzed the vibration signals of a piezoelectric accelerometer installed in the frame of the equipment. The Fast Fourier Transform (FFT) is applied to the acquired signals; thus, the high-frequency spectral analysis of the vibration provides a method to detect equipment failures. Otherwise, Venikar *et al.* (2015) present the use of harmonic analysis of no-load current to detect the presence of failures between turns in the incipient stage. In the steady state, the components of the third and fifth harmonics are predominant in the no-load current of the transformer. The onset of the failure produces a decrease in Total Harmonic Distortion (THD) and an increase in the primary current. Depending on the reduction of the THD of each phase and the simultaneous increase of the primary current, it is possible to detect the presence of a failure between turns. Wavelet analysis is also performed to validate the results. The proposed method offers greater sensitivity and precision since it can detect faults in an incipient stage that involves less than 2% of turns. The research of Masoum *et al.* (2017) is based on simulations by considering the correlation between the difference of input voltage and instantaneous output (ΔV) and the input current of a particular phase of the transformer that can be measured each cycle to identify and qualify incipient internal faults of the transformer under non-sinusoidal operating conditions. To accurately emulate the actual operation of the transformer, they develop a detailed finite element model.

In all the previous investigations, the use of variables that generate the information necessary to perform the analysis and diagnosis of the failures is required. Although the data generated from the monitoring of these variables already constitute a considerable number in terms of measurement, the use of more than one variable to allow a wide observation and real-time monitoring as outlined in this work implies new requirements related to computational intelligence and the ability to process data or signals and share information. Therefore, a considerable increase in the exchange and storage of data is very probable (Tcheou *et al.*, 2014). Currently, many researchers focused on various areas of engineering (electrical, electronic and mechatronic) are experimenting with different methodologies of analysis for both the treatment of large volumes of data and the monitoring and diagnosis of power transformer failures.

1.1.2 Data mining and data compression

Data mining is a method that plays an important role in building intelligent models to explore and identify hidden patterns and provide a better understanding of large volumes of data for the correct decision making, resulting in the generation of knowledge.

The application of data mining does not belong to an exclusive field, it has been used in different disciplines and areas of research such as astronomy, biology, chemistry, medicine, earth sciences and of course engineering (Grossman *et al.*, 2013).

Tang *et al.* (2013) establish remote telemetry into a diagnostic system for Electric Vehicles (EV). The developed system is capable of storing relevant and useful information for EV users, which can be used by engineers and researchers to carry out data mining for research, development, and problem diagnosis purposes. Razi-Kazemi *et al.* (2015) develop an online monitoring system to accurately assess the condition of Circuit Breaker (CB) during operation and anticipate any risk of failure using its current signals. They employed a data mining approach to extract and classify data and discover the distribution of related patterns. Subsequently, the probability distribution function of the data set associated with each group is determined to interpret the results of this process. By superimposing these probability distribution functions before the probability of CB failure, a classifier model that uses diagnostic parameters is developed. Consequently, it would be possible to plan the maintenance required in the CB in advance of the occurrence of a failure. In order to solve multi-source failure problems and the complex fault diagnosis process of the integrated aircraft transmission generator, Jing & Shi (2016) explore the relationship between generator output voltage, output frequency, current of the generator stator, excitation current and harmonics of excitation current using a decision table and the theory of rough assemblies of varying precision to reduce the decision rules, thus obtaining the final decision results from the perspective of data mining. In a similar venue, Menezes *et al.* (2018) describe the use of a decision tree based on the Computational Intelligence methodology for the analysis and diagnosis of incipient failures in power transformers by using the concentrations in ppm (parts per million) of the combustible gases present in samples of transformer oils. Due to the growing importance of having large amounts of information for processing, the use of data compression techniques becomes essential. In this regard, Bhuiyan *et al.* (2017) perform a comparative study of performance using the discrete Wavelet Daubechies 2 (db2) and Coiflet 1 (coif1) transforms for data compression generated by fasorial measurement units. Elseways, Sarkar *et al.* (2017) designed the DBEA (Differential Binary Encoding Algorithm), which can reduce the volume of data by 90% or even more for most of the available data. This low computational load algorithm can be integrated at the microcontroller level to develop an autonomous and intelligent system. But DBEA can only be used for selected matrices, particularly the programming information matrix, which can be seen as one limitation of the algorithm. In a review work, Tcheou *et al.* (2014) present a compilation of the compression techniques found in the literature to compress electrical signals generated by power systems, analyzing their compression performance. Some of the techniques addressed are Discrete Wavelet Transform (DWT, Discrete Wavelet Transform), the WPT, the Enhanced Disturbance Compression Method (EDCM), the Fundamental, Harmonic and Transient Coding (FHTCM) method, Lempel-Ziv, Huffman, and arithmetic coding, among many others, where the implementation of each one of them will depend specifically on the application. Taking this review as a reference, the application and development of preventive tests and diagnostic techniques are growing due to the economic motivation with the advantages of predicting fault conditions, improving maintenance, and increasing the reliability of the power transformer. Therefore, the need of changing and/or combining the time-based maintenance programs with the diagnostic procedures based on the condition and the online monitoring approaches is introduced (AJ *et al.*, 2018).

Based on the previously reviewed, i.e., failures in transformers, short-circuit failures, data mining, and data compression, it is necessary to develop a methodology capable of improving the diagnostic procedure by implementing a data compression technique that will satisfy the need to process considerable amounts of data. In addition, it would be desirable to have a system based on data mining algorithms capable of analyzing different physical variables, since in previous works the monitoring of variables such as current, voltage, temperature, and vibrations has proved to be useful enough for the assessment of electrically charged transformers with short-circuit failure, which was identified as one of the most common transformer failures in different studies. Therefore it has been considered as one of the key approaches of this thesis work.

1.2 Description of the problem

Based on the situation presented in the background of this document, four essential problems to be solved in this thesis work have been detected. The first one is the timely detection of incipient short-circuit faults because their detection is highly complex. A small fault slightly modifies the behavior of the system, generating that their different physical variables present small or null variations, compromising negatively the diagnosis task.

The second problem is related to the physical variables used to diagnose these failures. The handling of data (acquisition, transmission, and storage) as well as its analysis (processing) becomes extremely complex for online diagnosis.

From this point of view, the third problem relies on contributing to the reduction of time focused on the monitoring the transformer with electric charge, i.e., an early diagnosis. This problem causes the primary need of having timely and online information, which allows knowing and evaluating the performance of the transformer.

Finally, the fourth problem is aimed at enabling storage and massive data analysis due to the online monitoring of single-phase transformers.

1.3 Justification

In the context described above, the interest in researching, developing, and implementing projects concerning the electrical system has gained a greater attention because the reduction in operating costs of the elements that constitute the electrical energy system has been defined as a priority; among them, the power transformer is one of the most important and expensive devices. Therefore, transformer protection is an essential part of the maintenance system.

In this regard, it is imperative to detect incipient short-circuit faults in transformers since it offers multiple advantages when they are identified, e.g., avoid severe and costly damage to the equipment, allow minor and short-term repairs, minimize costs generated from both the supplier and the user, as well as improve decision making, quality, and continuity of the electricity supply.

In other areas of engineering, different works have improved the effectiveness in the control and diagnosis systems using multiple physical variables, such as collaborative controllers, MIMO systems, sensor fusion, etc. Therefore, it is suggested that the analysis of multiple physical variables such as temperature, vibrations, current, and voltage (variables related to the performance of the transformer) could lead to the proposal of diagnostic solutions for incipient short-circuit faults in transformers.

It is necessary to have the online data for operational improvement; hence , the development of a methodology that allows monitoring of the data received from the electrically charged transformer and that is capable of diagnosing a fault is needed. To contribute to this solution, it is proposed to implement data mining algorithms that allow diagnosis in a more holistic and automated way, which offers a wide range of operational benefits such as the increase in operational efficiency, monitoring, and control of results, and reduction of both processing times and maintenance costs.

The described scenario and the solution to these concerns impact on topics of energy efficiency and reliability, which justifies the need to make changes in the infrastructure of maintenance, quality and safety of electricity for smart grid technology approaches.

1.4 Hypotheses and objectives

1.4.1 Hypotheses

The development of a diagnostic methodology based on compression and data mining of multiple signals such as temperature, vibration, current, and voltage as well as their signal processing will allow the detection of incipient short-circuit faults in transformers with electric load by considering different levels of severity.

1.4.2 Objectives

1.4.2.1 General objective

Develop a methodology and a monitoring system focused on compression and data mining techniques for the analysis of current, voltage, temperature, and vibration signals, which allow the diagnosis of incipient short-circuit failures in transformers with electric load by considering different levelst of fault severity.

1.4.2.2 Particular objectives

1) Implement a system to obtain the current, voltage, temperature, and vibration signals through the integration of different sensors and a data acquisition system

2) Enable the connection of a transformer and induce short-circuit faults with different severity levels in order to have a test repository.

3) Evaluate different data compression techniques for the analyisis of the transformer signals with electric charge.

4) Code and test predictive data mining techniques to obtain classification rules for local detection of incipient short-circuit faults.

5) Analyze the behavior of the energizing current and other variables in both a healthy transformer and a damaged transformer by short-circuit faut in order to extract patterns and, consequently, detect the fault.

1.5 General approach

The general approach of this research work consists of the development and implementation of a monitoring and analysis system of current, voltage, temperature and vibrations of a single-phase transformer with electric charge, whose objective is to detect faults at an incipient level. A block diagram of the components to be used is shown in Figure 1.2. The five modules that compound the general approach are described below:

- *Single-phase transformer.* A device composed of two windings with different taps to represent the object of study; in this equipment, the tests will be carried out for both the healthy condition and the short-circuit condition.

- *Data acquisition.* Its function is to perform the acquisition of the electrical current, voltage, temperature, and vibration signals obtained through the corresponding sensors for their processing.

- *Data compression.* The module is responsible to compress and decompress all data received by the acquisition module through a data compression technique.

- *Diagnostic system.* It is in charge of applying the data mining technique on the set of variables to be analyzed in order to obtain a diagnosis result of an incipient short-circuit fault (or others). This processing, as well as the signal storage, is done in a personal computer.

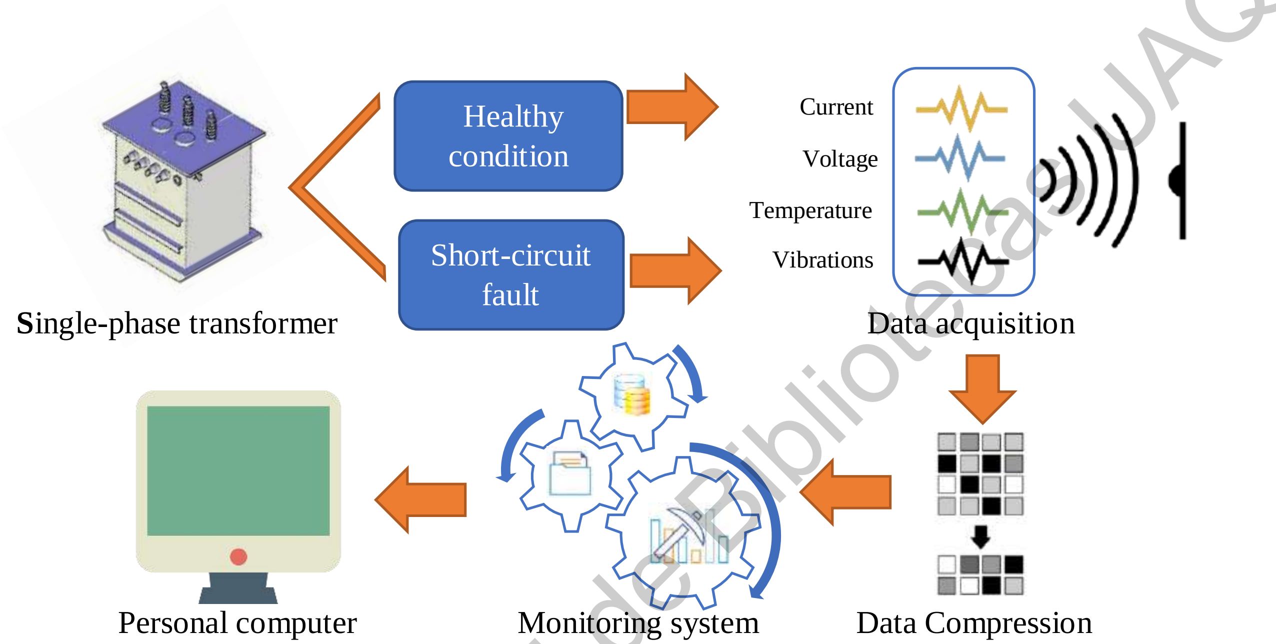


Figure 1.2 Block diagram for the analysis system of a single-phase transformer.

Chapter 2

2. Theoretical foundation

This chapter presents the theoretical and technological tools that support the development of this work. First, a general definition of the concepts corresponding to the four physical variables that are used to monitor the transformer's operating condition is presented. Next, the parts that constitute a transformer, its general operation and the causes of short-circuit generation in this device are exposed in addition to a review related to the severity levels of short-circuit failure. This review is an important aspect to be considered since the level of incipient failure used in the methodology of this project can be established. Then, an overview of the data compression techniques is presented and, then, the technique implemented in this research work is detailed. Finally, a theoretical framework on the data mining process is shown as well as the explanation of the method based on a vector support machine that will be used for the detection of incipient short-circuit faults in a single-phase transformer with electric charge by extracting the variance index.

2.1 Physical quantities

Nowadays, the electrical industry demands a complete monitoring approach to increase the reliability and availability of electrical equipment. For this reason, modern solutions allow the monitoring of various parameters to detect the existence of abnormal changes in the condition of the transformer and determine whether such changes could lead to a failure.

The behavior of the transformer is subject to changes that lead to an alteration of its physical process, such as the variations in intensity experienced by the electric current flowing through the transformer, the increase in temperature produced from an increase in the electric charge, voltage changes due to high current consumption, as well as disturbance in the mechanical vibrations of the transformer. Therefore, it is important to monitor the physical variables related to these changes.

The definitions for the physical quantities discussed in this work are described in the next subsections.

2.1.1 Electric current

The operation of the transformers would not be possible without the electric current flowing through it. The analysis of this fundamental unit of electricity is vital for the study of the occurrence of failures in electrical systems due to the trend for the implementation of much more robust and complex electrical networks.

Electric current is defined as the flow of electric charges through a cross-sectional area per unit of time (Tipler & Mosca, 2005). Figure 2.1 shows a segment of a wire in which the load carriers move. The current or intensity of the current I is given by:

$$I = \frac{\Delta Q}{\Delta t} \tag{1}$$

where ΔQ is the electric charge that flows through the cross-sectional area A at time Δt.

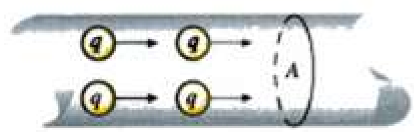


Figure 2.1 Segment of a current-carrying conductor wire (Tipler & Mosca, 2005).

The ninth General Conference of Weights and Measures (GFCM) adopted the Ampere, symbol A, as a unit of electric current intensity in 1948. The International Committee of Weights and Measures (CIPM) proposed the following definition: the Ampere is the intensity of a constant current that, keeping in two parallel, rectilinear, infinite length conductors, of negligible circular section and located at a distance of one meter from each other, in a vacuum, would produce between these conductors a force equal to 2×10^{-7} Newton per meter of length (Centro Español de Metrología, 2006).

When a charge of 63 trillion electrons (a coulomb) is transported in one second, it is said that a current intensity of one ampere (1 A) circulates in the circuit. Since this unit is large, a submultiple of the ampere is often used: the milliamp (mA), equivalent to one-thousandth of A as follows:

$$1\ A = 1000\ mA$$
$$1\ mA = 0.001\ A$$

2.1.2 Temperature

The life of the transformer depends, among other factors, on the temperature inside, especially that one located on its insulating materials. This variable represents one of the main points for the analysis of models and its measurement provides a thermal model to calculate the real dynamic capacity of a transformer.

It is said that temperature is a measure of the random translational movement of atoms and molecules of a body; more specifically, it is a measure of the average kinetic energy of the molecules and atoms of a body (Hewitt, 2006).

The definition of the thermodynamic temperature unit was established by the tenth CGPM (1954) that choses the triple point of water as the fundamental fixed point, assigning it the temperature of 273,16 K per definition. The thirteenth CGPM adopted the name "Kelvin", symbol K, instead of "Kelvin grade", symbol ºK and defined the unit of thermodynamic temperature as follows: the Kelvin, unit of thermodynamic temperature, is the fraction 1/273,16 of the thermodynamic temperature of the triple point of water (Centro Español de Metrología, 2006).

It follows that the thermodynamic temperature of the triple point of water is exactly 273,16 Kelvin, $T_{tpw} = 273{,}16\,K$. The thermodynamic temperature T continued to be expressed according to its difference with respect to the reference temperature $T_0 = 273{,}15\ K$, freezing point of water, due to the way in which temperature scales were usually defined. This temperature difference is called the Celsius temperature, symbol t, and is defined by the equation between magnitudes (Centro Español de Metrología, 2006):

$$t = T - T_0 \tag{2}$$

The unit of temperature Celsius is the degree Celsius, symbol ºC, whose magnitude is related by definition to the one of Kelvin. A difference or a temperature range can be expressed in both Kelvin and Celsius degrees (13th CGPM, 1967/68, Resolution 3, mentioned below). However, the numerical value of the Celsius temperature expressed in degrees Celsius is linked to the numerical value of the thermodynamic temperature expressed in Kelvin (Centro Español de Metrología, 2006) by the relationship:

$$\frac{t}{^{\circ}C} = \frac{T}{K} - 273.15 \tag{3}$$

2.1.3 Voltage

The crucial function of the transformer is to raise or reduce the voltage of the electricity generated in a power plant to the levels necessary to transmit electricity efficiently according to the final destination.

The electrical voltage, also referred to as voltage or potential difference, is attributed to the difference in electrical charges between the positive and negative poles of the circuit generator. This magnitude indicates the amount of energy that the electron current will be able to develop for the same current intensity (Tipler & Mosca, 2005).

Its unit of measurement is the Volt (*V*), mathematically the voltage between two points can be expressed as the change in energy that a load experiences:

$$V = \frac{\Delta U}{q} \tag{4}$$

where ΔU is equal to the change in potential energy and q is a charged particle (Tipler & Mosca, 2005).

2.1.4 Vibration

Transformers experience vibrations as they constitute one of the mechanical properties of the transformer. Therefore, its monitoring is used to detect failures in these devices.

A mechanical vibration usually occurs when a system moves from a stable equilibrium position. This is interpreted as the movement of a particle or body that oscillates around an equilibrium position. The system tends to return to its position under the action of restorative forces (either elastic forces or gravitational forces). But the system usually reaches its original position with some acquired speed that takes it beyond that position. Since the process can be repeated indefinitely, the system keeps moving from side to side of its equilibrium position. The time interval required for the system to perform a complete movement cycle is called the period of vibration. The number of cycles per unit of time defines the frequency and maximum displacement of the system from its equilibrium position, which is known as the amplitude of vibration (Beer *et al.*, 2010).

The period of free vibration τ_n measured in seconds is shown in equation (5):

$$\tau_n = \frac{2\pi}{\omega_n} \tag{5}$$

where the natural circular frequency of the vibration is denoted by ω_n.

The natural frequency of the vibration is known as the number of cycles described per unit of time and is denoted by f_n as follows:

$$f_n = \frac{1}{\tau_n} = \frac{\omega_n}{2\pi} \tag{6}$$

The unit of frequency represents 1 cycle per second, corresponding to a period of 1 s. In terms of fundamental units, the unit of frequency is consequently 1/s or s^{-1}. It is called Hertz (Hz) in the International System of Units (Beer *et al.,* 2010).

Vibration-based control and vibration-based diagnosis are different practical problems. In the diagnosis of vibrations, the oscillation force that is applied to the defective zone defines the fault and this force is linearly related to the acceleration of the oscillation. For diagnosis, often both vibration acceleration and vibration speed are measured in restricted low-frequency ranges. Most vibration measurements use vibration-acceleration sensors that are based on the piezoelectric effect. For this type of sensors, the electrical output load is proportional to the force applied to the sensor, that is, the vibration signal is converted into electrical signals (Beer *et al.*, 2010).

2.2 Transformer

Of the inventions that have been in the branch of electricity, the transformer is probably one of the most useful. This device can raise or reduce voltages or currents in alternating current circuits, isolate circuits from each other, and modify (increasing or decreasing) values of capacitors, resistors, or inductors in electrical circuits. Finally, the transformer allows to the electricity to be transmitted over long distances and distributed safely to factories and homes (Harper, 2005).

In general, from the point of view of the application and design of the transformer, these can be single-phase and three-phase (Harper, 2005). For the purposes of this research work, only the single-phase transformer will be considered; an example of this type of transformer can be seen in Figure 2.2.

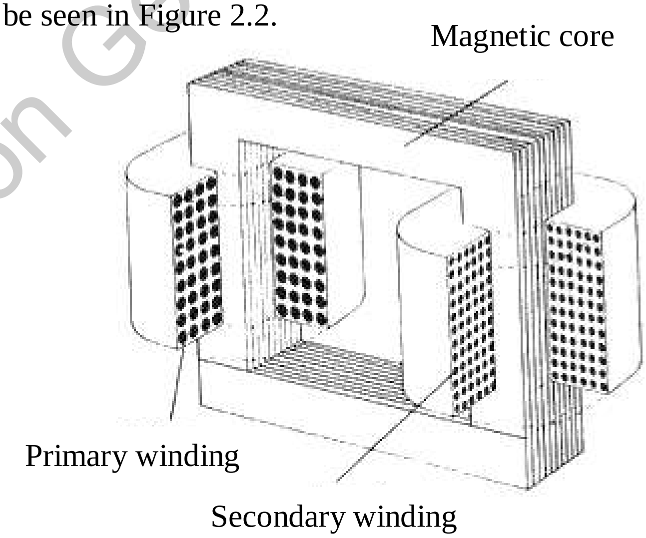


Figure 2.2 Single-phase transformer (Harper, 2005).

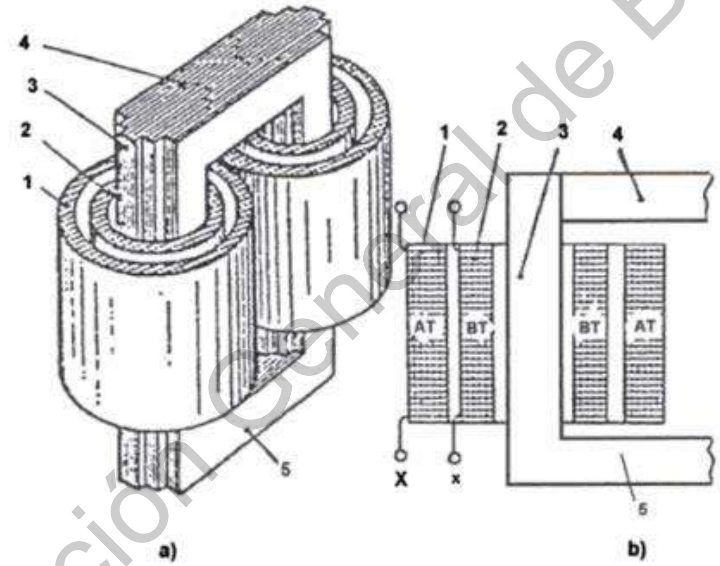
2.2.1 Transformer parts

The transformer, in its simplest form, consists of two stationary coils (two windings) coupled with a reciprocal magnetic flux. It is said that the coils are mutually coupled since the flow that links one coil interacts in the same way with the other one, or most of it (Harper, 2005).

From the point of view of its construction, there are basically two types of iron cores (Harper, 2005):

- The core type
- The shell type

These designs differ from each other in the way the core is constructed to cover the coils. Electrically, there is not much difference between the two types of construction. In fact, in both the coils are placed concentrically, the low voltage being the closest one to the core for insulation reasons and the high voltage on the external part as shown in Figure 2.3 (Harper, 2005).



1. High voltage winding
2. Low voltage winding
3. Core
4. Upper yoke
5. Lower yoke

Figure 2.3 Arrangement of high and low voltage coils: a) external view and b) schematic diagram (Harper, 2005).

2.2.2 Transformer operation

To better understand the operation of the transformers, in Figure 2.4 it can be seen the basic scheme of one of them.

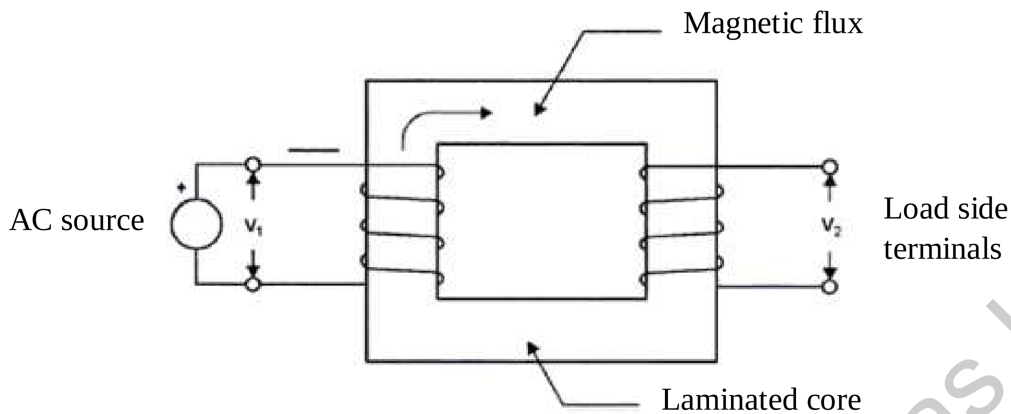


Figure 2.4 Basic scheme of a transformer (Harper, 2005).

The principle of operation is based on the concepts of the voltage induced in a coil, which are basically those derived from Faraday's Law. The coil of Figure 2.5 is crawled or surrounded by a variable magnetic flux (Φ). The flow is of the alternate sinusoidal type at a frequency *f* that periodically reaches positive and negative peaks Φmax. The alternating flow induces a sinusoidal alternating current voltage in the *E* value coil (Harper, 2005).

The induced voltage is given by the following expression:

$$E = 4.44 \; x \; f \; x \; N \; x \; \Phi\text{max} \tag{7}$$

where *E* is equal to the voltage induced in volts, *f* is the frequency of the flow in Hertz, *N* represents the number of turns in the coil, Φmax the peak value of the flow in Weber (Wb) and the constant that represents the exact value of $2\pi/\sqrt{2}$ is 4.44.

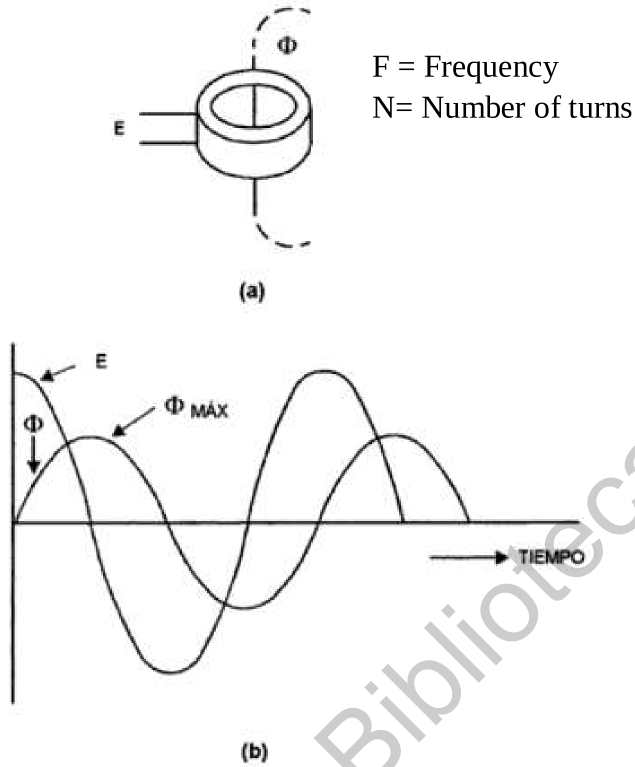


Figure 2.5 a) A voltage E is induced in a coil when it is linked with magnetic flux. b) A sinusoidal flow Φ induces a sinusoidal voltage E (Harper, 2005).

The induced voltage is given regardless the way in which the alternating current flow is present, that is, it can be created according to the basic principles of electromagnetism, by means of a magnet in movement in the vicinity of a coil, or through of an alternating current that circulates through the coil itself (Harper, 2005).

2.2.3 Causes of short-circuit fault generation

As mentioned in Chapter 1, the short-circuit fault in a transformer is one of the most common faults. This failure develops as a consequence of the rupture of the solid insulation that causes a sudden rise in temperature in the winding.

The breakdown of the solid insulation could be due to the natural wear of the insulation, repeated overload (current demand) or a deficiency in the cooling system (temperature changes), excessive vibrations associated with environmental conditions or inherent to the operation of the transformer, as well as variations in the voltage lines of the electricity supplier, which often results in severe aging of winding insulation. If the transformer operates abnormally, that is, if it has overheated and/or delivers a lower output voltage in relation to normal, the possibility of short-circuited turns can be asumed (Harper, 2005).

2.3 Severity levels of short-circuit faults in transformers

In the literature reviewed, there is no standard to measure the severity levels of failure in the transformers, however, these levels can be classified according to the turns that a particular transformer has.

Different authors handle different percentages for the failures treated in their research papers as can be seen in Table 2-1.

Table 2-1 Failure severity levels associated with percentages.

Fault severity	**Percentage**	**Author**
Incipient	2%	Venikar *et al*. (2015)
Incipient	3 – 5%	Mago et al. (2011)
Incipient	< 5%	Aljohani & Abu-Siada (2015)
Short-circuit	5%	Wiszniewski et al. (2018)
Short-circuit	5%	Gutten et al. (2018)
Short-circuit	5%	Mostafaei et al. (2018)
Short-circuit	5%	Masoum et al. (2017)
Short-circuit	5%	Ballal et al. (2016)
Short-circuit	10%	Mohseni et al. (2016)

Based on the information in Table 1, this work will consider a percentage of less than 4% to refer to incipient faults.

2.4 Data compression techniques

When talking about a compression technique or compression algorithm, it actually refers to two algorithms. There is the compression algorithm that takes an X input and generates an X_c representation that requires fewer bits or less data, and there is a reconstruction algorithm that operates on the compressed representation X_c to generate the Y reconstruction (Sayood, 2017). These operations are shown schematically in Figure 2.6. This image is only intended to show that the size decreases with quality reduction effects.

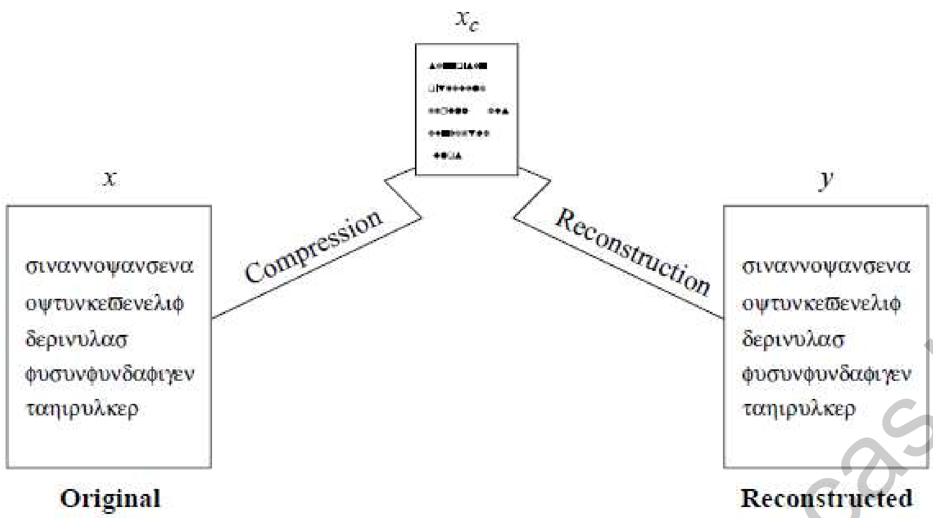


Figure 2.6 Compression and reconstruction (Sayood, 2017).

Based on the reconstruction requirements, data compression schemes can be divided into two broad classes: lossless and lossy compression schemes (Sayood, 2017).

2.4.1 Lossless compression

Lossless compression techniques, as the name implies, do not imply loss of information. The reconstructed signal is identical to the original signal and the performance is measured by the compression ratio (Tcheou *et al.*, 2014).

There are many situations where the reconstruction is required to be identical to the original. There are also many others where it is possible to omit this requirement in order to obtain greater compression.

2.4.2 Lossy compression

This technique implies some loss of information and the data that has been compressed by this technique cannot be recovered or reconstructed accurately. The reconstructed signal may differ from the original and the performance must now be measured by the compromise between the compression ratio and the distortion. When distortion is introduced in a controllable manner without compromising its use or application, it is possible to achieve higher compression ratios than lossless compression approaches (Tcheou *et al.*, 2014).

Once the data compression scheme has been developed, it is necessary to be able to measure its performance. Due to the different number of application areas, various terms have been developed to describe and measure performance. Subsequently, the performance measures used in this research work will be described.

2.4.3 Compression algorithm development

While the reconstruction requirements may force the decision of whether a compression scheme should be lossless or lossy, the exact compression scheme used will depend on a number of different factors. Some of the most important factors are the characteristics of the data that must be compressed. A compression technique that will work well for text compression may not work well for compressing images. Each application presents a different set of challenges.

The development of data compression algorithms for a variety of data can be divided into two phases. The first phase is generally known as modeling. This phase is about extracting information about any redundancy that exists in the data and describing the redundancy in the form of a model. The second phase is called coding. A description of the model and a "description" of how the model data differ, usually using a binary alphabet, are encoded. The difference between the data and the model is often called residue (Sayood, 2017).

2.4.4 Mathematical preliminaries for lossy compression

The idea of a quantitative measure of information has been around for a while. Claude Elwood Shannon defined a quantity called self-information. Suppose it has an event A, which is a set of outcomes of some random experiment. If $P(A)$ is the probability that the event A will occur, then the self-information associated with A is given by:

$$i(A) = log_b \frac{1}{P(A)} = -log_b P(A) \tag{8}$$

The use of the logarithm to obtain a measure of information was not an arbitrary choice. Recall that, $log(1) = 0$ and $-log(x)$ increases as x decreases from one to zero. Therefore, if the probability of an event is low, the amount of self-information associated with it is high; if the probability of an event is high, the information associated with it is low (Sayood, 2017).

Another property of this mathematical definition of information that makes intuitive sense is that the information obtained from the occurrence of two independent events is the sum of the information obtained from the occurrence of the individual events. Suppose *A* and *B* are two independent events. The self-information associated with the occurrence of both event *A* and event *B* is:

$$i(AB) = log_b \frac{1}{P(AB)} \tag{9}$$

As *A* and *B* are independent,

$$P(AB) = P(A)P(B) \tag{10}$$

and

$$\begin{aligned} i(AB) &= log_b \frac{1}{P(A)P(B)} \\ &= log_b \frac{1}{P(A)} + log_b \frac{1}{P(B)} \\ &= i(A) + i(B) \end{aligned} \tag{11}$$

The unit of information depends on the base of the log. If log base 2 is used, the unit is bits. The concepts introduced above allow estimating the number of bits needed to represent the result of a source given its probability model.

A natural thing to do when looking at the fidelity of a reconstructed sequence is to look at the differences between the original and reconstructed values, in other words, the distortion introduced in the compression process. Two popular measures of distortion or difference between the original and reconstructed sequences are the squared error measure and the absolute difference measure. These are called difference distortion measures (Sayood, 2017). If $\{x_n\}$ is the source output and $\{y_n\}$ is the reconstructed sequence, then the squared error measure is given by:

$$d(x_n, y_n) = (x_n - y_n)^2 \tag{12}$$

and the absolute difference measure is given by:

$$d(x_n, y_n) = |x_n - y_n| \tag{13}$$

In general, it is difficult to examine the difference on a term-by-term basis. Therefore, a number of average measures are used to summarize the information in a different sequence. The most often used average measure is the average of the squared error measure. This is called the mean squared error (mse) and is often represented by the symbol σ^2 or σ_d^2 (Sayood, 2017):

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - y_n)^2 \tag{14}$$

2.4.5 Compression and decompression algorithm of electrical signals with loss based on Discrete Wavelet Transform

In order to increase the compression capabilities, a lossy algorithm is used. The compression algorithm implemented in this research work consists of three essential steps which are described below.

1. Approximation

The Cohen-Daubechies-Feauveau 9/7 DWT is applied to convert the signal *f* into a vector *w* whose components are the wavelet coefficients $(c(1), c(2), \ldots, c(N))$ where *N* is the total number of components of *w*.

2. Quantization

The wavelet coefficients are transformed into integers by a mid-tread uniform quantizer in the following way:

$$c^{\Delta}(i) = \left\lfloor \frac{c(i)}{\Delta} + \frac{1}{2} \right\rfloor, i = 1, 2, \ldots, N \tag{15}$$

where *c* are the wavelet coefficients, Δ is the quantization parameter, $\lfloor x \rfloor$ indicates the largest integer number smaller or equal to *x*. After quantification, the coefficients and indices are further reduced by the elimination of the coefficients equal to 0. The signs of the coefficients are encoded separately in an array $(s(1), s(2), \ldots, s(K))$ using a binary alphabet, 1 for the + sign and 0 for the - sign.

3. Organization and Storage

In this step, the indices are re-ordered in ascending order $(\tilde{I})$, as are the coefficients $(\tilde{c}^{\Delta})$ and their corresponding signs $(\tilde{s})$. Reordered indexes are stored as smaller positive numbers taking differences between two consecutive values producing a single recovery.

Finally, the size of the *K* signal, the quantization parameter Δ, and the arrays $\tilde{I}, \tilde{c}^{\Delta}$ and $\tilde{s}$ are saved in Hierarchical Data Format (HDF). HDF operates through a fragmented storage mechanism. The data matrix is divided into fragments of equal size, each of which is stored separately in the file.

In order to recover the compressed signal, a decoding step is necessary. It can be described as follows:

1. The compressed file is read and the data contained in it is extracted (size of signal K, quantization parameter Δ, and matrices $\tilde{l}, \tilde{c}^{\Delta}$ and $\tilde{s}$).
2. Indexes are retrieved from the array $\tilde{l}$.
3. The signs of the wavelet coefficients are recovered as:

$$\tilde{s}^{\mathrm{r}} = 2\tilde{s} - 1 \tag{16}$$

4. The entire matrix of wavelet coefficients is completed as follows:

$$w^{r}(\tilde{l}) = \tilde{s}^{r} \cdot \tilde{c}^{r} \tag{17}$$

   where $\tilde{c}^{\mathrm{r}}$ represents the magnitude of the coefficients of its quantized version obtained as $\tilde{c}^{\mathrm{r}} = \Delta\tilde{c}^{\Delta}$.

5. The wavelet transform is inverted to recover the approximate signal f^{r}.

Figure 2.7 describes the compression and decompression procedure described above in diagram form.

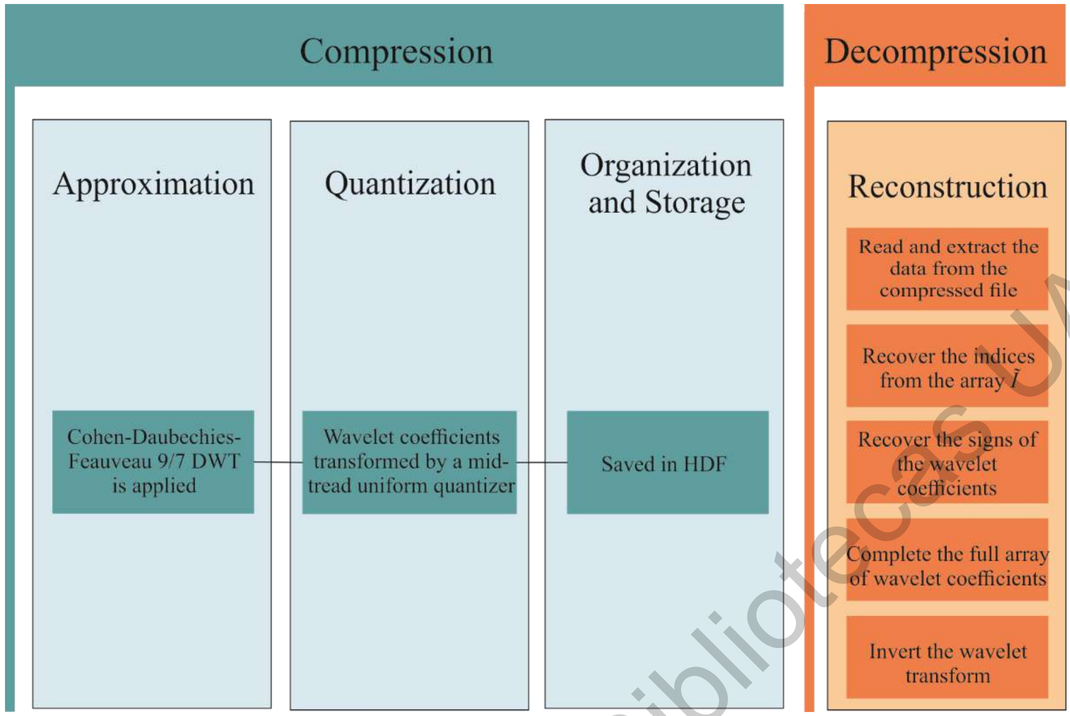


Figure 2.7 Compression and decompression algorithm steps.

2.4.6 Measures of performance

A compression algorithm can be evaluated in several ways. The relative complexity of the algorithm, the memory required to implement the algorithm, the speed with which the algorithm works on a given machine, the amount of compression and how close the reconstruction to the original could be measured. A very logical way to measure how well a compression algorithm compresses a given set of data is to observe the relationship between the number of bits required to represent the data before compression and the number of bits needed to represent the data after the compression. This relationship is called compression index. Another way to report compression performance is to provide the average number of bits required to represent a single sample. This is generally known as the compression rate.

In lossy compression, reconstruction differs from the original data. Therefore, to determine the efficiency of a compression algorithm, there must be some way to quantify the difference between the original and the reconstruction, this is often called distortion. Loss techniques are generally used for compression of data that originates as analog signals, such as voice and video. In voice and video compression, the final arbiter of quality is the human. Because human responses are difficult to model mathematically, many approximate distortion measures are used to determine the quality of the reconstructed waveforms.

Other terms that are also used when talking about differences between reconstruction and the original are fidelity and quality. When it is said that the fidelity or quality of reconstruction is high, it is said that the difference between reconstruction and the original is small. If this difference is a mathematical difference or a perceptual difference it must be evident from the context (Sayood, 2017). Since the lossy compression performance must be considered in relation to the quality of the recovered signals, the measures to evaluate the results of the procedure proposed in section 2.4.5 are introduced at this point.

Compression performance is evaluated by the Compression Ratio (*CR*) given by:

$$CR = \frac{Size\ in\ bytes\ of\ the\ original\ file}{Size\ in\ bytes\ of\ the\ compressed\ file} \tag{18}$$

The quality of a recovered signal is evaluated with respect to the Percentage Root-mean-square Difference (*PRD*) that is calculated as follows:

$$PRD = \frac{\|f - f^r\|}{\|f\|} x100\% \tag{19}$$

where *f* is the original signal, f^r is the signal reconstructed from the compressed file and $\|\cdot\|$ indicates the 2-Norm, which is the largest singular value.

The Quality Score (*QS*), which reflects the compromise between compression performance and reconstruction quality, is computed as:

$$QS = \frac{CR}{PRD} \tag{20}$$

2.5 Data mining

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. Data sources may include databases, data warehouses (data warehouse or homogeneous and reliable information), the Web, other repositories of information or data that are transmitted to the system dynamically.

To distinguish more clearly the area of application of data mining, its two main objectives tend to be classified as predictive and descriptive (Gorunescu, 2011):

- Predictive objectives: achieved by using a part of the variables to predict one or more other variables (e.g., classification, regression, detection of anomalies/outliers).

- Descriptive objectives: achieved by identifying patterns that describe data and that the user can easily understand (e.g., clustering, discovery of association rules, sequential pattern discovery).

In this work, a classification analysis technique is used according to the predictive objective that follows the line of this research.

2.5.1 Classification and regression for predictive analysis

Classification is the process of finding a model or function that describes and distinguishes kinds of data or concepts. The model is derived based on the analysis of a set of training data (that is, data objects for which class labels are known). The model is used to predict the class tag of the objects for which the tag is unknown (Han *et al.*, 2011).

The derived model can be represented in several ways, such as classification rules (i.e. IF-THEN rules), decision trees, mathematical formulas or neural networks (Figure 2.8). The decision diagram is a tree structure type flow chart, where each node denotes a test in an attribute value, each branch represents a test result, and the leaves of the tree represent classes or class distributions. Decision trees can easily be converted into classification rules. When a neural network is used for classification, it is typically a collection of neuron-type processing units with weighted connections between the units. There are many other methods for constructing classification models, such as Bayesian naive classification, support vector machines, and the nearest neighbor classification (Han *et al.*, 2011), among others.

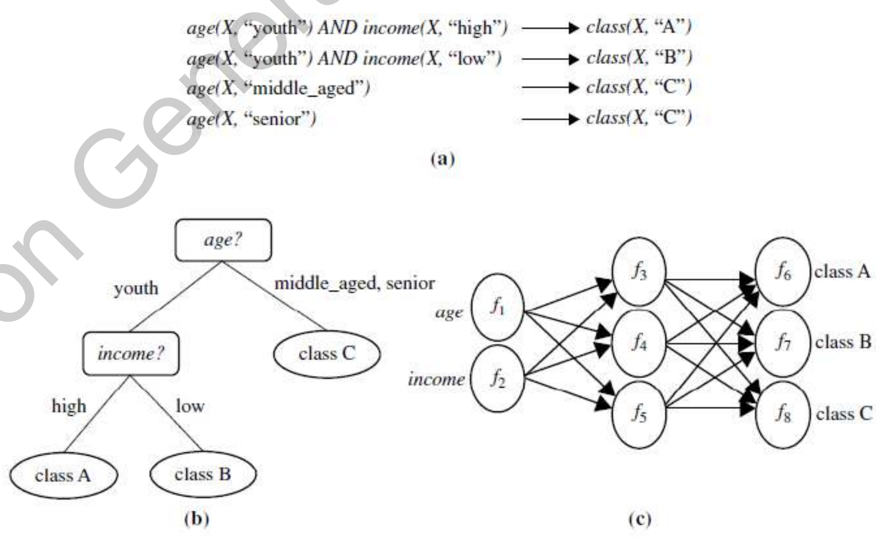


Figure 2.8. A classification model can be represented in various forms: (a) IF-THEN rules, (b) a decision tree, or (c) a neural network (Han et al., 2011).

While the classification predicts categorical labels (discrete, disordered), the regression models represents continuous value functions. That is, regression is used to predict missing or unavailable numerical data values instead of class (discrete) tags. The term prediction refers to both numerical prediction and class tag prediction. Regression analysis is a statistical methodology that is most frequently used for numerical prediction, although there are other methods. The regression also covers the identification of distribution trends based on the available data (Han *et al.*, 2011).

Classification and regression may need to be preceded by relevance analysis, which attempts to identify the attributes that are significantly relevant to the classification and regression process. Such attributes will be selected for the classification and regression process. Other attributes, which are irrelevant, can then be excluded from consideration (Han *et al.*, 2011).

2.5.2 The data mining process

The general experimental procedure adapted to data mining problems involves the following steps (Han *et al.*, 2011):

1. *Data cleaning* (to eliminate noise and inconsistent data).
2. *Data integration* (where multiple data sources can be combined).
3. *Data selection* (where the relevant data for the analysis task is retrieved from the database).
4. *Data transformation* (where data is transformed and consolidated into appropriate forms for mining when performing summary or aggregation operations).
5. *Data mining* (an essential process where intelligent methods are applied to extract data patterns).
6. *Evaluation of patterns* (to identify really interesting patterns that represent knowledge based on measures of interest).
7. *Knowledge presentation* (where visualization and knowledge representation techniques are used to present mined knowledge to users).

A popular trend in the information industry is to perform data cleaning and data integration as a preprocess, where the resulting data is stored in a data warehouse (Han *et al.*, 2011).

Figure 2.9 shows the step-by-step process addressed.

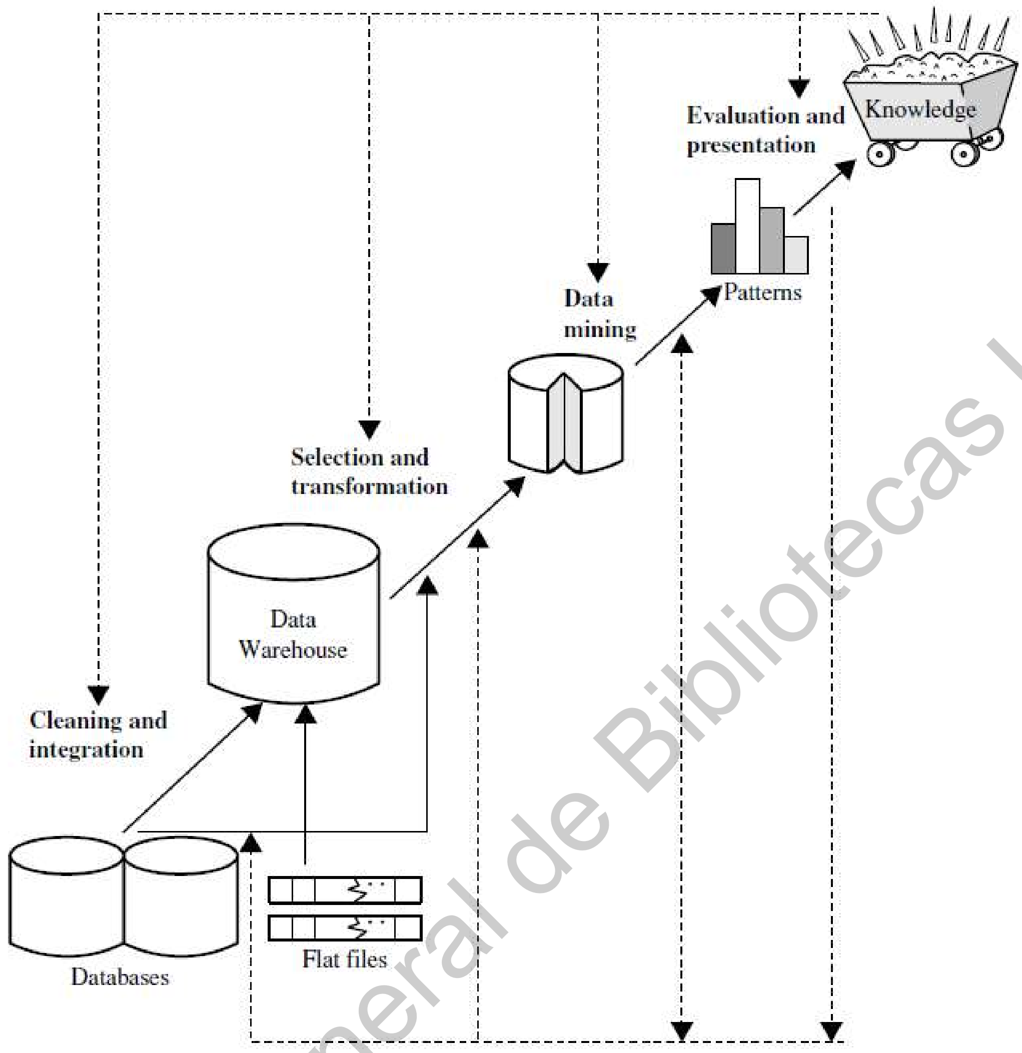


Figure 2.9 Data mining process (Han et al., 2011).

2.5.3 Data mining techniques

First, data mining techniques can be classified into modeling techniques originated by theory, modeling techniques originated by data, and auxiliary techniques (Pérez, 2014).

The modeling techniques originated by the theory specify the model for the data based on prior theoretical knowledge. The supposed model for the data must be checked after the data mining process before accepting it as valid. Formally, the application of any model must overcome the phases of:

- Objective identification. From the data, rules are applied to identify the best possible model that fits the data.
- Estimate. Process of calculating the parameters of the model chosen for the data in the identification phase.
- Diagnosis. Process of the contrast of the validity of the estimated model.
- Prediction. Process of using the identified, estimated and validated model to predict future values of the dependent variables.

These types can include all types of regression and association, analysis of variance and covariance, discriminant analysis and time series (Pérez, 2014).

In the modeling techniques caused by the data, no default role is assigned to the variables. The existence of dependent or independent variables and a previous model for the data is not assumed. Models are created automatically based on pattern recognition. The model is obtained as a mixture of the knowledge obtained before and after data mining and must also be checked before being accepted as valid. For example, complex models are discovered from neural networks and can be refined as data exploration progresses. Thanks to their learning capacity, they allow discovering complex relationships between variables without any external intervention. Auxiliary techniques are new methods based on descriptive statistical techniques and reports (more superficial and limited tools) (Pérez, 2014).

The diagram with the classification of the techniques mentioned above is summarized in Figure 2.10.

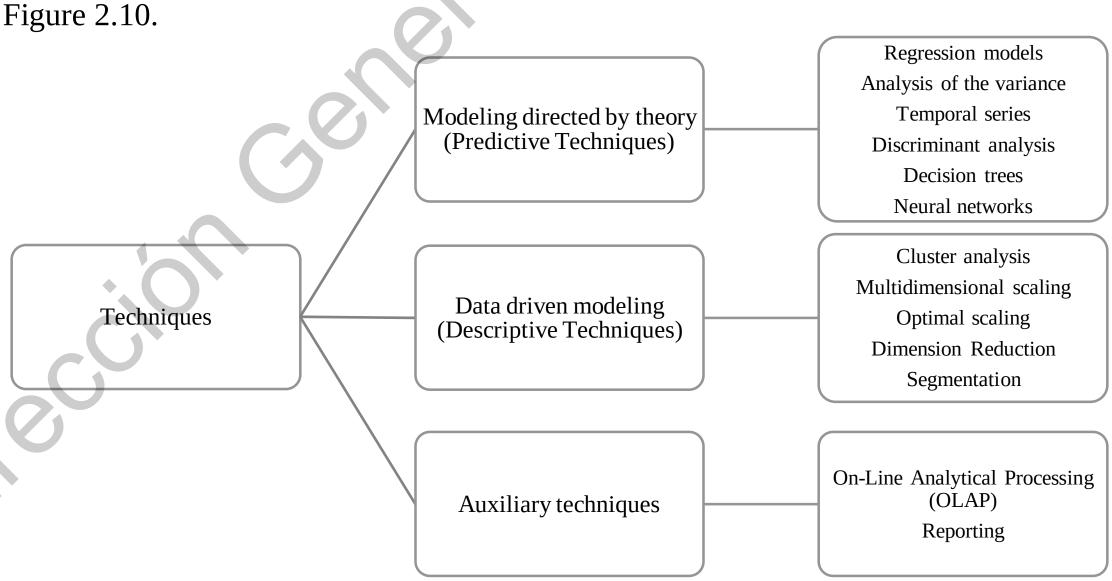


Figure 2.10 Classification of data mining techniques (Pérez, 2014).

The data mining technique that reports better percentages of effectiveness in terms of classification of failures according to the research of Contreras et al. (2019) is the Support Vector Machine (SVM) versus two other widely used techniques that are the J48 decision tree algorithm and neural networks.

2.5.4 Support Vector Machine

In the monitoring of the state of the machines and the problem of fault diagnosis, the SVM is used to recognize special patterns of the acquired signal, and then these patterns are classified according to the occurrence of failures in the machine. SVM has the potential to handle spaces of very large characteristics; therefore, dimension of the classified vectors does not have a clear influence on the SVM performance as on the performance of the conventional classifier. That is why it is observed that it is especially efficient in large classification problems. This will also benefit in the classification of failures, since the number of features that will be the basis of the diagnosis of failures may not be limited (Widodo & Yang, 2007).

Given the data input $x_i(i = 1,2, \ldots, M), M$ where is the number of samples. It is assumed that the samples have two classes, positive class and negative class. Each of the classes is associated with the labels by $y_i = 1$ for the positive class and $y_i = -1$ for the negative class, respectively. In the case of linear data, it is possible to determine the hyperplane $f(x) = 0$ that separates the given data:

$$f(x) = w^T x + b = \sum_{j=1}^{M} w_j x_j + b = 0 \tag{21}$$

where *w* is an *M*-dimensional vector and *b* is a scalar.

The vector *w* and the scalar *b* are used to define the position of separation of the hyperplane. The decision function is performed using the $f(x)$ sign to create a separation hyperplane that classifies the input data into positive and negative class. A different separation hyperplane must meet the restrictions:

$$\begin{gathered} f(x) = 1 \; si \; y_i = 1, \\ f(x) = -1 \; si \; y_i = -1 \end{gathered} \tag{22}$$

or it can be presented in the complete equation:

$$y_i f(x_i) = y_i(w^T x + b) \geq 1 \; para \; i = 1,2,\ldots,M.$$

(23)

The separation hyperplane that creates the maximum distance between the plane and the nearest data, that is, the maximum margin, is called the optimal separation hyperplane. An example of the optimal hyperplane of two data sets is presented in Figure 2.11. In this figure, a series of data points are shown for two different kinds of data, black squares for the negative class and white circles for the positive class. The SVM tries to place a linear boundary between the two different classes and orients it in such a way that the margin represented by the dotted line is maximized. In addition, SVM attempts to target the limit to ensure that the distance between the limit and the nearest data point in each class is maximum. Then, the limit is placed in the middle of this margin between two points. The closest data points that used to define the margin are called support vectors, represented by gray circles and squares. When the support vectors have been selected, the rest of the feature set is not required, since the support vectors can contain all the information needed to define the classifier. From geometry, it is found that the geometric margin is $\|w\|^{-2}$ (Widodo & Yang, 2007).

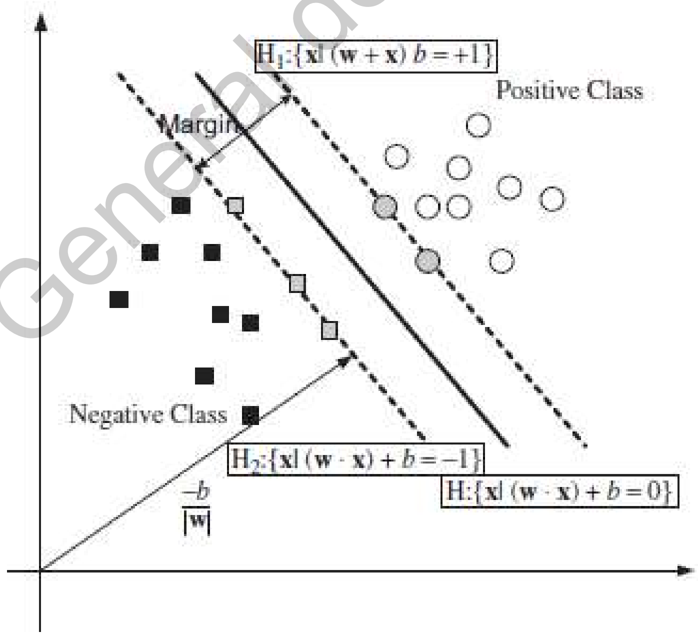


Figure 2.11 Classification of two classes using SVM (Widodo & Yang, 2007).

Taking into account the noise with slack variables ξ_i and the penalty for error C, the optimal hyperplane that separates the data can be obtained as a solution to the following optimization problem:

minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}\xi_i$

(24)

subject to
$$y_i(w^T x + b) \geq 1 - \xi_i, \quad i = 1, \dots, M,$$
$$\xi_i \geq 0, \quad i = 1, \dots, M,$$

(25)

where ξ_i is measuring the distance between the margin and the examples x_i that lying on the wrong side of the margin.

The calculation can be simplified by converting the problem with Kuhn–Tucker condition into the equivalent Lagrangian dual problem, which will be:

minimize $L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{M}\alpha_i\, y_i(w \cdot x_i + b) + \sum_{i=1}^{M}\alpha_i$

(26)

The task is minimizing equation (26) with respect to *w* and *b*, while requiring the derivatives of *L* to α to vanish. At the optimal point, it has the following saddle-point equations:

$$\frac{\partial \mathrm{L}}{\partial w} = 0, \quad \frac{\partial \mathrm{L}}{\partial b} = 0$$

(27)

which replace into the form:

$$w = \sum_{i=1}^{M}\alpha_i y_i x_i, \quad \sum_{i=1}^{M}\alpha_i\, y_i = 0$$

(28)

From equation (28), it is found that *w* is contained in the subspace spanned by x_i. By substituting equation (28) into equation (26), the dual quadratic optimization problem is described as follows:

maximize $L(\alpha) = \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i=1}^{M}\alpha_i\alpha_j\, y_i y_j x_i \cdot x_j$

(29)

subject to $\alpha_i \geq 0, \;\; i = 1, \dots, M, \;\; \sum_{i=1}^{M}\alpha_i\, y_i = 0$

(30)

Therefore, by solving the dual optimization problem, the coefficients α_i that are required to express the *w* are obtained to solve equation (24). This leads to a non-linear decision function:

$$f(x) = sign\left(\sum_{i,j=1}^{M} \alpha_i\, y_i\left(x_i x_j\right) + b\right) \tag{31}$$

SVM can also be used in non-linear classification tasks with application of kernel functions. The data to be classified is mapped onto a high-dimensional feature space, where the linear classification is possible. Using the non-linear vector function $\Phi(x) = (\varphi_1(x), \dots \varphi_1(x))$ to map the *n*-dimensional input vector *x* onto *l* dimensional feature space, the linear decision function in dual form is given by:

$$f(x) = sign\left(\sum_{i,j=1}^{M} \alpha_i\, y_i\left(\Phi^T(x_i) \cdot \Phi(x_j)\right) + b\right) \tag{32}$$

Working in the high-dimensional feature space enables the expression of complex functions, but it also generates the computational problem. This occurs due to the large vectors and the overfitting also exists due to the high-dimensionality. The latter problem can be solved by using the kernel function. The kernel is a function that returns a dot product of the feature space mappings of the original data points, stated as $K(x_i, x_j) = (\Phi^T(x_i) \cdot \Phi(x_j))$. When applying a kernel function, the learning in the feature space does not require explicit evaluation of Φ and the decision function will be:

$$f(x) = sign\left(\sum_{i,j=1}^{M} \alpha_i\, y_i K\left(x_i, x_j\right) + b\right) \tag{33}$$

There are different kernel functions used in SVM, such as linear, polynomial and Gaussian radial basis function. The selection of the appropriate core function is very important since the core defines the space of characteristics in which the training set examples will be classified. In this work, the linear function is evaluated and formulated as:

$$x^T \cdot x_j \tag{34}$$

SVM performs the monitoring and diagnosis of the state of the machine using its ability in the classification process. Based on the input data vectors that consist of the representation of machine failures, SVM will recognize these patterns. Usually, each fault produces special characteristics that can be considered patterns. So, the main task of SVM here is to recognize and classify these patterns as accurately as possibly related to the fault. SVM is motivated to represent patterns in a high dimension, typically much higher than the original feature space. With an appropriate nonlinear mapping that uses the kernel function to a sufficiently high dimension, data from two or more categories can always be separated by a hyperplane (Widodo & Yang, 2007).

For a good classification, data preprocessing is an important step. Good data preprocessing will reduce data noise and retain as much information as possible (Marshall, 1995). Based on clean data, the characteristics (condition indicators) can be calculated and considered as standards for fault diagnosis purposes.

The SVM that uses feature-based diagnostics has been widely used in many machine condition monitoring and fault diagnosis applications. Most of the results of the technique based on characteristics, for example, statistical indicators, were relatively satisfied according to various published articles. It means that the feature-based procedure is a recommended method when the recognition and classification process is performed (Widodo & Yang, 2007).

2.6 Statistical features

After the acquisition of the signal, a characteristic representation method can be performed to define the characteristics, for example, the calculation of statistical indicators for classification purposes. These characteristics can be considered as patterns that must be recognized by SVM.

In the process of diagnosis based on characteristics, in general, a great dimensionality in the representation of characteristics will be obtained. Unfortunately, not all features are significant and contain high information about the state of the machine. Some of them are useless and irrelevant features. These features should be removed to increase the accuracy of the classifier. Therefore, the extraction and selection of characteristics are necessary to produce good characteristics for the classification process (Barzilay & Brailovsky, 1999).

Among several statistical features such as mean, kurtosis, standard deviation, skewness and RMS, the one selected to represent the most significant characteristics of the signals acquired in this thesis work is the variance according to the research of Contreras *et.*

al., 2019. The calculation of the variance is done in order to benefit the performance and accuracy of the SVM.

2.6.1 Variance

The most used measure to estimate the dispersion of the data is the standard deviation. This is especially advisable when using the arithmetic mean as a measure of central tendency (Gorgas *et al.*, 2011). Like the average deviation, it is based on an average value of the deviations from the average. In this case, instead of taking absolute values of the deviations, to avoid thus compensating positive and negative deviations, the squares of the deviations are used. This also makes the data with large deviations greatly influence the final result. Variance is then a way of measuring data variability. Its construction is done on the basis of deviations from the arithmetic mean (Rustom *et al.*, 2012) and is defined as:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N} \tag{35}$$

where x_i represents each data, $\bar{x}$ the average of the data and *N* the total number of the data.

The most important properties of the variance and standard deviation are explained below:

- $P_1: y_i = x_i + k \rightarrow \sigma_y^2 = \sigma_x^2 \; and \; \sigma_y = \sigma_x$ states that the variance and standard deviation are not altered by adding a constant to the data.
- $P_2: y_i = k * x_i \rightarrow \sigma_x^2 = k^2\sigma_x^2 \; and \; \sigma_y = k * \sigma_x$ specifies that, by multiplying the data by a constant, the variance is amplified by the constant squared and the standard deviation only by the constant.
- $P_3: x_i = k \rightarrow \sigma_x^2 = \sigma_x = 0$, that is, the variability of a constant is zero.

Chapter 3

3. Methodology

This chapter describes in detail the methodology followed for the detection of incipient short-circuit faults in a single-phase transformer under three operating scenarios, as well as the load circuit for these scenarios. Similarly, the configuration of the single-phase transformer used is specified. The induced faults (incipient, light damage, and moderate damage) are controlled in order to monitor the device in a healthy condition and, then, do so with the fault conditions. The next step is to make an exhaustive analysis with which can be determined if there is a fault present in the transformer. Likewise, the technical aspects of the sensors and the acquisition board used are detailed. Finally, the interface for the acquisition signals, and the graphical user interface for analysis and monitoring are presented.

3.1 General description of the methodology

The methodology consists of the acquisition of signals from a single-phase transformer which is modified to simulate four different operating conditions: healthy, incipient, light damage, and moderate damage. The acquisition of the physical quantities described in section 2.1 is carried out under three different scenarios: transformer without load, with linear load, and with non-linear load. It is very important to consider these scenarios in order to evaluate the transformer in situations closer to a real context.

The signals are detected by different sensors according to the physical variable involved and each sensor is located at strategic points of the transformer. In the case of vibrations, an analog three-axis Kistler accelerometer is used; for the temperature, an LM35 sensor is used; Fluke i200s current clamps are used for the current and in the case of the voltage signal a 6 V 1 A transformer is used. The four types of signals are acquired by means of a data acquisition system constituted by the board of NI-USB 6211 acquisition of National Instruments and an interface developed in LabVIEW. This interface is responsible for showing to the user the behavior of the online signals and has the option of storing the acquired data in files.

After the data acquisition, a loss compression technique based on a DWT is applied. After the variance is extracted, a step of standardization of the data is carried out and, finally, the results will be processed by a SVM which classifies each transformer condition. With the exception of data acquisition, the rest of the methodology is implemented in the MATLAB software.

The sequence of the steps for the above-mentioned methodology can be found in the diagram in Figure 3.1. These steps will be described in more detail in the following subsections.

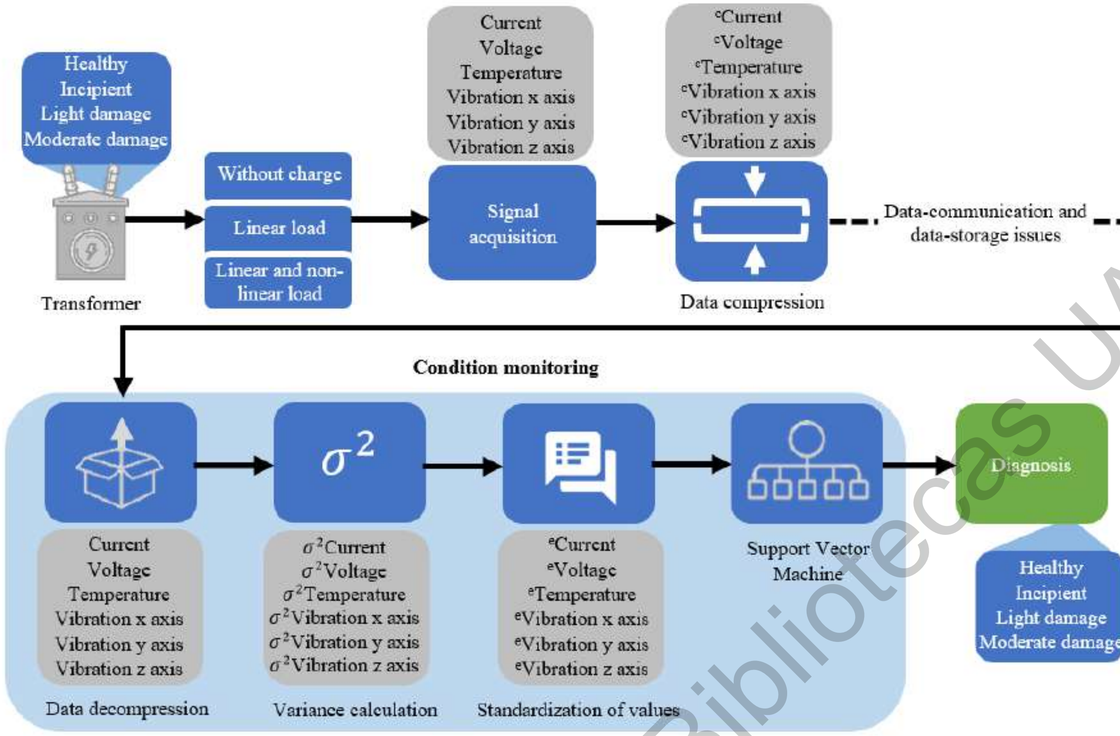


Figure 3.1 Scheme for the proposed methodology.

To perform the experimentation under the three scenarios, the circuit of Figure 3.2 will be carried out using three SAP4050D solid state relays (see Figure 3.3). The linear load module will have a resistance of 100 Ω and the non-linear module will have a rectifier bridge MB3505, a capacitance of 14 uF, and a resistance of 100 Ω. These values are chosen for infrastructure availability.

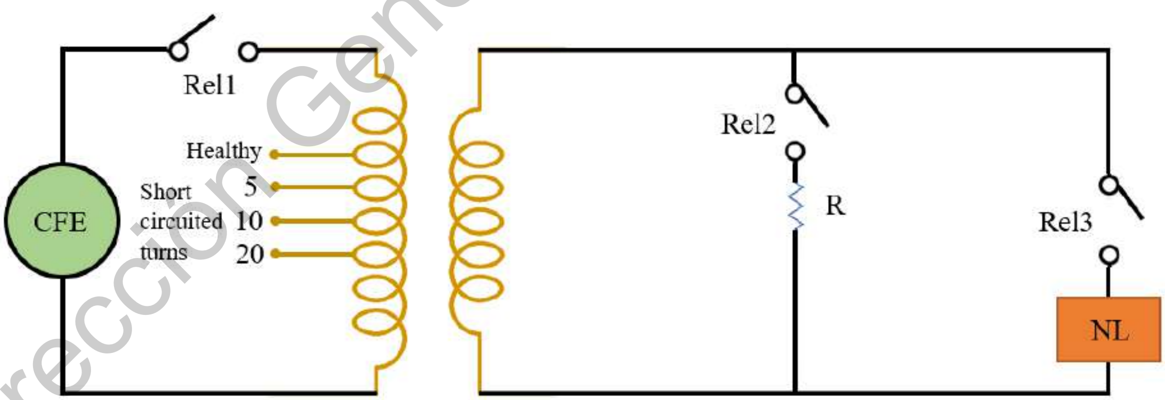


Figure 3.2 Circuit implementation scheme.

Figure 3.3 SAP4050D solid state relay.

The graph in Figure 3.3 exemplifies the operation times for the relays. During 1 s the three relays will remain low; at the end of this time, the first relay (Rel1) will be activated and remain in that state until the end of the test. The activation of Rel1 represents the no-load scenario. Meanwhile, the second relay (Rel2) will not be activated until after 3 s and will remain in this condition until the end of the test, representing the linear load scenario. Finally, the third relay will be activated after 5 s to represent the linear and non-linear load scenario until the end of the test (corresponding to 7 s of acquisition).

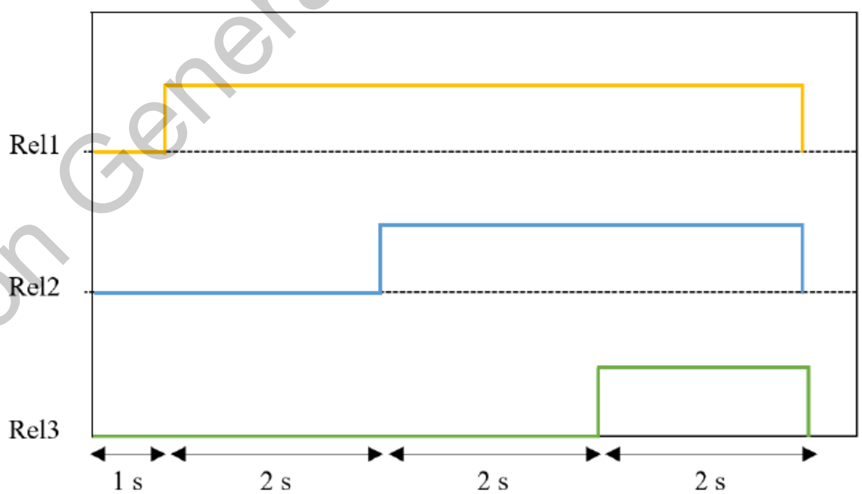


Figure 3.4 Relay activation chart.

3.2 Transformer

A single-phase transformer is used for the experimentation. It has a total of 135 turns and a power of 1.5 kVA, operating at 120 V. The transformer used for the tests is shown in Figure 3.5. The device was modified to represent the healthy condition and the three short-circuit conditions.

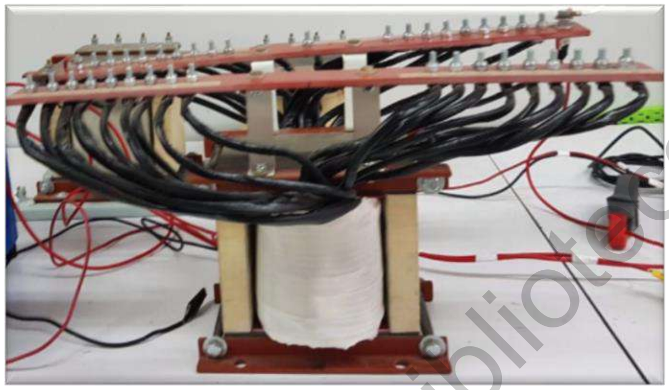


Figure 3.5 Transformer used.

3.3 Induction of faults

In this research work, three fault conditions will be considered in addition to the transformer's healthy condition, providing four conditions in total. As discussed in section 2.3, the first condition corresponds to the incipient fault which is considered with a percentage of 4% of the transformer's short-circuited turns, followed by the condition of 11%, and 18% of short-circuited turns which corresponds to light damage and moderate damage, respectively.

The fault induction process consists of generating the short-circuit fault in a synthetic way by isolating a certain number of turns of the transformer according to the in-test fault, see Table 3-1.

Table 3-1 Condition relationship according to fault type.

Condition	**Percentage of short-circuit turns**	**Number of short-circuits turns**
Healthy	0%	0
Incipient	4%	5
Light damage	11%	15
Moderate damage	18%	25

3.4 Data acquisition system

3.4.1 Sensors

A general description of the specifications of each sensor used in the tests of this work is presented.

For the capture of current signals and with the aim at making accurate readings without opening the circuit, the Fluke i200s current clamps for double range alternating current (20 and 200 A) with voltage output through a BNC connector and safety isolation are used. This clamp can be seen in Figure 3.6 and its speciafications can be consulted in Table 3-2.

Figure 3.6 Fluke i200s current clamps.

Table 3-2 Fluke i200s current clamps specifications.

Specification	Value
Rated current range	200 A
Direct current range	0.5 A to 200 A
Maximum non-destructive	240 A
Minimum measurement current	0.5 A
Basic accuracy	1% + 0.5 A (48 Hz a 65 Hz)
Wearable frequency	40 Hz a 10 kHz
Output levels	1 mA/A
Max voltage	600 V

A conventional transformer with 6 V as output was used to obtain the voltage signals as shown in Figure 3.7. This transformer has an input voltage of 127 V – 220 V and an intensity of 1 A. It is used to convert the measured input voltage into a proportional output voltage of up to 6 V so that this voltage can be read directly by the acquisition board.

Figure 3.7 6 V output transformer.

The accelerometer used for the vibration measurement can be seen in Figure 3.8. The Kistler brand 8395A10 triaxial accelerometer is of the IEPE type, has a measuring range of ± 10 *G* with a resolution of 500 *mV/G* and a bandwidth 1000 Hz. This type of accelerometer has a low level of noise, which allows accurate measurements in addition to withstanding temperatures from -54 ºC to 125 ºC. It has a magnetic assembly that facilitates instrumentation and ensures that the position in which it is placed remains stable, avoiding other disturbances. Besides, it has the advantage of having a voltage output proportional to the acceleration that is applied to it, which allows more efficient handling of the acquired data.

Figure 3.8 Triaxial accelerometer model 8395A10.

The LM35 sensor shown in Figure 3.9 was chosen to obtain temperature signals from the transformer. This sensor has a calibrated accuracy of 1 °C. Its measuring range is from -55 ºC to 150 ºC. The output is linear and each Celsius degree is equal to 10 *mV*. The low output impedance, its linear output, and its calibration make it suitable. Due to its low supply current, a very low self-heating effect occurs. Its main characteristics can be consulted in Table 3-3.

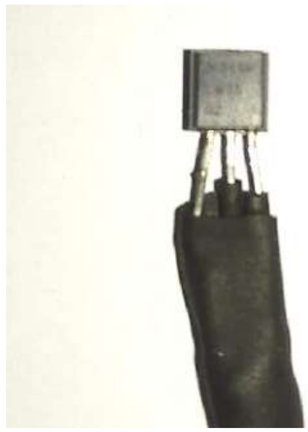


Figure 3.9 LM35 temperature sensor.

Table 3-3 LM35 sensor characteristics.

Specification	Value
Resolution	10 *mV* for ºC
Supply voltage	4 *V* to 20 *V*
Type of measurement	Analog output
Number of pins	3 pins (GND, VCC and Voutput)
Precision	± ¼ ºC
Current consumption	60 *mA*

3.4.2 Acquisition board

The data are acquired with the National Instruments board DAQ NI-USB 6211 model, which can be seen in Figure 3.10. This board is a multifunction Data Acquisition (DAQ) device that offers both analog and digital inputs and outputs and two 32-bit counters. It provides an integrated amplifier designed for fast settings at high scanning speeds. In addition, it allows two-way data transfer at high speed through the USB bus. It has 16 analog input channels, 2 for analog outputs, 4 input channels, and 4 digital outputs, all with a resolution of up to 16 bits and a maximum sampling frequency of 250 kS/s (kilo samples per second). The details of the NI-USB 6211 board can be found in Table 3-4.

Figure 3.10 Data acquisition board model NI-USB 6211.

Table 3-4 Details of the NI-USB 6211 board.

Specification	**Value**
Bus connector	USB 2.0
Maximum number of analog channels of single terminal	16
Maximum number of differential analog input channels	8
Sampling rate	250 kS/s
Analog Input Resolution	16 bits
Number of analog output channel	2

Being an instrument of National Instruments, distributor of LabVIEW software, communication between the software and the board is achieved by installing the NIDAQ 1451f1 library which is also provided by National Instruments through its website and direct download.

3.4.3 Control software for signal acquisition

The graphic interface designed in LabVIEW for the acquisition of signals corresponding to the four types of variables involved is displayed in Figure 3.11.

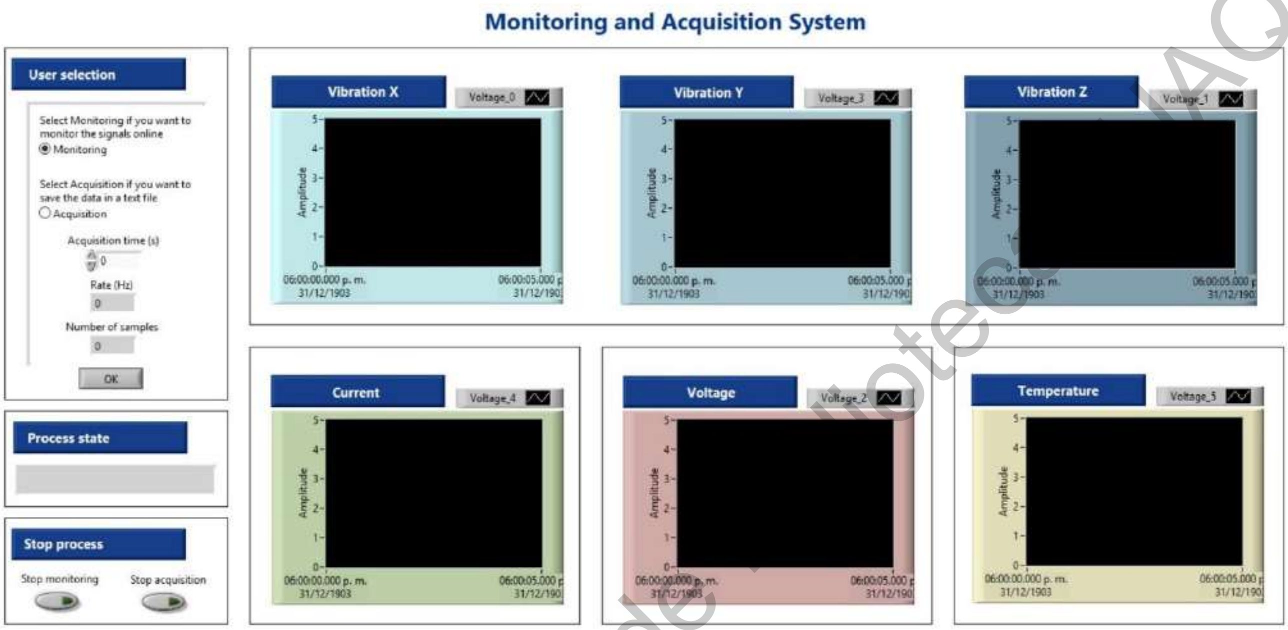
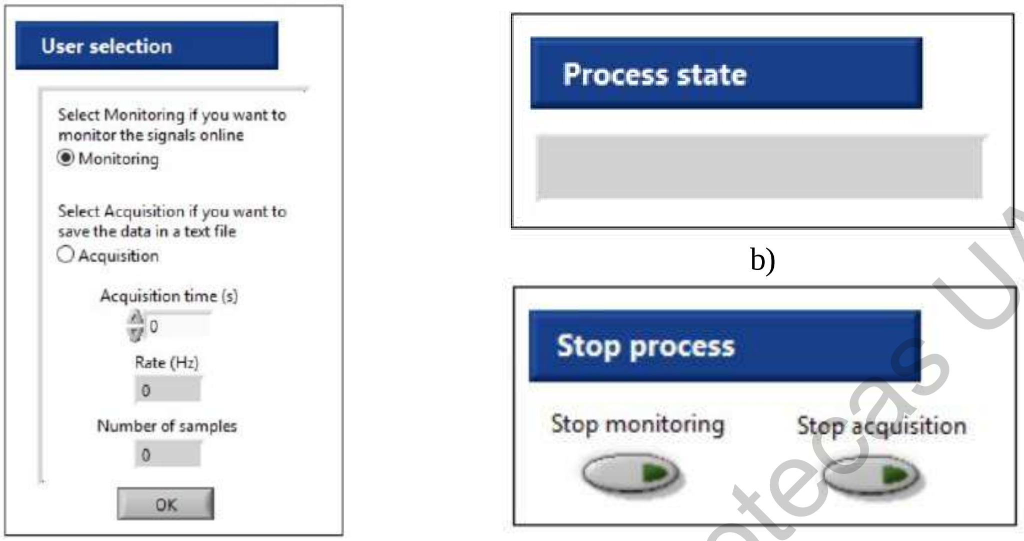


Figure 3.11 Graphical interface for signal acquisition.

The interface of Figure 3.11 consists of four sections. The first one, User selection, allows the user to choose the online signal monitoring mode and the signal acquisition mode to be stored in files with mat extension. The format in which the data is stored is as a column for each channel (vibration on the x-axis, vibration on the y-axis, vibration on the z-axis, current, voltage, and temperature). In this option, the user will have to indicate the acquisition time in seconds and the system will automatically calculate the total samples to be acquired based on the sampling frequency that is 600 Hz (Figure 3.12 a). The second section consists of six graphs in where it is possible to visualize each of the signals coming from the different sensors. The next section, Process state (Figure 3.12 b), has a text box that is updated with the legend "Online monitoring" in case the user has selected the Monitoring option and the legend "Data acquisition is complete" when the user select the Acquisition option and the data acquisition is complete. The last section, stop process, allows the user to stop the process by means of the corresponding switch (Figure 3.12 c).

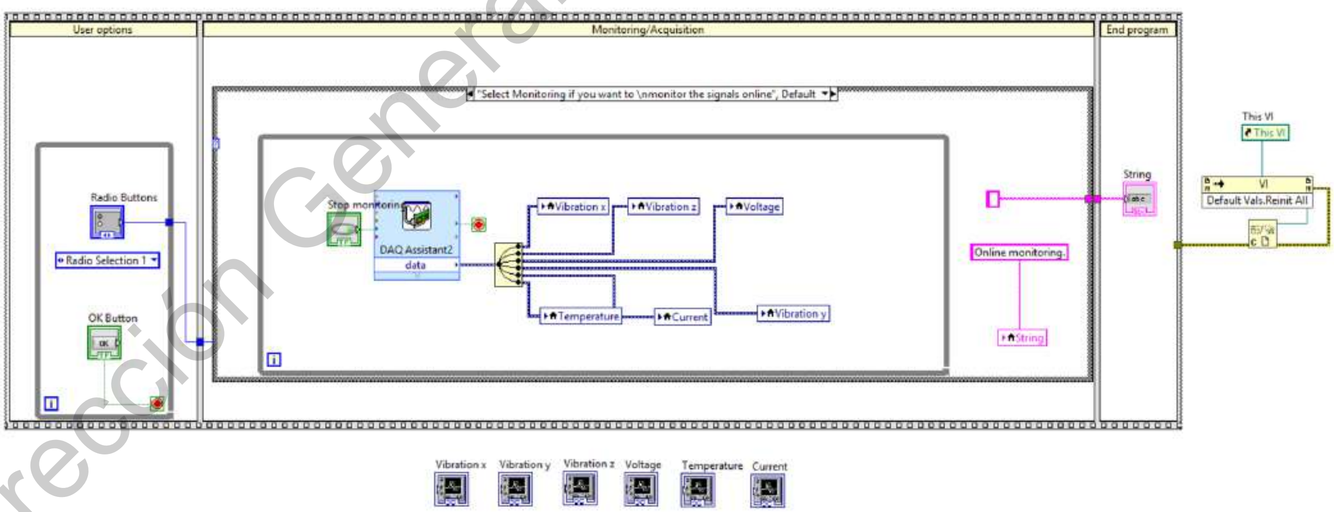


b)

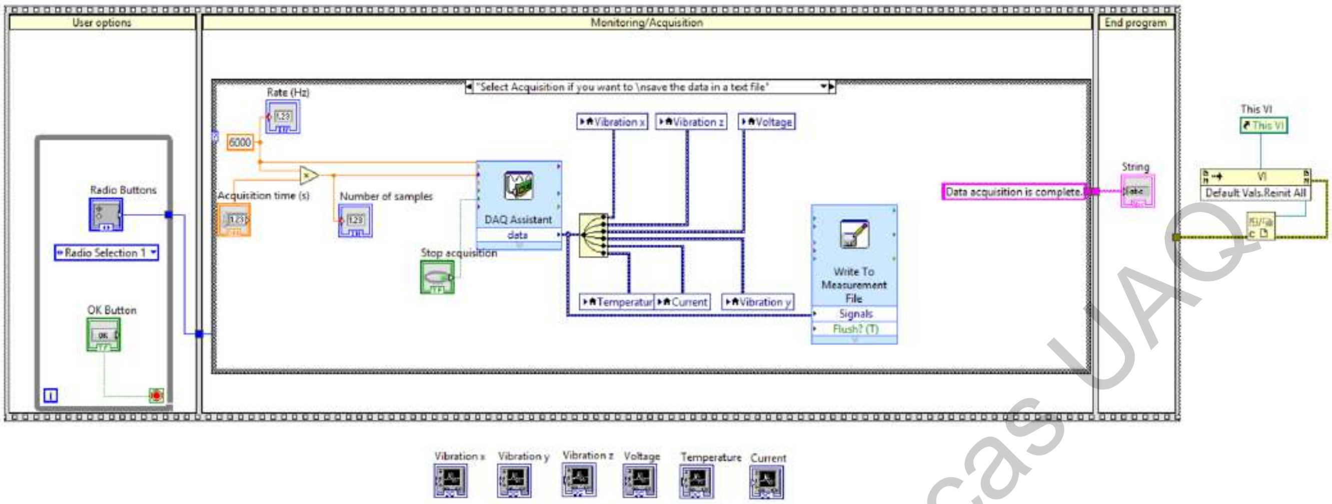
a) c)

Figure 3.12 a) User selection section, b) Process state section, c) Stop process section.

Figure 3.13 shows the LabVIEW block diagram for user selection Monitoring and Acquisition.



a)



b)

Figure 3.13 a) LabVIEW monitoring block diagram, b) LabVIEW acquisition block diagram.

3.5 Graphical user interface for monitoring

To monitor the condition of the transformer, the graphic user interface designed in MATLAB was created. In Figure 3.14, it is possible to see this interface developed in order to visualize the results obtained from the proposed methodology.

The window has five main sections:

1. *User selection.* In this part the user will enter the quantization parameter to start with the signal processing. This value, presented in section 2.4.5, is used for the compression of data. A default value of 0.5 is proposed according to the work of Contreras et al. (2019). This value clearly represents a good compromise between compression performance and the original signal.

2. *Compression.* In this section, the results obtained from the compression of the signals will be displayed. The criteria shown to the user are: Average compression time per record represented in seconds and the Average CR.

3. *Decompression*. The results obtained from the decompression step are shown in this section. The user can display the Average recover time per record shown in seconds and the results of equations 18, 19, and 20 of section 2.4.6, referring to Average recover CR, Average PRD, and Average recover QS, respectively.

4. *Variance.* In the Variance section, the equation 35 of section 2.6.1 is computed. Six graphs of the variance calculated for each of the variables studied will be displayed by considering the four operating conditions of the transformer.

5. *SVM Results.* This section provides the user with a visual representation of the scope of the classification performance, resulting from the computation of the SVM algorithm presented in section 2.5.4. The contour plot shown to the user is made up of the resulting classification regions and the marks (·) that represent the SVM measurements.

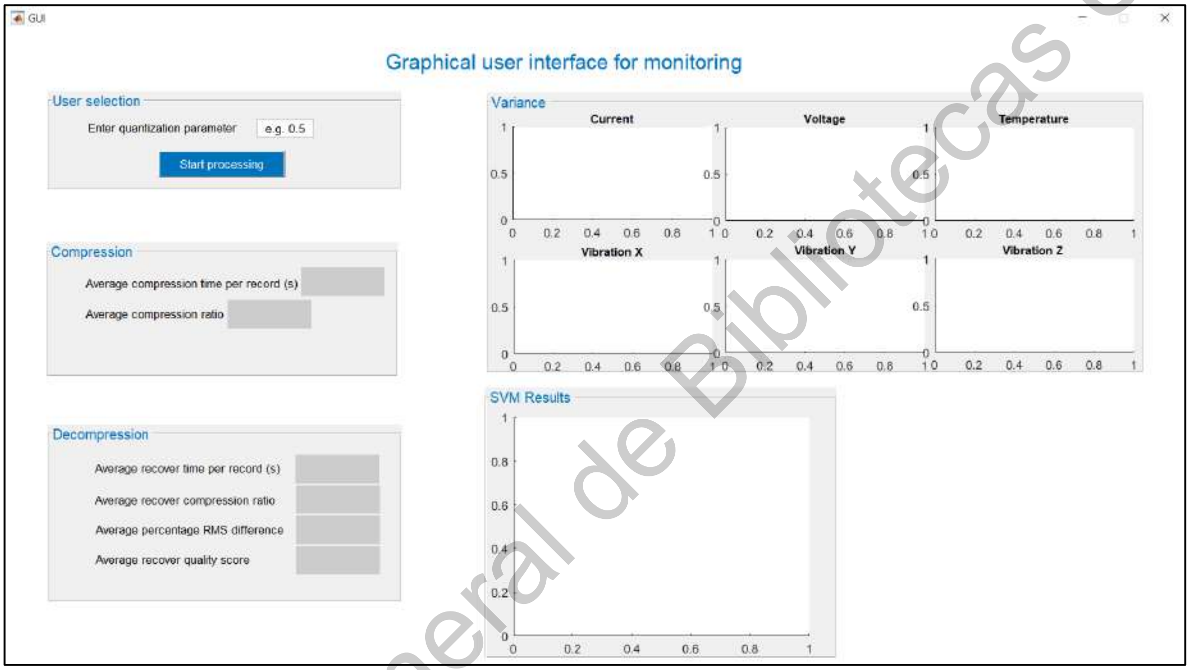


Figure 3.14 Graphical user interface for monitoring.

Chapter 4

4. Experimentation and results

4.1 Experimental setup

The experimental setup used to develop and test the proposed methodology is shown in Figure 4.1. The signals have a duration of 7 s with a sampling frequency of 6000 samples / s. The transformer under test operates in a healthy condition and under three different fault conditions: incipient (4% short-circuited turns), light damage (11% short-circuited turns), and moderate damage (18% short-circuited turns). For each condition, 20 tests or acquisitions are made. An autotransformer is used to de-energize the transformer after each test. All processing is carried out using a personal computer (PC).

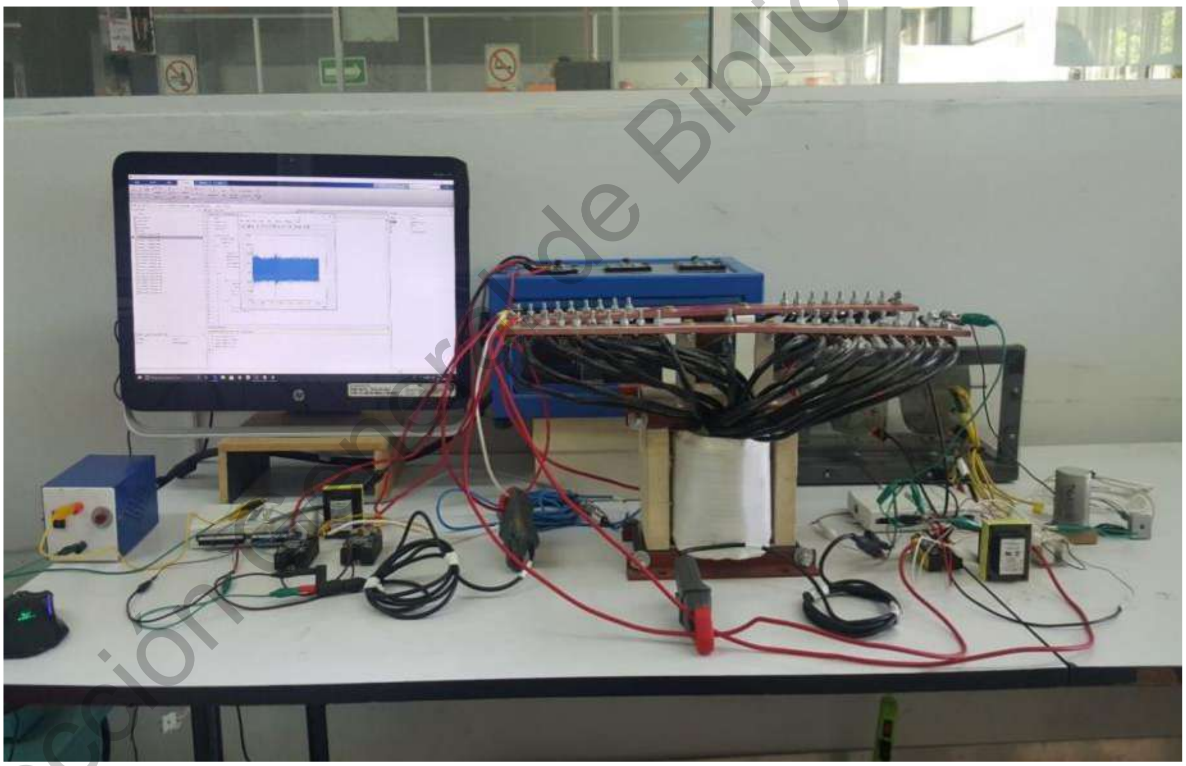


Figure 4.1 Experimental setup.

4.2 Fault matrix

In Table 4-1, it is possible to consult the matrix of failures according to the methodology, as well as the amount of tests that will be performed for each condition of the transformer, which are healthy condition, incipient failure, light damage, and moderate damage, under the three aforementioned scenarios (without electric load, linear electric load, and non-linear electric load), obtaining a total of 240 tests.

Table 4-1 Matrix of induced failures.

Scenario	Condition	Number of test	Acquisition time	Sampling rate
Without load	Healthy Incipient Light damage Moderate damage	80 (20 for each condition)	7 s	6000 Hz
Linear load	Healthy Incipient Light damage Moderate damage	80 (20 for each condition)	7 s	6000 Hz
Linear and non-linear load	Healthy Incipient Light damage Moderate damage	80 (20 for each condition)	7 s	6000 Hz

4.3 Signal acquisition results

As explained in section 3.1, the first step of the methodology consists of the acquisition of signals from the four variables studied (also considering the three vibration axes) by using the acquisition board and the sensors corresponding to each physical quantity. Figures 4.2 and 4.3 show the evidence of signal acquisition through the graphical interface developed in the LabVIEW software under the two available modalities: online monitoring and acquisition of the signals to be stored in files.

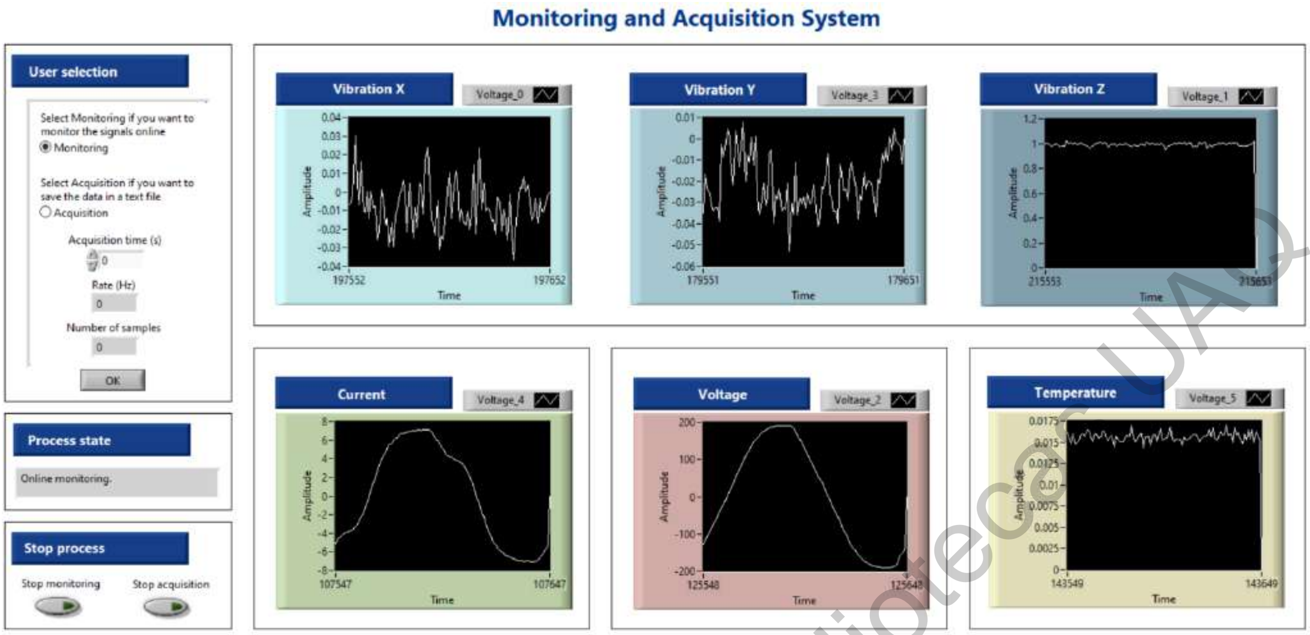


Figure 4.2 Graphical interface for signal acquisition monitoring mode.

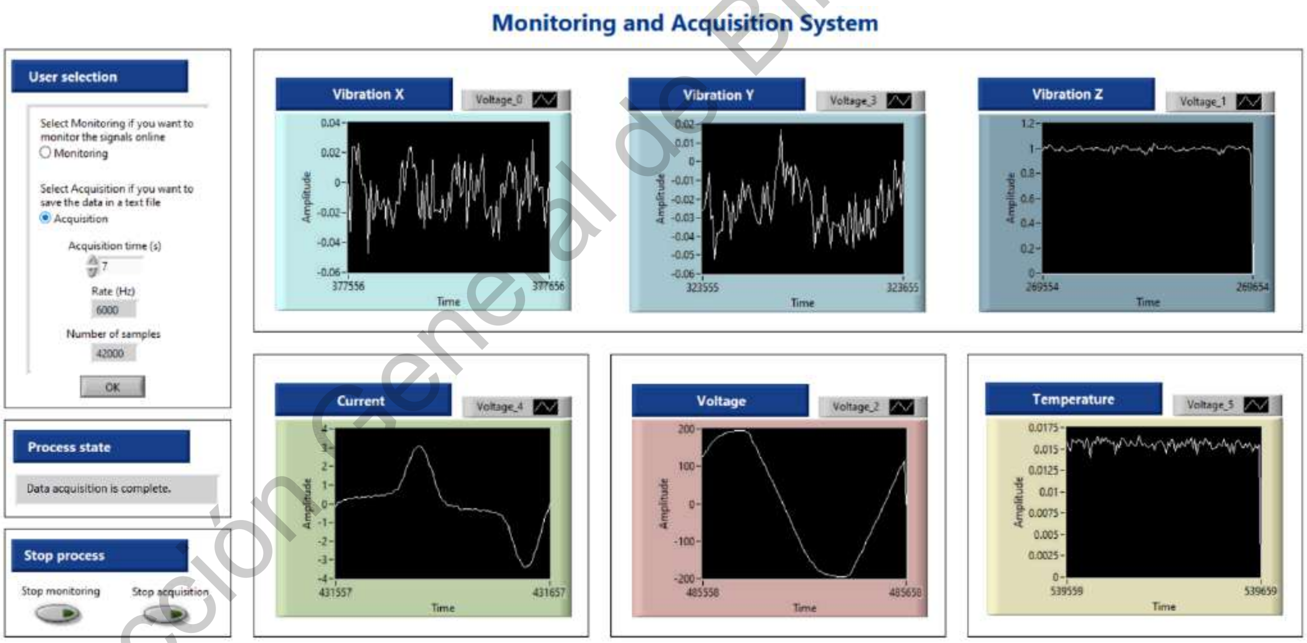


Figure 4.3 Graphical interface for acquisition of signals acquisition mode.

4.4 Data compression and decompression results

After the acquisition of the signals, the compression and decompression processes are executed. The compression results obtained from the current, voltage, temperature, and vibration signals (x, y, and z axes) are presented in Table 4-2. This table contains the quantification parameter (Δ) proposed for each variable, the average time of compression per test given in seconds, and the average CR for the 20 tests corresponding to each condition. The level of decomposition Wavelet used is level 4. This level was selected because it is the one with the highest success rates in terms of compression according to Contreras *et al.* (2019).

It is worth noting that different Δ values were tested. The second column of Table 4-2 shows the final Δ values. These values reflect a trade-off between compression performance and reconstruction quality. It should be noted that different parameters for Δ were used according to the in-test variable since they all have different characteristics from each other, e.g., amplitude, unit of measurement, among others.

Table 4-2 Compression results.

Compression				
Variable	**Δ**	**Condition**	**Average time per record (s)**	**Average CR**
Current	0.5	Healthy	0.0365	162.3703
		Incipient	0.0225	138.7834
		Light damage	0.0229	147.6528
		Moderate damage	0.0238	111.5360
Voltage	0.5	Healthy	0.0295	48.2334
		Incipient	0.0290	49.6718
		Light damage	0.0592	57.5713
		Moderate damage	0.0293	52.2812
Temperature	0.00001	Healthy	0.0350	38.0293
		Incipient	0.1005	17.9775
		Light damage	0.0883	17.5585
		Moderate damage	0.0940	18.3057
Vibration x	0.005	Healthy	0.0311	44.3742
		Incipient	0.0759	41.3827
		Light damage	0.0732	38.7574
		Moderate damage	0.0733	38.8161
Vibration y	0.005	Healthy	0.0946	47.3550
		Incipient	0.0977	45.9261
		Light damage	0.0753	40.5482
		Moderate damage	0.0525	43.6276

Vibration z	0.005	Healthy	0.0400	41.5727
		Incipient	0.0409	38.0366
		Light damage	0.0394	31.4623
		Moderate damage	0.0518	29.9164

The decompression results can be consulted in Table 4-3 where the time parameters per test in seconds, CR, PRD, and QS of average decompression are analyzed.

Table 4-3 Decompression results.

Decompression					
Variable	**Condition**	**Average time per record (s)**	**Average CR**	**Average PRD**	**Average QS**
Current	Healthy	0.0140	162.3703	687.8835	0.2360
	Incipient	0.0257	138.7834	576.1427	0.2409
	Light damage	0.0144	147.6528	411.1734	0.3591
	Moderate damage	0.0155	111.5360	323.5488	0.3447
Voltage	Healthy	0.0149	48.2334	635.4766	0.0759
	Incipient	0.0154	49.6718	508.9320	0.0976
	Light damage	0.0167	57.5713	446.6184	0.1289
	Moderate damage	0.0169	52.2812	290.9133	0.1797
Temperature	Healthy	0.0164	38.0293	94.5642	0.4022
	Incipient	0.0173	17.9775	113.2352	0.1588
	Light damage	0.0173	17.5585	89.1463	0.1970
	Moderate damage	0.0182	18.3057	66.4639	0.2754
Vibration x	Healthy	0.0168	44.3742	327.5888	0.1355
	Incipient	0.0170	41.3827	279.2048	0.1482
	Light damage	0.0173	38.7574	336.5877	0.1151
	Moderate damage	0.0177	38.8161	438.8015	0.0885
Vibration y	Healthy	0.0169	47.3550	241.3699	0.1962
	Incipient	0.0171	45.9261	405.0521	0.1134
	Light damage	0.0199	40.5482	449.9902	0.0901
	Moderate damage	0.0168	43.6276	443.0095	0.0985
Vibration z	Healthy	0.0172	41.5727	4.9341	0.4257
	Incipient	0.0203	38.0366	6.2107	0.1244
	Light damage	0.0176	31.4623	14.1939	0.2166
	Moderate damage	0.0189	29.9164	34.6791	0.8627

In order to schematically exemplify the above results, the graph of Figure 4.4 shows the compression performance in relation to the quality of the recovered signals evaluated by the PRD.

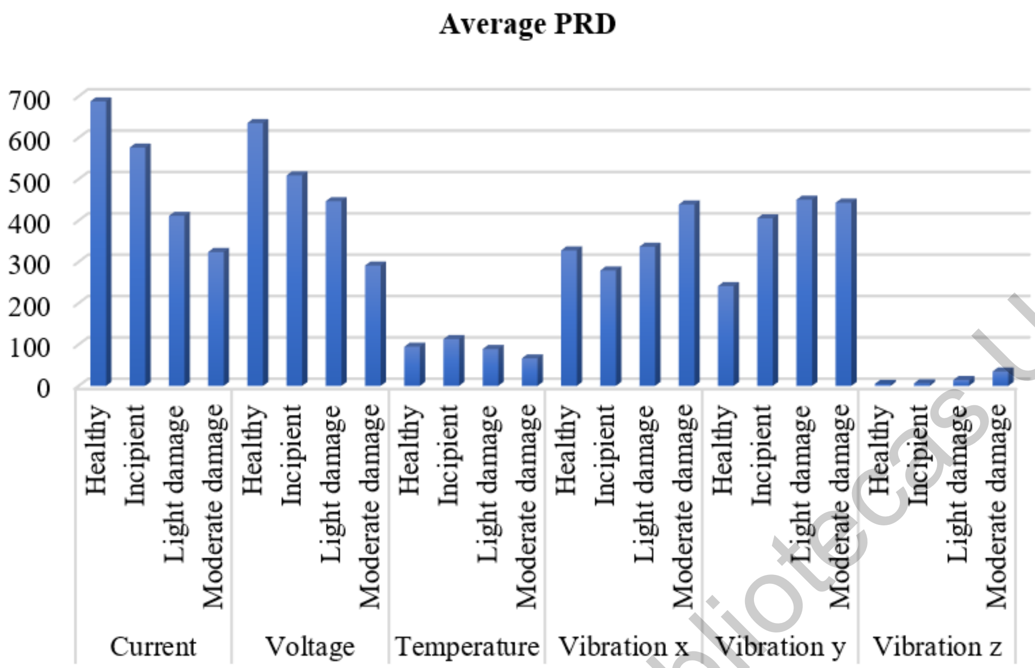
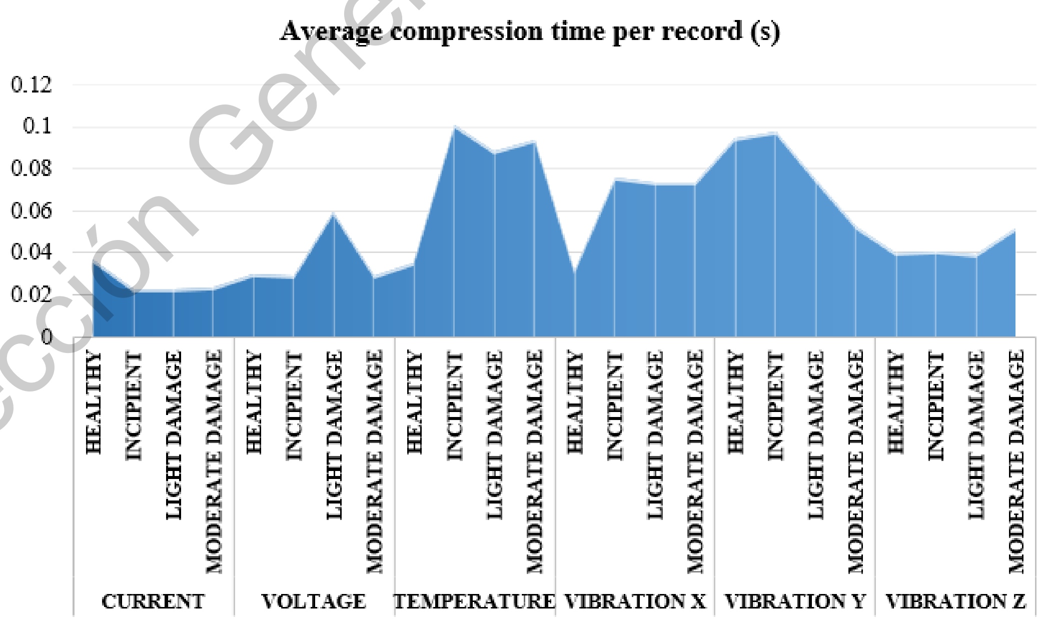
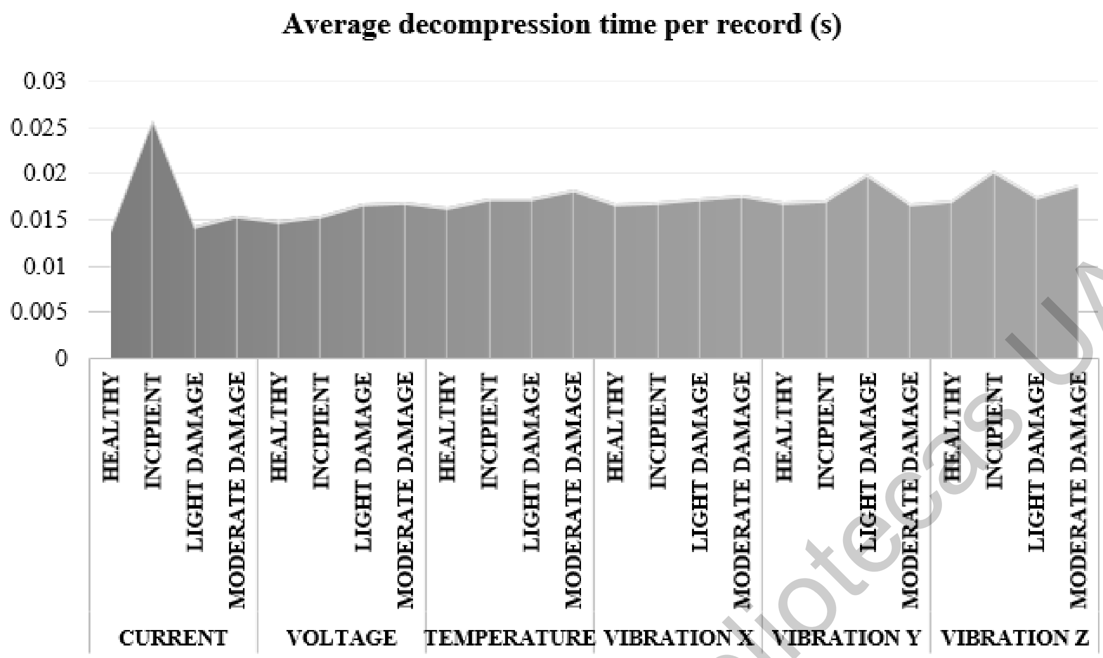


Figure 4.4 Average PRD.

The averages of compression and decompression time per record present results that are worth noting. The time to process a total of 42,000 data (corresponding to the data contained in a single record) does not exceed 0.1 s in terms of time of compression (see Figure 4.5 a). With respect to the decompression processing time, the maximum time does not exceed 0.03 s (see Figure 4.5 b); therefore, the technique can be considered faster.



a)



b)

Figure 4.5 a) Average compression time per record, b) Average decompression time per record.

Moreover, when plotting the size in bytes of the original files vs. the files that contain the recovered signal, an outstanding decrease in the size of the recovered signal is observed, moving from an average range of 1.3 Mb to a range of 0.03 Mb. Likewise, the CR values obtained in the compression stage are kept in the decompression step, which guarantees the proper execution of the data decompression algorithm.

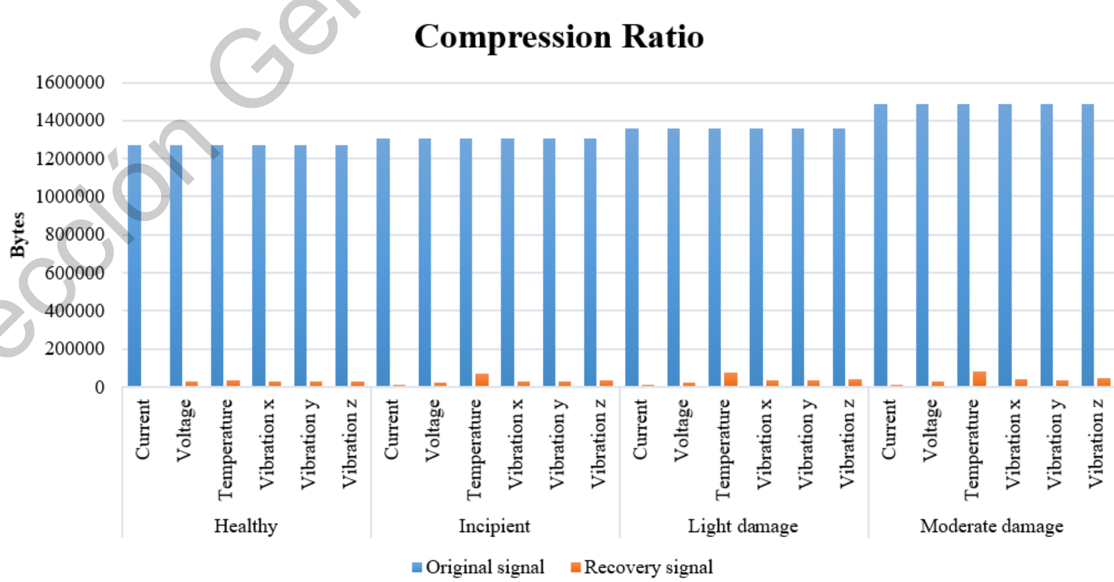


Figure 4.6 Compression Ratio.

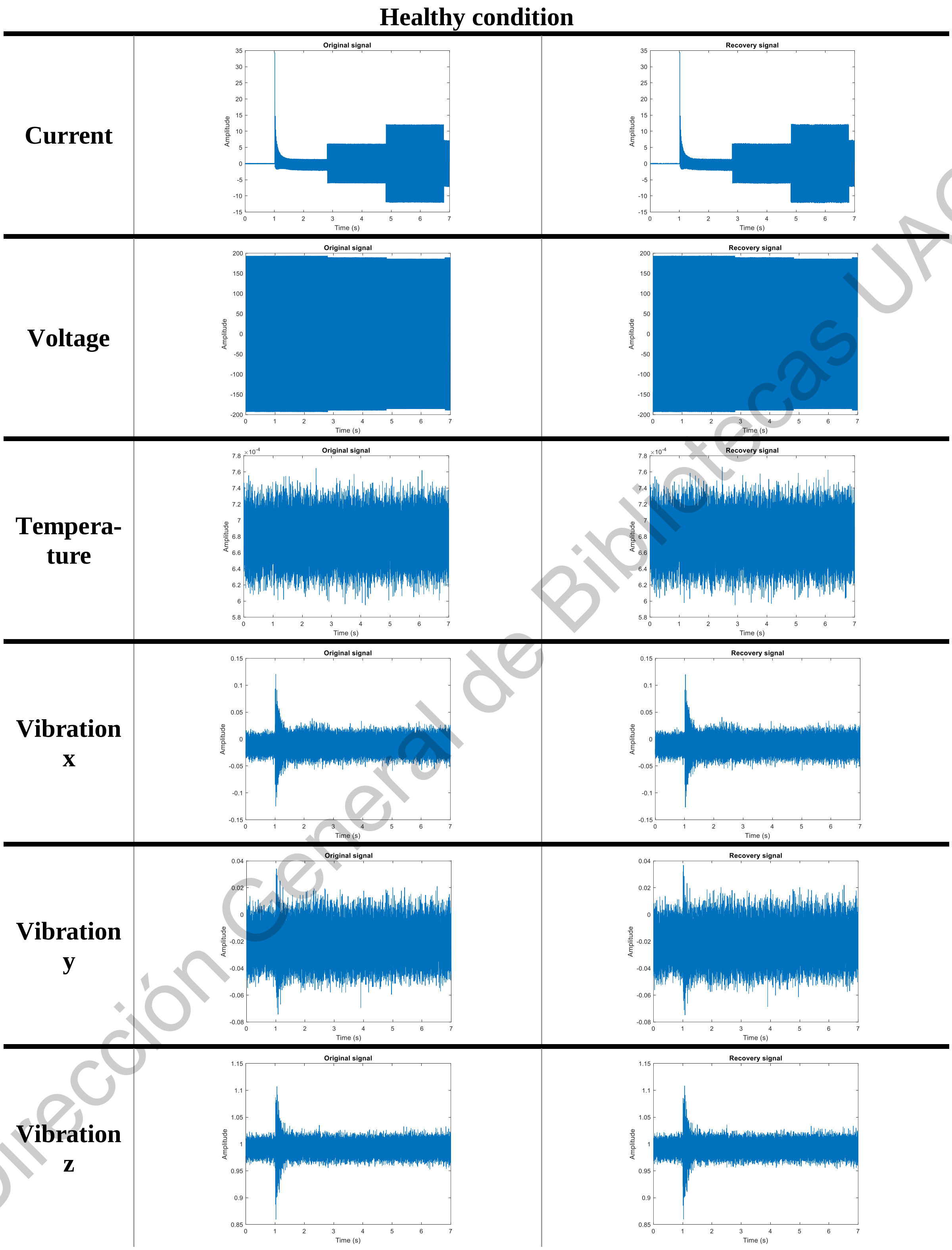


Figure 4.7 Visual quality of the original signal against the reconstructed signals: Healthy condition.

Incipient condition

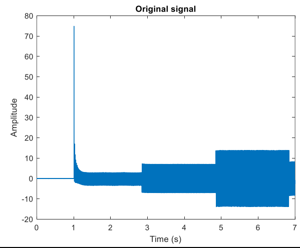
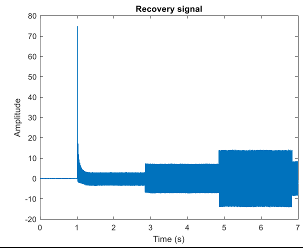
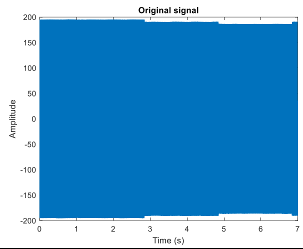
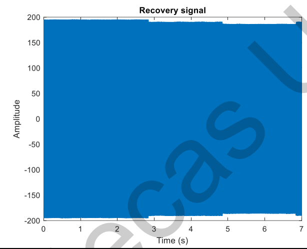
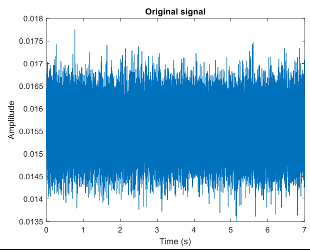
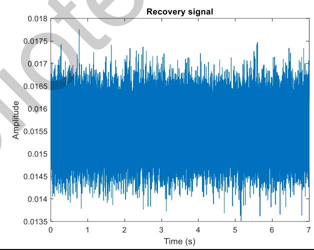
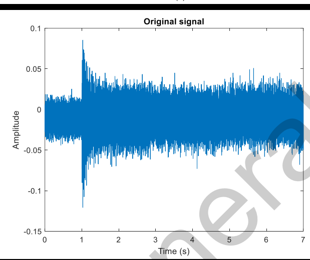
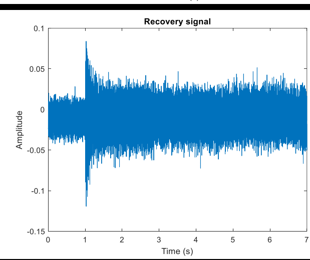
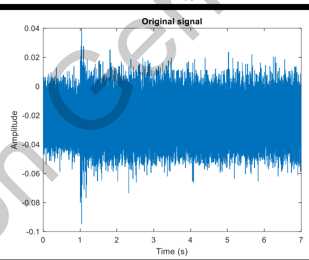
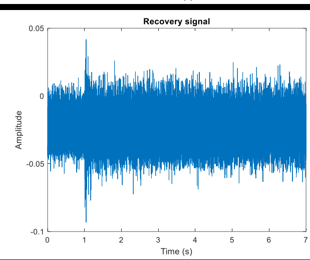
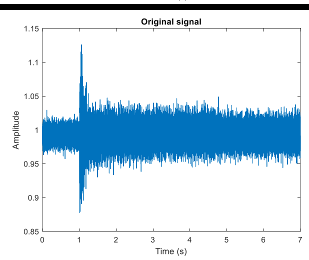
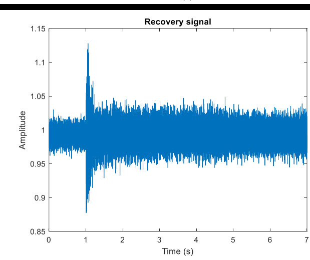
Current		
Voltage		
Tempera-ture		
Vibration x		
Vibration y		
Vibration z		

Figure 4.8 Visual quality of the original signal against the reconstructed signals: Incipient condition.

Light damage condition

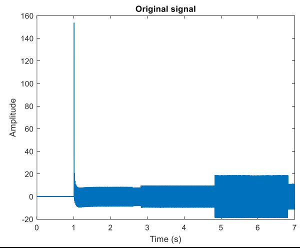
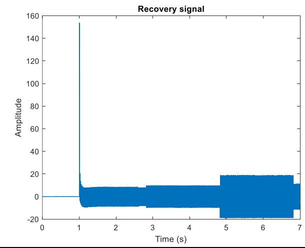
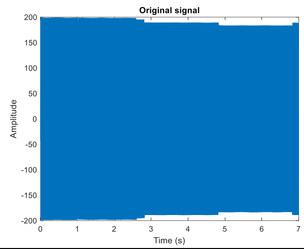
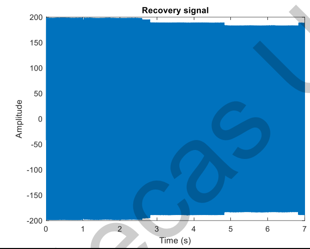
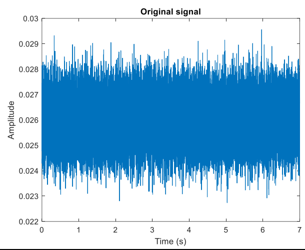
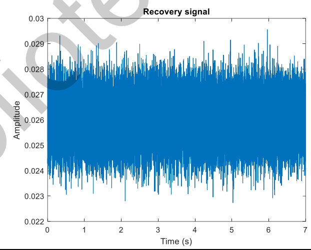
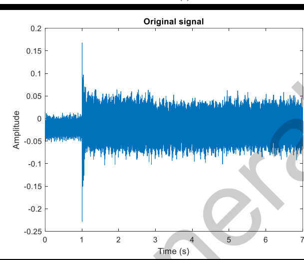
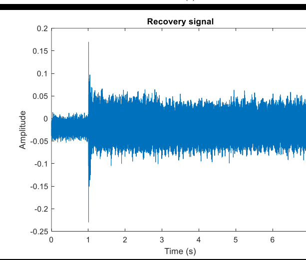
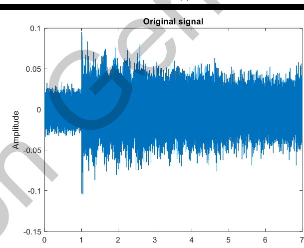
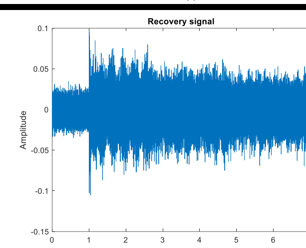
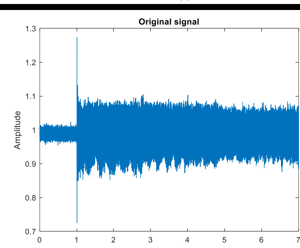
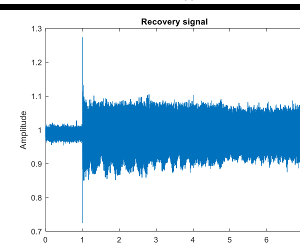
Current		
Voltage		
Tempera-ture		
Vibration x		
Vibration y		
Vibration z		

Figure 4.9 Visual quality of the original signal against the reconstructed signals: Light damage condition.

Moderate damage condition

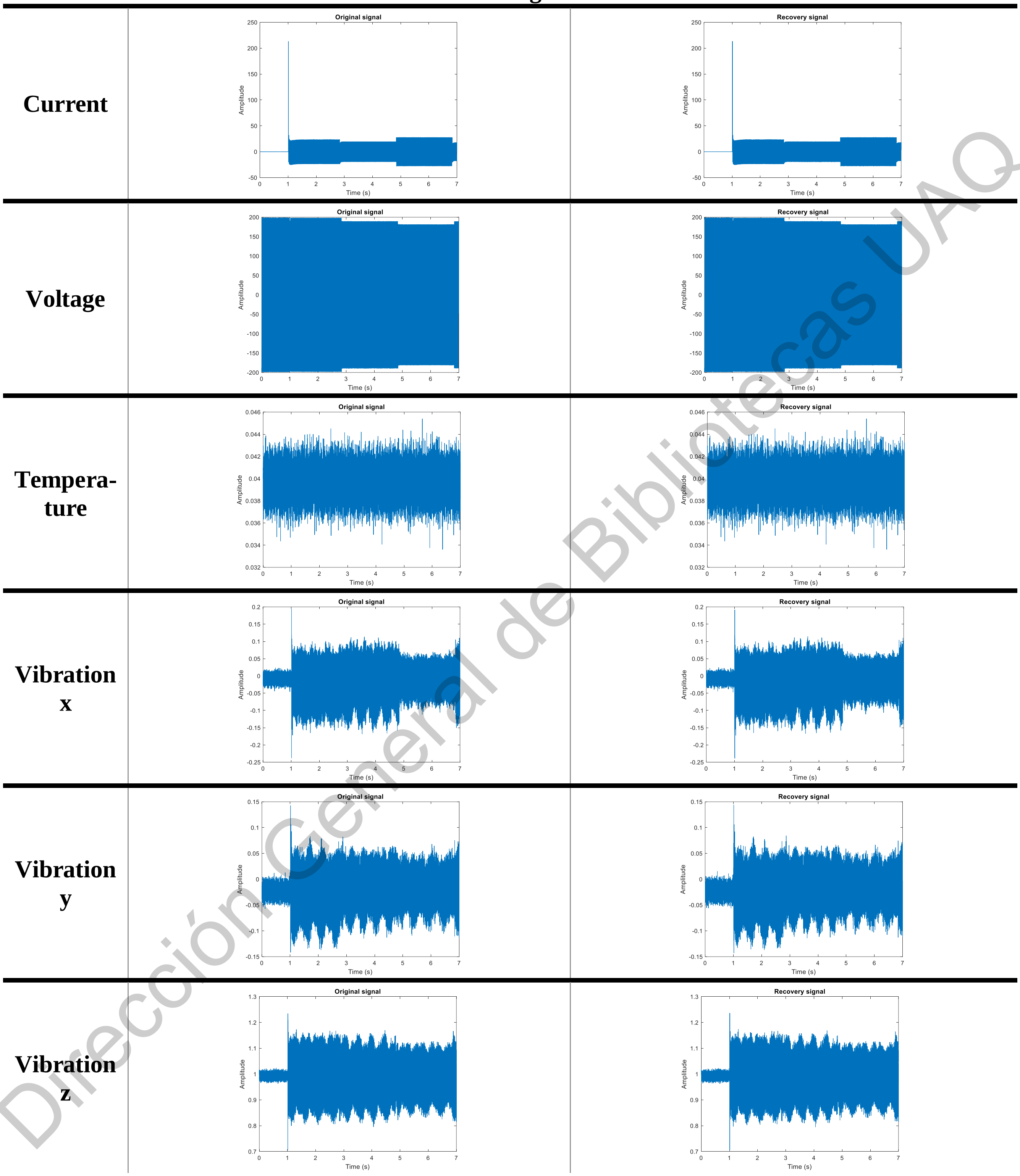


Figure 4.10 Visual quality of the original signal against the reconstructed signals: Moderate damage condition.

Although the QS values shown in Table 4-3 can be interpreted to conclude the good level of reconstruction quality obtained, the graphs in Figures 4.7, 4.8, 4.9 and 4.10 provide a much clearer picture of the data, where the visual quality of the reconstructed signals for each condition is compared.

4.5 Results of the variance calculation

After completing the process of compression and decompression of the signals, the calculation of the variance for each signal has to be carried out, where a standardization procedure for the values obtained is included in order to have a mean of 0 and a standard deviation of 1. This step provides, a common basis by taking as reference the healthy condition of the transformer. The results of the variance are presented in this section as graphs. In addition, the calculation of the average of the standardized data is presented in a table format under the three study scenarios. It is worth mentioning that due to standardization some results have negative values.

4.5.1 Scenario: without load

Figure 4.11 shows the plotted results of the variance obtained for each variable and in Table 4-4 the average of the standardized results is presented.

As can be observed in Figure 4.11, the variance results for the vibration on the x-axis show the clearest separation between the four classes by considering the 20 tests; in the rest of the variables, an overlap is noted between the healthy and incipient conditions mainly. In the temperature variable, the overlap occurs between light and moderate damage conditions.

Table 4-4 Mean variance for the scenario without load.

	Variable					
Condition	**Current**	**Voltage**	**Temperature**	**Vibration x**	**Vibration y**	**Vibration z**
Healthy	-1.0436	-7.9603	-4.7184	5.4401	-3.5749	1.2879
Incipient	6.4605	-1.2477	921.1392	9.7603	9.0066	10.3767
Light damage	58.0582	0.3304	2981.2593	37.1121	106.4429	105.7928
Moderate damage	214.3695	-3.6138	4684.8555	171.5593	269.5869	730.3983

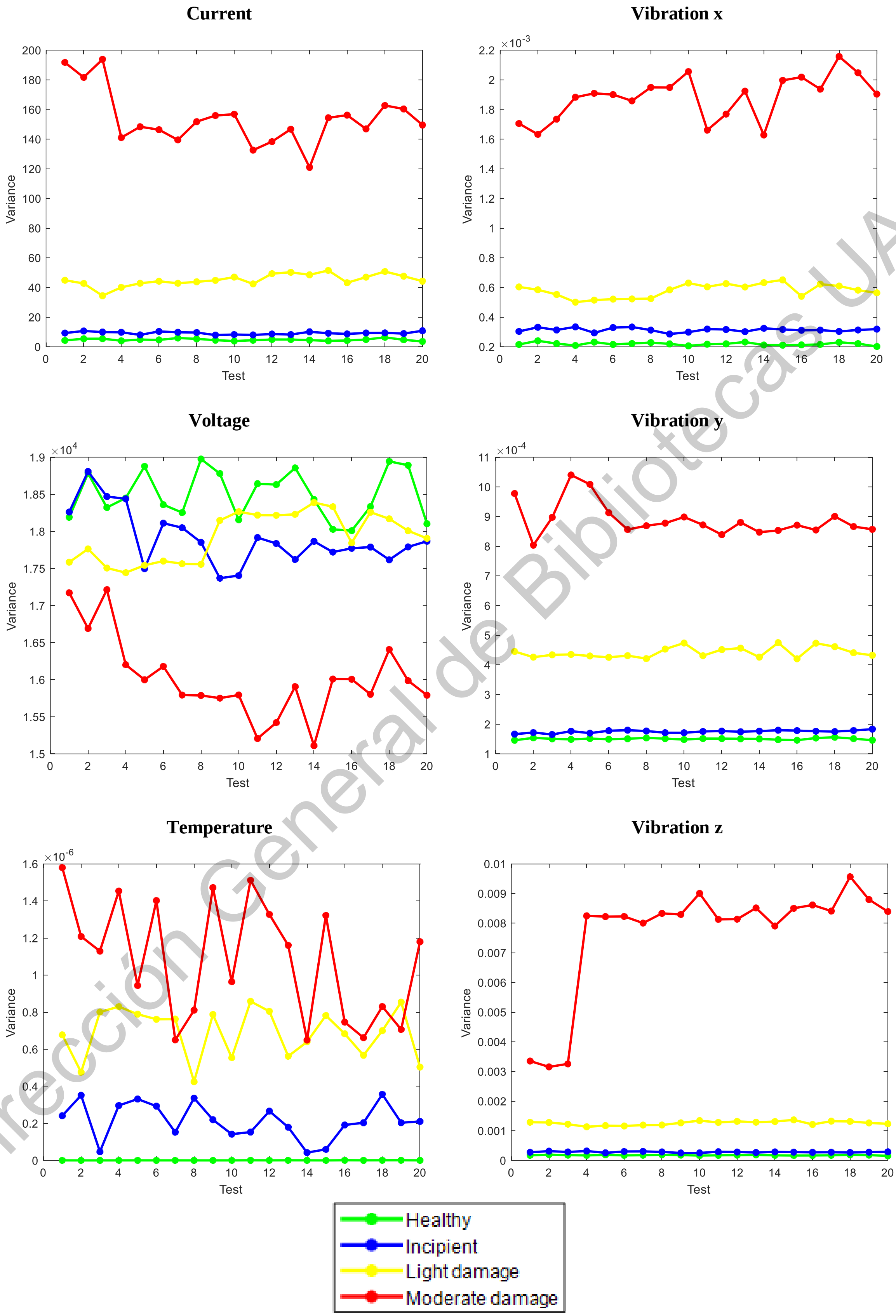


Figure 4.11 Variance results plotted for the scenario without electric load.

4.5.2 Scenario: linear load

According to Figure 4.12, the variance results in the scenario with linear electric load for the current and vibration variables on the x-axis indicate the separation for the four classes. In the case of vibration on the y-axis, the classes that overlap are healthy and incipient; on the other hand, in the case of temperature, the overlapped conditions are light and moderate damage. In the case of voltage, no class remains isolated from each other.

Table 4-5 shows the average value of the standardized variance of the four transformer conditions under the linear electric load scenario. These values show the general behavior for the four conditions and all the physical variables.

Table 4-5 Mean variance for the scenario with linear load.

	Variable					
Condition	**Current**	**Voltage**	**Temperature**	**Vibration x**	**Vibration y**	**Vibration z**
Healthy	-3.0930	-1.7986	3.9829	8.6486	-3.9080	-5.0904
Incipient	12.8307	-1.5497	916.5996	13.4262	7.8478	12.3067
Light damage	77.6123	-0.8589	2955.1043	69.3407	95.8409	221.4849
Moderate damage	254.7971	-6.0207	4621.5176	288.7851	239.6199	1214.6809

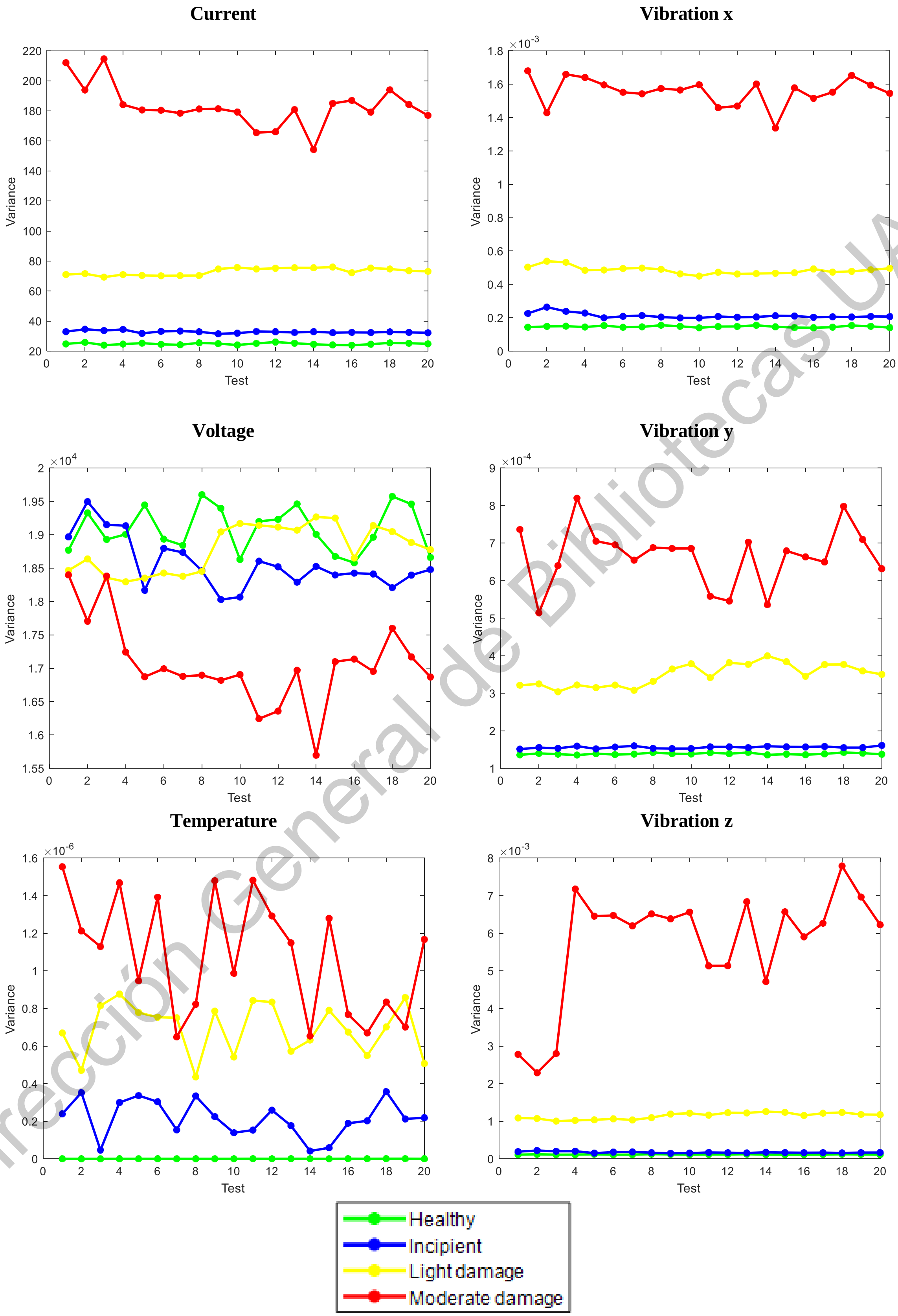


Figure 4.12 Variance results plotted for the scenario with linear electric load.

4.5.3 Scenario: linear and non-linear load

Regarding the variance results obtained for the scenario with linear and non-linear electric load, it is observed in Figure 4.13 that the fault conditions have a high sensitivity in the variables of current and vibrations in x and y since they are well defined (non-overlapped). The graphs corresponding to the variables of voltage, temperature and vibration in the z-axis show the same behavior referred to the scenario with linear load.

The product of the standardization of the variance results is reflected in Table 4-6 where the average of the 20 tests performed is presented. These values allow observing the general trends between the numerical values and the fault severity. Similar to Fig. 4.13, the values for current and vibrations do not show an overlap.

Table 4-6 Mean variance for the scenario with linear and non-linear load.

	Variable					
Condition	**Current**	**Voltage**	**Temperature**	**Vibration x**	**Vibration y**	**Vibration z**
Healthy	2.0872	-7.4718	-4.3021	4.8739	6.5503	2.9976
Incipient	15.9830	-1.8322	913.7344	11.5130	8.4262	15.3740
Light damage	81.5168	-1.7565	2952.6666	63.1803	64.2479	256.3168
Moderate damage	220.3974	-7.6266	4637.0067	237.4453	197.2661	1606.8282

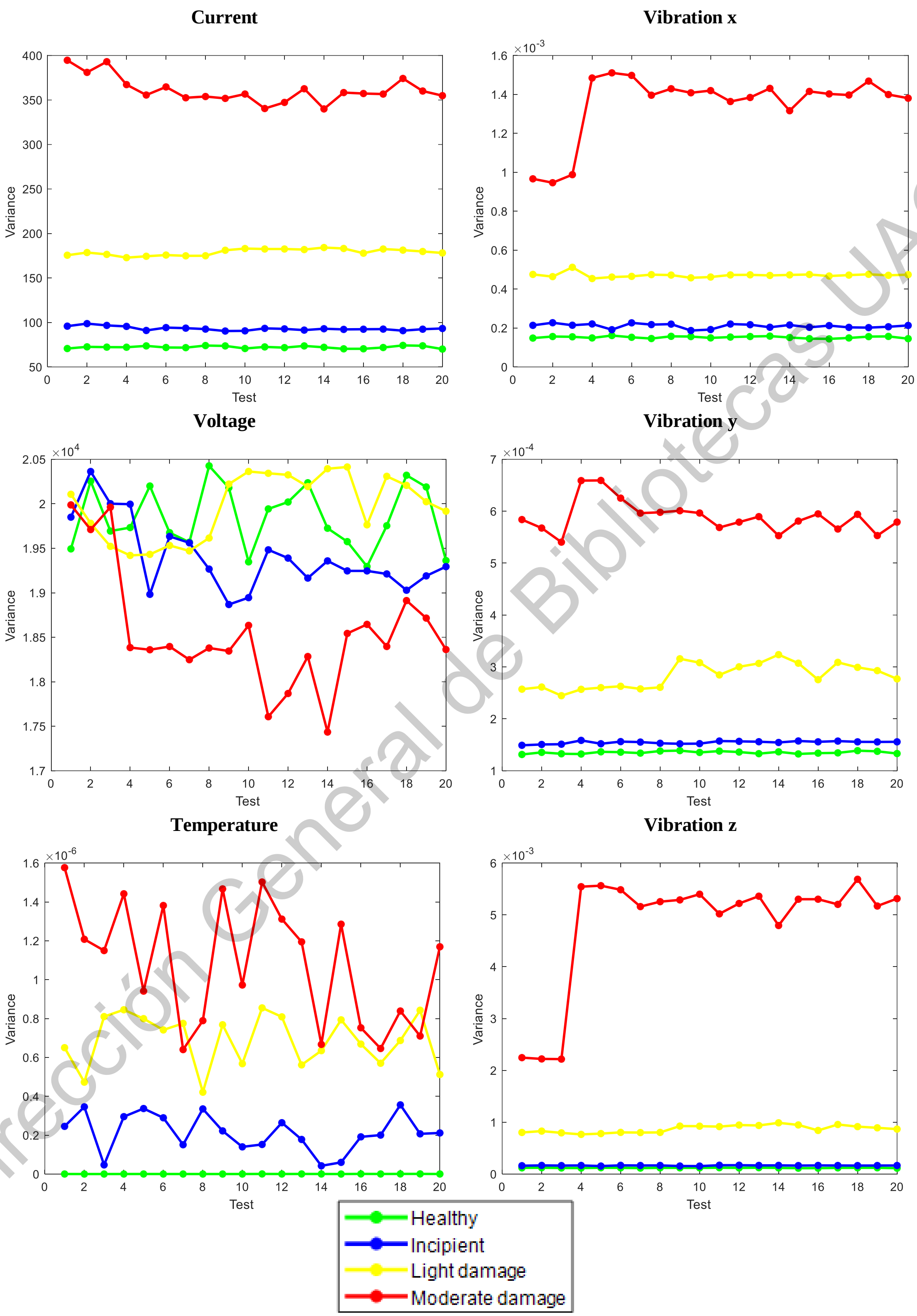


Figure 4.13 Variance results plotted for the scenario with linear and non-linear load.

4.6 Fault classification results by Support Vector Machine

Up to this point, four of the seven steps involved in the data mining process have been carried out. This procedure was previously detailed in subsection 2.5.1. First, the integration, the cleaning of the data by acquiring the four types of transformer signals, and the subsequent application of the data compression and decompression method were performed. These steps guarantee the combination of multiple data sources and their elimination of noise. Then, the data selection and transformation steps were carried out when the variance followed by the standardization of the results was calculated. Once the relevant data in the analysis task have been recovered and appropriately transformed, the step five which corresponds to the application of the data mining algorithm has to be implemented; in this case, the SVM is used. To obtain satisfactory predictive accuracy, the kernel with radial base function was used and two parameters of the kernel functions were adjusted: BoxConstraint argument equal to 2.5 and KernelScale equal to 1. These values were used for all scenarios.

The BoxConstraint argument can adjust the limits (i.e., the number of support vectors) by setting a frame constraint during training, while the kernel scale parameter (KernelScale) divides all the elements of the predictor matrix by the specified value.

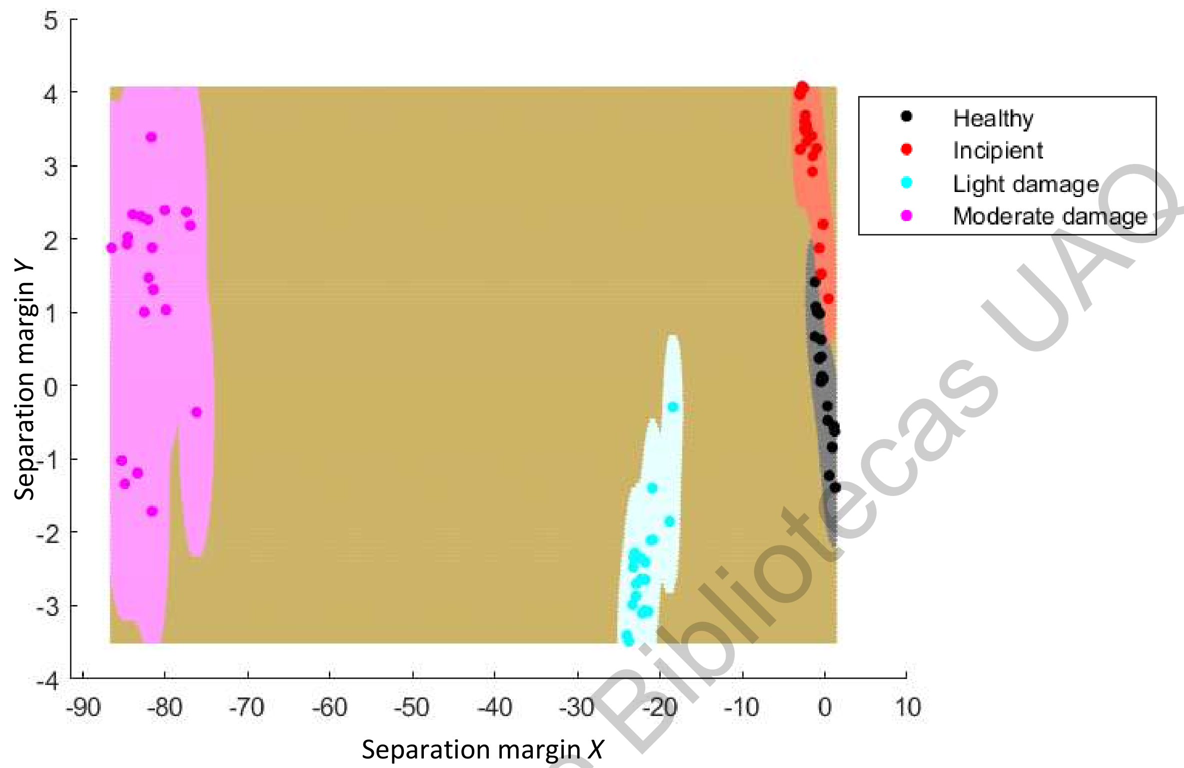
The algorithm input is an array of predictive data, where each row is an observation (corresponding to the 20 tests processed), and each column is a predictor, in this case, the variables of current, voltage, temperature, and vibrations (x-axis, y-axis, and z-axis).

The remaining steps of the data mining process that correspond to the evaluation and presentation of knowledge are detailed in the subsequent subsections. The technique of representation and visualization of the results corresponding to the graphic user interface designed in MATLAB is used to present the knowledge to the users.

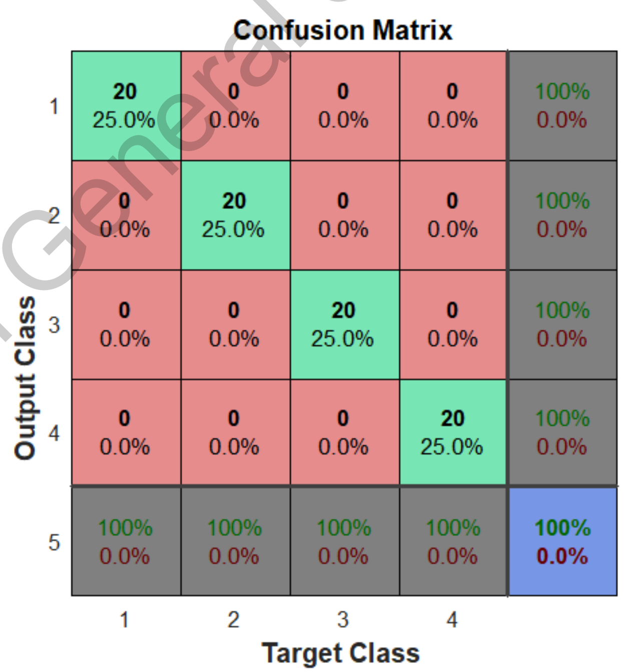
4.6.1 Scenario: without load

Figure 4.14 shows the SVM classification results for the conditions: healthy, incipient, light damage, and moderate damage, respectively, as well as the confusion matrix obtained.

Although the healthy and incipient conditions of the transformer seem to have an overlap in previous results, the SVM separates one class from the other one as can be seen in Figure 4.14 a). The results obtained indicate a classification effectiveness of 100% (Figure 4.14 b).



a)



b)

Figure 4.14. Stage diagnosis without load a) Contour plot of SVM classification, b) Confusion matrix.

Figure 4.15 shows the monitoring interface with the results obtained in the scenario without electric load.

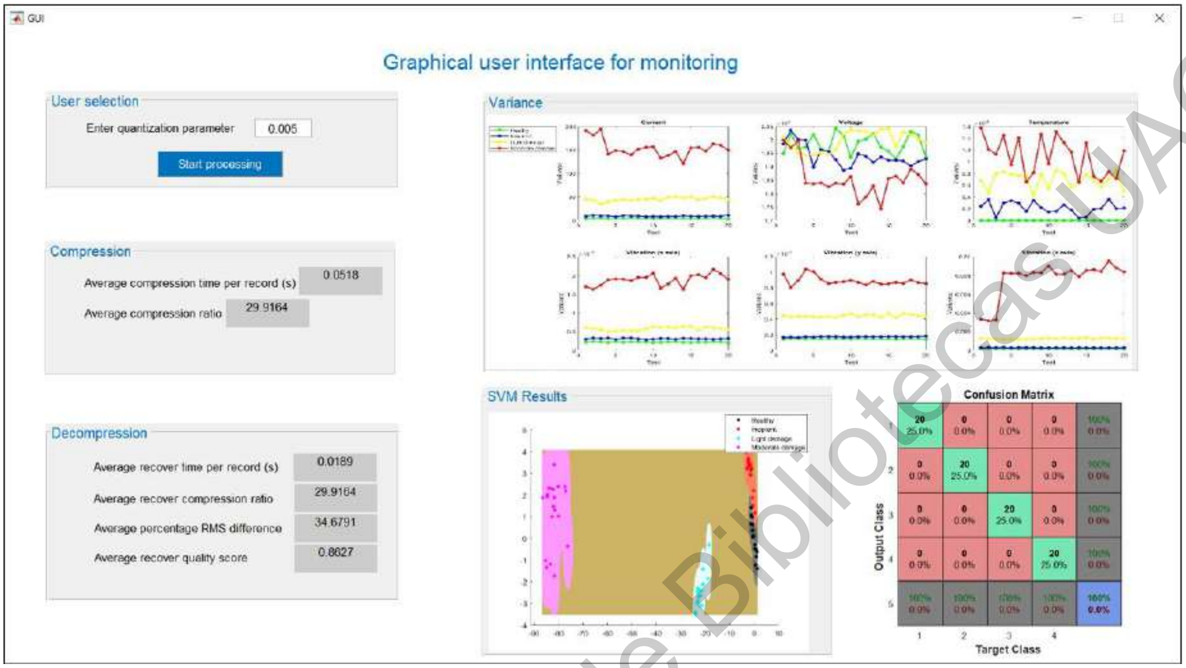
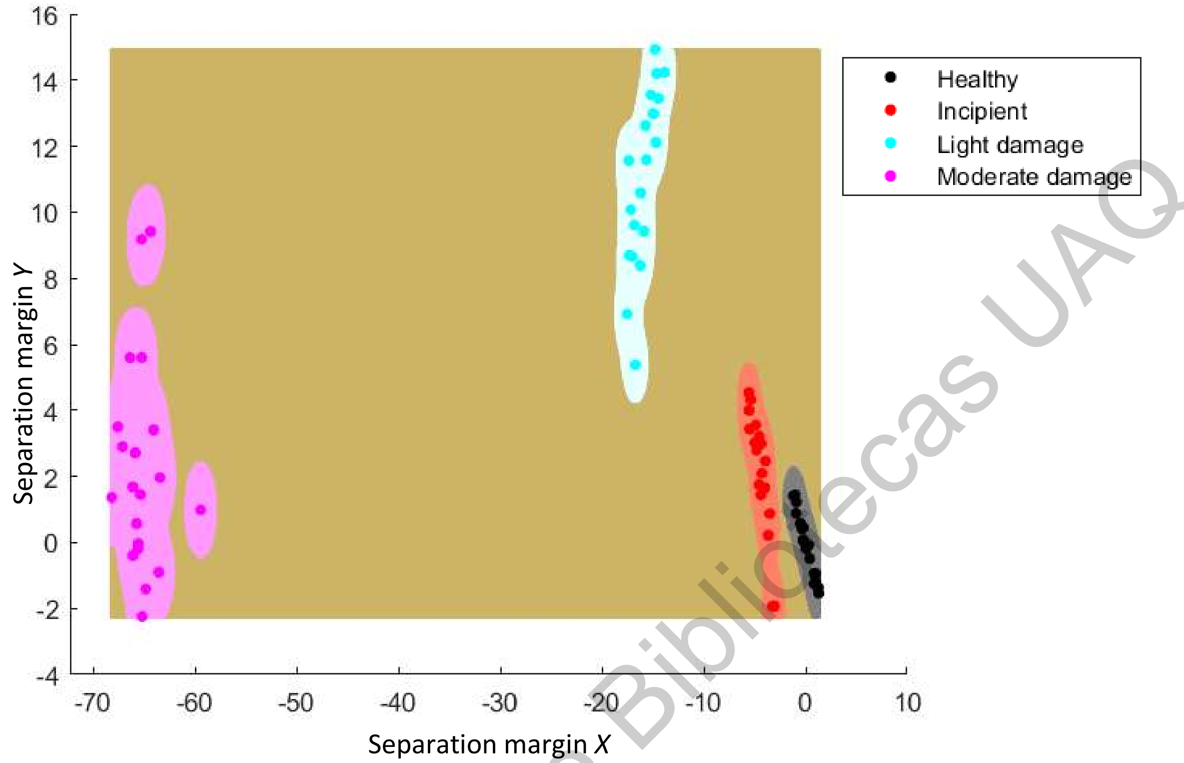


Figure 4.15 Graphical interface with the final results of the scenario without load.

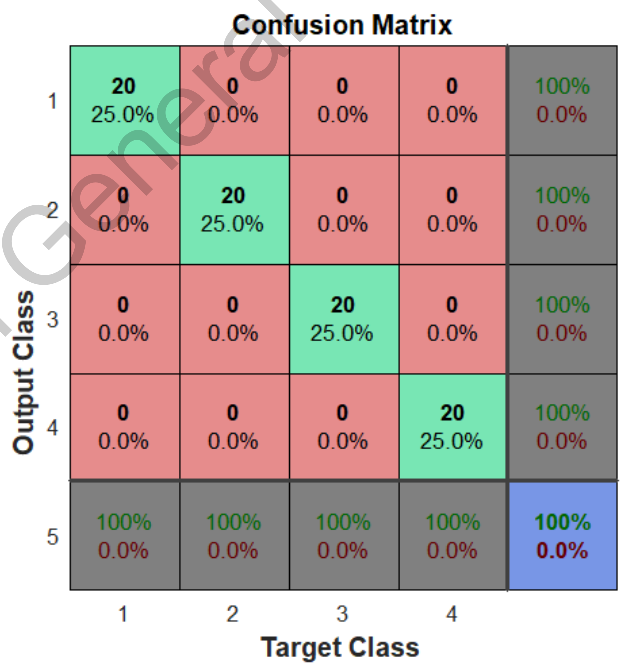
4.6.2 Scenario: linear load

In the following figures, it can be seen that the conditions plotted in Figure 4.16 a) show a considerable separation among the four conditions studied, allowing their correct identification and obtaining a classification percentage of 100% according to the confusion matrix (Figure 4.16 b).

The interface that contains the final results of the linear load scenario is shown in Figure 4.17.



a)



b)

Figure 4.16 Stage diagnosis with linear load a) Contour plot of SVM classification, b) Confusion matrix.

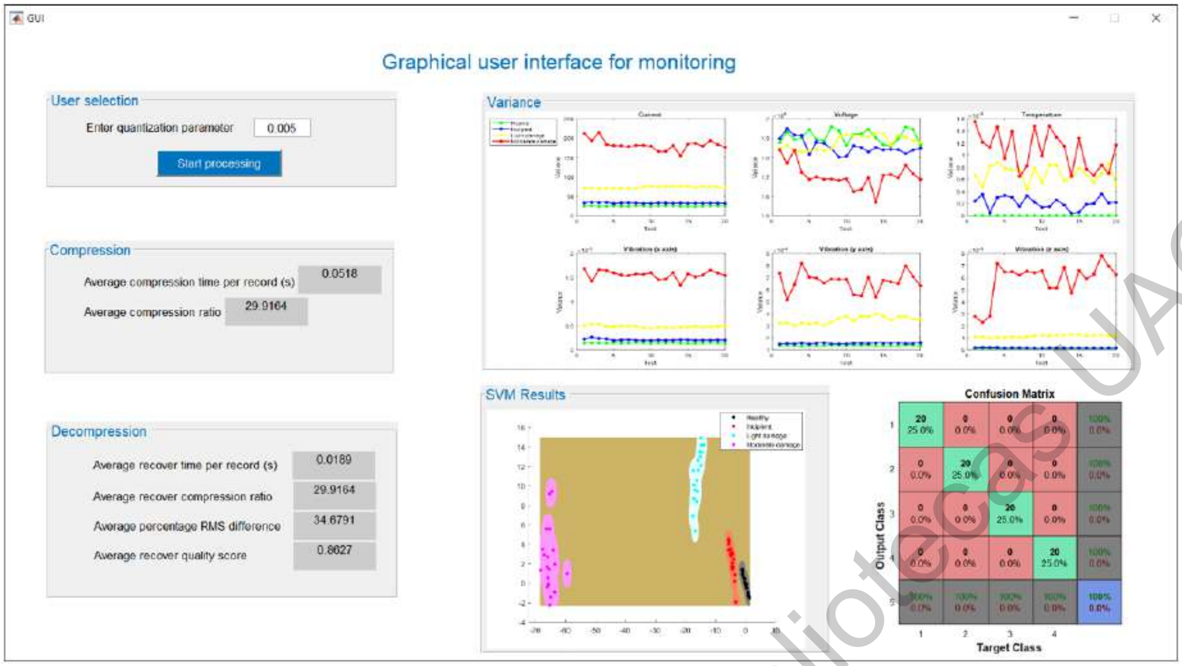
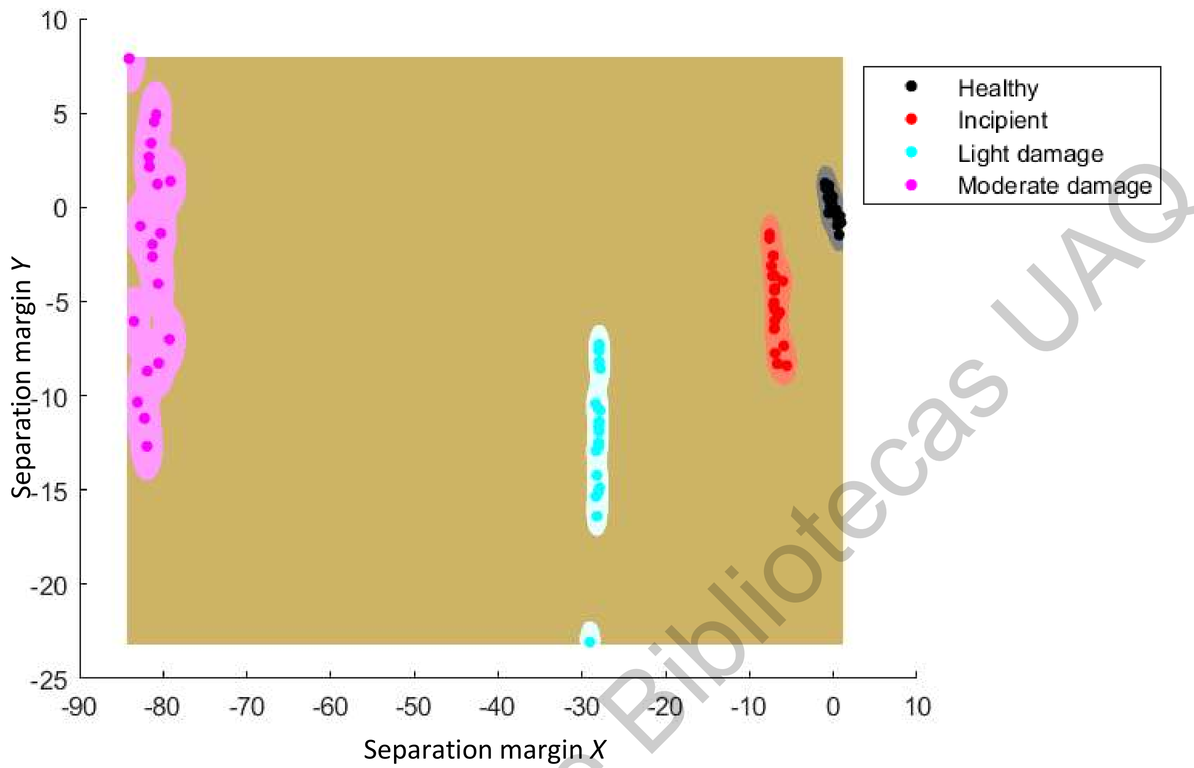


Figure 4.17 Graphical interface with the final results of the scenario with linear load.

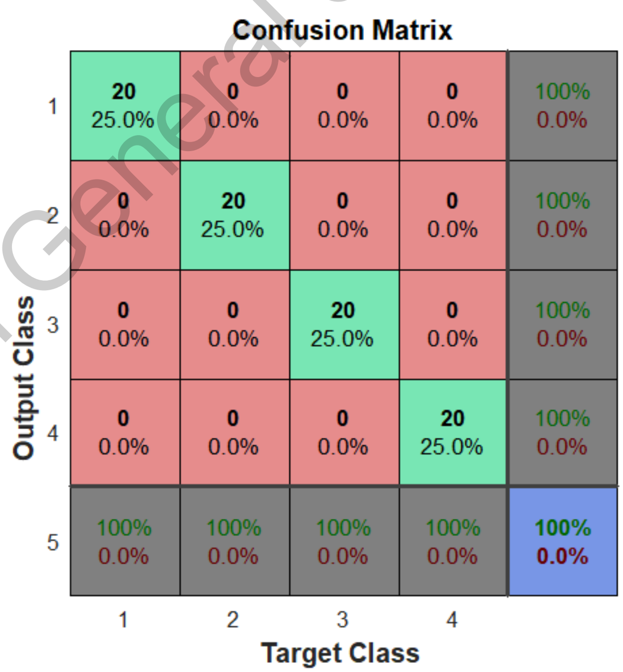
4.6.3 Scenario: linear and non-linear load

The diagnosis obtained by the SVM for the scenario with linear and non-linear load is quite clear as can be seen in the graph of Figure 4.18 a), where the SVM achieves an optimal separation margin between the healthy, incipient damage, light damage, and moderate damage of the transformer. Derived from the previous observation, it can be inferred that the percentage of classification effectiveness for this scenario is 100%. This statement can be verified by consulting Figure 4.18 b).

The final results of compression, decompression, and classification corresponding to the scenario with linear and non-linear load from the monitoring interface are shown in Figure 4.19.



a)



b)

Figure 4.18 Stage diagnosis with linear and non-linear load a) Contour plot of SVM classification, b) Confusion matrix.

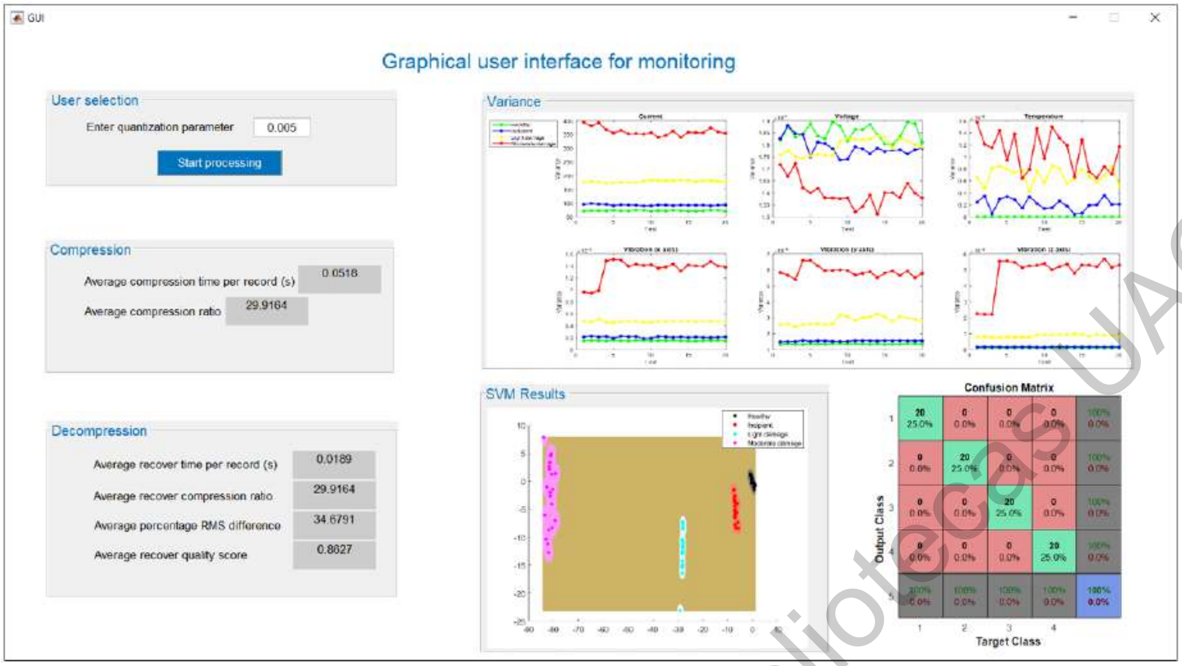


Figure 4.19 Graphical interface with the final results of the scenario with linear and non-linear load.

Figure 4.20 shows the graphical user interface when only one test is analyzed. The signals for an unknown condition of the transformer with linear load are acquired, the variance is calculated, the results are standardized, they enter to the designed SVM and, finally, a diagnostic label is obtained.

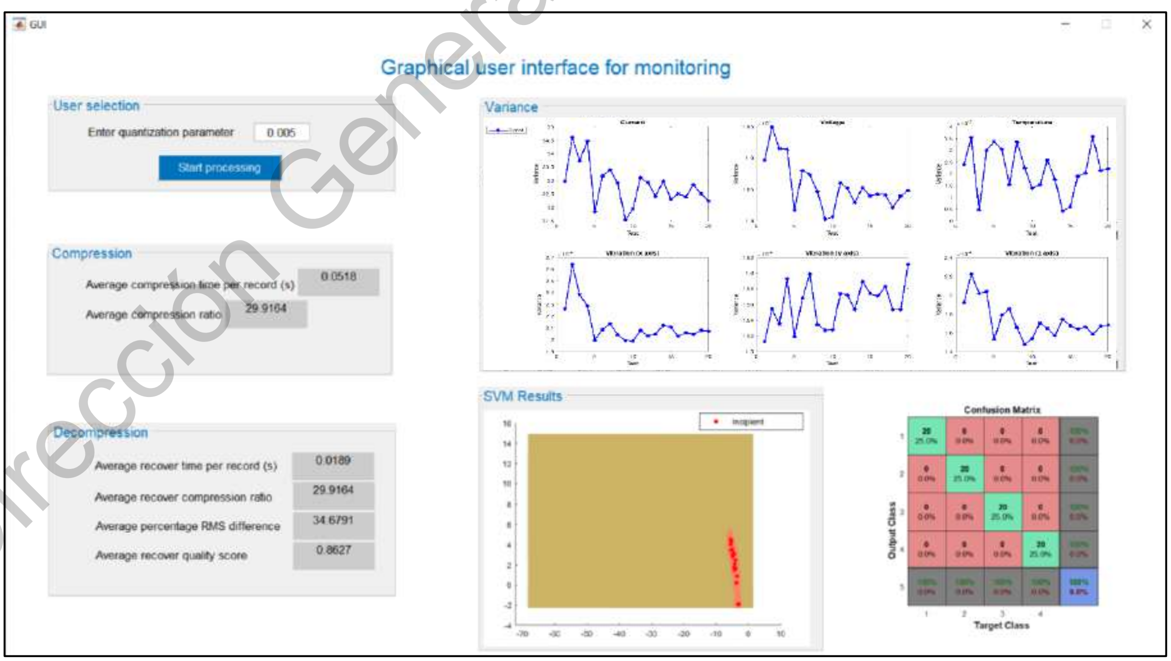


Figure 4.20 Graphical interface with the final results of an unknown condition.

Chapter 5

5. Conclusions

5.1 Conclusions

From this research work, the integration of a data compression algorithm and a data mining classification method was obtained to form a methodology for the monitoring and diagnosis of the operation status of a single-phase transformer. The methodology is based on the acquisition of current, voltage, temperature, and vibration signals that are generated in a transformer under conditions of short-circuit failure (incipient, light damage, and moderate damage) in addition to the healthy state which serves as a reference for detection of the fault. From the graphical interface developed in the LabVIEW software, the data acquisition process of the studied signals was achieved for a future analysis in the Matlab software.

Several tests were performed using three different types of loads. i.e., a stage without load, a scenario with simulated linear load from a resistor, and a scenario with linear load plus non-linear load using a rectifier bridge.

In this study, the implementation of a data compression technique with loss based on a DWT allowed obtaining excellent classification results, since this step worked somehow as a filter to eliminate the noise of the signals, cancelling irrelevant characteristics of them. To achieve a high similarity of the reconstructed signal against the original signal, a decomposition level 4 was considered for the DWT. Additionally, it is recommended to use a quantifying parameter less than 1, otherwise the technique becomes less effective. The ideal case is to find a suitable parameter for each type of signal since their characteristics differ from each other. The variation of this parameter will affect the resulting values of the reconstructed signal and, consequently, its representation in the dimensional space. It is concluded that the quantification parameters that indicate the best compression and decompression results are 0.5 for current and voltage, 0.00001 for temperature, and 0.005 for vibrations.

The extraction of characteristics by calculating the variance for each signal and its subsequent standardization was carried out in order to perform a preprocessing and reduce the variability of the data before being analyzed by the SVM, which resulted in an increase in the performance of the technique. From the use of this algorithm, the detection of the induced faults to the transformer was achieved, achieving an effectiveness percentage in the classification of 100% for the three scenarios.

The classification approach obtained excellent results for successfully detecting the incipient level of failure in addition to the other transformer conditions thanks to the adequate pre-treatment of the signals, since it facilitated the SVM-based classification task by maximizing the margin of separation between the four classes studied. Likewise, the model implied the integration of multiple physical quantities to improve the effectiveness in the diagnosis of failures, which represents a suitable approach to demonstrate the advantages and strengths of the proposed methodology.

5.2 Future work

The methodology can be expanded and improved for further development. This improvement could include a diagnostic method not only to detect short-circuit faults in the transformer, but also the identification of other types of failures that cause the abnormal behavior of the transformer. In this regard, future work will include a much more robust proposal in terms of the ability to detect and classify different types of failures.

In order to make the methodology more accessible and affordable for the users, one aspect to consider is to use another acquisition and processing cards to make the implementation more economical and generic.

On the other hand, it is important to consider the possibility of migrating the methodology code to a programming language that has an open source license, so that the part of the acquisition corresponding to LabVIEW and the part of the data analysis corresponding to Matlab can be integrated.

Bibliography

AJ, C., Salam, M. A., Rahman, Q. M., Wen, F., Ang, S. P., Voon, W. (2018). Causes of transformer failures and diagnostic methods – A review. Renewable and Sustainable Energy Reviews, 82, 1442–1456.

Aljohani, O., Abu-Siada, A. (2015). Minimum detection of power transformer short circuit fault using frequency response analysis. 2015 Australasian Universities Power Engineering Conference (AUPEC).

Álvarez-Monroy, G., Mejía-Barrón, A., Valtierra-Rodríguez, M., Amézquita-Sánchez, J. P., Granados-Lieberman, D., Olivares-Galván, J. C., Escarela-Pérez, R. (2018). Analysis of current signals for diagnostics of short-circuited turns in a transformer based on Wavelet transform. Perspectivas de la Ciencia y la Tecnología, Año 1, Num 001, 88-98.

Bagheri, M., Phung, B. T. (2016). Frequency response and vibration analysis in transformer winding turn-to-turn fault recognition. 2016 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS).

Bagheri, S., Moravej, Z., Gharehpetian, G. B. (2018). Classification and Discrimination Among Winding Mechanical Defects, Internal and External Electrical Faults, and Inrush Current of Transformer. IEEE Transactions on Industrial Informatics, 14(2), 484–493.

Ballal, M. S., Suryawanshi, H. M., Mishra, M. K., Chaudhari, B. N. (2016). Interturn Faults Detection of Transformers by Diagnosis of Neutral Current. IEEE Transactions on Power Delivery, 31(3), 1096–1105.

Barzilay, O., Brailovsky, V. L. (1999). On domain knowledge and feature selection using a support vector machine. Pattern Recognition Letters, 20(5), 475–484.

Beer, F. P., Johnnston, E. R., Cornwell P. J. (2010). Mecánica vectorial para ingenieros: Dinámica, 9 ed. Mc Graw Hill Educación.

Bhuiyan, B. A., Absar, M. W., Roy, A. (2017). Performance comparison of various wavelets in compression of PMU generated data in smart grid. 2017 3rd International Conference on Electrical Information and Communication Technology (EICT).

Centro Español de Metrología. (2006). El Sistema Internacional de Unidades SI. Oficina Internacional de Pesas y Medidas BIPM.

Contreras-Valdes, A., Amezquita-Sanchez, J. P., Granados-Lieberman, D., Valtierra-Rodriguez, M. (2019). Data compression based on discrete Wavelet transform and fault detection of short-circuit faults in transformers. 2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC).

Contreras-Valdes, A., Huerta-Rosales, J. R., Amezquita-Sanchez, J. P., Morales-Velazquez, L., Valtierra-Rodriguez, M. (2019). Statistical Features and Data Mining Techniques for Detection of Short-Circuit Faults in Transformers. 2019 XV International Engineering Congress (CONIIN).

Gómez Sarduy, J. R., Viego Felipe, P. R. (2014). La Universidad en la transformación hacia las Redes Eléctricas Inteligentes en América Latina. Universidad y Sociedad [seriada en línea], 6 (2). pp. 59-65.

Gorgas, G. G., Cardiel, L. N., Zamorano, C. J. (2011). Estadística básica para estudiantes de ciencias. Departamento de Astrofísica y Ciencias de la Atmósfera Facultad de Ciencias Físicas Universidad Complutense de Madrid.

Gorunescu, F. (2011). Data Mining: Concepts, Models and Techniques, Vol 12. Intelligent Systems Reference Library. Springer Science & Business Media.

Grossman, R. L., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R. (2013). Data Mining for Scientific and Engineering Applications. Springer US.

Gutten, M., Janura, R., Šebök, M., Korenčiak, D., Kučera, M. (2016). Measurement of Short-Circuit Effects on Transformer Winding with SFRA Method and Impact Test, Metrology and Measurement Systems, 23(4), 521-529.

Han, J., Pei, J., Kamber, M. (2011). Data Mining: Concepts and Techniques, 3 th ed. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.

Harper, G. E. (2005). El libro práctico de los generadores, transformadores y motores eléctricos. Limusa.

Hewitt, P. G. (2006). Conceptos de física. Limusa.

Iorgulescu, M., Beloiu, R., Popescu, M. O. (2010). Vibration monitoring for diagnosis of electrical equipment's faults. 2010 12th International Conference on Optimization of Electrical and Electronic Equipment.

Jing, T., Long, F., Shi, X. (2016). Research of Aircraft Integrated Drive Generator fault diagnostic decision based on attribute reduction in rough sets. 2016 IEEE International Conference on Aircraft Utility Systems (AUS).

Karimifard, P., Gharehpetian, G. B., Javid Ghanizadeh, A., Tenbohlen, S. (2012). Estimation of simulated transfer function to discriminate axial displacement and radial deformation of transformer winding. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 31(4), 1277–1292.

Li, X., Zhong, H., Liu, H., Sun, Y., Chen, A., Qiao, D. (2019). Transformer Fault Simulation and Analysis Based on Fractional Calculus. 2019 22nd International Conference on Electrical Machines and Systems (ICEMS).

Mago, M., Vallés, L., & Olaya, J., Zequera, M. (2011). Análisis de fallas en transformadores de distribución utilizando ensayos no destructivos y pruebas de tensión mecánicas. Revista INGENIERÍA UC, 18 (2), 15-26.

Malik, H., Jarial, R. K., Azeem, A., Yadav, A. K. (2011). Application of Modern Technology for Fault Diagnosis in Power Transformers Energy Management. 2011 International Conference on Communication Systems and Network Technologies.

Marshall, J. A. (1995). Neural networks for pattern recognition. Neural Networks, 8(3), 493–494.

Masoum, A. S., Hashemnia, N., Abu-Siada, A., Masoum, M. A. S., Islam, S. M. (2017). Online Transformer Internal Fault Detection Based on Instantaneous Voltage and Current Measurements Considering Impact of Harmonics. IEEE Transactions on Power Delivery, 32(2), 587–598.

Mejia-Barron, A., Valtierra-Rodriguez, M., Granados-Lieberman, D., Olivares-Galvan, J. C., 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.

Menezes, A. G. C., Almeida, O. M., Barbosa, F. R. (2018). Use of decision tree algorithms to diagnose incipient faults in power transformers. 2018 Simposio Brasileiro de Sistemas Eletricos (SBSE).

Mohammadpour, H., Dashti, R. (2011). Localization of short circuit faults in transformers using harmonic analysis and pattern recognition. 2011 International Conference on Advanced Power System Automation and Protection.

Mohseni, B., Hashemnia, N., Islam, S., Zhao, Z. (2016). Application of online impulse technique to diagnose inter-turn short circuit in transformer windings. 2016 Australasian Universities Power Engineering Conference (AUPEC).

Pérez, M. M. (2014). Minería de datos a través de ejemplos. RC Libros.

Rahman, M. F., Nirgude, P. (2019). A New Approach to Study Partial Discharge Inception due to Particles in Transformer Oil Using Electric Field Analysis. 2019 International Conference on High Voltage Engineering and Technology (ICHVET).

Rahimpour, E., Jabbari, M., Tenbohlen, S. (2010). Mathematical Comparison Methods to Assess Transfer Functions of Transformers to Detect Different Types of Mechanical Faults. IEEE Transactions on Power Delivery, 25(4), 2544–2555.

Rajamani, P., Chakravorti, S. (2012). Identification of simultaneously occurring dynamic disc-to-disc insulation failures in transformer winding under impulse excitation. IEEE Transactions on Dielectrics and Electrical Insulation, 19(2), 443–453.

Razi-Kazemi, A. A., Vakilian, M., Niayesh, K., Lehtonen, M. (2015). Data Mining of Online Diagnosed Waveforms for Probabilistic Condition Assessment of SF6 Circuit Breakers. IEEE Transactions on Power Delivery, 30(3), 1354–1362.

Rigatos, G., Siano, P., Piccolo, A. (2012). Incipient fault detection for electric power transformers using neural modeling and the local statistical approach to fault diagnosis. 2012 IEEE Sensors Applications Symposium Proceedings.

Rustom, J. A., Fernández, L. C., Mansilla M. A. (2012). Estadística descriptiva, probabilidad e inferencia: una visión conceptual y aplicada. Departamento de Economía Agraria Facultad de Ciencias Agronómicas Universidad de Chile.

Sarkar, S. J., Kundu, P. K., Sarkar, G. (2017). DBEA: A novel approach of repetitive data array compression for power system application. 2017 2nd International Conference for Convergence in Technology (I2CT).

Sayood, K. (2017). Introduction to Data Compression, 5 th ed. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufman.

Secretaría de Energía (SENER). (2016). Programa de Redes Eléctricas Inteligentes. Recuperado de: https://www.gob.mx/sener/acciones-y-programas

Shuin, V., Lebedev, V., Filatova, G. (2016). Study of possibility of digital voltage transformers usage in fault location determination. 2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM).

Tang, K. Z., Tang, S., Kusumadi, N. P., Chuan, S. H. (2013). Development of a remote telemetry and diagnostic system for electric vehicles and electric vehicle supply equipment. 2013 10th IEEE International Conference on Control and Automation (ICCA).

Tcheou, M. P., Lovisolo, L., Ribeiro, M. V., da Silva, E. A. B., Rodrigues, M. A. M., Romano, J. M. T., Diniz, P. S. R. (2014). The Compression of Electric Signal Waveforms for Smart Grids: State of the Art and Future Trends. IEEE Transactions on Smart Grid, 5(1), 291–302.

Tipler, P. A., Mosca, G. (2005). Física para la ciencia y la tecnología: Electricidad y magnetismo. Vol. 2A. Reverté.

Venikar, P. A., Ballal, M. S., Umre, B. S., Suryawanshi, H. M. (2015). Transformer incipient inter-turn fault detection based on no-load current harmonic analysis. 2015 IEEE Eindhoven PowerTech.

Widodo, A., Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21(6), 2560–2574.

Wiszniewski, A., Solak, K., Rebizant, W., Schiel, L. (2018). Calculation of the lowest currents caused by turn-to-turn short-circuits in power transformers. International Journal of Electrical Power & Energy Systems, 95, 301–306.

Xuewei, Z., Hanshan, L. (2018). Research on Transformer Fault Diagnosis Method and Calculation Model by using Fuzzy Data Fusion in Multi sensor Detection System. Optik.

Yang, R., Zhang, D., Li, Z., Yang, K., Mo, S., Li, L. (2019). Mechanical Fault Diagnostics of Power Transformer On-Load Tap Changers Using Dynamic Time Warping. IEEE Transactions on Instrumentation and Measurement, 1–9.

Zhan, C., Shengchang, J., Young, L., Lingyu, Z., Yuhang, S., Fuquiang, R. (2018). Winding Mechanical Fault Diagnosis Technique of Power Transformer Based on Time-Frequency Vibration Analysis. 2018 International Conference on Condition Monitoring and Diagnosis (CMD).

Zheng, T., Huang, T., Ma, Y., Zhang, Z., Liu, L. (2018). Histogram-Based Method to Avoid Maloperation of Transformer Differential Protection Due to Current-Transformer Saturation Under External Faults. IEEE Transactions on Power Delivery, 33(2), 610–619.

Appendix

applied sciences

MDPI

Review

Predictive Data Mining Techniques for Fault Diagnosis of Electric Equipment: A Review

Arantxa Contreras-Valdes [1], Juan P. Amezquita-Sanchez [1], David Granados-Lieberman [2] and Martin Valtierra-Rodriguez [1,*]

1 ENAP-Research Group, CA-Sistemas Dinámicos, Facultad de Ingeniería, Universidad Autónoma de Querétaro (UAQ), Campus San Juan del Río, Río Moctezuma 249, Col. San Cayetano, San Juan del Río, Qro. C. P. 76807, Mexico; acontreras01@alumnos.uaq.mx (A.C.-V.); juan.amezquita@enap-rg.org (J.P.A.-S.)

2 ENAP-Research Group, CA-Fuentes Alternas y Calidad de la Energía Eléctrica, Departamento de Ingeniería Electromecánica, Tecnológico Nacional de México, Instituto Tecnológico Superior de Irapuato (ITESI), Carr. Irapuato-Silao km 12.5, Colonia El Copal, Irapuato, Guanajuato C. P. 36821, Mexico; david.granados@enap-rg.org

* Correspondence: martin.valtierra@enap-rg.org

Received: 29 December 2019; Accepted: 27 January 2020; Published: 1 February 2020

check for updates

Abstract: Data mining is a technological and scientific field that, over the years, has been gaining more importance in many areas, attracting scientists, developers, and researchers around the world. The reason for this enthusiasm derives from the remarkable benefits of its usefulness, such as the exploitation of large databases and the use of the information extracted from them in an intelligent way through the analysis and discovery of knowledge. This document provides a review of the predictive data mining techniques used for the diagnosis and detection of faults in electric equipment, which constitutes the pillar of any industrialized country. Starting from the year 2000 to the present, a revision of the methods used in the tasks of classification and regression for the diagnosis of electric equipment is carried out. Current research on data mining techniques is also listed and discussed according to the results obtained by different authors.

Keywords: data classification; data mining; data regression; electric equipment; fault diagnosis

1. Introduction

Over the past few years, the number and diversity of electrical equipment, such as motors, transformers, generators, electric vehicles, and energy transmission and distribution systems, among many others, are getting bigger [1–6]. Their exponential growth is due to the need of people to perform a number of different activities, ranging from industrial processes to everyday activities such as charging the cell phone battery or starting the car to go to work. Due to their paramount importance in any facet of society, their safety and correct operation is vital, even more so when considering that a failure in one of its components can produce (1) high economical losses derived from its partial or total repair, (2) degradation and poor quality on its performance, (3) outages in the production process, (4) damages to other equipment, and (5) conditions that put in risk the physical integrity of people, among others.

In this regard, the application and development of new techniques and methods to monitor the condition of electric machines and systems are important topics of research. In general, a condition monitoring strategy consists of the following steps (see Figure 1): Data collection through different types of sensors, data processing and feature extraction, and data analysis for condition assessment. The latter can be seen as the process of exploring, finding, selecting, and using specific data to solve the given problem, e.g., a diagnosis problem; however, it is not an easy and straightforward process since the data analyst has to deal with different volumes and varieties of data, as well as redundant and unneeded data, which can compromise and difficult the solution of the assigned task; in fact, the reality is that, in many cases, only a small part of the dataset is used because its volume is simply too large to be used and processed effectively. One solution to this problem has been the use of data mining (DM) techniques. DM is one of the fastest growing fields at both the computational and industrial levels. Its main characteristic involves the search of patterns through the handling of different sets of data to discover the available knowledge. Kantardzic [7] calls DM to the process of applying a computer-based methodology for discovering knowledge from data. Although DM is based on computational algorithms, best results can be obtained by balancing the knowledge of human experts about the problem under study with the advantages and operating modes of different algorithms [4].

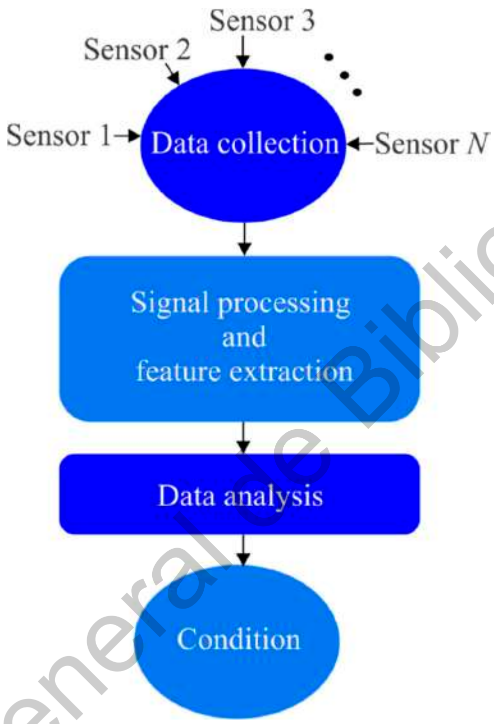


Figure 1. General elements of a condition monitoring strategy.

In general, DM functionalities can be divided into two categories: Predictive and descriptive. The former is used to construct models that allow the prediction of unknown or future values, whereas the latter is in charge of finding new information that allows the description of the dataset. In this regard, the prediction functionalities become the most suitable option to perform the condition monitoring since a new and unknown equipment condition can be determined or predicted from a specific input information. Therefore, this manuscript is aimed at reviewing the classification and regression tasks that fall within the predictive category of DM, as well as hybrid techniques that combine more than one prediction method. Specifically, classification techniques attempt to find a function or model that distinguishes or predicts the class of unknown data by analyzing a data training set [8]. The regression analysis is used for numerical prediction, i.e., to predict missing or new numerical data values [8].

In the literature, two main groups of research works related to DM and electric equipment are found. On the one hand, there are different reviews about DM applications, e.g., diagnosis in health [9–11], marketing [12], industrial [13,14], climatological [15], and financial [16] issues, among others. On the other hand, there are also reviews related to diagnosis methods for specific machines such as transformers [17–19], estimation strategies in electric vehicles [20], mathematical models used to study induction motors in defective conditions [21], or, in a more general sense, methodologies of fault classification in transmission systems [22,23] and distribution of energy [24]. There is also the work of Hare et al. [25], where they present a study of modern diagnosis methods in smart micro grids. Although there are specialized reviews on topics of either DM or electric equipment and systems, none of these works have been specifically focused on reviewing the research that has been carried out about the applications of DM techniques for condition monitoring of electric equipment and systems, which is very important in order to highlight the algorithms that have been used in specific equipment but can be applied to other machines since the application core is similar. In this regard, this manuscript provides a review of DM techniques focused particularly on the tasks of classification and regression within the category of predictive analysis applied to various electric machines and systems such as transformers, electric vehicles, heating, ventilation, and air conditioning (HVAC) systems, airplane, automotive, three-phase and multi-phase induction motors, centrifugal pumps, generators, distribution systems, and transmission lines, among others.

The rest of this manuscript is prepared as follows. Section 2 deals with the classification, regression, and hybrid techniques used for the detection of faults and the diagnosis of electrical systems. In Section 3, recent research works on these topics and the latest contributions on DM techniques that can be explored in fault diagnosis methodologies are presented. Finally, Section 4 shows the conclusions of this work.

2. Predictive Model

DM tasks can be conducted through prediction and description models [7,8] (see Figure 2a). In general, the prediction models are constructed through the learning of known data classes, whereas the description models arise from the findings obtained in a dataset [8]. In this regard, predictive DM techniques are the straightforward option to perform the diagnosis of equipment and systems since their different operating conditions can be learned and determined by a prediction model. In this model (see Figure 2b), its learning is carried out by means of the analysis of a data training set (input data with known outputs) and then used to predict the unknown output (class or value) of new input data. According to the nature of data (discrete or continuous), the prediction model can be used to perform either classification or regression functionalities [7] (see Figure 2a). A classification procedure consists of the assignment of an object (unknown class) into one of several predefined classes (predicted class) according to its properties or features. In a different way, a regression procedure involves the modeling of continuous functions to determine new numerical values (predicted values) according to specific inputs. Classification and regression techniques used for fault detection in electric equipment and systems are presented in the following sections.

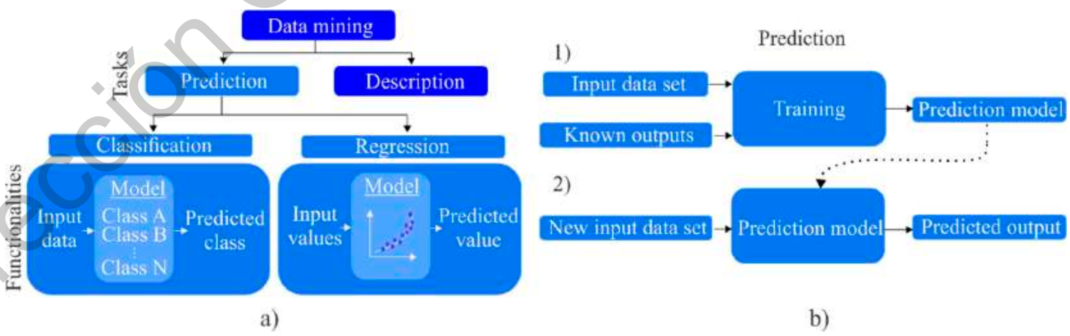


Figure 2. (**a**) Schematic diagram for tasks and functionalities of data mining (DM) and (**b**) prediction model.

2.1. Classification-Based Methods

In a condition monitoring context, a classification model can be constructed from a given system and used to provide warnings and predict certain failures in early stages. In this regard, researchers around the world have proposed and used different classification methods in machine learning, pattern recognition, and statistics to perform faults diagnosis.

Recent research on DM has been focused on developing classification techniques capable of handling datasets with different features, e.g., imbalance of proportionally, and large amounts of data. In the latter, this capacity is strongly required because, on the one hand, the availability of data is growing and, on the other hand, their performance can be compromised if limited datasets are analyzed. In fact, the amount of available data during the training of a neural network (NN) plays an important issue in its performance. For instance, Taylor et al. [26] contrasted three different techniques: Neural networks trained by using a hybrid of evolutionary search and backpropagation, neural networks trained by straightforward backpropagation, and simple predictive rulesets trained by evolutionary algorithms. Results indicate that evolved NNs outperform backpropagation trained NNs. However, the results are slightly unsatisfactory from a business viewpoint, obtaining a maximum accuracy of 77.9%, which can be somehow expected due to the small amount of training data, highlighting the need of additional data to establish a better reference during the pattern recognition task. Fortunately, there are many works in which the authors also use NNs as the basis of their investigations and promising results are obtained. In an energy consumption context, Magoulès et al. [27] diagnose different electrical equipment of an office building, including fans, pumps, cooling equipment, and chillers. They use a recursive deterministic perceptron NN to distinguish between normal and defective datasets, where an effectiveness percentage higher than 97% is obtained. Similarly, the use of NNs for fault detection on induction motors are presented [28–35]. Tallam et al. [28] presented an on-line diagnostic scheme to alert the engine protection system of an incipient failure. This scheme consists of a feed-forward NN with a self-organized feature map to display the operating conditions of the in-test machine. An interesting feature offered by the results is that the method is not sensitive to unbalanced supply voltages or asymmetries in the machine. Martins et al. [29] use the alpha-beta stator currents of a three-phase induction motor as input variables to diagnose stator faults. In their proposal, an unsupervised Hebbian-based NN is used to extract the main components of the stator current data. Other proposals combine NNs with fuzzy logic systems (FLSs) to detect inter-turn faults [30,31]. In particular, Ballal et al. [31] developed an ANFIS (Adaptive Neural Fuzzy Inference System) for the detection of stator inter-turn insulation and bearing wear faults, where five input parameters, i.e., current, bearing temperature, winding temperature, speed, and the noise of the machine are used to construct the model. For the inter-turn insulation fault, they obtain an effectiveness of 94.03% using two inputs and 96.67% using five inputs. For the bearing wear fault, the accuracy rate with two inputs is 90.5%, and 98.7% with five inputs. These results demonstrate the importance of an information-rich dataset. In [32–34], several NNs are implemented in field programmable gate arrays (FPGAs) to diagnose different faults in induction motors. The diagnosis of broken rotor bars is presented by Zolfaghari et al. [35], where the multi-layer perceptron NN used is able to detect the faults in the rotor with a classification effectiveness of 98.80%. Furthermore, modular NNs are used to diagnose transmission lines from the voltage and current signals of their elements (busses, transmission lines, and transformers). Given its modular nature, the diagnosis can be carried out by element, by area, or for the entire context of the electrical system [36]. In [37], adaptive linear neural networks and feed forward neural networks are combined to classify electrical disturbances that affect the electric equipment. The best classification results are obtained when only a single disturbance appears; when more disturbances are combined, the effectiveness is reduced, but it is worth noticing that the effectiveness percentage obtained exceeds 90% for a noiseless condition and exceeds 77% for a noisy condition in the presence of six combined disturbances. In addition, the overall methodology takes 46.5 milliseconds per half cycle analyzed. Hare et al. [25] present a survey for fault diagnostics in smart micro grids, in which they discuss the faults within various components of a micro network, e.g., photovoltaic panels, wind turbines, conventional generation systems, as well as cables and transmission lines, etc., where several classification algorithms such as NNs, decision trees, and FLSs, among others, are presented.

Regarding the transformers, Rigatos and Siano [38] propose the neural-fuzzy network modeling and the local statistical approach for the detection of incipient faults in power transformers. Another technique commonly used for diagnosis of transformers is the decision tree [39–41]. Menezes et al. [39] and Han et al. [40] used experimental data from a dissolved gas analysis (DGA) to illustrate the performance of their decision tree-based models. In [39], they present a comparison between the method based on the algorithm C4.5 and other methods used in DGA. They use only 162 samples for the analysis, obtaining the following accuracy: 99.38% for the proposed method, 98.15% for the rules extracted, 88.03% for Duval Triangle, 63.25% for Dornenburg IEC C57.104, and 56.41% for Rogers IEC C57.104. In [40], a decision-tree C4.5 algorithm obtained an effectiveness of 86% for a thermoelectric fault in oil-immersed transformer. Samantaray and Dash [41] analyze the current of a power transformer to discriminate between the current signals generated by the inrush effect and the ones generated by its internal faults. The processing time of the proposed approach is 0.12s and provides an accuracy greater than 96%, exceeding the accuracy of the support vector machine (93.33%). As can be noted, the type of variable to be analyzed by the decision tree-based methods is not restricted; in fact, the use of vibration signals for the diagnosis of faults in monoblock centrifugal pumps [42] and motors of internal combustion [43] are also presented. The latter also compares the classification accuracy obtained by the J48 algorithm, random forest tree algorithm, linear model tree algorithm, best first tree algorithm, and functional tree algorithm, where the linear model tree algorithm provides the best results, offering classification accuracy of 100% using statistical features. In general, it can note that the decision tree algorithms are a practical, economical, and very effective approach. In addition to these different types of decision trees, the fault tree is another alternative used for the diagnosis of systems. For example, Volkanovski et al. [44] evaluate the reliability of a power system for energy delivery by constructing a fault tree structure, which represents the system configuration and includes all the possible flow routes of interruption of the power supply from the generators to the loads, including energy transfer limitations, common cause failure of power lines, energy flows and the capacity of generators, and loads in the power system. Duan and Zhou [45] also use the fault tree analysis and Bayesian networks for fault detection of a system for oil pressure warning instructions in an aircraft engine, where a diagnostic decision tree to guide maintenance personnel to make more efficient decisions when attempting to repair the system is obtained. An advanced Bayesian non-linear state estimation technique called Unscented Kalman Filtering to detect faults in HVAC (heating, ventilation, and air conditioning) components is presented by Bonvini et al. [46]. This algorithm can detect common faults in a chiller plant and functional failures caused by problems in the compressor and occlusions in the valves with a computational performance of 0.25s using Intel Xeon (R) 2.67 GHz–19 Cores and 0.52s using Intel Core i7 2.8 GHz–1 Core. Another tree-based method is the tree-structured fault dependence kernel developed by Li et al. [47]. It implements a structured labeling to include dependency information and describe severity levels in a high-margin learning framework for fault detection of building cooling systems. It is important to highlight that the testing accuracy increases or decreases accordingly with the change of training samples. For instance, in [47], the testing accuracy of the proposed strategy boosted from 69.64% (six training samples) to 99.12% (180 training samples). That is, accumulating more training data is beneficial for the fault detection and diagnosis.

Other classification method that has been widely used is the FLS. In general, it uses knowledge-based reasoning to construct logical rules and, thus, diagnose faults. In this type of algorithms, the designer knowledge about the in-test equipment, e.g., operating conditions, nominal parameters, overall performance, etc., plays a fundamental role. In [48], an FLS is designed to diagnose stator winding faults in induction motors. Similar results are obtained under noisy and noiseless conditions. Therefore, FLS is a good option because there is no general and accurate analytical model that describes completely the induction motor under fault conditions, leaving the open doors to uncertainties or noisy conditions. Amezquita-Sanchez et al. [49] present two FLSs to detect broken rotor bars (BRB) in both regimes of operating conditions, i.e., transient and stationary. The combination of fractal dimension analysis and FL system demonstrated to be highly effective on identifying half-BRB, one BRB, and two BRB, as well as healthy condition, since an effectiveness of 95% and 100% for start-up transient and steady state is obtained. For transformers, Islam et al. [50] present the diagnosis of several transformer faults using dissolved gas in oil analysis (DGA) and an FLS for its interpretation. An overview of different FLSs for DGA is presented in [51], where it is indicated that there is not a single technique that can enable the detection of the full range of faults, therefore the combination of different methods has to be explored as a promising solution. Although promising results have been obtained using FLSs, a relatively high superiority of an adaptive neuro fuzzy inference system for DGA is presented in [52], obtaining an accuracy of 98% for all the 100 fault cases under study, while FL obtained 95%. Regarding other electric systems an equipment, the fault diagnosis of the power system using fuzzy logic is presented in [53]. An online monitoring system of voltage variations in electric systems is presented in [54], where an FLS is used to diagnose and classify instantaneously, i.e., sample to sample, the severity of the electric variation. Their proposal is a suitable tool for analyzing stored data; furthermore, it provides phase information unlike the conventional root mean square technique; moreover, it gives results sample to sample, which is better for nonstationary signals. Lauro et al. [55] diagnose a fan coil electric and Zio et al. [56] classify the faults of a steam generator of a pressurized water reactor. In the latter, a fuzzy clustering-based classification model is transformed into a fuzzy logic inference model, allowing its direct interpretation and inspection; also, improvements in the obtaining of the model are presented to allow the treatment of more complicated scenarios.

Table 1 shows a summary for the above reviewed works, where the used techniques and conventional applications, along with the physical variables that have been analyzed by them, are presented. As can be observed, NNs, decision trees, and FLSs are the most commonly used methods for fault detection. Although NNs can be more suitable for fault detection from a generalization viewpoint, decision trees have been preferred in many cases because of the clarity in their interpretation (human friendly) and their low computation burden, which are desirable features in online condition monitoring systems. Also, if the amount of data is limited, a simple decision tree can be used; yet, other aspects of such small dataset have to be taken into account, for instance: redundancy of data, data imbalance, information contained, data type (continuous or discrete), range, time dependency, etc. Regarding the physical variable measured from the in-test equipment, the current signals show to be a powerful and representative source of information for fault detection; although promising results are obtained, the combination of multiple physical variables, e.g., current and vibrations signals, should be explored in order to improve the reliability of new classification schemes and expand the number of fault conditions that can be determined by a single classification algorithm, exploiting the information that each signal can provide, e.g., current signals can provide information to diagnose electrical faults and vibration signals can provide information to diagnose mechanical faults.

Table 1. Classification methods and their applications.

Classification Methods	Equipment Under test	Physical Variable Used as Information Source
■ NNs and rule sets [26] ■ Recursive deterministic perceptron NN [27] ■ Feedforward NN [28] ■ Hebbian NN [29] ■ B-spline membership fuzzy NN [30] ■ ANFIS [31,52] ■ Feed forward NN [32–34] ■ Multi-layer NN [35] ■ Modular NNs [36] ■ Adaptive linear NN and feed forward NN [37] ■ Neural-fuzzy network and statistical analysis [38]	■ Refrigeration systems ■ Electric equipment (fan, coil, pump, and chiller) ■ Stator winding of an induction motor ■ Bearings in induction motors ■ Broken rotor bars ■ Electric transmission lines ■ Transformer	■ Temperature ■ Energy consumption ■ Voltage ■ Data from DGA ■ Current ■ Vibrations ■ Speed

Table 1. *Cont.*

Classification Methods	Equipment Under test	Physical Variable Used as Information Source
■ Decision tree: C4.5 algorithm [39,40,42] ■ Decision tree: CART algorithm [41] ■ Decision trees: J48 algorithm, best first algorithm, random forest algorithm, functional algorithm, and linear model algorithm (a comparison) [43] ■ Fault tree analysis [44] ■ Fault tree analysis and Bayesian networks [45]	■ Transformer ■ Monoblock centrifugal pump ■ Internal combustion engine ■ Power system ■ System of oil pressure warning instructions in aircraft engines	■ Data from DGA ■ Current ■ Vibration ■ Power flows
■ Bayesian non-linear state estimation technique (detection based on a threshold value) [46] ■ Tree-structured fault dependence kernel [47]	■ Chiller plant ■ Building cooling system	■ Pressure ■ Temperature
■ Fuzzy logic system [48–51,53,54] ■ Fuzzy sets and fuzzy logic [55] ■ Fuzzy decision tree [56]	■ Induction motor ■ Transformer ■ Power systems ■ Fan coil electric consumption ■ Steam generator of a pressurized water reactor	■ Current ■ Voltage ■ Electric power ■ Temperature ■ Pressure ■ Flow

2.2. *Regression-Based Methods*

In general, the use of regression techniques consists of numerical prediction, i.e., a methodology to generate a methatical function or model to predict missing or new numerical data values; but it also covers the identification of distribution trends based on the available data. For the latter, the support vector machine (SVM) has been widely used since a regression function is found from the training dataset.

Among the available research works, SVMs have been presented in the literature as one of the most promising methods to diagnose faults in power transformers [57]. Lv et al. [58] and Bacha et al. [59] implement SVM-based strategies to establish the classification of faults in power transformers by using the gases available from the DGA. Both works present an interesting performance comparison among different methods. In [58], five artificial intelligence methods are presented. It is found that the SVM is the most effective and fastest method, obtaining an accuracy of 100% and a training time less than 1s, NN (92.76% accuracy and 81s training time), expert system (89.34% accuracy and training time no mentioned), FL (92.32% accuracy and 82s training time), and combined NN and expert system (93.54% accuracy and 44s training time). In [59], the classification accuracy of FL (86.7%), multi-layer perceptron (80%), radial basis function (86.7%), and SVM (90%) is presented. In a similar venue, SVMs are also explored for the detection and localization of faults in transmission lines, where Johnson and Yadav [60] and Parikh et al. [61] conclude that the SVMs are a highly accurate method for these tasks. Zhang et al. [62] present a SVM-based methodology for data-based line trip fault prediction in power systems, where long-term memory networks are used to capture time series characteristics from multiple sources in large systems. The accuracy of the line trip fault prediction can reach about 97%. SVMs have been also employed in the diagnosis of induction motors, e.g., Gangsar and Tiwari [63] carry out a comparative investigation to predict mechanical and electrical faults in induction motors from the analysis of vibration and current signals and the use of multiclass SVM methods. Zhang et al. [64] propose a method based on the robust local linear embedding algorithm and an SVM for the diagnosis the gear fault from an experimental setup composed by a motor, a torque transducer/encoder, and a dynamometer. The diagnosis of fault severity in the stator winding of induction motors using SVM in regression mode is presented by Das et al. [65]. In their research, they analyze the current signals for different levels of short circuit fault, different unbalance conditions in the voltage supply, and different load levels. In the methodology, they use recursive feature elimination to select the optimum number of features and an SVM as a load-immune classifier, demonstrating the high capabilities of SVMs. Among other electric machines where the SVM has been applied, heating, ventilation, and air conditioning (HVAC) systems [66,67] and the steam generator and pressure boundary of the Chinese CNP300 PWR (Qinshan I NPP) reactor coolant system [68] are included. For the latter, a specialized SVM module monitors the subunits of the reactor coolant system and is capable of making fault diagnosis at the component level. Finally, Lai et al. [69] investigate partial discharge activities for online monitoring of power equipment. They use back-propagation NN, self-organizing map, and SVM for classification and comparison, concluding that SVM is the best method in terms of classification accuracy and processing speed.

Some other approaches related to regression models include Poisson regression, least-square regression, and logistic regression [70]. Publications such as Jena and Bhalja [71] use a logistic regression binary classifier for the development of a new fault zone identification scheme for busbar verified by modeling an existing power generation station in a design software package. The proposed scheme is able to identify the fault zone with an accuracy of 99% when it is tested on a large dataset (28,800) by using a small training dataset (9600 cases). In the diagnosis of power systems, Xu and Chow [72] report the results obtained after using two different techniques, i.e., logistic regression and artificial NN, for the identification of the cause of faults in the power distribution systems. Logistic regression is a parametric model that is rarely used in power system fault diagnosis, while artificial NN is a nonparametric method that has been extensively used in this field. Logistic regression as a conventional statistical method has formalized models to exhibit the nonlinear relationship between the independent and dependent variables, while artificial NN can increase its flexibility by including hidden layers, which is often regarded as a substantial advantage. They conclude that both can be easily implemented. As seen from the results, artificial NN can achieve higher balanced accuracy than logistic regression; however, logistic regression is much faster because the artificial NN requires a relatively long training time and cross-validation requires an even longer computation time. Regarding the linear regression-base methods, the work of Cha et al. [73] presents the diagnosis and detection of faults in the main engine of a space shuttle during a stable state. Within the automotive industry, Jiang and Yin [74] present a new design and implementation approach based on recursive total principle component regression for efficient data-driven fault detection in automobile cyber-physical systems. Meanwhile, Bolovinou et al. [75] solve the problem of predicting the distance at which an electric vehicle can be driven before the energy recharge is required. The fact that the model is online implies that the prediction is made at any distance traveled from the beginning of the trip, which is achieved from a regression analysis. Using square linear regressions, Cappiello et al. [76] present a statistical model to predict the instantaneous emissions and fuel consumption of light-duty vehicles. Yu et al. [77] provide theoretical support for the prediction of faults in highway electromechanical equipment through a panel data model-based multi-factor predictive model. This model is characterized by a two-dimensional multivariate regression analysis based on and individuals and time. Emphasizing the intelligent diagnosis of faults, the classification and regression tree (CART) is used by Gopinath et al. [78] as a back-end classifier to diagnose synchronous generators. The statistical characteristics of the frequency domain are extracted from the current signals of the in-test generators. According to the work presented by Bangura et al. [79], the hidden patterns and nuances of differences between healthy performance firms and several fault signatures using time-series DM for the diagnosis of eccentricities and bar/end-ring connector breakages in polyphase induction motors can be identified. In a more general scenario, Wang and Jiao [80] propose a method of failure prediction related to quality by constructing a total principal component regression model, which can divide the space of the variables into two subspaces, and only one of them will be related to the quality fault.

Table 2 summarizes the above-reviewed information, where the effectiveness percentage of each method is also presented; from this information, it is evident that the SVMs are one of the most used methods for fault detection. Many authors agree that SVMs are more robust than other algorithms and satisfy the minimization of structural risk; yet, its effectiveness relies on the features and preparation of data. In addition, they have a high correct identification relationship according to the reported effectiveness percentages. In several works, SVMs have presented a better performance than NNs. These works highlight that SVMs reach the global optimum in a more direct way, are less prone to overfitting, present a smaller computational model, etc. Similar to other algorithms such as NNs, once the training stage has been carried out, the computational time to perform a SVM-based diagnosis is relatively short, making it a suitable for online and continuous diagnosis of electric equipment. Although in several works SVMs have presented a low computation cost/time, it cannot be suitably compared if aspects such as effectiveness reached, overfitting issues, robustness, number of hidden layers and neurons per layer, number of nodes, model complexity, activation and kernel functions used, and training algorithms, among others, are not taken into account.

Table 2. Regression-based methods and their applications.

Regression Methods	Equipment Under Test	Type of Fault	Effectiveness Percentage
■ Multilayer SVM [57]	■ Transformer	■ Partial discharge and arcing	100
■ Multilayer SVM [58]	■ Transformer	■ Discharge and thermal faults	100
■ Multilayer SVM [59]	■ Transformer	■ Discharge and thermal faults	90
■ SVM [60]	■ High-voltage, direct current transmission lines	■ Pole-1 to ground, pole-2 to ground and pole-1 to pole-2	100
■ SVM [61]	■ Series compensated transmission line	■ Line-to-ground, line-to-line fault involving ground and line-to-line	98.703
■ LSTM and SVM [62]	■ Power systems	■ Data-based line trip fault	97.7
■ Multiclass SVM [63]	■ Induction motor	■ Mechanical fault	97.48
■ Robust locally linear embedding algorithm and SVM [64]	■ Motor, torque transducer/encoder and dynamometer	■ Gear fault	90–100

Table 2. *Cont.*

Regression Methods	Equipment Under Test	Type of Fault	Effectiveness Percentage
■ SVM in regression mode [65]	■ Stator winding of an induction motor	■ Stator winding short circuit faults	95.1, 80.8, and 92.7
■ Autoregressive time series model with exogenous variables and SVM [66]	■ HVAC systems	■ Air handling unit faults	92.3
■ Autoregressive time series model with exogenous variables and SVM [67]	■ Chillers	■ Reduced condenser and evaporator water flow, condenser fouling, non-condensable in refrigerant and refrigerant leak	90.31
■ Multiclass SVM [68]	■ Steam generator and pressure boundary of a reactor coolant system	■ Incipient faults	100
■ Logistic regression [71]	■ Busbar	■ Internal and external faults	99.69
■ CART [78]	■ Synchronous generators	■ Inter-turn fault	95.58–98.15
■ Time-series data mining [79]	■ Polyphase induction motors	■ Eccentricities and bar/end-ring connector breakages	100

2.3. Hybrid Techniques

It is common to find research where the authors decide to use not only predictive techniques, but also to combine different algorithms that lead them to obtain models or methods that offer better results, including greater precision and efficiency, as well as better handling of data. This section deals with those works whose authors use more than one method, combining classification techniques and regression techniques, as well as other methods that do not belong to the predictive modeling of DM. The use of hybrid techniques, i.e., techniques that combine different methods, is frequently observed in the diagnosis of equipment such as motors, transformers, and electric vehicles mainly, as stated below.

In the extensive field of motors, Seera et al. [81] use the hybrid fuzzy min–max (FMM) neural network and classification and regression tree (CART), which is known as FMM–CART, to perform rule extraction and data classification in order to detect and classify faults in different motor conditions. They show the overall accuracy rates of five motor conditions (healthy, broken rotor bars, unbalanced voltages, eccentricity, and stator winding faults). FMM presented the lowest accuracy, 93.62%, while CART and FMM–CART achieved 98.11% and 98.25%, respectively, for multiple motor conditions in a time of 0.21s, 0.92s, and 0.96s, respectively. Two years later, in 2014, Seera and Lim [82] implement this hybrid model and conclude that it can produce accurate predictions of motor failures in an online learning environment. In addition, the results of the model are better than those compared with CART, FMM, and multi-layer perceptron. At the noisy test, multi-layer perceptron and FMM presented 78.39% and 94.88% accuracy, whereas FMM–CART and CART achieved stable results with 96.54% and 97.82% accuracy. The multilayer perceptron structure was the most complex with 30 hidden nodes, whereas FMM produced 12 nodes (hyperboxes). FMM–CART and CART created eight and six leaves, respectively. The computational time of FMM was only 0.13s. Multilayer perceptron consumed the longest time (2.08s), whereas FMM–CART and CART used almost 1s. The CART method combined with adaptive neuro-fuzzy inference system is presented by Tran et al. [83]. They use current and vibration signals from the induction motor for fault diagnosis; additionally, the hybrid of back-propagation and least-squares algorithm is used to adjust the parameters of membership functions. The total classification accuracy was 91.11% and 76.67% for vibration and current signals, respectively. Other works such as the one presented by Pramesti et al. [84] involve the identification of stator failures in induction motors using the multinomial logistic regression analysis and the Wavelet Transform (WT). Júnior et al. [85] use a multiple linear regression modeling technique along with the analysis of variance and the genetic algorithm optimization to obtain classification models to diagnose three-phase induction motors under normal and short-circuit conditions. The method presents percentages of hits greater than 95% in the diagnosis of the normal and incipient short-circuit fault condition, even at different motor load levels. In addition to the low cost and simplicity, this method does not require physical access to the machine because the current and voltage can be measured from the motor control board. Thus, the probability of the occurrence of human accident is reduced significantly. Unlike several other reported methods for fault diagnosis, the proposed approach requires few data and only uses simulation data to construct the expert system. Seshadrinath et al. [86] propose an algorithm based on two parts: In the first one, the optimal size of the structure of the Probabilistic Neural Network (PNN) is determined, using an orthogonal least-squares regression algorithm. In the second part, the fusion of a Bayesian classifier is recommended as an effective solution to diagnose incipient interturn fault in the machine. To track the health status of a degraded system and predict the remaining service life of a turbofan engine, Zhou et al. [87] propose a method that combines the echo state kernel recursive least-squares algorithm and a Bayesian technique, which demonstrates an excellent performance with respect to long-term prediction.

To obtain an effective diagnosis of faults in automotive systems, intelligent monitoring schemes of the vehicle's condition are needed. In this regard, Choi et al. [88] develop three new approaches for fusion of classifiers in order to reduce the error rate. These approaches are: Joint optimization of the fusion center and individual classifiers, class-specific Bayesian fusion, and dynamic fusion, demonstrating that the proposed techniques surpass the individual classifiers such as PNN, k-Nearest Neighbor (kNN), or principal components analysis. A fault detection scheme for applications in the automotive industry is presented by Jakubek and Strasser [89]. They achieve the detection of faults by using kernel regression techniques and a NN. The resulting network uses significantly less basis functions than a radial basis function network with the same accuracy. Oliva et al. [90] present a model-based approach to predict the remaining driving range by combining a particle filtering and Markov chains by implementing detailed models of the battery, electric motor, and vehicle dynamics. Tseng and Chau [91] and Grubwinkler and Lienkamp [92] study different methodologies for the prediction of electric vehicle energy consumption. In particular, Tseng and Chau compare three approaches that include (1) approaches based on driver/vehicle/environment dependent factors using speed profile matching and driving habit matching, (2) approach of comparison with the average using personalized adjustment, and (3) a collaborative filtering approach that uses matrix factorization; whereas Grubwinkler and Lienkamp use the least-mean square algorithm for the prediction of the mean energy consumption. To have a broad overview of the methodologies used in estimation strategies related to the battery, control, and energy management of both hybrid and electric vehicles, it is recommended to review the work of Cuma and Koroglu [93].

In transformers, their preventive maintenance is very often emphasized. Liao et al. [94] use least-square SVM (LS-SVM) and particle swarm optimization in order to optimize the regression parameters for the diagnosis of transformers immersed in oil by using dissolved gases. A comparison with back-propagation neural network, radial basis function neural network, generalized regression neural network, and support vector regression is carried out. Advantages of the regression model include those inherited from the support vector regression, i.e., a unique solution and support of statistical learning theory. In the next years, the wavelet technique is fused with the LS-SVM by Zheng et al. [95] and Zhang et al. [96] to diagnose transformers as well. From the analysis and interpretation of the data generated by the concentration of dissolved gases, Yang and Hu [97] propose a fault diagnosis system, which combines back-propagation NN and a multinomial logistic regression model. Al-Janabi et al. [98] also propose a hybrid system to diagnose transformers. The proposal is based on genetic algorithms and neural networks; in general, it provides information to identify the exact fault in the transformer and its fault state. Fei and Zhang [99] also make use of genetic algorithms along with SVM for fault detection in transformers. Unlike the abovementioned works, Koley et al. [100] use the WT for the extraction of characteristics of the impulse test response of a transformer in the time and frequency domains and the SVM in regression mode to classify transformer faults. It should be pointed out that the SVM tool trained with only simulated data was capable of predicting fault classes accurately when the analog data were presented to the trained SVM for fault prediction.

For the diagnosis of faults in centrifugal pumps, Yunlong and Peng [101] present a new method based on LS-SVM and the empirical mode decomposition. In the case of monoblock centrifugal pumps, Sakthivel et al. [102] use a decision tree-fuzzy hybrid system. In the test dataset, the classification accuracy was 99.3% in decision tree-fuzzy method, 97.50% in rough set-fuzzy method, and 96.67% in case of PCA-based decision tree fuzzy method. For the same task, in [103], Muralidharan and Sugumaran use Wavelet analysis, the Naïve Bayes (NB) algorithm, and the Bayes network algorithm. In [104], they apply the J48 algorithm and the continuous Wavelet transform (CWT). The sym3, rbio2.6, and coif1 mother wavelets are the most suitable for fault diagnosis of centrifugal pumps, reaching a classification accuracy of 100%. Finally, in [105], they use the SVM and the CWT. In this case, bior3.7_17 is the wavelet that gives maximum classification accuracy (99.76%). Hence, it can be considered as the best wavelet as it has the maximum fault discriminating capability for the system under study. Other works that have used the WT are the systems for electric power distribution. Jamil et al. [106] implement an algorithm based on fuzzy logic that uses the DWT to identify 10 different types of faults in an electrical power distribution system. For high impedance fault detection in electrical distribution networks, the WT extracts dynamic characteristics to feed a decision-making system based on SVM [107]. The SVM is also used along with the Hilbert Huang transform to decompose the voltages of transmission lines into intrinsic mode functions [108] for fault classification in power systems. The main contribution of the proposed algorithm is the possibility of its application to any transmission line, no matter the line configuration, with no need for re-training at different load values, voltage levels, and fault resistances. In 2018, Singh and Vishwakarma [109] present a methodology to classify cross-country faults in series-compensated double circuit transmission lines. This method is based on EMD and three different classifiers: SVM, NB, and PNN. The effectiveness is 95% for SVM, 91.66% for Naïve-Bayes, and 96.7% for PNN, where their response times are 0.03s, 0.012s, and 0.016s respectively. Da Silva et al. [110] apply qualitative trend analysis and NB for the diagnosis of multiple failures in transmission lines. This hybrid diagnosis system can be generalized to deal with other types of faults along the transmission line.

Regarding other machines, Lin and Horng [111] use a scheme of classification and detection of faults in an ion implanter, proposing a hybrid classification tree, i.e., they combine a grouping algorithm with CART. They indicate that their methodology is general and can be applied to other machines by simply modifying the warning generation criteria. For the fault detection in components of nuclear power plants components, statistical methods have been used. Di Maio et al. [112] used a set of auto-associative kernel regression models, a hybrid approach based on correlation analysis, a genetic algorithm, and a sequential probability ratio test to detect faults by taking as a case study a coolant pump of a typical pressurized water reactor. Liangyu et al. [113] propose an artificial NN combined with optimal zoom search to recognize various degrees of failure in a high-pressure feedwater heater system. The classification of the healthy and defective conditions of a face milling tool is done through the acquisition of sound signals using the discrete WT (DWT) and the J48 algorithm, which is a decision tree technique [114]. On the other hand, to detect and diagnose faults in HVAC systems, Du et al. [115] combine NNs and clustering analysis.

Table 3 lists the works that have presented hybrid techniques for the detection of faults and diagnosis of the abovementioned electric equipment and systems. According to the information shown in Table 3, two different are combined on average to perform the diagnosis, where not only DM techniques are implemented, but other signal processing algorithms are used to extract or highlight features contained into the analyzed signals in order to simplify the fault classification task. As main techniques, the WT and the EMD are found. While the works that use WT exploit its capability for time frequency decomposition in a symmetric way, the works that use EMD exploit its capability to decompose a signal in an adaptive way. In this regard, EMD has been preferred in many works since a-priori information for the analyzed signal is not needed. From this point of view, other recent schemes based on EMD such as down-sampling EMD [116], which is a method that provides specific advantages over EMD, should be explored in the field of fault detection in electric equipment. WT has been also widely used to remove the unwanted noise in an electrical signal. This noise is generated by acquisition systems, sensors, or any electronic device. Regarding DM techniques, FLSs have presented suitable results under noisy conditions in the input signals, since this noise is somehow compared with the uncertainties of the input data, which is an inherent ability of FLSs.

Table 3. Summary for hybrid techniques used.

Data Mining Techniques	Other Techniques	Application
■ Fuzzy min-max NN and CART [81,82]	-	■ Induction motors
■ ANFIS and CART [83]	-	■ Induction motors
■ Logistic regression [84]	WT	■ Stator winding of an induction motor
■ Multiple linear regression and genetic algorithms [85]	RMS	■ Three-phase induction motors
■ PNN and orthogonal least squares regression algorithm [86]	DWT	■ Induction machines
■ Kernel recursive least squares algorithm and Bayesian technique [87]	-	■ Turbofan engine
■ SVM, PNN, kNN, and Principal components analysis [88]	-	■ Engine system

Table 3. *Cont.*

Data Mining Techniques	Other Techniques	Application
■ Kernel regression techniques and NN [89]	-	■ Automotive industry
■ Particle Filtering and Markov Chains [90]	-	■ Electric ■ Vehicles
■ Average and collaborative filtering [91]	Similarity Matching	■ vehicle energy consumption
■ Mean and least-mean square algorithm [92]	-	■ vehicle energy consumption
■ LS-SVM regression [94]	Particle swarm optimization algorithm	■ Transformer
■ LS-SVM [95,96]	WT	■ Transformer
■ Multinomial logistic regression and NN [97]	-	■ Transformer
■ NN and genetic algorithm [98]	-	■ Transformer
■ SVM and genetic algorithm [99]	-	■ Transformer
■ SVM in regression mode [100]	WT	■ Transformer
■ LS-SVM [101]	EMD	■ Centrifugal pump
■ Decision tree-fuzzy and rough set-fuzzy methods [102]	-	■ Monoblock centrifugal pump
■ NB classifier and Bayes net classifier [103]	Wavelet analysis	■ Monoblock centrifugal pump
■ Decision tree [104]	Wavelet analysis	■ Monoblock centrifugal pump
■ SVM [105]	CWT	■ Monoblock centrifugal pump
■ Fuzzy logic [106]	WT	■ Electrical power distribution system
■ SVM [107]	WT	■ Distribution networks
■ SVM [108]	EMD	■ Power systems
■ SVM, NB and PNN [109]	EMD	■ Transmission Lines
■ Naïve Bayes [110]	Qualitative trend analysis	■ Transmission Lines

Table 3. *Cont.*

Data Mining Techniques	Other Techniques	Application
■ CART and clustering algorithm [111]	-	■ Ion implanter
■ Auto-associative kernel regression, correlation analysis, genetic algorithm, and probability ratio test [112]	-	■ Reactor Coolant Pump of a typical Pressurized Water Reactor
■ NN [113]	Optimal zoom search	■ High-pressure feedwater heater system
■ Decision tree [114]	DWT	■ Face milling tool
■ NN and subtractive clustering analysis [115]	-	■ HVAC systems

3. Recent Methods for General Applications

In the literature, various articles that involve the most recent research on classification and regression algorithms, which can be used in different areas of application, have been presented. Djeffal et al. [117] present a method based on filtering and revision stages to delete samples that have little influence on the learning results of a SVM, where the goal is to reduce training time without losing accuracy. This strategy could be used for handling and reducing huge databases before the application of any other algorithm. Zhao et al. [118] tackle the challenging problem of classification in the presence of label noise. In this regard, they propose a Markov chain sampling framework that robustly learns effective classifiers and accurately identifies mislabeled instances. Hwang and Son [119] propose a prototype-based classification to select some data from a dataset for development of learning rules and prediction, demonstrating that the proposed approach overcomes other classifiers such as the Bayes classifier and the nearest neighbor. Regarding the Bayesian approaches, Zhang et al. [120] present a probability density estimation approach based on the nonparametric kernel mixing model in order to estimate reliable class-conditional probability functions; in general, the proposed Bayesian classifier consists of three steps: Partitioning, structure learning, and estimation of probability density functions. Zhang et al. [121] propose a learning scheme that offers a recursive algorithm to explore the distribution of class density for the Bayesian estimation and an automated approach to select powerful discriminant functions for the classification of high-dimensional data, while Celotto [122] proposes a unified visual approach to compare and classify a large subset of Bayesian confirmation measures. In the work of Becker et al. [123], analytical and approximate inference methods are discussed to calculate the marginal probabilities of Bayes factors, providing guidance on the interpretation of results and offering new types of analysis to study sequential data in many application areas.

Regarding regression analysis, Le et al. [124] present the geometric-based online Gaussian process that could scale with massive datasets, guaranteeing that the proposed algorithm produces a good enough solution (close to the optimal one) and a fast-online regression. Marx and Vreeken [125] present an information theory-based approach using the Kolmogorov complexity and the principle of minimum description length to provide a practical solution to the problem of inferring the direction of causal dependence of observational data. Rudaś and Jaroszewicz [126] analyze two uplift modeling approaches for linear regression and identify the situations in which each model works best; in fact, they propose a third model that combines the benefits of both approaches. Liang et al. [70] propose the model called heterogeneous-target robust mixture regression that addresses the challenges and practical concerns of joint learning for multiple objectives/multi-tasking learning by managing mixed types of objectives simultaneously, imposing structural constraints on each component of the mixture and adopting robustness strategies.

On the other hand, Chen and Guestrin [127] describe a scalable end-to-end tree boosting system (XGBoost) and propose a new algorithm based on data dispersion providing information on cache access patterns, data compression, and fragmentation to construct a scalable tree boosting system. They claim that XGBoost is widely used by data scientists to achieve cutting-edge results in many machine learning challenges. Teinemaa et al. [128] evaluate the temporal stability and prediction accuracy of different existing predictive process monitoring methods, finding that the methods based on the XGBoost and LSTM exhibit the highest temporal stability. In relation to NN, Baldi [129] studies the internal and external approaches for the design of recursive neural architectures. Zhang et al. [130] address the problems of intelligent fault diagnosis when the data at the time of training and testing does not come from the same distribution by using domain adaptive convolutional NNs. Bouguelia et al. [131] propose an adaptive algorithm to continuously update a system of neurons through the extension of the growing neural gas algorithm with three complementary mechanisms, which allows one to closely monitor the gradual and sudden changes in the distribution of data. The imbalance data problem is addressed by Xi et al. [132]. They propose the least-squares support vector machine for class imbalance learning by evaluating two parameters of misclassification costs; also, the Cholesky factorization is used to enhance computational stability. In order to reduce the estimation error in online sequential extreme learning machine systems, Lu et al. [133] present a new training approach based on Kalman filter. Although the two last works have been applied to fault detection in aircraft engines, they can be used in other machines.

Table 4 shows a compendium of the abovementioned methods. They are grouped by year in order to show their chronological appearance and highlight which ones are the latest algorithms or strategies proposed in the literature to solve DM issues or improve DM tasks. As these methods can address general applications, it is recommended their research and integration in fault detection methodologies of electric equipment and systems. For instance, the least-squares SVM is useful for imbalance data, i.e., when there is a disproportionate ratio of observations in each class, and the Markov chain sampling is a useful tool for mislabel data.

Table 4. Recent methods for general applications.

Year	Methods	Usage
2016	Naïve Bayes and feature weighting approaches [121]	■ High-dimensional massive data classification
	Visual approach to represent Bayesian confirmation measures (BCMs) [122]	■ Visualize the behavior and symmetry properties of BCMs
	XGBoost [127]	■ Scalable machine learning system for tree boosting
2017	Covering-based samples reduction [117]	■ Fast binary support vector machine learning method
	MixedTrails, a Bayesian approach [123]	■ Types of analysis to study sequential data
	Geometric-based Online Gaussian Process for fast regression [124]	■ Handling of large-scale datasets
	Kolmogorov complexity and use the Minimum Description Length [125]	■ Solution to the problem of inferring the direction of causal dependence of observational data
	Recursive neural architectures [129]	■ Design of recursive architectures for numerical data of variable size

Table 4. *Cont.*

Year	Methods	Usage
2018	Heterogeneous-target robust mixture regression (HERMIT) [70]	■ Handling of heterogeneous data
	Markov chain sampling [118]	■ Classification of data in presence of label noise
	Prototype-based classification [119]	■ Learning and prediction based on the selection of handfuls of class data
	Kernel mixture model [120]	■ Probability density estimation in Bayesian classifiers
	Linear regression [126]	■ Uplift modeling
	Random forest, XGBoost, and LSTM [128]	■ Analysis of temporal stability and accuracy for binary classification
	Convolutional NN [130]	■ Intelligent fault diagnosis when the data at training and testing time does not come from the same distribution
	Growing Neural Gas Algorithm [131]	■ Adaptive algorithm got evolving data streams
2019	Least squares support vector machine [132]	■ Imbalance of data
	Logistic regression and Kalman filter [133]	■ Estimation error reduction in online sequential extreme learning machine systems

4. Conclusions

The development of efficient and reliable methodologies represents an extremely important task for researchers and developers of diagnosis systems; in order to contribute to the solution of this task, DM techniques have been widely used. To offer the reader an overview of DM techniques used in the detection of faults and diagnosis of electrical equipment and systems in recent years, this paper provides a general review that can facilitate informed decision-making for specific applications. All the details and results obtained by the authors cited here can be consulted directly from the bibliography of each research.

Although certain techniques have been constantly used for specific applications, e.g., SVMs in transformers, the selection of an appropriate DM technique for either classification or regression will depend on many factors, e.g., monitoring technology, features of data, and knowledge about the in-test system operation. However, it is important to take into account that the more complex and robust the systems, the greater the amount and variety of data produced, and the more difficult the detection of faults and the diagnosis. Additionally, the researcher has to be informed about the features of specific DM algorithms so that, through its implementation, the information contained in the acquired data can be exploited.

It is extremely difficult for a single technique to detect the full range of faults of a system in a 100% reliable way. Each method has its own strengths and weaknesses. Outputs from various diagnostic methods must be aggregated into an overall evaluation system; thus, instead of using one diagnostic method, intelligent hybrid methods that combine the strengths of each method can be developed. In the literature, there are many articles using hybrid techniques in order to increase the percentage of efficiency, accuracy, reliability, and speed of their models. On the one hand, different signal processing techniques have been used for pre-processing of data. This pre-processing allows highlighting and extracting information from raw data. Typical operations are denoising, frequency or mode decomposition, and space transformation, where the WT- and EMD-based methods have demonstrated promising results. On the other hand, the combination of different DT algorithms has been also explored in order to take advantage of their individual benefits. Wu et al. [134] present an important analysis on the 10 most influential DM algorithms in the research community, being C4.5, CART, PageRank, k-Means, kNN, Apriori, AdaBoost, Expectation-Maximization, NB, and SVM. It should be noted that these algorithms cover statistical learning, classification, clustering, link mining, and association analysis. Despite obtaining promising results in many works and having knowledge about both the in-test system and analyzed data, it is difficult or impossible to conceive a perfect algorithm in terms of accuracy, velocity, or complexity for specific applications, mainly considering that even similar applications can have many different requirements; in this regard, the design and development of new algorithms and methods are still of paramount importance.

Also, special attention has to be given to equipment related to renewable energy sources such as wind turbines, photovoltaic systems, power converters, energy storage systems, among others, due to their rapid development and growth [135,136]. For instance, the void defects evolving into damage in wind turbine blades are investigated in [137]. The improvement of photovoltaic and wind power storage systems based on the prediction of battery life and its faults using SVMs is presented in [138]; in these systems, the correct operation of batteries is fundamental. It is clear that all the elements of a system are important and the research of specialized fault detection methodologies for the individual elements and the system as a whole are critical for the maintenance and repair of the system.

Some recommended directions for future research are: (i) Fusion and analysis of multiple physical variables as source of information of a specific equipment, (ii) exploration and integration of recent algorithms to improve the quality of data before the application of a DT-based algorithm, (iii) development of practical hardware solutions for online and real-time fault diagnosis, and (iv) detection of incipient faults.

Author Contributions: Conceptualization, A.C.-V. and M.V.-R.; investigation, resources and visualization, A.C.-V., J.P.A.-S. and D.G.-L.; funding acquisition, J.P.A.-S. and M.V.-R.; Writing—original draft, review & editing, all the Authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the "Consejo Nacional de Ciencia y Tecnología (CONACYT)" under the scholarship 892305.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adil, A.M.; Ko, Y. Socio-technical evolution of Decentralized Energy Systems: A critical review and implications for urban planning and policy. *Renew. Sustain. Energy Rev.* **2016**, *57*, 1025–1037. [CrossRef]
2. García-Villalobos, J.; Zamora, I.; San Martín, J.I.; Asensio, F.J.; Aperribay, V. Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches. *Renew. Sustain. Energy Rev.* **2014**, *38*, 717–731. [CrossRef]
3. Valtierra-Rodriguez, M. Fractal dimension and data mining for detection of short-circuited turns in transformers from vibration signals. *Meas. Sci. Technol.* **2019**, *31*, 025902. [CrossRef]
4. Mejia-Barron, A.; Valtierra-Rodriguez, M.; Granados-Lieberman, D.; Olivares-Galvan, J.C.; Escarela-Perez, R. Experimental data-based transient-stationary current model for inter-turn fault diagnostics in a transformer. *Electr. Power Syst. Res.* **2017**, *152*, 306–315. [CrossRef]
5. Mejia-Barron, A.; de Santiago-Perez, J.J.; Granados-Lieberman, D.; Amezquita-Sanchez, J.P.; Valtierra-Rodriguez, M. Shannon Entropy Index and a Fuzzy Logic System for the Assessment of Stator Winding Short-Circuit Faults in Induction Motors. *Electronics* **2019**, *8*, 90. [CrossRef]
6. Lee, J.H.; Pack, J.H.; Lee, I.S. Fault Diagnosis of Induction Motor Using Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 2950. [CrossRef]
7. Kantardzic, M. *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed.; Wiley—IEEE Press: Piscataway, NJ, USA, 2011.

8. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*, 3rd ed.; The Morgan Kaufmann Series in Data Management Systems; Elsevier Science: Waltham, MA, USA, 2011.
9. Jothi, N.; Rashid, N.A.; Husain, W. Data Mining in Healthcare—A Review. *Procedia Comput. Sci.* **2015**, *72*, 306–313. [CrossRef]
10. Paramasivam, V.; Yee, T.S.; Dhillon, S.K.; Sidhu, A.S. A methodological review of data mining techniques in predictive medicine: An application in hemodynamic prediction for abdominal aortic aneurysm disease. *Biocybern. Biomed. Eng.* **2014**, *34*, 139–145. [CrossRef]
11. Bellazzi, R.; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inf.* **2008**, *77*, 81–97. [CrossRef]
12. Ngai, E.W.T.; Xiu, L.; Chau, D.C.K. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst. Appl.* **2009**, *36*, 2592–2602. [CrossRef]
13. Ge, Z.; Song, Z.; Ding, S.X.; Huang, B. Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access* **2017**, *5*, 20590–20616. [CrossRef]
14. Xu, Y.; Sun, Y.; Wan, J.; Liu, X.; Song, Z. Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications. *IEEE Access* **2017**, *5*, 17368–17380. [CrossRef]
15. Odell, S.D.; Bebbington, A.; Frey, K.E. Mining and climate change: A review and framework for analysis. *Extr. Ind. Soc.* **2018**, *5*, 201–214. [CrossRef]
16. Ngai, E.W.T.; Hu, Y.; Wong, Y.H.; Chen, Y.; Sun, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decis. Support Syst.* **2011**, *50*, 559–569. [CrossRef]
17. Christina, A.J.; Salam, M.A.; Rahman, Q.M.; Wen, F.; Ang, S.P.; Voon, W. Causes of transformer failures and diagnostic methods—A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1442–1456.
18. De Faria, H.; Costa, J.G.S.; Olivas, J.L.M. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renew. Sustain. Energy Rev.* **2015**, *46*, 201–209. [CrossRef]
19. Sun, H.C.; Huang, Y.C.; Huang, C.M. Fault Diagnosis of Power Transformers Using Computational Intelligence: A Review. *Energy Procedia* **2012**, *14*, 1226–1231. [CrossRef]
20. Zheng, Y.; Ouyang, M.; Han, X.; Lu, L.; Li, J. Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles. *J. Power Sources* **2018**, *377*, 161–188. [CrossRef]
21. Liu, Y.; Bazzi, A.M. A review and comparison of fault detection and diagnosis methods for squirrel-cage induction motors: State of the art. *ISA Trans.* **2017**, *70*, 400–409. [CrossRef]
22. Prasad, A.; Belwin, E.J.; Ravi, K. A review on fault classification methodologies in power transmission systems: Part—I. *J. Electr. Syst. Inf. Technol.* **2018**, *5*, 48–60.
23. Prasad, A.; Belwin, E.J.; Ravi, K. A review on fault classification methodologies in power transmission systems: Part-II. *J. Electr. Syst. Inf. Technol.* **2018**, *5*, 61–67. [CrossRef]
24. Gururajapathy, S.S.; Mokhlis, H.; Illias, H.A. Fault location and detection techniques in power distribution systems with distributed generation: A review. *Renew. Sustain. Energy Rev.* **2017**, *74*, 949–958. [CrossRef]
25. Hare, J.; Shi, X.; Gupta, S.; Bazzi, A. Fault diagnostics in smart micro-grids: A survey. *Renew. Sustain. Energy Rev.* **2016**, *60*, 1114–1124. [CrossRef]
26. Taylor, D.W.; Corne, D.W.; Taylor, D.L.; Harkness, J. Predicting alarms in supermarket refrigeration systems using evolved neural networks and evolved rulesets. In Proceedings of the 2002 Congress on Evolutionary Computation, Honolulu, HI, USA, 12–17 May 2002.
27. Magoulès, F.; Zhao, H.; Elizondo, D. Development of an RDP neural network for building energy consumption fault detection and diagnosis. *Energy Build.* **2013**, *62*, 133–138. [CrossRef]
28. Tallam, R.M.; Habetler, T.G.; Harley, R.G.; Gritter, D.J.; Burton, B.H. Neural network based on-line stator winding turn fault detection for induction motors. In Proceedings of the Conference Record of the 2000 IEEE Industry Applications Conference, Thirty-Fifth IAS Annual Meeting and World Conference on Industrial Applications of Electrical Energy, Rome, Italy, 8–12 October 2000.
29. Martins, J.F.; Pires, V.F.; Pires, A.J. Unsupervised Neural-Network-Based Algorithm for an On-Line Diagnosis of Three-Phase Induction Motor Stator Fault. *IEEE Trans. Ind. Electron.* **2007**, *54*, 259–264. [CrossRef]
30. Xu-hong, W.; Yi-gang, H. Fuzzy Neural Network based On-line Stator Winding Turn Fault Detection for Induction Motors. In Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, China, 23–25 May 2007.

31. Ballal, M.S.; Khan, Z.J.; Suryawanshi, H.M.; Sonolikar, R.L. Adaptive Neural Fuzzy Inference System for the Detection of Inter-Turn Insulation and Bearing Wear Faults in Induction Motor. *IEEE Trans. Ind. Electron.* **2007**, *54*, 250–258. [CrossRef]
32. Cabal-Yepez, E.; Valtierra-Rodriguez, M.; Romero-Troncoso, R.J.; Garcia-Perez, A.; Osornio-Rios, R.A.; Miranda-Vidales, H.; Alvarez-Salas, R. FPGA-based entropy neural processor for online detection of multiple combined faults on induction motors. *Mech. Syst. Signal Process.* **2012**, *30*, 123–130. [CrossRef]
33. Camarena-Martinez, D.; Valtierra-Rodriguez, M.; Garcia-Perez, A.; Osornio-Rios, R.A.; Romero-Troncoso, R.J. Empirical mode decomposition and neural networks on FPGA for fault diagnosis in induction motors. *Sci. World J.* **2014**, *2014*, 908140. [CrossRef]
34. Amezquita-Sanchez, J.P.; Valtierra-Rodriguez, M.; Camarena-Martinez, D.; Granados-Lieberman, D.; Romero-Troncoso, R.J.; Dominguez-Gonzalez, A. Fractal dimension-based approach for detection of multiple combined faults on induction motors. *J. Vib. Control* **2016**, *22*, 3638–3648. [CrossRef]
35. Zolfaghari, S.; Noor, S.B.M.; Rezazadeh, M.M.; Marhaban, M.H.; Mariun, N. Broken rotor bar fault detection and classification using wavelet packet signature analysis based on fourier transform and multi-layer perceptron neural network. *Appl. Sci.* **2018**, *8*, 25. [CrossRef]
36. Flores, A.; Quiles, E.; Garcia, E.; Morant, F. Fault Diagnosis of Electric Transmission Lines using Modular Neural Networks. *IEEE Lat. Am. Trans.* **2016**, *14*, 3663–3668. [CrossRef]
37. Valtierra-Rodriguez, M.; Romero-Troncoso, R.J.; Osornio-Rios, R.A.; Garcia-Perez, A. Detection and classification of single and combined power quality disturbances using neural networks. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2473–2482. [CrossRef]
38. Rigatos, G.; Siano, P. Power transformers' condition monitoring using neural modeling and the local statistical approach to fault diagnosis. *Int. J. Electr. Power Energy Syst.* **2016**, *80*, 150–159. [CrossRef]
39. Menezes, A.G.C.; Almeida, O.M.; Barbosa, F.R. Use of decision tree algorithms to diagnose incipient faults in power transformers. In Proceedings of the Simposio Brasileiro de Sistemas Eletricos (SBSE), Niteroi, Brazil, 12–16 May 2018.
40. Han, Y.; Zhao, D.; Hou, H. Oil-immersed Transformer Internal Thermoelectric Potential Fault Diagnosis Based on Decision-tree of KNIME Platform. *Procedia Comput. Sci.* **2016**, *83*, 1321–1326. [CrossRef]
41. Samantaray, S.R.; Dash, P.K. Decision Tree based discrimination between inrush currents and internal faults in power transformer. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 1043–1048. [CrossRef]
42. Sakthivel, N.R.; Sugumaran, V.; Babudevasenapati, S. Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. *Expert Syst. Appl.* **2010**, *37*, 4040–4049. [CrossRef]
43. Sharma, A.; Sugumaran, V.; Babu, S. Misfire detection in an IC engine using vibration signal and decision tree algorithms. *Measurement* **2014**, *50*, 370–380. [CrossRef]
44. Volkanovski, A.; Čepin, M.; Mavko, B. Application of the fault tree analysis for assessment of power system reliability. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 1116–1127. [CrossRef]
45. Duan, R.; Zhou, H. A New Fault Diagnosis Method Based on Fault Tree and Bayesian Networks. *Energy Procedia* **2012**, *17*, 1376–1382. [CrossRef]
46. Bonvini, M.; Sohn, M.D.; Granderson, J.; Wetter, M.; Piette, M.A. Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques. *Appl. Energy* **2014**, *124*, 156–166. [CrossRef]
47. Li, D.; Zhou, Y.; Hu, G.; Spanos, C.J. Fault detection and diagnosis for building cooling system with a tree-structured learning method. *Energy Build.* **2016**, *127*, 540–551. [CrossRef]
48. Rodríguez, P.V.J.; Arkkio, A. Detection of stator winding fault in induction motor using fuzzy logic. *Appl. Soft Comput.* **2008**, *8*, 1112–1120. [CrossRef]
49. Amezquita-Sanchez, J.P.; Valtierra-Rodriguez, M.; Perez-Ramirez, C.A.; Camarena-Martinez, D.; Garcia-Perez, A.; Romero-Troncoso, R.J. Fractal dimension and fuzzy logic systems for broken rotor bar detection in induction motors at start-up and steady-state regimes. *Meas. Sci. Technol.* **2017**, *28*, 075001. [CrossRef]
50. Islam, S.M.; Wu, T.; Ledwich, G. A novel fuzzy logic approach to transformer fault diagnosis. *IEEE Trans. Dielectr. Electr. Insul.* **2000**, *7*, 177–186. [CrossRef]
51. Huang, Y.C.; Sun, H.C. Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 974–981. [CrossRef]

52. Khan, S.A.; Equbal, M.D.; Islam, T. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 590–596. [CrossRef]
53. Chin, H.C. Fault section diagnosis of power system using fuzzy logic. *IEEE Trans. Power Syst.* **2003**, *18*, 245–250. [CrossRef]
54. Valtierra-Rodriguez, M.; Granados-Lieberman, D.; Torres-Fernandez, J.E.; Rodríguez-Rodríguez, J.R.; Gómez-Aguilar, J.F. A new methodology for tracking and instantaneous characterization of voltage variations. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 1596–1604. [CrossRef]
55. Lauro, F.; Moretti, F.; Capozzoli, A.; Khan, I.; Pizzuti, S.; Macas, M.; Panzieri, S. Building Fan Coil Electric Consumption Analysis with Fuzzy Approaches for Fault Detection and Diagnosis. *Energy Procedia* **2014**, *62*, 411–420. [CrossRef]
56. Zio, E.; Baraldi, P.; Popescu, I.C. A fuzzy decision tree method for fault classification in the steam generator of a pressurized water reactor. *Ann. Nucl. Energy* **2009**, *36*, 1159–1169. [CrossRef]
57. Mittal, M.; Bhushan, M.; Patil, S.; Chaudhari, S. Optimal Feature Selection for SVM Based Fault Diagnosis in Power Transformers. *IFAC Proc. Vol.* **2013**, *46*, 809–814. [CrossRef]
58. Lv, G.; Cheng, H.; Zhai, H.; Dong, L. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr. Power Syst. Res.* **2005**, *75*, 9–15. [CrossRef]
59. Bacha, K.; Souahlia, S.; Gossa, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electr. Power Syst. Res.* **2012**, *83*, 73–79. [CrossRef]
60. Johnson, J.M.; Yadav, A. Complete protection scheme for fault detection, classification and location estimation in HVDC transmission lines using support vector machines. *IET Sci. Meas. Technol.* **2017**, *11*, 279–287. [CrossRef]
61. Parikh, U.B.; Das, B.; Maheshwari, R. Fault classification technique for series compensated transmission line using support vector machine. *Int. J. Electr. Power Energy Syst.* **2010**, *32*, 629–636. [CrossRef]
62. Zhang, S.; Wang, Y.; Liu, M.; Bao, Z. Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM. *IEEE Access* **2018**, *6*, 7675–7686. [CrossRef]
63. Gangsar, P.; Tiwari, R. Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mech. Syst. Signal Process.* **2017**, *94*, 464–481. [CrossRef]
64. Zhang, Y.; Ye, D.; Liu, Y. Robust locally linear embedding algorithm for machinery fault diagnosis. *Neurocomputing* **2018**, *273*, 323–332. [CrossRef]
65. Das, S.; Purkait, P.; Koley, C.; Chakravorti, S. Performance of a load-immune classifier for robust identification of minor faults in induction motor stator winding. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 33–44. [CrossRef]
66. Mulumba, T.; Afshari, A.; Yan, K.; Shen, W.; Norford, L.K. Robust model-based fault diagnosis for air handling units. *Energy Build.* **2015**, *86*, 698–707. [CrossRef]
67. Yan, K.; Shen, W.; Mulumba, T.; Afshari, A. ARX model based fault detection and diagnosis for chillers using support vector machines. *Energy Build.* **2014**, *81*, 287–295. [CrossRef]
68. Ayodeji, A.; Liu, Y. Support vector ensemble for incipient fault diagnosis in nuclear plant components. *Nucl. Eng. Technol.* **2018**, *50*, 1306–1313. [CrossRef]
69. Lai, K.; Phung, B.; Blackburn, T. Application of data mining on partial discharge part I: Predictive modelling classification. *IEEE Trans. Dielectr. Electr. Insul.* **2010**, *17*, 846–854. [CrossRef]
70. Liang, J.; Chen, K.; Lin, M.; Zhang, C.; Wang, F. Robust finite mixture regression for heterogeneous targets. *Data Min. Knowl. Discov.* **2018**, *32*, 1509–1560. [CrossRef]
71. Jena, S.; Bhalja, B.R. Development of a new fault zone identification scheme for busbar using logistic regression classifier. *IET Gener. Transm. Distrib.* **2017**, *11*, 174–184. [CrossRef]
72. Xu, L.; Chow, M.Y. A Classification Approach for Power Distribution Systems Fault Cause Identification. *IEEE Trans. Power Syst.* **2006**, *21*, 53–60. [CrossRef]
73. Cha, J.; Ha, C.; Ko, S.; Koo, J. Application of fault factor method to fault detection and diagnosis for space shuttle main engine. *Acta Astronaut.* **2016**, *126*, 517–527. [CrossRef]
74. Jiang, Y.; Yin, S. Recursive Total Principle Component Regression Based Fault Detection and Its Application to Vehicular Cyber-Physical Systems. *IEEE Trans. Ind. Inf.* **2018**, *14*, 1415–1423. [CrossRef]

75. Bolovinou, A.; Bakas, I.; Amditis, A.; Mastrandrea, F.; Vinciotti, W. Online prediction of an electric vehicle remaining range based on regression analysis. In Proceedings of the IEEE International Electric Vehicle Conference (IEVC), Florence, Italy, 17–19 December 2014.
76. Cappiello, A.; Chabini, I.; Nam, E.K.; Lue, A.; Abou, Z.M. A statistical model of vehicle emissions and fuel consumption. In Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems, Singapore, 3–6 September 2002.
77. Yu, Q.; Qin, Y.; Liu, P.; Ren, G. A Panel Data Model-Based Multi-Factor Predictive Model of Highway Electromechanical Equipment Faults. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1–7. [CrossRef]
78. Gopinath, R.; Santhosh, K.C.; Ramachandran, K.I.; Upendranath, V.; Sai Kiran, P.V.R. Intelligent fault diagnosis of synchronous generators. *Expert Syst. Appl.* **2016**, *45*, 142–149. [CrossRef]
79. Bangura, J.F.; Povinelli, R.J.; Demerdash, N.A.O.; Brown, R.H. Diagnostics of eccentricities and bar/end-ring connector breakages in polyphase induction motors through a combination of time-series data mining and time-stepping coupled FE-state space techniques. In Proceedings of the Conference Record of the 2001 IEEE Industry Applications Conference, 36th IAS Annual Meeting, Chicago, IL, USA, 30 September–4 October 2001.
80. Wang, G.; Jiao, J. Quality-Related Fault Detection and Diagnosis Based on Total Principal Component Regression Model. *IEEE Access* **2018**, *6*, 10341–10347. [CrossRef]
81. Seera, M.; Chee, P.L.; Ishak, D.; Singh, H. Fault Detection and Diagnosis of Induction Motors Using Motor Current Signature Analysis and a Hybrid FMM–CART Model. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 97–108. [CrossRef] [PubMed]
82. Seera, M.; Chee, P.L. Online Motor Fault Detection and Diagnosis Using a Hybrid FMM-CART Model. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 806–812. [CrossRef] [PubMed]
83. Tran, V.T.; Yang, B.S.; Oh, M.S.; Tan, A.C.C. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **2009**, *36*, 1840–1849. [CrossRef]
84. Pramesti, W.; Damayanti, I.; Asfani, D.A. Stator fault identification analysis in induction motor using multinomial logistic regression. In Proceedings of the International Seminar on Intelligent Technology and Its Applications (ISITIA), Lombok, Indonesia, 28–30 June 2016.
85. Júnior, A.M.G.; Silva, V.V.R.; Baccarini, L.M.R.; Mendes, L.F.S. The design of multiple linear regression models using a genetic algorithm to diagnose initial short-circuit faults in 3-phase induction motors. *Appl. Soft Comput.* **2018**, *63*, 50–58. [CrossRef]
86. Seshadrinath, J.; Singh, B.; Panigrahi, B.K. Incipient Interturn Fault Diagnosis in Induction Machines Using an Analytic Wavelet-Based Optimized Bayesian Inference. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 990–1001. [CrossRef] [PubMed]
87. Zhou, H.; Huang, J.; Lu, F.; Thiyagalingam, J.; Kirubarajan, T. Echo state kernel recursive least squares algorithm for machine condition prediction. *Mech. Syst. Signal Process.* **2018**, *111*, 68–86. [CrossRef]
88. Choi, K.; Singh, S.; Kodali, A.; Pattipati, K.R.; Sheppard, J.W.; Namburu, S.M.; Chigusa, S.; Prokhorov, D.V.; Qiao, L. Novel Classifier Fusion Approaches for Fault Diagnosis in Automotive Systems. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 602–611. [CrossRef]
89. Jakubek, S.; Strasser, T. Fault-diagnosis using neural networks with ellipsoidal basis functions. In Proceedings of the American Control Conference, Anchorage, AK, USA, 8–10 May 2002.
90. Oliva, J.A.; Weihrauch, C.; Bertram, T. Model-Based Remaining Driving Range Prediction in Electric Vehicles by using Particle Filtering and Markov Chains. *World Electr. Veh. J.* **2013**, *6*, 204–213. [CrossRef]
91. Tseng, C.M.; Chau, C.K. Personalized Prediction of Vehicle Energy Consumption Based on Participatory Sensing. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3103–3113. [CrossRef]
92. Grubwinkler, S.; Lienkamp, M. A modular and dynamic approach to predict the energy consumption of electric vehicles. In Proceedings of the Conference on Future Automotive Technology, Munich, Germany, 18–19 March 2013.
93. Cuma, M.U.; Koroglu, T. A comprehensive review on estimation strategies used in hybrid and battery electric vehicles. *Renew. Sustain. Energy Rev.* **2015**, *42*, 517–531. [CrossRef]
94. Liao, R.; Zheng, H.; Grzybowski, S.; Yang, L. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electr. Power Syst. Res.* **2011**, *81*, 2074–2080. [CrossRef]

95. Zheng, H.; Zhang, Y.; Liu, J.; Wei, H.; Zhao, J.; Liao, R. A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. *Electr. Power Syst. Res.* **2018**, *155*, 196–205. [CrossRef]
96. Zhang, Y.Y.; Wei, H.; Yang, Y.D.; Zheng, H.B.; Zhou, T.; Jiao, J. Forecasting of Dissolved Gases in Oil-immersed Transformers Based upon Wavelet LS-SVM Regression and PSO with Mutation. *Energy Procedia* **2016**, *104*, 38–43. [CrossRef]
97. Yang, M.T.; Hu, L.S. Intelligent fault types diagnostic system for dissolved gas analysis of oil-immersed power transformer. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 2317–2324. [CrossRef]
98. Al-Janabi, S.; Rawat, S.; Patel, A.; Al-Shourbaji, I. Design and evaluation of a hybrid system for detection and prediction of faults in electrical transformers. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 324–335. [CrossRef]
99. Fei, S.; Zhang, X. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Syst. Appl.* **2009**, *36*, 11352–11357. [CrossRef]
100. Koley, C.; Purkait, P.; Chakravorti, S. Wavelet-Aided SVM Tool for Impulse Fault Identification in Transformers. *IEEE Trans. Power Deliv.* **2006**, *21*, 1283–1290. [CrossRef]
101. Yunlong, Z.; Peng, Z. Vibration Fault Diagnosis Method of Centrifugal Pump Based on EMD Complexity Feature and Least Square Support Vector Machine. *Energy Procedia* **2012**, *17*, 939–945. [CrossRef]
102. Sakthivel, N.R.; Sugumaran, V.; Nair, B.B. Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. *Mech. Syst. Signal Process.* **2010**, *24*, 1887–1906. [CrossRef]
103. Muralidharan, V.; Sugumaran, V.; Indira, V. Fault diagnosis of monoblock centrifugal pump using SVM. *Eng. Sci. Technol. Int. J.* **2014**, *17*, 152–157. [CrossRef]
104. Muralidharan, V.; Sugumaran, V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft Comput.* **2012**, *12*, 2023–2029. [CrossRef]
105. Muralidharan, V.; Sugumaran, V. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement* **2013**, *46*, 353–359. [CrossRef]
106. Jamil, M.; Singh, R.; Sharma, S.K. Fault identification in electrical power distribution system using combined discrete wavelet transform and fuzzy logic. *J. Electr. Syst. Inf. Technol.* **2015**, *2*, 257–267. [CrossRef]
107. Mortazavi, S.H.; Moravej, Z.; Shahrtash, S.M. A hybrid method for arcing faults detection in large distribution networks. *Int. J. Electr. Power Energy Syst.* **2018**, *94*, 141–150. [CrossRef]
108. Ramesh, B.N.; Jagan, M.B. Fault classification in power systems using EMD and SVM. *Ain Shams Eng. J.* **2017**, *8*, 103–111. [CrossRef]
109. Singh, S.; Vishwakarma, D.N. A Novel Methodology for Identifying Cross-Country Faults in Series-Compensated Double Circuit Transmission Lines. *Procedia Comput. Sci.* **2018**, *125*, 427–433. [CrossRef]
110. Da Silva, P.R.N.; Gabbar, H.A.; Vieira Junior, P.; da Costa Junior, C.T. A new methodology for multiple incipient fault diagnosis in transmission lines using QTA and Naïve Bayes classifier. *Int. J. Electr. Power Energy Syst.* **2018**, *103*, 326–346. [CrossRef]
111. Lin, S.; Horng, S. A Classification-Based Fault Detection and Isolation Scheme for the Ion Implanter. *IEEE Trans. Semicond. Manuf.* **2006**, *19*, 411–424. [CrossRef]
112. Di Maio, F.; Baraldi, P.; Zio, E.; Seraoui, R. Fault Detection in Nuclear Power Plants Components by a Combination of Statistical Methods. *IEEE Trans. Reliab.* **2013**, *62*, 833–845. [CrossRef]
113. Liangyu, M.; Yongguang, M.; Lee, K.Y. An Intelligent Power Plant Fault Diagnostics for Varying Degree of Severity and Loading Conditions. *IEEE Trans. Energy Convers.* **2010**, *25*, 546–554. [CrossRef]
114. Madhusudana, C.K.; Kumar, H.; Narendranath, S. Fault diagnosis of face milling tool using decision tree and sound signal. *Mater. Today Proc.* **2018**, *5*, 12035–12044. [CrossRef]
115. Du, Z.; Fan, B.; Jin, X.; Chi, J. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Build. Environ.* **2014**, *73*, 1–11. [CrossRef]
116. Camarena-Martinez, D.; Valtierra-Rodriguez, M.; Perez-Ramirez, C.A.; Amezquita-Sanchez, J.P.; Romero-Troncoso, R.J.; Garcia-Perez, A. Novel downsampling empirical mode decomposition approach for power quality analysis. *IEEE Trans. Ind. Electron.* **2015**, *63*, 2369–2378. [CrossRef]
117. Djeffal, A.; Babahenini, M.C.; Ahmed, A.T. Fast binary support vector machine learning method by samples reduction. *Int. J. Data Min. Model. Manag.* **2017**, *9*, 1. [CrossRef]

118. Zhao, Z.; Chu, L.; Tao, D.; Pei, J. Classification with label noise: A Markov chain sampling framework. *Data Min. Knowl. Discov.* **2018**, *33*, 1468–1504. [CrossRef]
119. Hwang, D.; Son, Y. Prototype-based classification and error analysis under bootstrapping strategy. *Int. J. Data Min. Model. Manag.* **2018**, *10*, 293. [CrossRef]
120. Zhang, W.; Zhang, Z.; Chao, H.C.; Tseng, F.H. Kernel mixture model for probability density estimation in Bayesian classifiers. *Data Min. Knowl. Discov.* **2018**, *32*, 675–707. [CrossRef]
121. Zhang, J.; Wang, S.; Chen, L.; Gallinari, P. Multiple Bayesian discriminant functions for high-dimensional massive data classification. *Data Min. Knowl. Discov.* **2016**, *31*, 465–501. [CrossRef]
122. Celotto, E. Visualizing the behavior and some symmetry properties of Bayesian confirmation measures. *Data Min. Knowl. Discov.* **2016**, *31*, 739–773. [CrossRef]
123. Becker, M.; Lemmerich, F.; Singer, P.; Strohmaier, M.; Hotho, A. MixedTrails: Bayesian hypothesis comparison on heterogeneous sequential data. *Data Min. Knowl. Discov.* **2017**, *31*, 1359–1390. [CrossRef]
124. Le, T.; Nguyen, K.; Nguyen, V.; Nguyen, T.D.; Phung, D. GoGP: Fast Online Regression with Gaussian Processes. In Proceedings of the IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017.
125. Marx, A.; Vreeken, J. Telling Cause from Effect Using MDL-Based Local and Global Regression. In Proceedings of the IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017.
126. Rudaś, K.; Jaroszewicz, S. Linear regression for uplift modeling. *Data Min. Knowl. Discov.* **2018**, *32*, 1275–1305. [CrossRef]
127. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
128. Teinemaa, I.; Dumas, M.; Leontjeva, A.; Maggi, F.M. Temporal stability in predictive process monitoring. *Data Min. Knowl. Discov.* **2018**, *32*, 1306–1338. [CrossRef]
129. Baldi, P. The inner and outer approaches to the design of recursive neural architectures. *Data Min. Knowl. Discov.* **2017**, *32*, 218–230. [CrossRef]
130. Zhang, B.; Li, W.; Li, X.L.; Ng, S.K. Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks. *IEEE Access* **2018**, *6*, 66367–66384. [CrossRef]
131. Bouguelia, M.R.; Nowaczyk, S.; Payberah, A.H. An adaptive algorithm for anomaly and novelty detection in evolving data streams. *Data Min. Knowl. Discov.* **2018**, *32*, 1597–1633. [CrossRef]
132. Xi, P.P.; Zhao, Y.P.; Wang, P.X.; Li, Z.Q.; Pan, Y.T.; Song, F.Q. Least squares support vector machine for class imbalance learning and their applications to fault detection of aircraft engine. *Aerosp. Sci. Technol.* **2018**, *84*, 56–74. [CrossRef]
133. Lu, F.; Wu, J.; Huang, J.; Qiu, X. Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm. *Aerosp. Sci. Technol.* **2018**, *84*, 661–671. [CrossRef]
134. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.F.M.; Liu, B.; Yu, P.S.; et al. Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
135. Zhao, H.; Wu, Q.; Hu, S.; Xu, H.; Rasmussen, C.N. Review of energy storage system for wind power integration support. *Appl. Energy* **2015**, *137*, 545–553. [CrossRef]
136. Saponara, S.; Saletti, R.; Mihet-Popa, L. Hybrid micro-grids exploiting renewables sources, battery energy storages, and bi-directional converters. *Appl. Sci.* **2019**, *9*, 4973. [CrossRef]
137. Zhou, B.; Yu, F.; Li, H.; Xin, W. A Quantitative Study on the Void Defects Evolving into Damage in Wind Turbine Blade Based on Internal Energy Storage. *Appl. Sci.* **2020**, *10*, 491. [CrossRef]
138. Liu, Z.F.; Li, L.L.; Tseng, M.L.; Tan, R.R.; Aviso, K.B. Improving the reliability of photovoltaic and wind power storage systems using least squares support vector machine optimized by improved chicken swarm algorithm. *Appl. Sci.* **2019**, *9*, 3788. [CrossRef]

Statistical Features and Data Mining Techniques for Detection of Short-Circuit Faults in Transformers

Arantxa Contreras-Valdes, Jose R. Huerta-Rosales, Juan P. Amezquita-Sanchez, Luis Morales-Velazquez, Martin Valtierra-Rodriguez*
Universidad Autónoma de Querétaro, Facultad de Ingeniería, ENAP-RG-CA Sistemas dinámicos
Campus San Juan del Río, Querétaro, México.
* martin.valtierra@enap-rg.org

***Abstract*—Accurate diagnosis of electrical systems still involves a great challenge for developers and researchers from different areas. The use of data mining techniques emerges as an outstanding option for this purpose. In this work, a new methodology to detect short-circuit faults in transformers from current signals is presented. Firstly, statistical parameters are extracted from both the transient state and the steady state of the current signals of a single-phase transformer of 1.5 kVA with no load, where three different conditions, i.e., healthy, 10 short-circuited turns (SCTs), and 20 SCTs, are analyzed. For automatic diagnosis, three common techniques for data mining, i.e., decision trees, neural networks, and support vector machines, are applied and their performance is compared. Results show that the proposal can achieve a classification effectiveness of 96.6%.**

Keywords—decision trees, fault diagnosis, neural networks, support vector machines, short-circuit currents, transformers

I. INTRODUCTION

The reliability and safety of the electric systems are aspects required in all the sectors, e.g., energetic, industrial, and automotive, among others. In particular, the transformer due to its essential function within the electrical network has motivated the development and application of new fault detection methods [1], which provide multiple benefits such as the detection of incipient problems (those that are in their initial phase and whose progressive development can result in more significant faults), reduction of unplanned events, increase of both equipment lifetime and system performance, among others. Although several techniques and methodologies have been presented in literature for this task, increasing requirements to achieve a more efficient and reliable analysis in terms of performance and processing still demand a further research.

The predictive techniques of data mining represent an important alternative to this need. Three of the techniques most used by researchers for the detection and diagnosis of faults in different electrical systems are decision trees (DT), neural networks (NN), and support vector machines (SVM). Having as a study case a transformer, in [1] a DT is used to discriminate the internal faults of a power transformer from the input currents. In [2], a DT is also used for the diagnosis of incipient faults in transformers but the analysis is carried out using the concentrations in parts per million of the combustible gases present in the samples of oil. Based on the analysis of dissolved gases, the algorithm of C4.5 for DT is used to solve the internal diagnosis of thermoelectric potential faults [3]. In [4], this algorithm is also used to determine the good and defective conditions of a monoblock centrifugal pump. Regarding the NN-based methodologies, the online diagnosis of stator faults in induction motors are presented [5-6]. On the other hand, in [7] they propose a neural-fuzzy scheme along with a statistical approach for the diagnosis of incipient faults in power transformers. The diagnosis by assigning a neuronal module for each type of component present in an electrical power system, i.e., a transformer, a bus, or a transmission line, is presented in [8]. The use of SVM to diagnose faults in power transformers from dissolved gas analysis is presented in [9-11]. Other variables considered to train SVM and detect several faults in different systems are the signals of vibration and current [12]. In [13], the results obtained for the location of faults in transmission lines using SVM and current signals show an average error of 0.03% while the detection and classification modules have an effectiveness of 100%.

In general, the monitoring of current signals in the diagnosis of electrical systems has been an attractive option since it offers the possibility of remote monitoring and low cost of implementation, and allows the on-line monitoring and extensive fault coverage. In the case of transformers, the current signal can be divided into two parts [14-16], the first one corresponding to the transient part of the signal and the second one corresponding to the stable state. The study of both parts offers somehow a more complete analysis in a fault diagnosis context. In order to extract the information provided by these types of signals, the statistical features of low order and superior order have been used in different works [17-22]. In specific, the mean [17], variance [18], standard deviation [19], kurtosis [20], skewness [21], and RMS (root mean square) [22], can allow the obtaining of relevant aspects to be interpreted by means of data mining techniques, satisfying the criteria of representativeness and reliability regarding the condition of the transformer.

Based on the results published in the abovementioned literature and taking into account the advantages of working with current signals and statistical features, the contribution of this work is a new methodology to detect the short-circuit fault using the current of a single phase transformer. This methodology is based on the extraction of statistical parameters, i.e., mean, variance, standard deviation, kurtosis, skewness, and RMS, and the use of an automatic detection algorithm. For the

This work was supported by the "Consejo Nacional de Ciencia y Tecnología (CONACYT)" under project number 892305.

latter, the performance of DT, NN, and SVM, is compared by evaluating the percentage of accuracy of each of the suggested algorithms.

II. Mathematical Background

The definitions and mathematical concepts used in this article are briefly described in this section.

A. *Statistical features*

The calculation of statistical parameters of a signal can provide the necessary information to represent it; if the signals are different, their statistical parameters will change and allow their distinction in a classification context. For this study, parameters of first order, second order, and superior order are considered. They are summarized in Table I.

As each parameter can contribute in a different way in the classification process, it is necessary to consider them all. Also, as different data mining techniques can lead to different results, three different methods, i.e., J48 algorithm for DT, multilayer perceptron, and SVM, are used to process the above-mentioned parameters in order to compare their performance and obtain the best performance.

B. *Data mining algorithms*

The methods used for the classification of future/unknown events constitute an important knowledge structure. The role that data mining algorithms play is to determine and use the characteristics of each data set to perform a specific task. The J48 algorithm for DT, the multilayer perceptron, and the SVM are considered in this study as they are highly used in literature.

- Decision tree

The J48 algorithm of DT consists of two phases: building and pruning [20]. For the building, the set of characteristics that are treated as inputs to the algorithm and the corresponding outputs are presented to the algorithm. The resulting DT consists of a root node, a number of leaves that indicate the class labels, a number of nodes related to the classes and a number of branches that display each predictive value of the final node. In each decision node in the tree, the most useful characteristic is chosen. However, the need of a maximum precision in the training data through this approach may result in a classifier with excessive rules that are highly selective to the training data. Pruning is used to correct the overfitting of experimental data, produce fewer and simpler results, and reduce computational time.

- Neural network

The architecture of an NN is characterized by having layers with individual or multiple neurons in each layer. The mathematical function given in (1) which describes each neuron consists of the sum of the multiplications between the inputs and the associated multipliers commonly called weights for each input plus a bias; then, this result is evaluated with a non-linear function to provide the NN with the ability to model non-linear relationships. To find the weights of the network, pairs of input-output data are presented; then, a training rule is used to adjust these weights. Thus, the error between the desired and calculated outputs is minimized. Finally, the training sets are repeatedly presented to the NN until the general error is reasonable.

$$y = f(\textstyle\sum_{i=I}^{I} w_i x_i + b) \tag{1}$$

Where y is the output, $f(\bullet)$ the activation function, I the total number of inputs, w_i are the weights, x_i the inputs and b is the bias [21].

A multilayer perceptron classifier is used to classify instances through backpropagation. This network can be monitored and modified during training time. The nodes in this network are all sigmoid (except for when the class is numeric and the output nodes become unthresholded linear units).

- Support vector machine

Given a set of points (subset of a space), in which each of them belongs to one of two possible categories, an algorithm based on SVM constructs a model capable of predicting whether a new point (whose category is unknown) belongs to one category or to another.

TABLE I. Statistical Features

Reference	Statistical feature	Order	Equation	Description
[17]	Mean	1	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	The mean is equal to the sum of all the numbers in a given set and then divided by the number of values in the set.
[18]	Standard deviation	2	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$	It is a measure that is used to quantify the variation or dispersion of a set of numerical data.
[19]	Variance	2	σ^2	The variance is defined as the squared deviation standard of a variable/number with respect to the mean of a data set.
[20]	Kurtosis	Superior	$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$	It is a characteristic of the frequency distribution of a specific data set.
[21]	Skewness	Superior	$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$	It allows to establish the degree of symmetry that has a probability distribution of a random variable.
[22]	RMS	Superior	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$	RMS is the square root of the arithmetic mean of the squares of a data set.

The SVM looks for a hyperplane that has the maximum distance (margin) with the points that are closest to it in such a way that it separates optimally the points of one class from the other, which could have been previously projected to a space of superior dimensionality. In this way, the points of the vector that are labeled with a category will be on one side of the hyperplane and the cases that are in the other category will be on the other side.

The hyperplane, which successfully separates the points according to their classes, can be given by [22]:

$$w^T x_i + b = 0 \tag{2}$$

where w denotes the weight vector, x_i is ith real valued n-dimensional input vector and b is the term bias. The separation margin (m) is given by:

$$m = \frac{2}{\|w\|} \tag{3}$$

The choice of the most appropriate representation of the studied universe is done through a process called feature selection, in this study the John Platt's sequential minimal optimization algorithm for training the support vector classifier is used [23]. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default.

III. Experimentation

The development of the experiment of this work is guided by the flowchart presented in Fig. 1. The first step consists of the acquisition of the current signal data for the transformer operating at different conditions. In order to provide a fault detection method during both the transient state and the steady state, the acquired signals are divided into two parts. Then, the six parameters presented in Table I are computed for each part of the signal using Matlab software. Finally, the three data mining algorithms described in section II are applied using the WEKA (Waikato Environment for Knowledge Analysis) software in order to offer an automatic diagnosis.

The experimental setup used to develop and test the proposed methodology is shown in Fig. 2. For the acquisition of the current signals, a current clamp Fluke i200s and a National Instruments NI-USB 6211 board are used. The signals have a duration of 6s with a sampling frequency of 6000 samples/s. The single-phase transformer under test has 135 turns and a power of 1.5 kVA, operating at 120V in three different conditions: healthy, 10 SCTs (F1) and 20 SCTs (F2). For each condition, 20 tests/acquisitions are performed. The fault conditions are simulated using the taps generated synthetically into the transformer. All the processing is carried out using a personal computer (PC). An autotransformer is used to de-energize the transformer after each test.

Once the current signal is acquired, it is segmented into two stages: the transient state from 1s to 1.5 s and the steady state from 1.5s to 6s as shown in Fig. 3. Then, the six statistical features addressed in the Mathematical background section are calculated for each part of the signal using Matlab software. It is

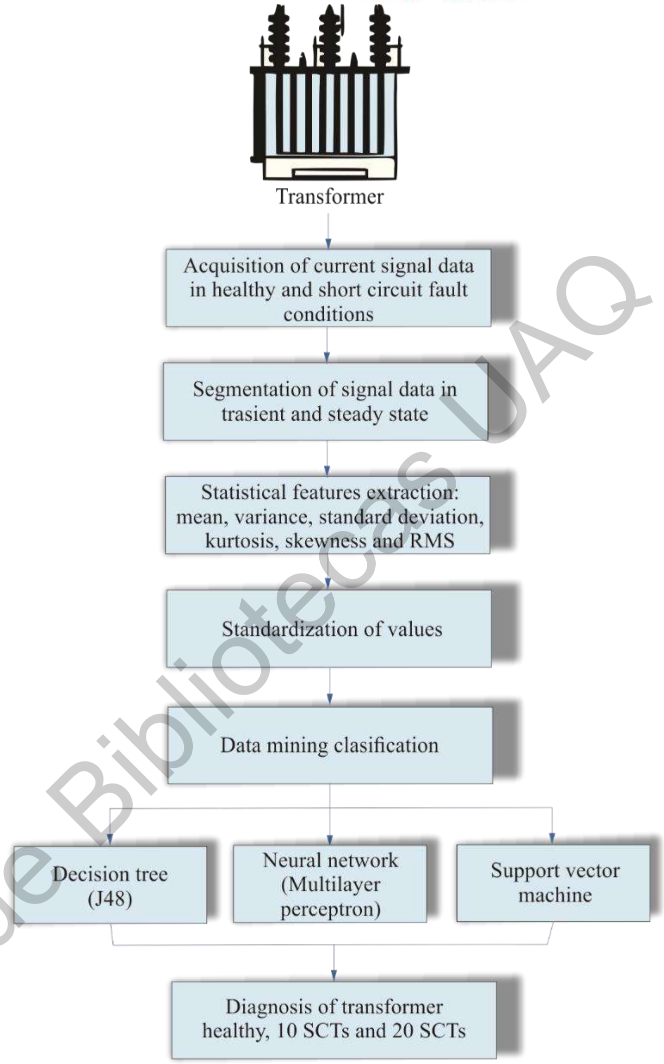


Fig. 1. Proposed methodology flowchart.

important to mention that the obtained values are standardized to have a mean of 0 and a standard deviation of 1 and, therefore, a common basis. Then, each algorithm (J48 algorithm, multilayer perceptron, and SVM) is applied using the WEKA (Waikato Environment for Knowledge Analysis) data mining software, version 3.6.15. The obtained results are exposed and discussed in the next section.

IV. Results

The results of the statistical features and the classification accuracy obtained from each algorithm are presented for each condition of the current signal. All the experiments are carried out under the test option of cross-validation considering 10 folds.

A. Transient state

Table II shows the mean of the six indicators calculated for the inrush current (transient stage) by considering the 20 tests. The following subsections show the results obtained for the data mining algorithms treated in this work.

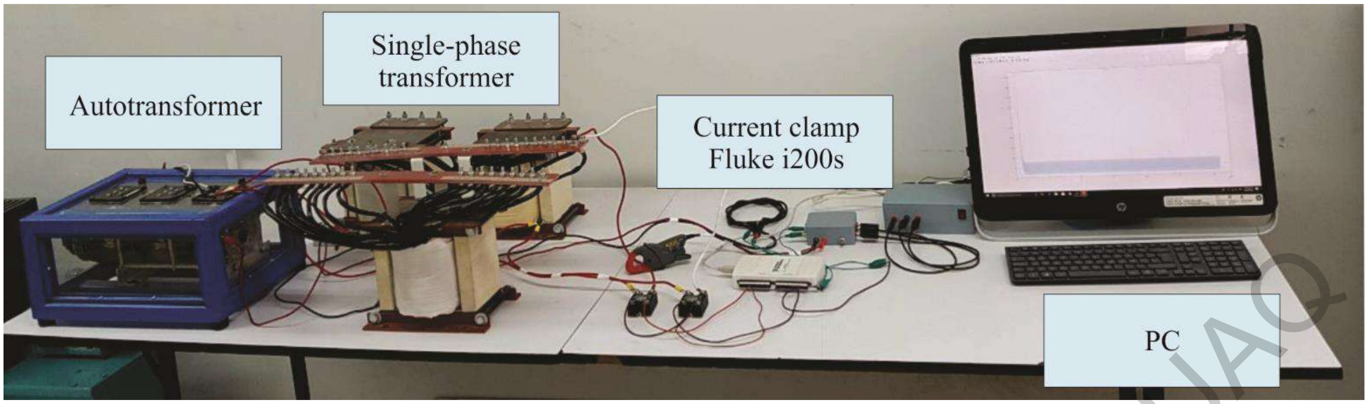


Fig. 2. Experimental setup.

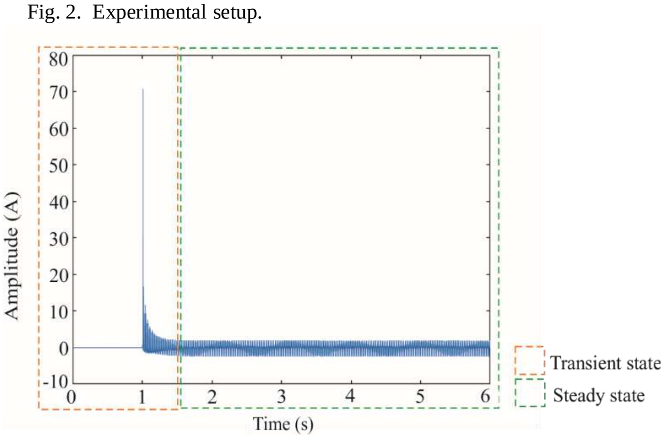


Fig. 3. Graphic representation of the signal segmentation in transient state and steady state.

- Decision tree

The effectiveness obtained after the application of the J48 algorithm is 90%. The confusion matrix that arises when applying said algorithm is shown in Table III. The most effective DT is shown in Fig. 4. Based on the calculation of the variance, the DT consists of a main node from which two branches arise, classifying first the F2 (variance greater than 25,425) from the other two conditions. Then, it is divided into other two branches, classifying as Healthy a variance less than or equal to 0.90726 and as F1 a variance greater than 0.90726.

TABLE II. MEAN FOR STATISTICAL FEATURES IN TRANSIENT STATE

Statistical feature	Mean Healthy	Mean 10 SCTs	Mean 20 SCTs
Mean	-0.0375	-0.0876	2.2307
Variance	-0.0827	20.8864	70.2881
Standard deviation	-0.0820	13.1742	30.074
Skewness	-0.0719	3.0959	0.9660
Kurtosis	-0.0719	2.8778	0.8485
RMS	-0.0823	13.1914	30.1262

TABLE III. J48 ALGORITHM: CONFUSION MATRIX FOR TRANSIENT STATE

a	b	c	Classification
17	2	1	a = Healthy
1	17	2	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

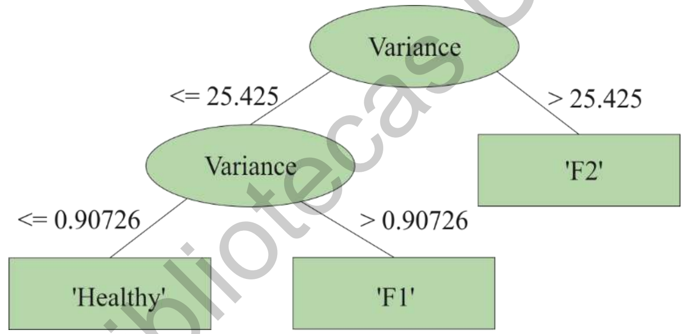


Fig. 4. Decision tree obtained for transient state.

- Neural network

The confusion matrix depicted in Table IV, showing a 95% of effectiveness, is the result of the implementation of the multilayer perceptron. The sigmoid activation function is used. The architecture of the neural network consists of 7 sigmoid nodes including the 3 input nodes corresponding to the classes 'Healthy, F1, and F2'.

TABLE IV. MULTILAYER PERCEPTRON: CONFUSION MATRIX FOR TRANSIENT STATE

a	b	c	Classification
18	1	1	a = Healthy
1	19	0	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

- Support vector machine

The percentage of effectiveness reached by the SVM is 95% and the corresponding confusion matrix can be consulted in Table V. The linear kernel is used. For each classifier the number of kernel evaluations is calculated, resulting in the first classifier 'Healthy, F1' a total of 154 evaluations, for 'Healthy, F2' 261 evaluations, and for the 'F1, F2' 265 evaluations.

TABLE V. SVM: CONFUSION MATRIX FOR TRANSIENT STATE

a	b	c	Classification
18	1	1	a = Healthy
1	19	0	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

B. *Steady state*

Table VI shows the mean of the six indicators calculated for the steady state current by considering the 20 tests. The following subsections show the results obtained for the data mining algorithms treated in this work.

TABLE VI. MEAN FOR STATISTICAL FEATURES IN STEADY STATE

Statistical feature	Mean Healthy	Mean 10 SCTs	Mean 20 SCTs
Mean	0.0259	-4.6530	-4.6489
Variance	0.0026	58.5697	430.3674
Standard deviation	0.0011	30.1556	103.9022
Skewness	-0.0377	7.1180	8.8029
Kurtosis	0.0065	-14.6765	-29.0954
RMS	0.0011	30.1556	103.9022

- Decision tree

The most effective DT, based on the calculation of the variance, is shown in Fig. 5. It consists of a main node from which two branches arise, classifying first the healthy state (variance less than or equal to 1.9541). Then, the fault conditions, F1 and F2, are divided into the following two branches considering F1 as a variance less than or equal to 75.803 and as F2 a variance greater than 75.803, obtaining a total of four branches. The effectiveness obtained after the application of the J48 algorithm is 93.3%. The confusion matrix that arises when applying this algorithm is observed in Table VII.

TABLE VII. J48 ALGORITHM: CONFUSION MATRIX FOR STEADY STATE

a	b	c	Classification
18	1	1	a = Healthy
1	18	1	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

- Neural network

An effectiveness percentage of 96.6% and the confusion matrix shown in Table VIII are the results of the implementation of the multilayer perceptron. The sigmoid activation function is used. The architecture of the neural network consists of 7 sigmoid nodes including the 3 input node corresponding to the classes 'Healthy, F1, and F2'.

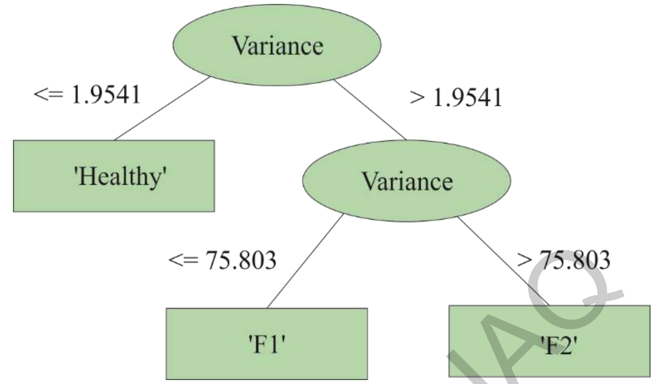


Fig. 5. Decision tree obtained for steady state.

TABLE VIII. MULTILAYER PERCEPTRON: CONFUSION MATRIX FOR STEADY STATE

a	b	c	Classification
19	1	0	a = Healthy
1	19	0	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

- Support vector machine

The percentage of effectiveness reached by the SVM for the steady state current is 96.6% and the corresponding confusion matrix can be consulted in Table IX. The linear kernel is used. The number of kernel evaluations for the classifier 'Healthy, F1' is 238 evaluations, for the second classifier 'Healthy, F2' is 79 evaluations, and for the last classifier 'F1, F2' is 206 evaluations.

TABLE IX. SVM: CONFUSION MATRIX FOR STEADY STATE

a	b	c	Classification
19	1	0	a = Healthy
1	19	0	b = 10 short-circuited turns
0	0	20	c = 20 short-circuited turns

Table X presents the summary of the percentages of effectiveness of each algorithm tested in this study. It is noted that the J48 algorithm reflects an effectiveness below 95% contrary to the other two methods that show better results.

The results of Table X show that the multilayer perceptron and SVM offer the best results; however, it is important to mention that the DT implies a lower computational load in its implementation. From the results of Figs. 5 and 4, it is also observed that the trees resulting from each processing use only the variance for the diagnosis of each condition, simplifying even more the computational load. Therefore, they can be a more suitable solution if their performance is improved though the addition of new features.

TABLE X. SUMMARY OF CLASSIFICATION EFECTIVENESS FOR BOTH STATES

State	Data Mining Algorithm		
	DT: J48 algorithm	*Multilayer perceptron*	*SVM*
Transient	90%	95%	95%
Steady	93.3%	96.6%	96.6%

V. CONCLUSIONS

In this work, a new methodology based on statistical parameters and a data mining algorithm for detection of short-circuited faults in transformers is presented. Results demonstrate that the proposal can diagnose the following conditions: Healthy, 10 SCTs, and 20 SCTs. Statistical features such as mean, variance, standard deviation, kurtosis, skewness, and RMS are used as features for each data mining algorithm. In the NN and SVM, all the statistical features are used; on the contrary, the variance values are sufficient to classify each condition in the DTs. From the results of both states, it is concluded that the J48 algorithm shows the lowest percentage of effectiveness (90% for the transient state and 93.3% for the steady state); on the other hand, both multilayer perceptron and SVM present the same percentages of effectiveness for transient state (95%) and for steady state (96.6%). The similarity in the results lies in the fact that the models based on SVMs are closely related to NN. The advantage presented by the SVM compared to the NN is the linear kernel used to calculate the weights, which implies less computational resources and, therefore, shorter processing time. As a future work, other levels of fault severity and other transformer operating conditions, e.g., the transformer with load, will be analyzed.

REFERENCES

[1] S. R. Samantaray and P. K. Dash, “Decision Tree based discrimination between inrush currents and internal faults in power transformer,” International Journal of Electrical Power & Energy Systems, vol. 33, no. 4, pp. 1043–1048, May 2011.

[2] A. G. C. Menezes, O. M. Almeida and F. R. Barbosa. “Use of decision tree algorithms to diagnose incipient faults in power transformers,” Simposio Brasileiro de Sistemas Eletricos (SBSE), 2018.

[3] Y. Han, D. Zhao and H. Hou, “Oil-immersed Transformer Internal Thermoelectric Potential Fault Diagnosis Based on Decision-tree of KNIME Platform,” Procedia Computer Science, vol. 83, pp. 1321–1326, 2016.

[4] N. R. Sakthivel, V. Sugumaran and S. Babudevasenapati, “Vibration based fault diagnosis of monoblock centrifugal pump using decision tree,” Expert Systems with Applications, vol. 37, pp. 4040–4049, June 2010.

[5] J. F. Martins, V. F. Pires and A. J. Pires “Unsupervised Neural-Network-Based Algorithm for an On-Line Diagnosis of Three-Phase Induction Motor Stator Fault,” IEEE Transactions on Industrial Electronics, vol. 54, no. 1, pp. 259–264, Feb 2007.

[6] R. M. Tallam, T. G. Habetler, R. G. Harley, D. J. Gritter and B. H. Burton, “Neural network based on-line stator winding turn fault detection for induction motors,” Conference Record of the 2000 IEEE Industry Applications Conference. Thirty-Fifth IAS Annual Meeting and World Conference on Industrial Applications of Electrical Energy (Cat. No. 00CH37129).

[7] G. Rigatos and P. Siano, “Power transformers’ condition monitoring using neural modeling and the local statistical approach to fault diagnosis,” International Journal of Electrical Power & Energy Systems, vol. 80, pp. 150–159, Sep 2016.

[8] A. Flores, E. Quiles, E. Garcia and F. Morant, “Fault Diagnosis of Electric Transmission Lines using Modular Neural Networks,” IEEE Latin America Transactions, vol. 14, no. 8, pp. 3663–3668, Aug. 2016.

[9] K. Bacha, S. Souahlia and M. Gossa, “Power transformer fault diagnosis based on dissolved gas analysis by support vector machine,” Electric Power Systems Research, vol. 83, no. 1, pp. 73–79, Feb 2012.

[10] G. Lv, H. Cheng, H. Zhai and L. Dong, “Fault diagnosis of power transformer based on multi-layer SVM classifier,” Electric Power Systems Research, vol. 74, no. 1, pp. 1–7, April 2005.

[11] M. Mittal, M. Bhushan, S. Patil and S. Chaudhari, “Optimal Feature Selection for SVM Based Fault Diagnosis in Power Transformers,” IFAC Proceedings Volumes, vol. 46, no. 32, pp. 809–814, Dec. 2013.

[12] P. Gangsar and R. Tiwari, “Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms,” Mechanical Systems and Signal Processing, vol. 94, no. 15, pp. 464–481, Sep. 2017.

[13] J. Johnson and A. Yadav, “Complete protection scheme for fault detection, classification and location estimation in HVDC transmission lines using support vector machines,” IET Science, Measurement & Technology, vol. 11, no. 3, pp. 279–287, Jan. 2017.

[14] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J.C. Olivares-Galvan and R. Escarela-Perez, “The application of EMD-based methods for diagnosis of winding faults in a transformer using transient and steady state currents,” Measurement, vol. 117, pp. 371-379, March 2018.

[15] G. Alvarez-Monroy, A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J.C. Olivares-Galvan and R. Escarela-Perez, “A new EMD-Shannon entropy-based methodology for detection of inter-turn faults in transformers,” 2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC).

[16] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J.C. Olivares-Galvan and R. Escarela-Perez, “Experimental data-based transient-stationary current model for inter-turn fault diagnostics in a transformer,” Electric Power Systems Research, vol. 152, pp. 306–315, Nov. 2017.

[17] A. Sharma, V. Sugumaran and S. Babu Devasenapati, “Misfire detection in an IC engine using vibration signal and decision tree algorithms,” Measurement, vol. 50, pp. 370–380, April 2014.

[18] V. Muralidharan, V. Sugumaran and V. Indira, “Fault diagnosis of monoblock centrifugal pump using SVM,” Engineering Science and Technology, an International Journal, vol. 17, no. 3, pp. 152–157, Sep. 2014.

[19] N. R. Sakthivel, V. Sugumaran and B. B. Nair, “Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump,” Mechanical Systems and Signal Processing, vol. 24, no. 6, pp. 1887–1906, Aug. 2010.

[20] M. Seera and Chee Peng Lim, “Online Motor Fault Detection and Diagnosis Using a Hybrid FMM-CART Model,” IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 4, pp. 806–812, April 2014.

[21] S. Das, P. Purkait, C. Koley and S. Chakravorti, “Performance of a load-immune classifier for robust identification of minor faults in induction motor stator winding,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 21, no. 1, pp. 33–44, Feb. 2014.

[22] V. T. Tran, B. S. Yang, M. S. Oh and A. C. C. Tan, “Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference,” Expert Systems with Applications, vol. 36, no. 2, pp. 1840–1849, March 2009.

[23] V. Muralidharan and V. Sugumaran, “Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump,” Measurement, vol. 46, no. 1, pp. 353–359, Jan. 2013.

[24] J. L. Gonzalez-Cordoba, A. Mejia-Barron and M. Valtierra-Rodriguez, “Automatic detection and classification of electrical disturbances by means of empirical mode decomposition method,” 2015 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC).

[25] U. B. Parikh, B. Das and R. Maheshwari, “Fault classification technique for series compensated transmission line using support vector machine,” International Journal of Electrical Power & Energy Systems, vol. 32, no. 6, pp. 629–636, July 2010.

[26] J. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, 1998.

2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC 2019). Ixtapa, Mexico

Data compression based on discrete Wavelet transform and fault detection of short-circuit faults in transformers

Arantxa Contreras-Valdes[1], Juan P. Amezquita-Sanchez[1], David Granados-Lieberman[2], Martin Valtierra-Rodriguez[1,*]
[1]Universidad Autónoma de Querétaro, Facultad de Ingeniería, ENAP-RG-CA Sistemas dinámicos
[2] Instituto Tecnológico Superior de Irapuato, ENAP-RG, Departamento de Ingeniería Electromecánica
* martin.valtierra@enap-rg.org

Abstract— **The large amount of data generated from the monitoring of electrical transformers generates the need to compress such information in order to store large volumes of data and transmit them at high speeds. The methodology proposed in this document focuses on the implementation of a data compression algorithm based on the Discrete Wavelet Transform. The algorithm is applied to a database of current signals demonstrating its suitability according to the high compression ratio achieved for transmission and storage purposes. The storage is carried out in hierarchical data format. Its high effectiveness is confirmed by demonstrating that the compressed files reduce their size considerably, presenting a performance superior to 78 in terms of compression ratio. In addition, it is demonstrated that the decompressed signal can be used to detect short circuit faults in a single phase transformer, obtaining percentages of classification effectiveness above 96.6%.**

Keywords—data compression, discrete wavelet transform, short-circuit currents, transformers, electrical fault detection.

I. INTRODUCTION

In the energy system, the transformer plays a fundamental role since it is considered as one of the most important elements [1], its constant monitoring implies having a high speed of data transmission, as well as a good storage capacity and a timely diagnosis of failures. The above considerations represent two of the main objectives of a smart grid since it seeks to increase the efficiency and reliability of the electrical system. According to these needs, the ability to compress the amount of data for further studies such as monitoring and fault detection is paramount.

There is an extensive amount of research on transformer failure detection [2-5], however, the data compression has not been considered from the point of view treated in this work. There are other works that specialize in studying different data compression algorithms such as auto regression compression method [6], wavelet analysis method [7] and delta modulation technique [8], which are applied in various areas of science ranging from medicine [9-11] to engineering [6-7], [12]. This is closely related to the increase of devices for the acquisition and processing of data that entails a heavy burden for transmission and storage derived from the generation of an increasing amount of data [8]. Recently, in [13], ECG (electrocardiogram) signals are used obtaining remarkable compression results; in this regard and from a condition monitoring viewpoint, it would be desirable to explore this algorithm to compress current signals of a transformer for storage and transmission issues and achieve the detection of faults on it.

The proposed methodology considers the identification of short-circuit faults of a single-phase transformer at different levels of severity. This methodology makes use of the algorithm of compression and decompression of data reported in [13]. The main idea of this algorithm is to extract the largest coefficients applying the fast Cohen-Daubechies-Feauveau 9/7 Discrete Wavelet Transform to the signal. In addition, it is possible to add a Huffman coding process to the algorithm which promises to further improve compression but at the expense of processing time. The compression procedure has the ability to encode a current signal into a file significantly smaller than that contained in the uncompressed record. This effectiveness is achieved directly by saving in the Hierarchical Data Format (HDF) [14]. The technique used is cataloged as a type of compression with loss, this refers to allowing a certain distortion which does not harm the information of the data. In this work, this feature works as a filter that cleans the noise of the signal, favoring the task of failure classification. In this regard, the signals that have gone through the compression and decompression processes are used to achieve an automatic diagnosis, where the performance of two classifiers, a neural network and a support vector machine, is compared in order to identify different short circuit fault conditions.

II. THEORETICAL BACKGROUND

In this section, the algorithm for compression and decompression of data using a DWT is introduced, as well as the measures to evaluate the results of said procedure. Also, the methods used for classification purposes are described.

A. Data compression and decompression algorithm

The compression algorithm consists of three essential steps which are described below [13].

1) Approximation

The Cohen-Daubechies-Feauveau 9/7 DWT is applied to convert the signal f into a vector w whose components are the wavelet coefficients $(c(1), c(2), ..., c(N))$ where N is the total number of components of w.

This work was supported by the Mexican Council of Science and Technology (CONACyT) under project number 892305.

978-1-7281-2898-6/19/$31.00 ©2019 IEEE

2) Quantization

The wavelet coefficients are transformed into integers by a mid-tread uniform quantizer in the following way:

$$c^{\Delta}(i) = \left\lfloor \frac{c(i)}{\Delta} + \frac{1}{2} \right\rfloor, i = 1, 2, \dots, N \tag{1}$$

where c represents the wavelet coefficients, Δ is the quantization parameter, $\lfloor x \rfloor$ indicates the largest integer number. After quantification, the coefficients and indices are further reduced by the elimination of the coefficients equal to 0. The signs of the coefficients are encoded separately in an array *(s(1), s(2), … , s(K))* by using a binary alphabet, 1 for the + sign and 0 for the - sign.

3) Organization and Storage

In this step the indices are re-ordered in ascending order ($\tilde{I}$). Reordered indexes are stored as smaller positive numbers, taking differences between two consecutive values and producing a single recovery.

At this point it is possible to add an additional step, which is the Huffman coding [14] before saving the arrays in order to further improve the compression as proposed in [13].

Finally, the size of the K signal, the quantization parameter Δ and the arrays $\tilde{I}$, $\tilde{c}^{\Delta}$ and $\tilde{s}$ are saved in HDF. HDF operates through a fragmented storage mechanism. The data matrix is divided into fragments of equal size, each of which is stored separately in the file.

In order to recover the compressed signal, a decoding stage is necessary. The steps are described as follows:

1) The compressed file is read and the data contained in it is extracted (size of signal K, quantization parameter Δ and matrices $\tilde{I}$, $\tilde{c}^{\Delta}$ and $\tilde{s}$)

2) Indexes are retrieved from the array $\tilde{I}$

3) The signs of the wavelet coefficients are recovered as

$$\tilde{s}^{\mathrm{r}} = 2\tilde{s} - 1 \tag{2}$$

4) The full array of wavelet coefficients is completed as

$$w^{r}(\tilde{l}) = \tilde{s}^{r} \cdot \tilde{c}^{r} \tag{3}$$

where $\tilde{c}^{\mathrm{r}}$ represents the magnitude of the coefficients of its quantized version, which is obtained as $\tilde{c}^{\mathrm{r}} = \Delta\tilde{c}^{\Delta}$

5) The wavelet transform is inverted to recover the approximate signal f^{r}.

Fig. 1 describes the compression and decompression procedure in a diagram form.

Different measures to evaluate the results of the procedure should be considered since the algorithm is of the lossy compression type. Compression performance is evaluated by the Compression Ratio (CR) given by

$$CR = \frac{\textit{Size in bytes of the original file}}{\textit{Size in bytes of the compressed file}} \tag{4}$$

The compression performance in relation to the quality of the recovered signals is evaluated with respect to the Percentage Root-mean-square Difference (PRD), which is calculated as follows:

$$PRD = \frac{\|f - f^{r}\|}{\|f\|} x100\% \tag{5}$$

where f is the original signal, f^{r} is the signal reconstructed from the compressed file and $\| x \|$ indicates the 2-Norm, which is the largest singular value.

The Quality Score (QS), which reflects the compromise between compression performance and reconstruction quality, is quantified as:

$$QS = \frac{CR}{PRD} \tag{6}$$

B. Classification algorithms

- Neural network

In general, a neural network (NN) is characterized by having a layered architecture which can be formed by individual or multiple neurons in each layer. Each neuron is described mathematically as the sum of the multiplications between the inputs and the associated multipliers for each input called regularly as weights plus a bias. The result of the summation is evaluated with a non-linear function to provide the NN with the ability to model non-linear relationships. The weights of the network are calculated using a training rule to adjust the input-output data pairs and minimize the error between the desired and calculated outputs by presenting the training sets repeatedly to the NN until the general error is admissible [15]. The algorithm used in this work is a multilayer perceptron that is a neural network that learns with backward propagation to classify instances. The nodes of this network are all sigmoidal functions.

- Support vector machine

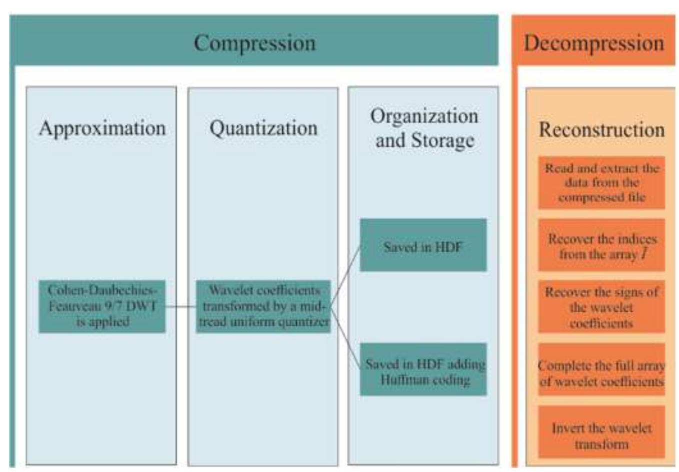


Fig. 1. Steps involved in the compression and decompression algorithm.

Support vector machine (SVM) is one of the algorithms mostly used in the classification of patterns due to its proven efficacy [16]. The SVM used in this paper implements the sequential minimal optimization algorithm to train a classifier of support vectors, using core functions such as polynomial or Gaussian cores. The missing values are replaced globally, the nominal attributes are transformed into binary values and the attributes are normalized by default, that is, the coefficients in the output are based on the normalized data, not on the original data. This is important to interpret the classifier. In the case of this work where a multiclass problem is treated, the predicted probabilities will be combined using the paired classification [17].

III. Methodology

In general, the methodology consists of acquiring current signals from a single-phase transformer in different conditions. Then, these signals are compressed by means of a lossy compression algorithm. Since the best results corresponded to the Cohen-Daubechies-Feauveau 9/7 family according to [13], in this approach DWT is applied at level 3 and 4. Two levels are analyzed because their study could offer the advantage of detecting the short-circuit failure in a better way since the frequency information associated with the fault could be better isolated at a certain level. The stage of monitoring the condition is composed of the decompression of the signal, where it is divided into two parts: transient and steady state. This segmentation allows a better analysis in the detection of the fault. Then, the mean, variance, standard deviation, kurtosis, skewness, and RMS are extracted for each part of the signal. Finally, two data mining classification algorithms are applied to detect the type of fault using the decompressed signals. The algorithms used are a neural network, specifically a multilayer perceptron (MP) and a Support Vector Machine (SVM), which are applied using the 3.6.15 version of the WEKA (Waikato Environment for Knowledge Analysis) data mining software. Finally, the classification effectiveness percentages are compared against the results obtained without applying the compression and decompression data algorithms. With the exception of the two classification algorithms, the implementation of the rest of the methodology is done in the MATLAB software; therefore, HDF storage is achieved by simply using the save function to store the data. The sequence of the steps of the methodology described above can be found in the diagram of Fig. 2.

IV. Experimentation and Test Results

This section presents the experimental setup, the results for the compression and decompression algorithm where two comparative tests are presented, and the results obtained of fault classification algorithms.

A. Experimental setup

The database used to test the compression algorithm is made up of current signals from a single-phase transformer operating at 120V. It has 135 turns and a power of 1.5 kVA. The signals are acquired with a current clamp Fluke i200s and the National Instruments NI-USB 6211 board in three different conditions: healthy, 10 short-circuited turns (SCTs) and 20 SCTs. The fault conditions are simulated using the taps generated synthetically into the transformer. An autotransformer is used to de-energize the transformer after each test. All the processing is carried out using a personal computer (PC). In Fig. 3, the experimental setup described above is shown. The current signals that are used have a sampling frequency of 6000 samples / s acquired in a period of time of 6s. 20 tests / acquisitions are carried out for each conditions obtaining a total of 36000 samples per test. Each signal is segmented in transient state from 1s to 1.5 s and steady state from 1.5s to 6s as shown in Fig. 4.

B. Compression and decompression algorithm performance

Two comparative tests are presented. They have in common that the DWT is decomposed in two different levels (3 and 4). Time and CR of average compression are evaluated for the 20

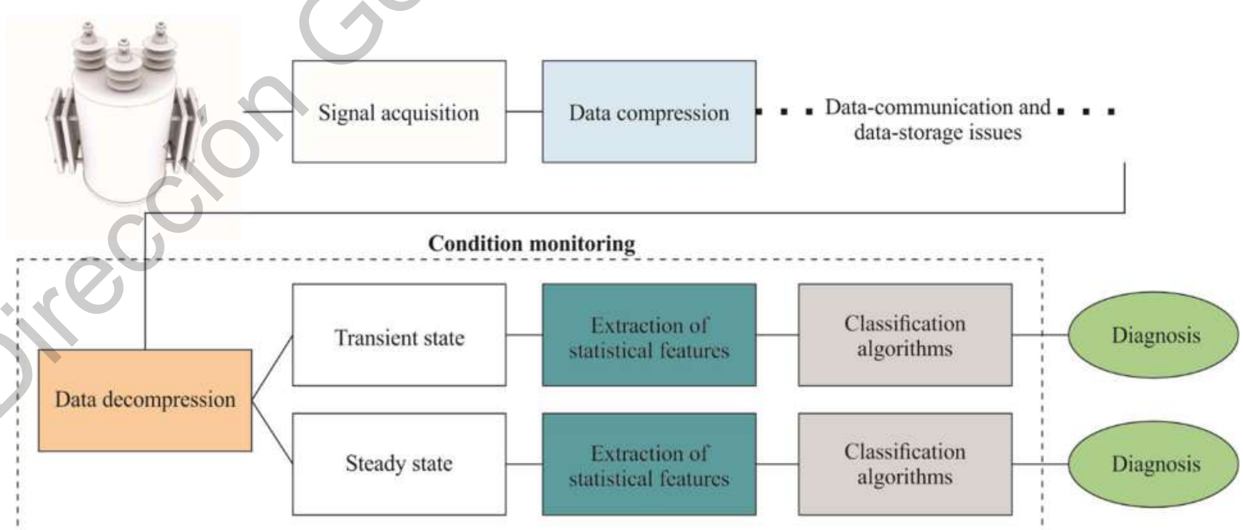


Fig. 2. Methodology.

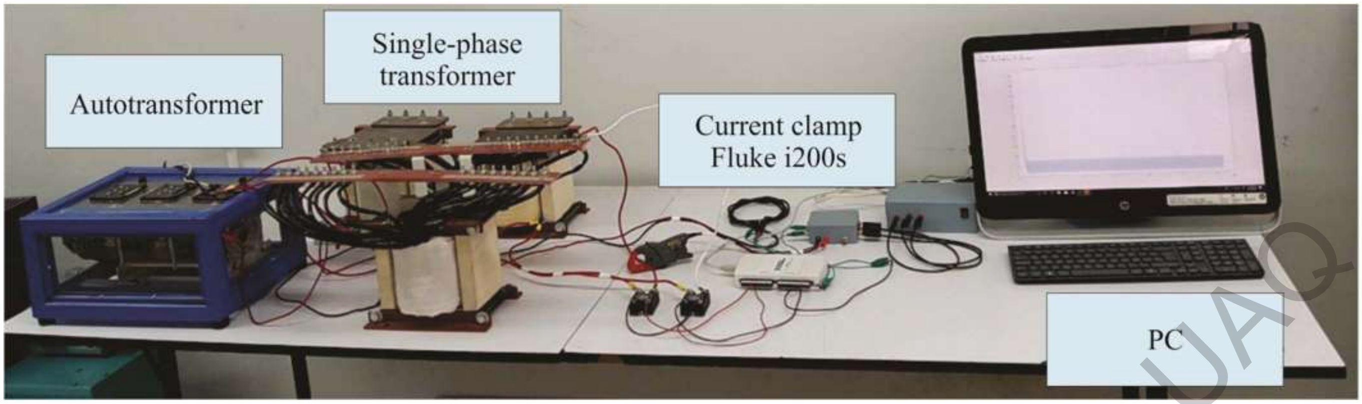


Fig. 3. Experimental setup.

records corresponding to each condition. The decompression algorithm and the average decompression time parameters, i.e., CR, PRD and QS, are analyzed.

C. Test 1. Different quantization parameters

The first test consists of proposing two different quantization parameters and analyzing the results obtained for both the compression and the decompression algorithms. Table I shows the results for Δ = 9 and Δ = 0.5. While the average compression time recorded for both Δ is not significant, the CR however represents a notable decrease for Δ = 0.5, equal for the values of QS. It is also observed that the difference of the values obtained for the levels 3 and 4 is not crucial. By taking into account the QS, the respective results at level 4 are those that reflect less compromise between the compression performance and the quality of reconstruction. Therefore, the values of the wavelet decomposition at level 4 are taken into consideration to have a better appreciation of said results as seen in Fig. 5 where the visual quality of the reconstructed signals for each condition is compared. The difference in visual quality between the decompressed signals with Δ = 9 and Δ = 0.5 is remarkable. The best approximation against the original signal is the reconstructed signal by using Δ = 0.5.

- Test 2. Adding a Huffman coding

Derived from Test 1 and according to the premise that the addition of Huffman coding step before saving the compressed data in HDF will improve the obtained results, the values of wavelet decomposition level 4 and Δ = 0.5 are taken as reference

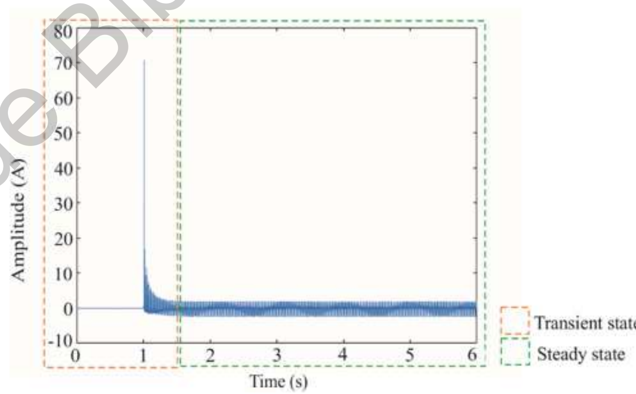


Fig. 4. Representation of the segmented signal in transient state and steady state.

TABLE I. COMPRESSION AND DECOMPRESSION RESULTS FOR Δ = 9 AND Δ = 0.5

			Compression		Decompression			
Conditions	**Δ**	**Wavelet level**	***Average time per record (s)***	***Average CR***	***Average time per record (s)***	***Average CR***	***Average PRD***	***Average QS***
Healthy	9	Level 3	0.0447	487.9211	0.1521	487.9211	466.1128	1.0468
		Level 4	0.0445	455.5727	0.1220	455.5727	481.5568	0.9460
	0.5	Level 3	0.0483	126.0836	0.2558	126.0836	488.3419	0.2582
		Level 4	0.0495	110.2389	0.2793	110.2389	488.4666	0.2257
10 SCTs	9	Level 3	0.0452	417.8253	0.1216	417.8253	496.1672	0.8421
		Level 4	0.0467	445.3591	0.1211	445.3591	501.7194	0.8877
	0.5	Level 3	0.0651	96.9631	0.2701	96.9631	494.5139	0.1961
		Level 4	0.0500	94.2584	0.2939	94.2584	494.4056	0.1906
20 SCTs	9	Level 3	0.0443	463.9189	0.1215	463.9189	330.5426	1.4035
		Level 4	0.0455	380.3349	0.1230	380.3349	329.4849	1.1543
	0.5	Level 3	0.0480	78.9104	0.2903	78.9104	325.6912	0.2423
		Level 4	0.0510	68.3330	0.3132	68.3330	325.6609	0.2098

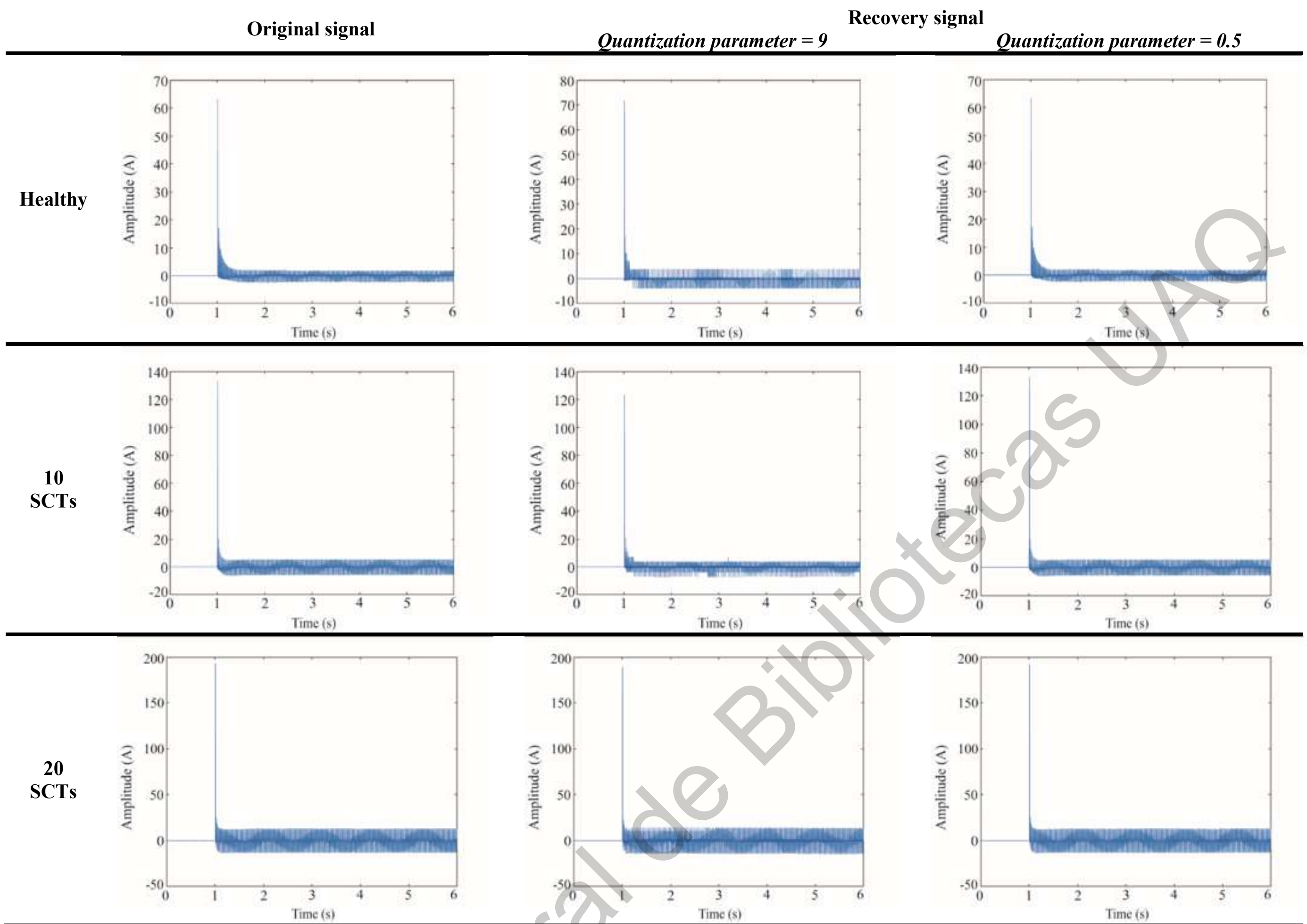


Fig. 5. Visual quality of the original signal against the reconstructed signals for $\Delta = 9$ and $\Delta = 0.5$

in order to analyze the impact of the Huffman coding in the compression quality.

Table II shows the results produced with the addition of the Huffman coding step and without its addition where the calculated PRD does not indicate any variation and the values of CR and QS do not reflect revealing changes. The most noticeable change is seen in the average compression and decompression time. The 3th and 5th columns show the compression and decompression time averages per record, respectively. They present a difference that is worth noting. While the average compression time of the three conditions is 0.05s, the average compression time adding Huffman is 0.44s. Although the time difference is minimal, it can be significant over time. The average decompression time of the three conditions without the Huffman addition is 0.29s and with the addition of the step it is 0.42s. The relationship between time and average performance for the same quantization value indicates that the addition of a Huffman coding step before saving the

TABLE II. COMPARISON OF THE RESULTS PRODUCED FOR $\Delta = 0.5$ WITH ADDITION OF HUFFMAN CODING AND WITHOUT THE SAME

		Compression		**Decompression**			
Conditions	**Huffman 'H'/ Without Huffman 'WH'**	***Average time per record (s)***	***Average CR***	***Average time per record (s)***	***Average CR***	***Average PRD***	***Average QS***
Healthy	H	0.2985	93.6321	0.3283	93.6321	488.4666	0.1917
	WH	0.0495	110.2389	0.2793	110.2389	488.4666	0.2257
10 SCTs	H	0.3901	80.3367	0.4116	80.3367	494.4056	0.1625
	WH	0.0500	94.2584	0.2939	94.2584	494.4056	0.1906
20 SCTs	H	0.6379	65.2702	0.5202	65.2702	325.6609	0.2334
	WH	0.0510	68.3330	0.3132	68.3330	325.6609	0.2098

compressed data in HDF format does not significantly support the compression quality.

D. Fault classification

Table III presents the effectiveness results of classification of the NN and SVM when the signals are studied in their original format and when they have been decompressed. The six statistical features for each segment of the signal (transient and steady state) are calculated and the values obtained are standardized to have a common base with a standard deviation of 1 and a mean of 0. The percentages in Table III are divided into two parts: transient state and the steady state. These percentages indicate that the compression and decompression process improve the capacity of the techniques for classifying the short circuit fault conditions in a single-phase transformer, since the loss compression type favors the elimination of noise over the entire signal, which intensifies the classification accuracy. In this way, a more complete analysis is offered in a fault diagnosis context.

TABLE III. CLASSIFICATION EFECTIVENESS

State	Original signals		Reconstructed signals	
	MP	***SVM***	***MP***	***SVM***
Transient	95%	95%	98.3 %	98.3 %
Steady	96.6%	96.6%	100%	100%

V. CONCLUSIONS

The results obtained from the implementation of a short-circuit fault classification methodology using a data compression algorithm based on DWT to compress current signals from a single-phase transformer are presented. The main characteristics of the compression algorithm are its simplicity and the fact of saving the algorithm outputs in HDF. Two levels of Wavelet decomposition have been considered for the realization of the tests; from these results, it is concluded that the selection of the level is not the crucial factor for the success of the compression technique since the results show insignificant differences between the levels; however, the level of decomposition could impact in the detection of the failure since the frequency information associated with the failure could be better isolated at a certain level, even more when levels of incipient severity or with greater sensitivity are required, e.g., turn to turn. Regarding the quantification parameters, it is found that for $\Delta>1$ the technique becomes less effective. Another feature that appears for $\Delta > 1$ is that the application of the Huffman coding step, before saving the data in HDF format, can be avoided because it does not cause a significant impact on the signals; on the contrary, this solution implies a time delay that could be important over a prolonged period of time. From the tested parameters, the best compression results were achieved with $\Delta = 0.5$ since this value clearly represents a good compromise between the compression performance and the high visual similarity between the recovered signal and the original signal. Additionally, it is demonstrated that a fault detection method can be applied after the decompression process; in fact, the percentages of effectiveness for the short circuit fault are 96.6% for the transient state and 100% for the steady state.

REFERENCES

[1] A. J. Christina, M. A. Salam, Q. M. Rahman, F. Wen, S. P. Ang, and W. Voon, "Causes of transformer failures and diagnostic methods – A review," Renewable and Sustainable Energy Reviews, 82, 1442–1456, 2018.

[2] G. Alvarez-Monroy, A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J. C. Olivares-Galvan, and R. Escarela-Perez, "A new EMD-Shannon entropy-based methodology for detection of inter-turn faults in transformers," 2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2017.

[3] M. S. Ballal, H. M. Suryawanshi, M. K. Mishra and B. N. Chaudhari, "Interturn faults detection of transformers by diagnosis of neutral current," IEEE Transactions on Power Delivery, 31(3), 1096–1105,2016.

[4] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J. C. Olivares-Galvan, and R. Escarela-Perez, "Experimental data-based transient-stationary current model for inter-turn fault diagnostics in a transformer," Electric Power Systems Research, 152, 306-315, 2017.

[5] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J. C. Olivares-Galvan, and R. Escarela-Perez, "The application of EMD-based methods for diagnosis of winding faults in a transformer using transient and steady state currents," Measurement, 117, 371-379, 2018.

[6] T. Pu and J. Bai, "An auto regression compression method for industrial real time data," The 26th Chinese Control and Decision Conference,2014.

[7] Z. Xueming and X. Lei, "Insulator leakage current data compression based wavelet analysis," 2010 International Conference On Computer Design and Applications, 2010.

[8] D. Zhang, Y. Bi and J. Zhao, "A new data compression algorithm for power quality online monitoring," 2009 International Conference on Sustainable Power Generation and Supply, 2009.

[9] C. Tan, L. Zhang and H. Wu, "A novel blaschke unwinding adaptive Fourier decomposition based signal compression algorithm with application on ECG Signals," IEEE Journal of Biomedical and Health Informatics, 2018.

[10] B. Hejrati, A. Fathi and F. Abdali-Mohammadi, "Efficient lossless multi-channel EEG compression based on channel clustering," Biomedical Signal Processing and Control 31, 295–300, 2017.

[11] M. S. Manikandan and S. Dandapat, "Wavelet-based electrocardiogram signal compression methods and their performances: A prospective review," Biomedical Signal Processing and Control 14, 73–107, 2014.

[12] S. Agarwal, E. E. Regentova, P. Kachroo and H. Verma, "Multidimensional compression of ITS data using wavelet-based compression techniques," IEEE Transactions on Intelligent Transportation Systems, 18(7), 1907–1917, 2017.

[13] L. Rebollo-Neira, "Effective high compression of ECG signals at low level distortion," Scientific Reports, 9(1), 2019.

[14] D. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," Proceedings of the IRE 9, 1098–1101, 1952.

[15] J. L. Gonzalez-Cordoba, A. Mejia-Barron and M. Valtierra-Rodriguez, "Automatic detection and classification of electrical disturbances by means of empirical mode decomposition method," 2015 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2015.

[16] J. M. Valencia-Ramirez, J. A. Raya, J. R. Cedeno, R. R. Suarez, H. J. Escalante and M. Graff, "Comparison between Genetic Programming and full model selection on classification problems," 2014 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2014.

[17] E. Frank, M. A. Hall and I. H. Witten, "The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.