DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

University Kasdi Merbah– OUARGLA

Faculty of New Technologies of Information and Communication

Department of Electronics and Telecommunications

Dissertation

Field: Sciences and technologies
Sector: Automatic
Specialty: Automatic and systems

Presented by:

KAFI Omar Rafik

ABDESSAMED Zakaria

Title:

Fault detection in photovoltaic power converter

Jury members:

Dr.KAFI Mohamed Redouane	Supervisor	UKM Ouargla
Dr.Mahboub abdelbasset	President	UKM Ouargla
Dr.Rouabeh Boubakeur	Examinator	UKM Ouargla
Dr.Bensid Khaled	Examinator	UKM Ouargla

Academic Year: 2021 – 2022

Kasdi Merbah University
Automatics And Systems Master's Degree (2022)

TITLE: fault detection in multicellular converter
AUTHOR: Kafi Omar Rafik - Abdessemed Zakaria (Kasdi Merbah University)
SUPERVISOR: Dr. kafi Mohamed Redouane
JURY:

Dr. Mahboub abdelbasset

Dr. Rouabeh Boubakeur

Dr. Bensid Khaled

NUMBER OF PAGES: 1, 61

Abstract

This work deals with fault diagnosis of power converter used in photovoltaic systems which supply and isolated electric load and how it transform the current safely to devices. this thesis treat the multicellular power converter describe how much the power converter failures influence to the load current the main focusing is in the capacitors faults which can affect badly to the load currents so it can result bad consequences on devices. which proposed a solution for this problem using (KNN) k-nearest neighbor machine learning algorithm to build a classification model for diagnosis. and using two modes of control in order to compare in term of the function under failures , load current preservation and the smoothest, and in term of the accuracy of the classification model built.as well as to using the sliding mode control mode and the exact linearization mode, this is for the purpose of comparison in terms of system performance during the occurrence of faults and the extent of their impact on the load current by examining the shape of its signal and analyzing the robustness of the two controls in not being greatly affected by the faults and showing it.

Keywords : Fault diagnosis based machine learning, multi cellular power converter, photovoltaic system, no linear control

ملخص

تتعامل هذه الدراسة مع تشخيص اخطاء محول الطاقة المستخدم في الأنظمة الكهروضوئية التي تزود الحمل الكهربائي المعزول وكيف يحول التيار بأمان إلى الأجهزة. هذه المذكرة تعالج محول الطاقة متعدد الخلايا وتصف مدى تأثير فشل محول الطاقة على تيار الحمل الرئيسي. التركيز هو في عيوب المكثفات التي يمكن أن تؤثر بشكل سيء على تيارات الحمل لذلك يمكن أن يؤدي ذلك إلى عواقب وخيمة على الأجهزة. حيث اقترح حلاً لهذه المشكلة باستخدام خوارزمية التعلم الآلي (الجار الاقرب) لبناء نموذج تصنيف لتشخيص الاعطال. واستخدام طريقتين للتحكم من أجل المقارنة من حيث الوظيفة تحت الإخفاقات، والحفاظ على تيار الحمل ولتحديد أي منهما الأكثر سلاسة، من حيث دقة نموذج التصنيف المبني. بالإضافة إلى استخدام وضع التحكم في الوضع المنزلق ووضع الخطي الدقيق، وذلك لغرض المقارنة من حيث أداء النظام أثناء حدوث الأعطال ومدى تأثيرها على تيار الحمل من خلال فحص شكل إشارته وتحليل قوة التحكمين في عدم التأثر بشكل كبير بالأعطال وإظهارها.

الكلمات المفتاحية : تشخيص الأعطال على أساس التعلم الآلي ، ومحول الطاقة متعدد الخلايا ، والنظام الكهروضوئي ، والتحكم غير الخطي

Résumé

Ce travail traite du diagnostic de panne du convertisseur de puissance utilisé dans les systèmes photovoltaïques qui alimentent et isolent la charge électrique et comment il transforme le courant en toute sécurité vers les appareils. Cette thèse traite du convertisseur de puissance multicellulaire et décrit à quel point les pannes du convertisseur de puissance influencent le courant de charge le principal la focalisation est dans les défauts des condensateurs qui peuvent affecter gravement les courants de charge, ce qui peut entraîner de mauvaises conséquences sur les appareils. qui a proposé une solution à ce problème en utilisant l'algorithme d'apprentissage automatique KNN pour construire un modèle de classification pour le diagnostic. En utilisant deux modes de contrôle afin de comparer en terme de fonction sous pannes, conservation du courant de charge et le plus lisse, et en terme de précision du modèle de classification construit. ainsi que d'utiliser le mode de commande en mode glissant et le mode de linéarisation exacte, ceci à des fins de comparaison en termes de performances du système lors de l'apparition de défauts et de l'ampleur de leur impact sur le courant de charge en examinant la forme de son signal et analyser la robustesse des deux contrôles à ne pas être fortement affectés par les défauts et à le montrer.

Mots clés : Diagnostic de panne basé sur l'apprentissage automatique, convertisseur de puissance multicellulaire, système photovoltaïque, contrôle non linéaire

Acknowledgements

I would like to express my deepest gratitude to my advisor, Dr. Kafi, whose sincerity and encouragement I will never forget. Dr. Kafi has been an inspiration as I hurdled through the path of this Masters degree. He is the true definition of a leader and the ultimate role model. This thesis would not have been possible without Dr. Rouabah, whose guidance from the initial step in research enabled me to develop an understanding of the subject. I am thankful for the extraordinary experiences He arranged for me and for providing opportunities for me to grow professionally. It is an honor to learn from Dr. Kafi and Dr. Rouabah.

I am grateful for my parents whose constant love and support keep me motivated and confident. My accomplishments and success are because they believed in me. Deepest thanks to my siblings, who keep me grounded, remind me of what is important in life, and are always supportive of my adventures. I am forever thankful for the unconditional love and support throughout the entire thesis process and every day.

Contents

List of Figures

List of Tables

Declaration of Authorship

We, Kafi Omar Rafik - Abdessemed Zakaria, declare that this thesis titled, fault detection in multicellular converter and the work presented in it are our own.

Introduction

Recently, the solar photovoltaic (PV) energy has seen a huge demand in the electric power production, because it is a clean energy source, environmentally friendly that produces no gas emitted [37] And it led to a wide development in many aspects, such as avoiding new project electric transportation lines and reducing the electric energy bill and used to provide with water in isolated areas. Among the solar photovoltaic component, the power converter is used as an interface between photovoltaic panels and electric load, the power converter constitutes between 43% and 70% of the PV power plant service requests [55] in order to convert the direct current (DC) voltage produced by photovoltaic panels to alternating current (AC) voltage, however when a failure occurs in power converter all advantages may affected negatively, such as no water extracted in isolated areas, high cost and increased time of maintenance intervention, as well as, it has a significant impact on the reliability of the PV systems [4] Therefore, the use of fault diagnosis of power converter used in photovoltaic system is considering as a big challenge task. several causes that prevent power converters to operates in health conditions specially in power switches and flying capacitors. To deal with this issue many research works have been developed in literature, such as in multicellular power converters control in [46] and [52],a solar PV system based on Step-Up Boost converter in order to extract the maximum power form solar panels [5], in addition of treating, in the wind energy as in [63],as well as in the wind turbine converters in [53] and [50] ,the evolution also included the controlling systems using exact linearization control as mentioned in [49] and active filtering in [71] in [47], including the optimization in conversion energy effeciency in [51], the contribution of multicellular power converter as mentioned,fault tolerant control in [48] ,also in the integration of PV systems characteristics in energy extraction [29] , as well as the classic converter using before and its limitations [4], [65].

This thesis include the fault diagnosis of multicellular power converter four level, focusing on the capacitor failures that make distortions in the load current shape, Including the appearance of harmonics which affecting electric loads, by reducing the lifetime of devices, increase mechanical vibrations and overheating in machine windings. These faults may also lead to catastrophic damages depending on the environment concerned. It can be said that the harmonics are the disturbances which disperses the signal waveform. Distortions of waveforms leading to malfunctions, increases in peak values causing dielectric breakdowns causing overheating and thus additional losses in current as well as in voltage, as well as a spectral spread causing vibrations and mechanical fatigue are just a few of the effects of harmonics on electrical installations and equipment. Because of the increased costs, energy efficiency degradation, oversizing, and productivity losses, all of these factors have a major economic impact. Fault detection methods is the proposed solution in this work for tracing how faults form in multicellular power converter systems, knowing the source of the fault, to be able to address faults directly without interfering with the healthy parts. Machine learning method topped the list of scientific research concerns,medical field [45], electrical [1]... It has received a lot of uses, due to its accuracy in automatic classification such as in [40] . It uses algorithms that are similar to human thinking in distinguishing between types in which classification is achieved by a series of logical processes and by following the decision tree [10] , it will be able to give us the correct classification through simple symbols that human understands.Given enough data while training the model [7] , these are capable of modeling even the most complex problems in high efficiency.

In order to exploiting the study also in choosing a control in line with the system specifications in terms of resistance to faults and their observation, there are many control systems for multicellular power converters, There will always be a discrepancy between the actual plant and its mathematical model , these discrepancies arise from unkown external disturbances, plant parameters, and parasitic/unmodeled dynamics, designing control laws that provide the desired performance in presence of these disturbances/uncertainties is a very challenging task [70]. This has led to intense interest in the development of the so-called robust control methods. One particular approach to robust controller design is the so-called sliding mode control technique. Sliding mode control is employed for nonlinear systems for controlling the power converter [61] , it is a technique that uses an intermittent control signal to change the dynamics of the system (or more strictly, a control signal of a specified value) [58]. Sliding mode control is a particular type of variable structure control ,the main advantages of sliding ode control are robustness, finite-time convergence, and reduced-order compensated dynamics.[27]. Exact linearization control is used to cancel the model's non linearity [39]. because the exact linearization control is vulnerable in the presence of unknown dynamics and uncertainties, and the gains of the linear stabilizing control are sensitive to output performance [28]. exact linearization is a common nonlinear control strategy used to control nonlinear systems [68] . The transformation must be a diffeomorphism to ensure that the transformed system with exact linearization is an equivalent representation of the original system as in [56]. That is, not only must the transformation be invertible (bijective), however, it must be smooth in order for the differentiability in the original coordinate system to be preserved in the new coordinate system, so chapter 1 represents multicell power converter types and modes of control mentioning the main faults that occurs the load current shape and the determination of the state representation of ML converter circuit, chapter 2 to remains the characteristics of Machine learning and the approaches used to process data as well as the describing of KNN algorithm and FFT spectral analysis, the chapter 3 contains the implementation and results of the study where all steps of work will be considered beginning with data collection and processing under the application of the two modes of control (sliding mode control and exact linearization control) moving to data preparation and feature selection then the application of KNN machine learning algorithm with the comparison between the two modes of control and determining the best accurate model.

Chapter 1

Multicell Power Converter

1.1 Introduction

This chapter is composed in 3 parts. The first part is dedicated to multi-cell (ML) energy converters and recollects their topologies and their usage, notably withinside the domain of renewable energies with a specific attention on solar applications and DC microgrids. The accompanying sections are linked with managing of energy electronics converters. We remember a few essential repercussions on classical control approaches dedicated to power converters and past efforts linked with control of interleaved multi-cell power converter.Finlay, we'll define different kinds of faults in ML converters and state representation of control circuit in ML power converter .

1.2 Topologies of Multilevel DC/DC Converters

There are a number of exceptional topologies of multilevel power converters reliant on the power range, form of conversion, their applications etc. however they've the identical basic properties. Such as apparent switching frequency and harmonic cancellation[35].

1.2.1 Series connections

There are more than one approach to construct a series multi-cell power converter depending on the kind of power supply.

1.2.1.1 Series connection with isolated power sources

Figure 1-1 indicates a multi-cell DC-DC power converter with series connections. In this shape in which the power supplies are isolated, the introduced power from every power supply, the duty cycles and the switching frequencies may be specific from each cell. This topology is used in DC-AC applications more than in DC-DC applications.

For DC applications they may be used with DC storage devices, which includes batteries in which every power converter manages the battery cell energy while providing, though the series connection, the capability to supply high voltages[66]. The DC sources also may be low voltage generators such as fuel cells or PV panels[15][13]. For PV panels in a identical string, all cells will have the identical design, switching frequency and duty cycles. It is likewise feasible to use an interleaved pulse width modulation (PWM) to this shape reduce the voltages and currents ripples on the input and output of the power converter[23].

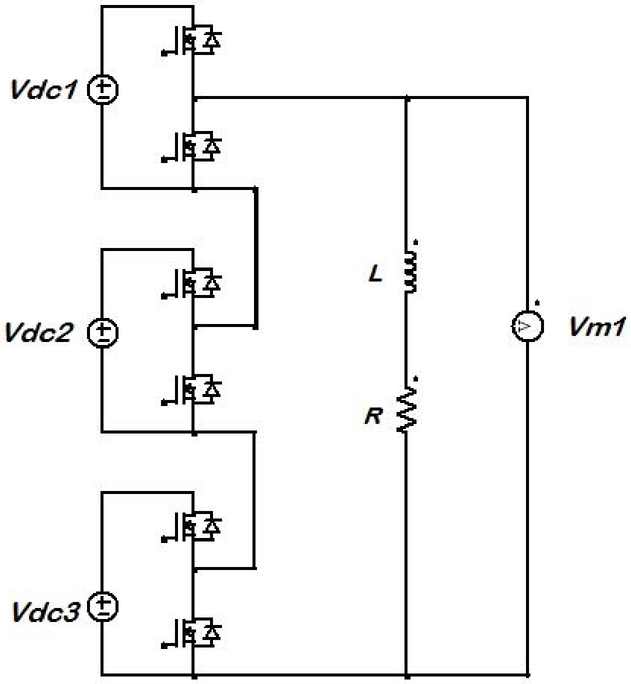


FIGURE 1.1: Classical Power converter: series connection with isolated power supply

1.2.1.2 Flying capacitor

This topology can be used with a single DC supply and may be used both for AC or DC applications. In case of unidirectional output current some switches may be changed by diodes as shown in Figure1.1 [34].

Such converters may be designed with any variety of cells as shown in Figure 1-2 b [33]. For AC sources, opposite blocking devices can replace bidirectional switches of Figure 1-2-b to build the current source inverter shown in Figure 1-2-c. This also can be executed by using a four-quadrant AC chopper as in [26]. In those to final configurations, the voltage source is AC, the flying capacitor voltages are AC too. The manipulate of such AC topologies are greater complicated to realize as the dynamic of active balancing of the capacitor voltages must be fast In order to follow the AC voltage of the supply. This includes a high switching frequency.

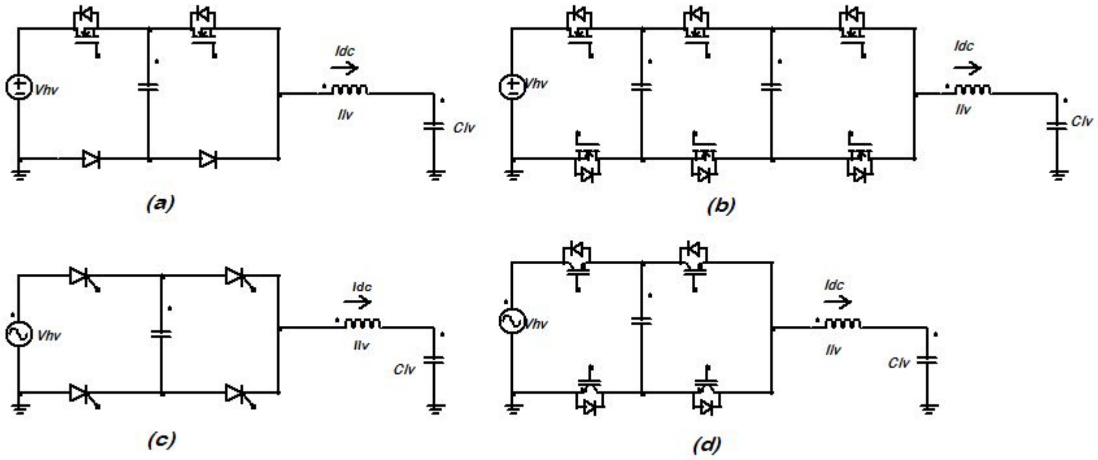


FIGURE 1.2: Different topologies of flying capacitor converters

1.2.2 Parallel connections

1.2.2.1 Star connection with interleaved PWM:

Figure 1-3 indicates a multicell boost converter with a star connection of inductors [35]. The 3 inductors on this connection have the identical size and the supply current is split into 3 equal parts through the 3 cells. The major advantage of the interleaved multicell power converter with a star connection is the reduction of the voltages and current ripples on the input and output of the converter. This provides the opportunity to reduce substantially the size of input and output filtering capacitors.

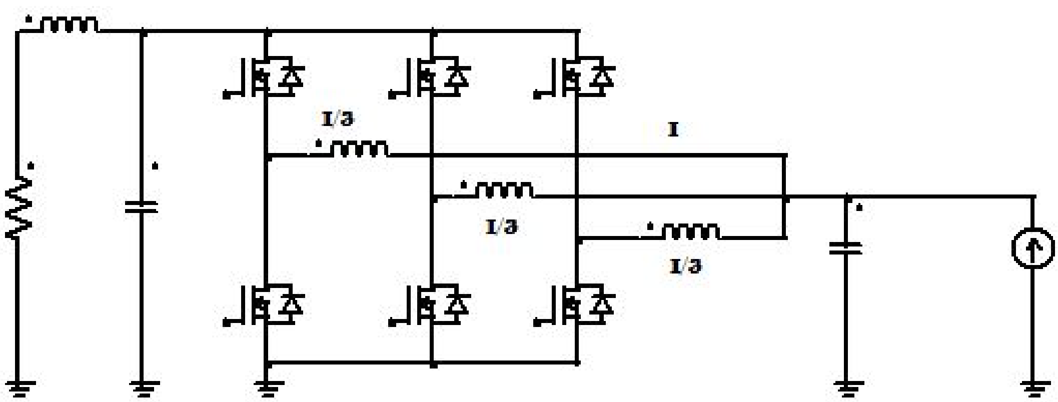


FIGURE 1.3: Parallel multi-cell converter with star-connected inductors

1.2.2.2 Interleaved multicell power converter using InterCell Transformers (ICTs):

The current ripple reduction in interleaved parallel star connection utilizing uncoupled inductors is only observable on the input and the output. In such topologies, there is no decrease in the rate of current ripple inside the cells. As a consequence, such a solution is contained to a small number of cells (3 or four). It is conceivable to overcome this issue by employing magnetically coupled inductors.magnetic coupling may greatly minimize the ripples within the cells.we can observe that the cell-current ripples are considerably decreased. These findings may be compared to the waveforms of cell currents for uncoupled inductors. This property allows for enormous reductions in the size of inductors.

The cells of a multi-cell parallel power converter may be magnetically connected in two ways. The cyclic cascade construction, depicted in Figure 1-4-a, in which winding transformers are utilised to link nearby cells. The second conceivable approach is to employ a monolithic magnetic device, as illustrated in Figure 1-4-b. In this architecture, all of the windings are coiled on the same magnetic core, enabling cells to react with one another.Different design topologies for linked inductors and intercell transformer (ICT) are shown in [54] Because of the subdivision of currents in numerous cells, such parallel topologies may be employed for huge current supply systems with very low voltage. The prospective decrease of the global output inductance is also a noteworthy component because it gives to such converter the capability to drive a highly dynamic load [35][36].

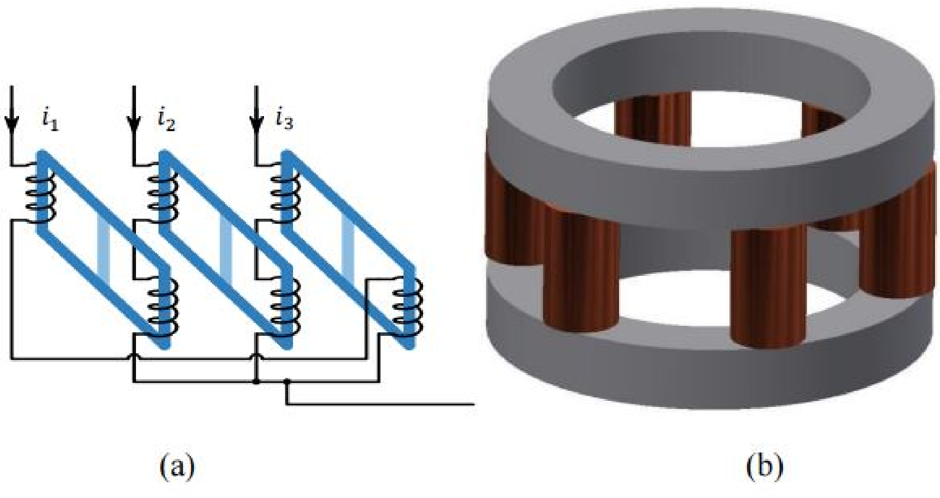


FIGURE 1.4: a) cyclic cascade configuration b) monolithic configuration

1.3 Faults In Multicellular Converters

Multicellular converters as all power converters consistof active elements of semiconductors (diodes, MOSFET,IGBT, GTO) and energy storage elements (inductances and capacitors). Although studies are rare, it is common that the firsts and especially the controlled switches are the most exposed to failure. This could happen from their control as a physical fault always related to a thermal problem caused by an overload due to a power surge causing an avalanche; which provokes the increasing of internal temperature of the power interrupter [32] [16].

1.4 Adding cells to multicellular converters

1.4.1 Advantages

The main advantage of the elevation of cells is to operate in health conditions especially with raising the number of power converter cells. Because as we elevate flying capacitors our signal's output load current distortions and harmonics appearance gets low .also the elevation of power switches makes the multicell converter work comfortably and with high efficiency owing to voltage division among the cells which increases the switching frequency.

1.4.2 Disadvantages

As the number of cells in power converter increases, the control algorithm gets more complicated due to the considered number of commutating switches.

1.5 Flying cpacitor converter Modeling:

Capacitors are employed in this sort of multilevel converter configuration. Figures 1.5 1.6 1.7 demonstrate 3,5,and 7-level FC (flying capacitor) converter types, respectively. Its maximum output voltage cannot be more than half of the input voltage. Capacitor may be used to regulate power flow. Due to the high frequency switching, switching losses will exist in this kind of converter [59].

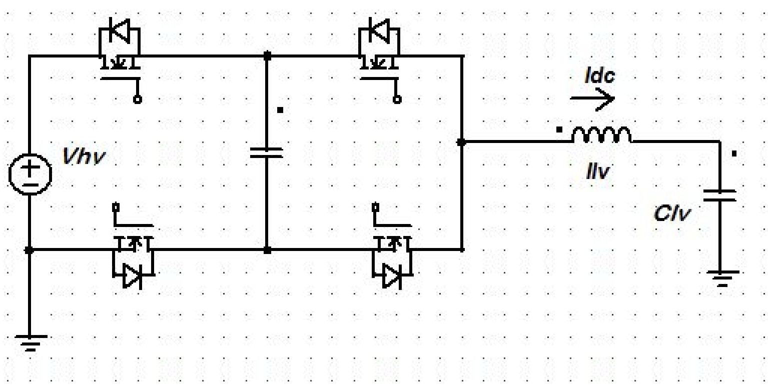


FIGURE 1.5: 3-level FC converter model

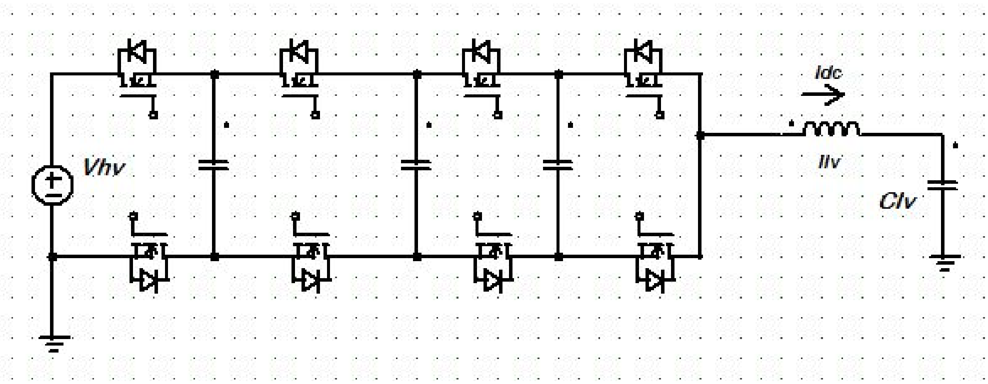


Figure 1.6: 5-level FC converter model

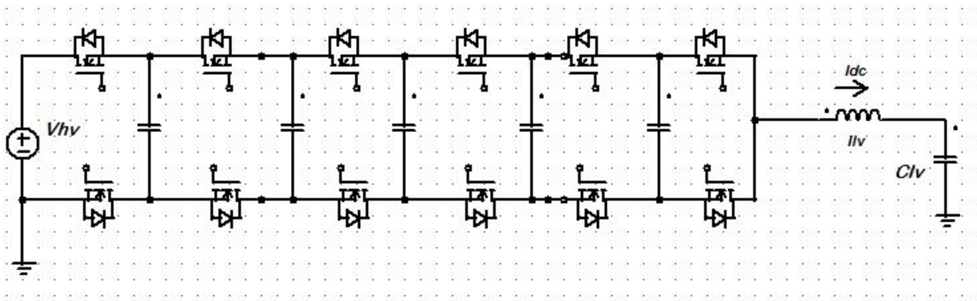


Figure 1.7: 7-level FC converter model

The output voltage of a multilevel converter is created at a high frequency, with a low switching frequency and minimal distortions. A hybrid multi-level converter is used to generate high voltage at the output, minimizing total harmonic distortion (THD) , dv/dt stress, and common mode voltage [42] [43].

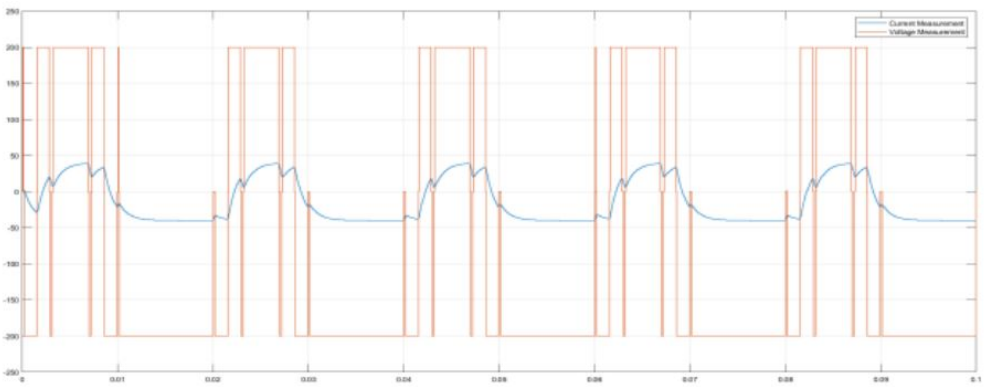


Figure 1.8: Output voltage and current waveform of 3-level FC inverter

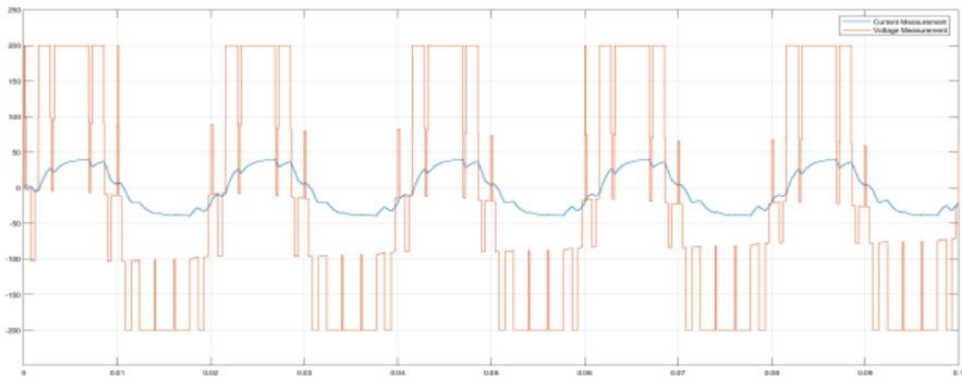


FIGURE 1.9: Output voltage and current waveform of 5-level FC inverter

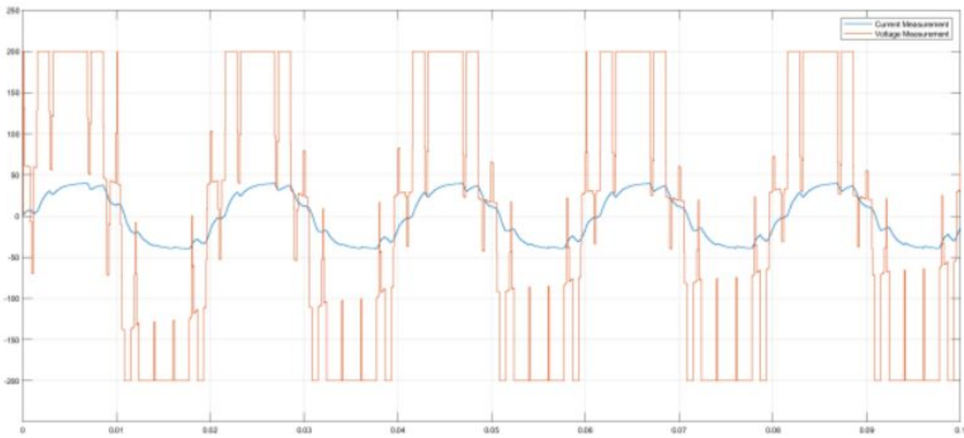


FIGURE 1.10: Output voltage and current waveform of 7-level FC inverter

Thus, greater levels of multilevel inverter perform better than lower levels of multilevel inverters(as shown in fig 1.8 1.9 1.10), and noise on the output side of the inverter decreases as the level of multilevel inverter increases[60]. leading to the conclusion that for dc to ac conversion in renewable energy applications, higher level flying capacitor multilevel inverters should be preferable.

1.6 Multicell Power Converter in Solar application and microgrids

Parallel multi-cell converters using ICTs are an interesting approach in the sector of low and medium voltage and high current power converters. These very versatile systems may be used in many varieties of power conversion systems including DC-to-DC power converters with boost or buck functions, as well as DC-to-AC inverters and AC-to-DC synchronized rectifiers. They are widely utilized in diverse applications and are especially useful in renewable energy systems such as photovoltaic inverters [12],[18] storage management systems [73], fuel cell converters [18] in addition to electric vehicles [14]. Although this achievement, future improvements, such as ICT design for fault-operation [54] and control enhancement [25],[17] may be possible. A simplified schematic of a microgrids structure is shown in Figure 1.11. The main player in such microgrids is the power converters. In microgrids, power converters are used to balance power, voltage, and current. The multicell power converters are appropriate for use in renewable energy resources because of their features and advantages [31].

A photovoltaic array feeds a load which might be likely a battery, directly some DC loads or a grid inverter [66]. As each PV's maximum power point and the load voltage can vary greatly, it is obligatory to interface a converter among the load and the source.

This may be executed for low voltages and low power with a multicell interleaved power converter. The key advantage of this power electronics arrangement over a classic single buck converter is that it ensures minimal current ripples on both the input and output sides. In fact, concerning the input stage, the input current ripple is reduced by an factor while the input current obvious frequency is elevated by an aspect of n. As a result,the C_i capacitance may be decreased by a significant n^2 factor leading to improve the system dynamics and specifically it's capacity to track faster the PV array's greatest power point. Similarly, the amplitude of segment modern-day ripples are decreased by a n^2 factor as compared to an uncoupled multi-cell converter (considering a comparable filtering inductance value), which reduces t+he limitations on the power semi-conductors and the associated losses. Moreover, the global power converter output current ripple is decreased as compared to a classical one-cell Buck DC-DC converter, similar to an interleaved multi-cell DC-DC Buck converter with uncoupled inductors This reduces the requirement to filter the output voltage: in certain conditions, no additional output capacitor is needed.

The control theory of nonlinear system has been attracting growing attention in recent years, both for its technical importance as well as for its influence in diverse sectors of application. In important fields, such as aerospace, chemical and petrochemical industries, biotechnology, and robotics, a new practical use for this instrument develops every day.

1.7 Classical control and LQR

There are some control strategies to control power converters and drives. Figure 1.12 demonstrates the most well-known control techniques.Some approaches, such as hysteresis control, are very well covered and simple, even though other control techniques are more sophisticated and need more calculation power, but lead to higher system performance.

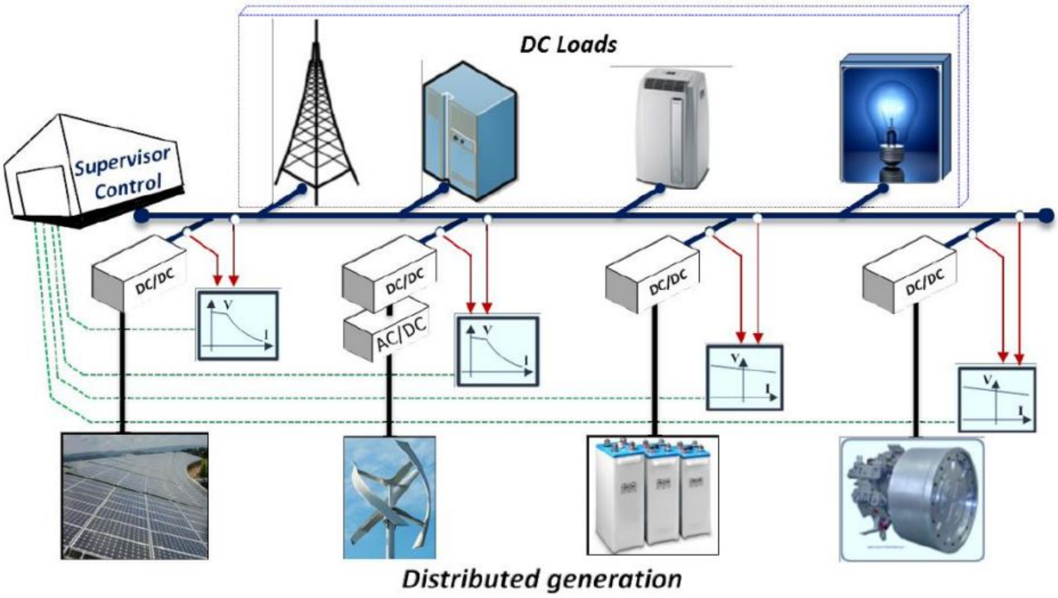


FIGURE 1.11: schematic diagram of DC microgrid structure [62]

Control in power converters is the most investigated subject. There are two basic classical approaches for such control that have been intensively investigated over the last decades: hysteresis control and linear control utilizing pulse width modulation (PWM).

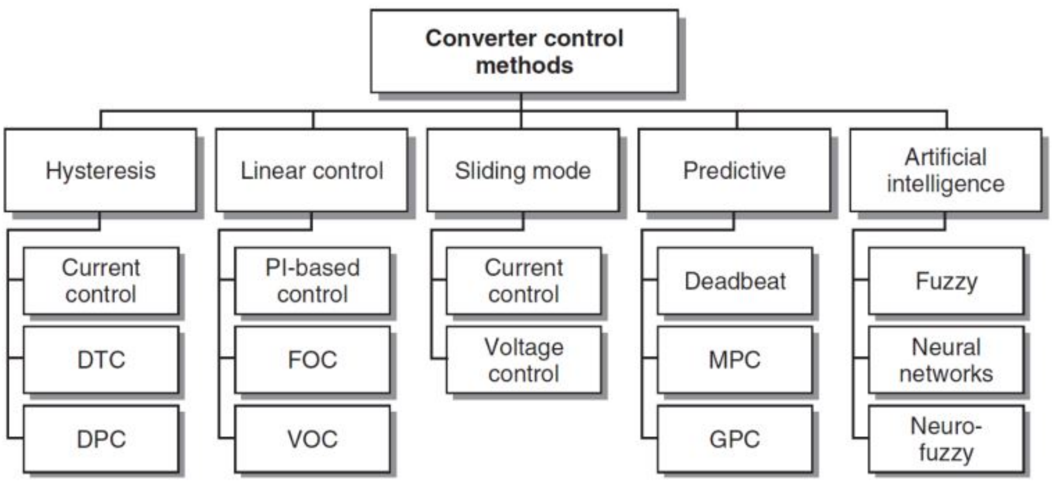


FIGURE 1.12: Different types of converter control for power converters and drives

1.7.1 Hysteresis control

Hysteresis control uses the non linearity of a power converter caused through the switching states of a power converter. The proper switching state may be described as a way to insure an oscillation of the controlled quantities around the specified one with a given hysteresis width. It may be used in simple applications, as an example in case of simple current control but also can be generalized to more complicated systems like in direct power control [38].

Usually this type of controller is carried out with analog electronic devices. Indeed, an implementation on a digital platform required a totally excessive sampling frequency to properly control variations of the controlled quantities. Furthermore, the switching frequency is variable in such control for a set hysteresis width but note that there are a few possible changes of this control approach to obtain a fixed switching frequency. Figure 1.13 indicates waveforms for an hysteresis current control applied to a single phase inverter. The controlled current is the load current (i_L). This cost is in comparison to the reference (i_L^*).The calculated difference (error ϵ) pass through an hysteretic comparator. If the error reached the upper limit ($\delta/2$), the controller activates T_1 and $T_4$and turns off the opposite switches. The contrary command is implemented while the error is much less than the lower limit $-\delta/2$ It may be observed from. Figure 1.13-b that the load current follows its reference with a pick to pick oscillation equal to the hysteresis width.

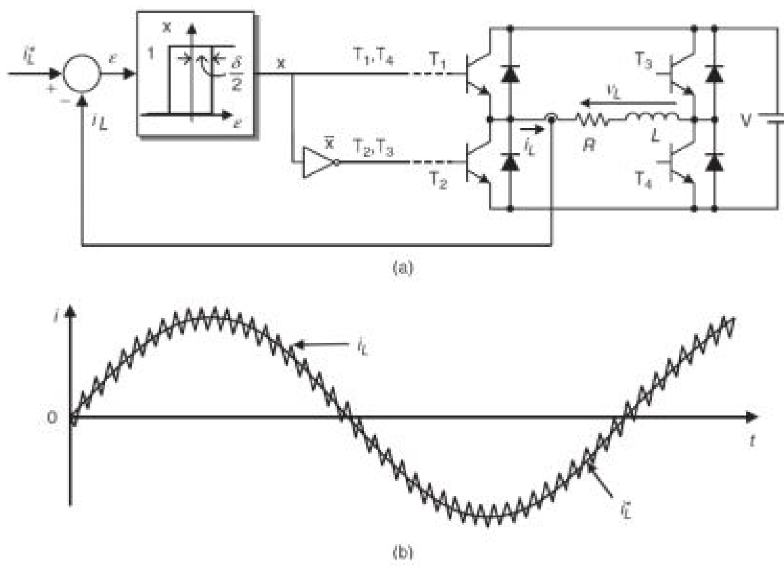


FIGURE 1.13: Hysteresis current control for a single-phase inverter. [62]
(a) Control scheme. (b) Load current

1.7.2 Linear control using PWM

Power converters are linear switched systems. From an average point of view, at the switching frequency degree, they may be linearized. With such an approach, it is conceivable to utilize any linear controller with a PWM or a space vector modulator and proportional-integral (PI) controllers are typically employed.

Nevertheless, for such power converters, the control complexity is also increased by increasing the number of cells.

1.8 Sliding mode control

Sliding mode control basically considers robustness issues as part of the design process this is the special thing about sliding mode control because although feedback control is robust by structure but sliding mode control is one of those feedback control techniques which also include the uncertainty as part of the design process so this is gives additional robustness properties compared to any usual feedback control technique in sliding mode control also has other names such as variable structure control this name is given to sliding mode control because of its switching nature the control law designed by sliding mode technique usually involves switching and that we will see in detail later on at the sliding mode control has been used in many applications including power systems and robotics one of the most research control design technique the sliding mode control is one of the most popular non linear control design techniques and due to its robustness properties there are many variations in sliding mode control it has been you know many modifications . sliding mode control have been proposed for example sting side sliding mode control super twisting sliding mode control terminal sliding mode control discrete event based sliding mode control adaptive sliding mode control.one problem however that has been there for with this sliding mode control is that chattering problem chattering problem occurs mostly due to the switching property of the sliding mode control laws but this has there been many techniques in the literature to avoid the chattering.

1.9 Concept of sliding mode control

In dynamic systems controlled by ordinary differential equations with discontinuous state functions on the right-hand sides, the phenomenon "Sliding Mode" may occur.

1.9.0.1 Sliding surface

In the context of model imprecision on f(x) and b, the control task is to get the state x to follow a certain time varying state .

The nonlinear system input:

$$x^{(n)} = f(x,t) + b(x,t)u(t) \tag{1.1}$$

$x(t)$: the state vector

$u(t)$: the input control vector

$f(x) and b(x)$: non linear functions of state and time

$x_d = [x_d, \dot{x_d}....., x_d^{(n-1)}]$ For the tracking job to be attainable with a finite control , the initial desired, state $x_d(0)$ must be such that:

$$x_d(0) = x_(0)$$

from the analysis of the state plane the trajectorie has a direction in runing which results sliding mode along this line (s=0) , as follow :

$$\dot{x} + cx = 0$$

the equation's order is lower than that of the original system,the sliding mode is specified simply by parameter c and is unaffected by plant dynamics.

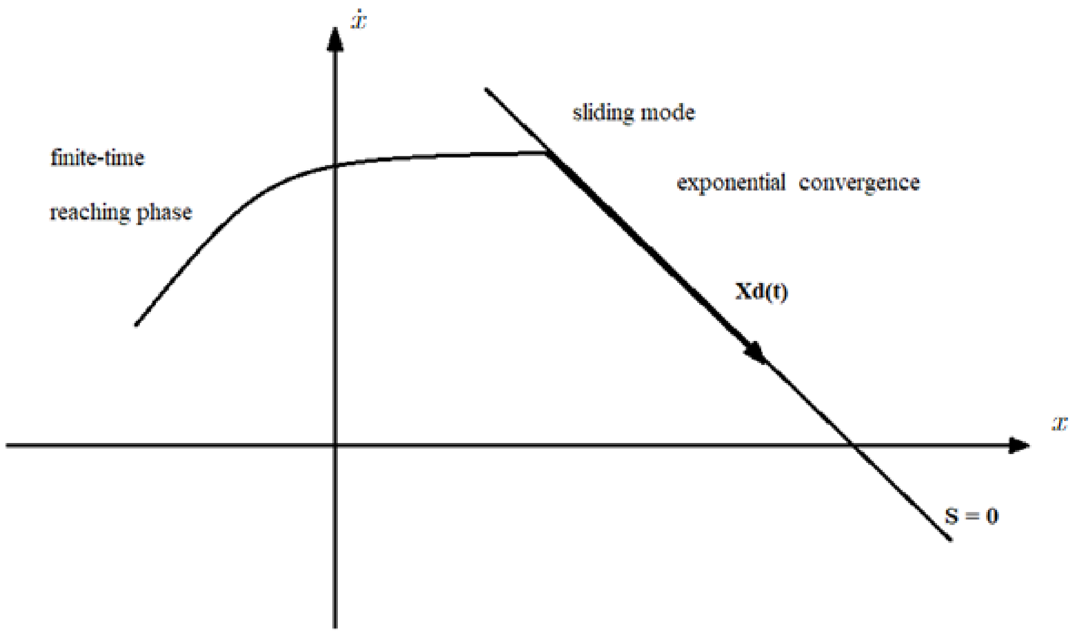


FIGURE 1.14: Sliding mode in a first order system

the characterization of dynamic mouvement of the state trajectory to the sliding surface performed by the generalized Lyaponov function ,is determined by the surface. One selects the "gains" for each switched control structure such that the Lyapunov function's derivative is negative definite., So it is guaranteed that the mouvement of the state path to the surface. After correct design of the surface, a switching controller is developed such that the tangent vectors of the state trajectory point towards the surface such that the state is pushed to and maintained in the sliding surface. Such controllers result in discontinuous closed-loop systems.

Let the tracking error in x be :

$$\tilde{x} = x - x_d \tag{1.2}$$

the surface of the time varying surface S(t) in the state space where :

$$s(x,t) = (\frac{d}{dt} + \lambda)^{n-1}\tilde{x} \tag{1.3}$$

λ strictly positive constant.

if $n = 2$

$$s = \dot{\tilde{x}} + \lambda\tilde{x} \tag{1.4}$$

Given initial condition , the issue of tracking x = xd is similar to that of staying on the surface $S(t)$ for all $t > 0$ surely s = 0 expresses a linear differential equation whose unique solution is $\tilde{x} = 0$, given initial conditions . Thus, the issue of monitoring the n-dimensional vector x_d may be simplified to that of maintaining the scalar quantity s at zero. More exactly, the issue of tracking the n-dimensional vector x_d (i.e. the original nth-order tracking problem in x) may in effect be replaced by a 1^{st}order stabilization problem in s. Indeed, the expression of s includes $\tilde{x}$, we only need to distinguish s once for the input u to appear.

1.10 Exact linearization control

System nonlinearity is described when at least one component or subsystem is nonlinear. The standard approaches employed in the study of nonlinear systems, notably frequency analysis, are not frequently applied to the nonlinear systems. It is required to employ additional methodologies to examine the control of these systems. When the nonlinear effects of systems grow considerable, nonlinear control approaches often fail to deliver the intended performance. In these instances, there are numerous nonlinear control strategies that have shown to give superior outcomes.

The methodology to study a nonlinear control for this type of system. The study of the different models has shown us that a multicellular system has non-linearities mainly due to the fact that the commands of the different cells are not independent. This does not represent a strong non-linearity but obliges us to consider a decoupling control for this system. To study a classical input/output linearization type control, several solutions are available . The first is to make an approximate linearization around a point of equilibrium of the system, ie a linear decoupling of a linearized model of the system. This approach is interesting and lends itself well to regulation problems. It nevertheless has the disadvantage of limiting the excursions of the point of equilibrium under penalty of losing the validity of the linearized model. The second, which is called exact input/output linearization, allows large transitions of the equilibrium point, but on the other hand this method can cause singularity problems to appear. It is important to note that regardless of the method chosen . the state of the system will have to be measured we have chosen to apply the exact method in this first study, which will allow us to have an independent control from the point of view of operation, even if certain problems of singularities will have to be solved. The interest is also the portability of the method according to the structure (converters). Moreover, the application of a classic method will allow us to clearly highlight all the problems relating to the control of multicellular systems.

from a nonlinear dynamical system, to a totally or partially linear system. Thus, after transformation, all the techniques of synthesis of linear correctors can be applied. The interest of this method is that it performs an exact transformation of the nonlinear state system, without going through a linear approximation like the classical linearization techniques (Jacobian). The exact linearization of a system can therefore be seen as a method for transforming a system model into a model with a simpler form. This method has been used successfully in many applications. Nevertheless, a large number of drawbacks and limitations are linked to this method as we will see during this study. After a few necessary reminders a multi-input multi-output nonlinear system which can be represented by:

$$\begin{cases} \dot{X} = f(X) + \sum_{m}^{k=1} g_k(X) u_k \\ \\ y_j = h_j(X) \end{cases} \tag{1.5}$$

with $1 \geq p$ the vector representation can be given by:

$$f(X) = \begin{pmatrix} f_1(X) \\ \vdots \\ f_n(X) \end{pmatrix}, g_k(X) = \begin{pmatrix} g_{1k}(X) \\ \vdots \\ g_{nk}(X) \end{pmatrix} \; and \;\; h(X) = \begin{pmatrix} h_1(X) \\ \vdots \\ h_n(X) \end{pmatrix} \tag{1.6}$$

f, g_k and h are smooth functions, that is to say that these functions are infinitely differentiable with respect to each of their arguments. after calculating the gradient of each element and the Jacobian, the general result will be :

$$\dot{y_j} = L_f h_j(X) + \sum_{k=1}^{m} L_{gk}(L_f^{(r_j-1)} h_j(X))u_k \tag{1.7}$$

if $L_{g_k} h_j(X) = 0$ the inputs u_k do not influence the outputs.

$$y_j^{(r_j)} = L_f^{(r_j)} h_j(X) + \sum_{k=1}^{m} L_{gk}(L_f^{(r_j-1)} h_j(X))u_k \tag{1.8}$$

the decoupling matrix is m x m, $\Delta(X)$:

$$\Delta(X) = \begin{pmatrix} L_{g1} L_f^{(r1-1)} h_1(X) & \cdots & L_{gm} L_f^{(r1-1)} h_1(X) \\ \vdots & \ddots & \vdots \\ L_{g1} L_f^{(rm-1)} h_m(X) & \cdots & L_{gm} L_f^{(rm-1)} h_1(X) \end{pmatrix} \tag{1.9}$$

The vector $\Delta_0(X)$:

$$\Delta(X) = \begin{pmatrix} L_f^{(r1)} h_1(X) \\ \vdots \\ L_f^{(rm)} h_m(X) \end{pmatrix} \tag{1.10}$$

So :

$$Y = \begin{pmatrix} y_1^{(r1)} \\ \vdots \\ y_m^{(rm)} \end{pmatrix} = \Delta_0(X) + \Delta(X) \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \tag{1.11}$$

1.11 State representation of control circuit in multicellular power converter DC-AC(inverter)

considering the system define by :

$$\dot{X} = A(X) + B(X) + H \tag{1.12}$$

$$\text{so the state vector} \qquad \dot{X} = \begin{pmatrix} \dot{Vc_1} \\ \dot{Vc_2} \\ \dot{i_{ch}} \end{pmatrix}$$

and $c1 = c2$ so $ic_1 = ic_2$, and ic can take three possible values depending on each switch case of $(S_1, S_2$ and $S3)$

$$\begin{pmatrix} ic = 0 \\ ic = i_{ch} \\ ic = i_{ch} \end{pmatrix} \tag{1.13}$$

$$ic = c\,\frac{dVc}{dt} \quad \Rightarrow \quad \frac{dVc}{dt} = \frac{1}{c}\,ic \tag{1.14}$$

$$(S_n - S_{n-1})\,ic_{n-1} = c_{n-1}\frac{dVc_{n-1}}{dt} \quad \Rightarrow \quad \frac{dVcn-1}{dt} = \frac{1}{c_{n-1}}\,(S_n - S_{n-1})ic_{n-1} \tag{1.15}$$

$$V_s - L\,\frac{d\,i_{ch}}{dt} - R_{ch}i_{ch} + \frac{V_{dc}}{2} = 0 \tag{1.16}$$

$$V_s = \frac{S_3}{2L_{ch}}V_{dc} + \frac{S_2 - S_3}{Lch}\,Vc_2 + \frac{S_1 - S_2}{Lch}\,Vc_1 - \frac{d\,i_{ch}}{dt} - \frac{R_{ch}}{L_{ch}}\,i_{ch} = 0 \tag{1.17}$$

$$\dot{i_c} = \frac{d\,i_{ch}}{dt} = \frac{S_3}{2L_{ch}}V_{dc} + \frac{S_2 - S_3}{Lch}\,Vc_2 + \frac{S_1 - S_2}{Lch}\,Vc_1 - \frac{R_{ch}}{L_{ch}}i_{ch} \tag{1.18}$$

$$\begin{pmatrix} \dot{\mathbf{Vc_1}} \\ \dot{\mathbf{Vc_2}} \\ \dot{\mathbf{i_{ch}}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{-\mathbf{R_f}}{\mathbf{L_f}} \end{pmatrix} \begin{pmatrix} \mathbf{Vc_1} \\ \mathbf{Vc_2} \\ \mathbf{i_{ch}} \end{pmatrix} + \begin{pmatrix} -\frac{\mathbf{1}}{\mathbf{c}}\mathbf{i_c} & \frac{\mathbf{1}}{\mathbf{c}}\mathbf{i_c} & \mathbf{0} \\ \mathbf{0} & -\frac{\mathbf{1}}{\mathbf{c}}\mathbf{i_c} & -\frac{\mathbf{1}}{\mathbf{c}}\mathbf{i_c} \\ \frac{\mathbf{Vc_1}}{\mathbf{L_{ch}}} & \frac{\mathbf{Vc_2 - Vc_1}}{\mathbf{L_{ch}}} & \frac{\mathbf{V_{dc} - 2Vc2}}{\mathbf{2L_{ch}}} \end{pmatrix} \begin{pmatrix} \mathbf{S_1} \\ \mathbf{S_2} \\ \mathbf{S_3} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{\mathbf{V_{dc}}}{\mathbf{2L_{ch}}} \end{pmatrix} \tag{1.19}$$

Chapter 2

Machine Learning Fault Detection Methodologies

2.1 Introduction

The fault diagnosis in electrical power systems has been a more essential topic lately, owing to the merger of renewable power sources with existing systems. Machine Learning is the latest approach that is taking attention to various power system applications. In this context, using machine learning tools comes out apparent and logical to deal with the challenges of diagnosis in these systems. In this chapter, we're going to show different advanced techniques that can automatically locate patterns in data and classify this data into different classes, and then use the uncovered patterns to predict future data.

2.2 Machine learning

Machine learning is one of the subfields of artificial intelligence that offers an efficient way to capture knowledge in data to incrementally improve the performance of classification models and make data-driven decisions. It has become a ubiquitous technology, and we are taking advantage of it in: email spam filters, self-driving cars, image and voice recognition, and world-class gamers.

There are a number of different algorithms that we can employ in machine learning. The required output is what decides which to use. we'll see the different algorithm types being put to work. Machine learning algorithms characteristically fall into one of two learning types: supervised or unsupervised learning.

2.3 Types of machine learning

Machine learning is usually divided into two main types.

2.3.1 Supervised learning

Supervised learning refers to a sort of machine learning model that can be trained on a set of samples where the desired outputs (or labels) are already known. The models learn from these later results and make modifications to their internal parameters to accommodate the input data. Once the model is well trained, it can make correct predictions approximately unseen or future data.There are two main applications of supervised learning: classification and regression.

Overview of the overall process:

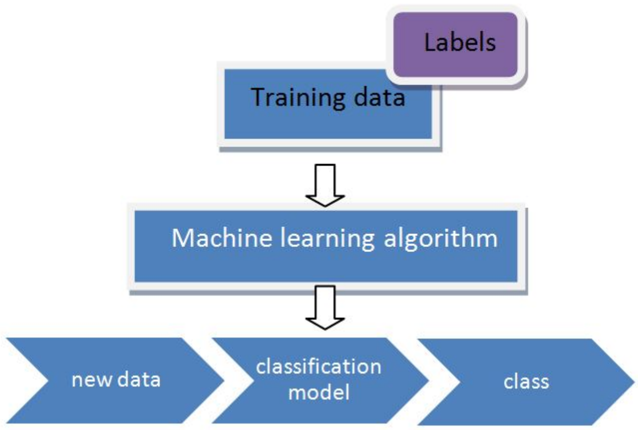


FIGURE 2.1: supervised learning process

2.3.2 Unsupervised learning

When the records used to train are neither classified nor labeled, unsupervised machine learning algorithms are used. Unsupervised learning investigates how systems can infer a function from unlabeled data to explain a hidden shape. The system does not recognize an appropriate output with certainty under any conditions. Instead, it draws inferences about what the output should be from datasets.

2.3.3 Semi-supervised learning

semi-supervised learning, where the teacher gives an incomplete training signal: a training set with some (often many) of the target outputs missing. We will focus on unsupervised learning and data clustering in this blog post.

2.3.4 Reinforcement learning

Reinforcement learning employs a positive incentive system for correct behavior and a negative reward system for incorrect behavior. As a result, the technique assigns positive values to desired activities and negative values to unwanted acts in order to influence the agent. This instructs our agent to seek the largest overall reward over the long term in order to arrive at the best option. These long-term objectives prevent the agent from stopping there. The system eventually learns to avoid bad acts and only do positive ones.

Trial and error is used to learn through interaction with the environment.

2.4 Classification

In Machine Learning and Statistics, Classification is the problem of identifying to which of a set of categories (subpopulations), a new observation belongs, on the basis of a training set of data containing observations and whose categories membership is known.

2.4.1 Types of Classification

Classification is of two types:

2.4.1.1 Binary Classification

When we have to categorize given data into 2 distinct classes. Example – On the basis of given health conditions of a person, we have to determine whether the person has a certain disease or not [2].

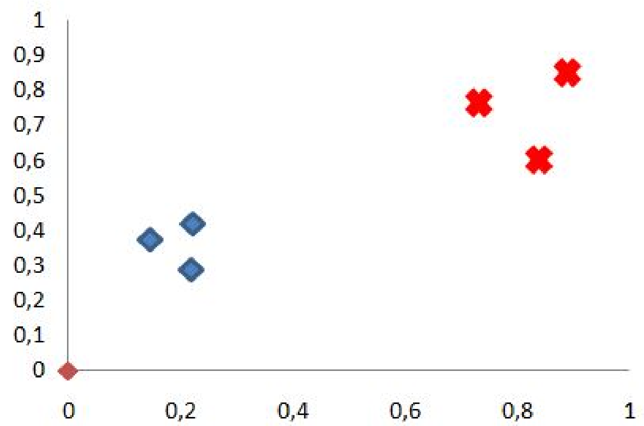


FIGURE 2.2: example of Binary classification representation

2.4.1.2 Multiclass Classification

The number of classes is more than 2. For Example – On the basis of data about different species of flowers, we have to determine which specie does our observation belongs to.

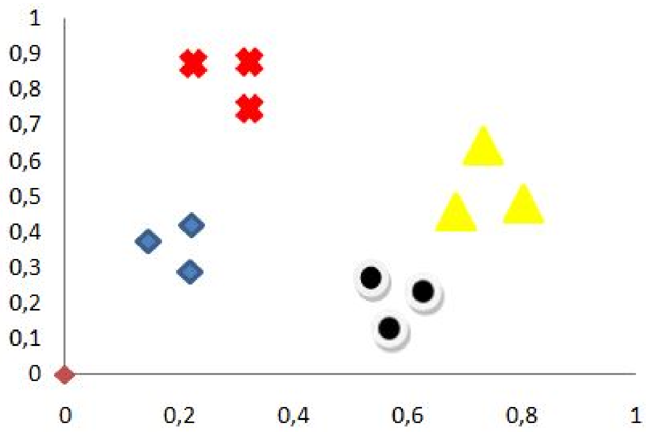


FIGURE 2.3: example of Multiclass classification representation

2.5 Dataset

2.5.1 Learning data in supervised and unsupervised learning

In supervised learning , experts classify the data and teach the model precisely what it needs to discover. For example, in the area of spam detection, the input is any text, and the label makes it apparent if the message is spam. Supervised learning is stronger since we don't allow the model draw its own inferences from data beyond the limitations labeled with our labels.

In unsupervised learning , people feed the model raw, unlabeled data, and the model finds patterns in the data. For example, recognizing the level of similarity or difference between two data samples based on common features extracted. This helps the model make inferences and come to conclusions, such as separating similar images or grouping them into clusters.

2.5.1.1 Training Data

Training data refers to the initial data set passed to the machine learning model, on which the model is trained. Humans learn best from examples, and machines also need a set of data to learn patterns from it. In most cases, the training data contains input:annotation pairs collected from various sources that are used to train the model to perform a specific task with a high level of accuracy.They may consist of raw data (images, texts, or sound) containing annotations such as bounding boxes , labels, or links. Machine learning algorithms learn annotations from training data to handle new, unlabeled data in the future. All training methods start with the collection of raw data from various sources. Raw data can be of any kind: text, images, sounds, videos, etc. However, to tell the model what to look for in this data, we must add .These annotations help us to regulate learning by verifying that the model focuses on the features we specify, instead of making conclusions from other related (but not conditioned) data items.

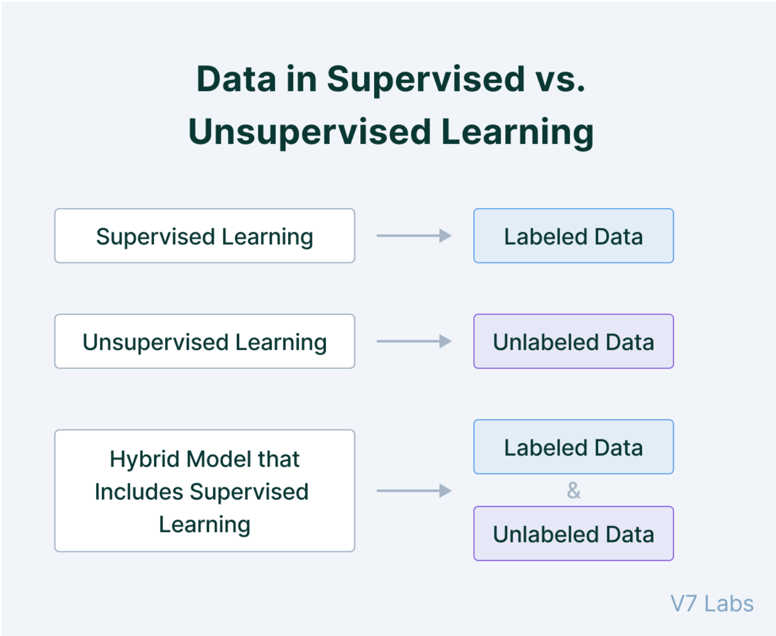


FIGURE 2.4: Data in supervised vs unsupervised learning

All incoming data must have an appropriate label to allow the machine to move in the direction of what the forecast should look like. Such a processed dataset can be obtained using humans, and sometimes other ML models that are accurate enough for reliable labeling.Once the labeled dataset is ready to be passed to the AI, the training phase begins .On it, the model tries to identify important features that are common to all the examples that we have assigned labels. For example, if we segmented several cars in the imagery, then she will understand that wheels, rear-view mirrors and door handles are features that correlate with a car. The models continually test themselves on the validation data set produced before the training process.After the model is completed, the last check is performed on the test dataset (a set that the model has never seen before); this gives us insight into the performance of the model on relevant new examples.

The training data includes the training, validation, and test datasets. The more training data we have, the more accurate the model will be.

2.5.1.2 Labeled Data

Labeled data is data supplemented with labels/classes containing meaningful information.Here are some examples of labeled data: images tagged with cat/dog, emails/messages tagged as spam, stock market price predictions (labeled as future state), nodule malignancy with polygon highlighting, audio files with information about what words are spoken in them. Accurately labeled data allows the machine to recognize patterns according to the task, so it is widely used in solving complex problems.

2.5.2 Datasets for training, validation and testing

No AI model can be trained and tested on the same data.the estimate of the model will be biased, because the model is being tested on what it already knows. This is analogous to giving students the same questions in an exam that they have already answered in class. This way we don't know if the student remembered the answers or really understood the topic.The same rules apply to machine learning models.

Here are their percentages of data volumes:

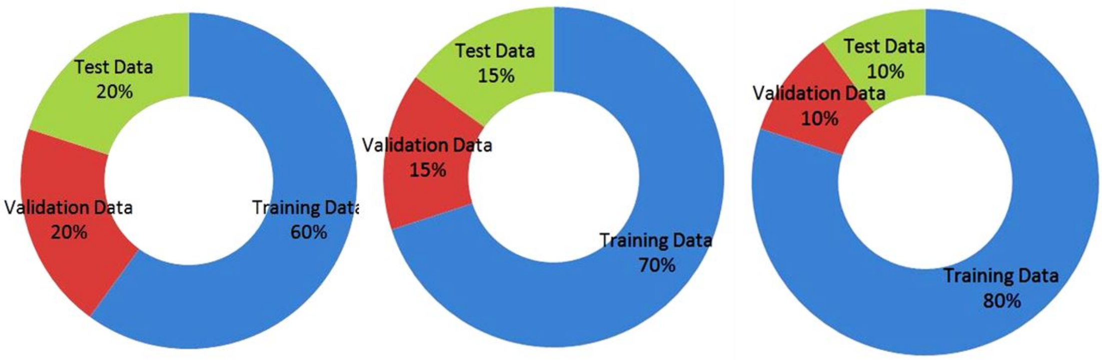


FIGURE 2.5: training data needs

Training data - at least 60% of the data must be used for training.

Validation data - a sample (10%-20%) of the total data set used for validation and periodically checked against the model during training. This validation dataset should be a representative sample of the training dataset.

Test data - This dataset is used to test the model after it has been fully trained. It is separated from both the training set and the validation set. After training and validation, the model is tested on the test set. The data in the test set should look exactly like the real data will look after the model is deployed.

There can be multiple test sets in a shared dataset Each test set can be used to check if the model has trained enough for a specific application scenario.

2.5.3 Data Quantitative Requirements

we need to have at least 1000 training examples for every possible class in our scenario . If we utilize 10% of the data as our test set, we should be able to assess class correctness with a margin of error of at least 1%.For reference,1000 examples is a sufficient data set. 10000 is a great data set,100 thousand-1 million is an excellent data set. High-quality models are trained on large amounts of training data, and for good reason - modern neural network architectures work great because they are able to store many weights (parameters) efficiently. However, if we don't have much training data, then we can only use a fraction of the potential of wer model.The size of the data set also depends on the scope of our problem and the variance of each class. However, if we require a generic human recognizer then a collection of 10,000 instances will only represent just a portion of the differences in sizes, positions, look, and clothing styles. Therefore, a class with high variance, such as "human", requires much more training data.

2.5.4 Improving the Quality of Training Data

The high quality of dataset markup is required for the machine learning model to operate correctly. The term “qualitative data” refers to cleaned data that contains all the attributes that model training depends on. The consistency and correctness of the labeled data may be used to rate the quality .here is 4 characteristics of qualitative data for training ML models:

1. Relevance - The data set should contain only those features that provide meaningful information to the model. Identifying important features is a complex task that requires knowledge of the area and a clear understanding of which features should be considered and which should be eliminated.

2. Consistency - Similar examples should have similar labels, ensuring that the dataset is homogeneous.

3. Uniformity - The values of all attributes must be comparable across all data. Irregularities or the presence of outliers in datasets adversely affect the quality of training data.

4. Comprehensiveness - The dataset must include enough parameters or characteristics to ensure that no edge situations are missing. The dataset should contain enough samples of these edge cases so that the model can learn them as well.

2.5.5 Metrics in classification problems

Metrics are used in machine learning tasks to measure the quality of models and compare different algorithms, and their selection and analysis is an essential element of a data scientist's job. We'll look at several quality criteria in classification issues and talk about what matters when picking a metric and what may go wrong.

2.5.5.1 Accuracy, precision, recall and specificity

An important concept needs to be introduced to describe these metrics in terms of classification errors, the confusion matrix .Suppose we have two classes and an algorithm that predicts whether each object belongs to one of the classes, then the classification error matrix will look like this:

Here , is the response of the algorithm on the object, and is the true label of the class on this object. Thus, there are two types of classification errors: False Negative (FN) and False Positive (FP).

	$y = 1$	$y = 0$
$\hat{y} = 1$	True Positive (TP)	False Positive (FP)
$\hat{y} = 0$	False Negative (FN)	True Negative (TN)

FIGURE 2.6: the classification error matrix

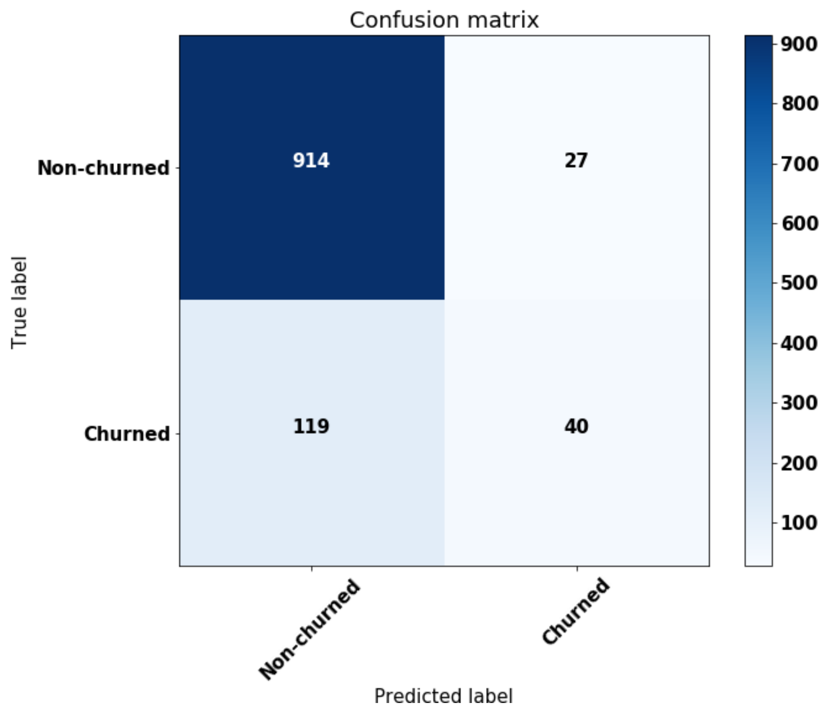


FIGURE 2.7: the response of the algorithm on the object

2.5.5.2 Accuracy:

An intuitive, obvious and almost unused metric is accuracy - the proportion of correct answers of the algorithm:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

This metric is useless in problems with unequal classes, and this is easy to show with an example.

Let's say we want to evaluate the performance of a mail spam filter. We have 100 non-spam emails, 90 of which our classifier determined correctly (True Negative = 90, False Positive = 10), and 10 spam emails, 5 of which were also correctly determined by the classifier (True Positive = 5, False Negative = 5). Then accuracy:

$$accuracy = \frac{5 + 90}{5 + 90 + 10 + 5}$$

However, if we just predict all emails as non-spam, we get a higher accuracy:

$$accuracy = \frac{0 + 100}{0 + 100 + 0 + 10}$$

At the same time, our model does not have any predictive power at all, since initially we wanted to identify spam emails. The transition from a common metric for all classes to individual indicators of class quality will help us overcome this.

2.5.5.3 Precision, recall and F-measure

To assess the quality of the algorithm on each of the classes separately, we introduce the metrics precision (accuracy) and recall (completeness).

$$precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$recall = \frac{TP}{TP + FN} \tag{2.3}$$

Precision can be interpreted as the proportion of objects called positive by the classifier and at the same time are really positive, and recall shows what proportion of objects of a positive class out of all objects of a positive class the algorithm found.

It is the introduction of precision that does not allow us to write all objects into one class, since in this case we get an increase in the False Positive level. Recall illustrates the algorithm's capacity to recognize a particular class in general, while precision demonstrates the ability to differentiate this class from other classes.

2.5.6 Training Data Preparation

2.5.6.1 Data cleaning

Raw data can be very messy and corrupted in many ways. If not properly cleaned, they can skew results and cause the AI model to produce erroneous results. Data cleaning is the process of repairing or deleting inaccurate, corrupted, or duplicated data from a dataset.. The steps in the data cleaning process depend on the particular data set.

1. Check for duplicates - the same sample data may occur several times in a dataset. This may be generated by gathering data from several sources, resulting in comparable data. They need to be deleted since they may cause the model to overfit certain patterns and provide incorrect predictions.

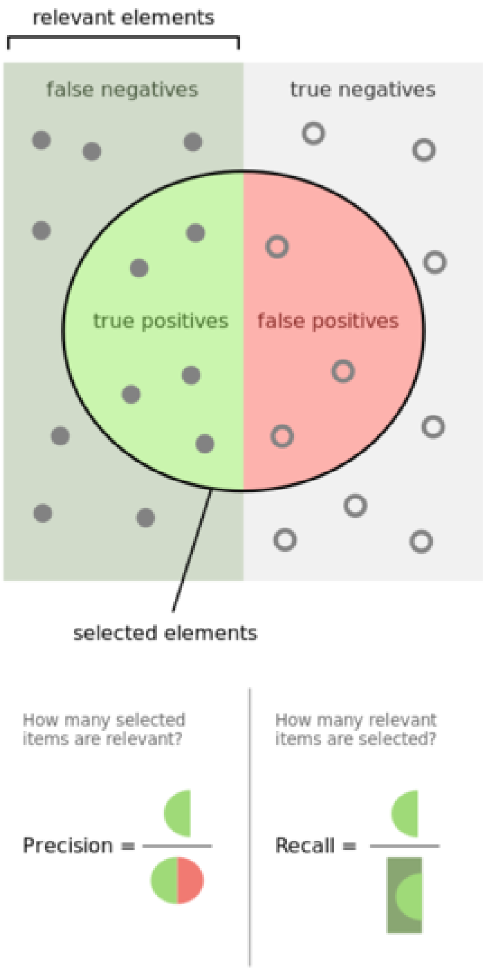


FIGURE 2.8: Precision and recall illustration

2. Eliminate outliers - some parts of the data behave differently from the rest of the data. An example would be the SessionID, which is constantly found in the weblog data. This may be due to some malicious activity that does not need to be passed to our model. As a result, monitoring emissions is one method of removing data that should not be supplied to the machine.

3. Fix Structural Errors - In certain conditions, the dataset may include faulty markup. For example, "Dog" and "dog" are regarded to be separate classes, although "doog" and "dog" are different according to a typo resulting in misclassification.

4. Check for missing values - there may be components in the dataset for which data examples are sorely missing attributes/features. We can overcome this issue by simply not include these components in the training dataset.

2.5.6.2 Data markup

Data markup is the process in which we assign a value to data in the form of a class or a label. Data labeling can be performed by employees, operators in the control loop, or any automated machine that speeds up the labeling process.

1. Establish the gold standard - Data scientists are considered the gold standard in data labeling, labeling raw data with maximum sensitivity and accuracy. Their markup is considered a guide for the annotation team and can be used as responses when screening annotation options.

2. Don't use too many labels - splitting a dataset into a large number of classes can confuse employees when annotating it. In addition, to select among the many labels, the analysis of more features will be required. For example, it will be difficult for annotators to mark up data with classes such as "Very Expensive", "Expensive", "Less Expensive".

3. Use multiple passes - data markup should be done by multiple annotators. This is necessary to improve the overall quality of the data. While this is more time consuming and resource intensive, this approach is used to build consensus within the team.

4. Establish a validation system - to reduce the chance of errors, ready-made data markup should be verified by another person or through self-improvement checks. This will allow any annotator to understand where he can improve, his level of accuracy, and the kind of training needed to improve his work.

2.5.6.3 standard deviation

Standard Deviation in simple terms, this is a measure of how scattered a set of data is.

By calculating it, you can find out whether the numbers are close to the average value or far from it. If the data points are far from the mean, then there is a large variance in the data set .When the data is more spread out, the standard deviation increases.

Standard Deviation Formula :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)}{n-1}} \tag{2.4}$$

Where:

σ standard deviation.

X_i are different sample values,

μ is the arithmetic mean of the sample, n is the sample size.

2.6 The k Nearest Neighbors (kNN)

The k Nearest Neighbors (kNN) method is a popular classification algorithm that is used in various types of machine learning problems. Along with the decision tree, this is one of the most understandable approaches to classification [57]. On an intuitive level, the essence of the method is simple: we look at the neighbors around, which of them prevail, that is what we are. Formally, the technique is based on the compactness hypothesis: if the space metric among examples is delivered successfully, then similar examples are much more likely to lie in the same class than in different ones.

- In the case of using the method for classification, the object is assigned to the class that is the most common among the k neighbors of this element, whose classes are already known.

- In the case of using the method for regression, the object is assigned the average value of the k objects closest to it, the values of which are already known.

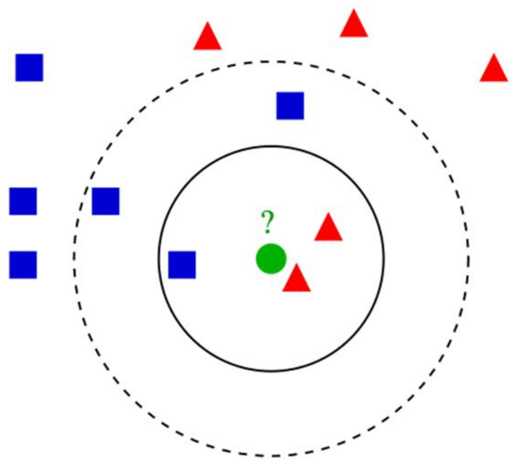


Figure 2.9: An example of k-nearest neighbor classification.

We have a test sample in the form of a green circle. The blue squares will be designated as class 1, the red triangles as class 2 .

- The green circle must be classified as class 1 or class 2 . If the area we are considering is a small circle, then the object is classified as 2nd class, because inside this circle there are 2 triangles and only 1 square.

- If we consider a large circle (with a dotted line), then the circle will be classified as 1st class, since there are 3 squares inside the circle as opposed to 2 triangles.

2.6.1 Theoretical component of the k-NN algorithm

2.6.1.1 Euclidean metric

Beyond the simple explanation, the expertise of the underlying mathematical components of the k-nearest neighbors algorithm is absolutely necessary Euclidean metric (Euclidean distance, or Euclidean distance) - a metric in Euclidean space, the distance between two points in Euclidean space, calculated by the Pythagorean theorem [67]. Simply expressed, this is the lowest feasible space across locations A and B. Although Euclidean distance is useful for small dimensions [3], it does not work for large dimensions and for categorical variables. The disadvantage of Euclidean distance is that it ignores the similarity between attributes. Each of them is seen as completely different from all the others.

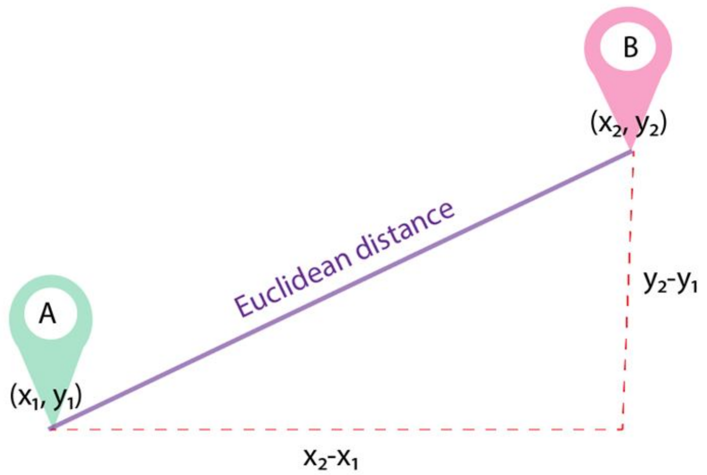


FIGURE 2.10: The Euclidean distance

The Euclidean distance is calculated using the following formula:

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{2.5}$$

Another important component of the method is normalization [20]. Different attributes usually have different ranges of values represented in the selection. For example, attribute A is represented in the range 0.01 to 0.05, and attribute B is represented in the range 500 to 1000). In this case, the distance values can be highly dependent on attributes with larger ranges. accordingly, the data in most cases is going through normalization. There are two essential approaches to normalize data in cluster analysis: Min-Max normalization and Z-normalization .

When undertaking some type of study in which we have multiple variables measured on different scales and we want each of the variables to have the same range, we often normalize variables.

2.6.1.2 Feature selection

Experts in feature selection face a variety of challenges, including extracting and decreasing features. The features are what make every class in system have a variance or significance in term of it, some times required a sophisticated features generating [30] with specialists in order to increase features so raising dimensions of the feature space, that is in the event of a shortage of the required features, One of the most common techniques to improve features is to work with statistical characteristics[9] such as spectral density, standard deviation, variance, and so o, using the available features. in the case of accumulation of the features, in order to reduce it , the features that ensure the best results and distinct differences in the distribution of situations must be identified. One of the most common ways to reduce the number of features and ensure the preservation of high accuracy is to use principle component analysis(PCA)[24], which allows you to reduce the number of features while also creating new features with different significance for each class.

2.6.1.3 Z-normalization

which normalize every value in data set so that the mean of every data set is 0 and the standard deviation is 1.

$$Nv = (x - M[x])/\sigma[x] \tag{2.6}$$

where:
σ is the standard deviation.
x is the original value.
$M[x]$ is the mean of data.

In this case, most of the values fall within the range.

2.6.2 Min-Max normalization

Both values will be changed to the same scale/range when the data of the various scales is normalized. Both values, for example, will be in the range of 0 to 1.

The data will have a value of 0 for the lowest value and a value of 1 for the greatest value, with all other values falling between 0 and 1.

$$z = \frac{x_i - min(x)}{max(x) - min(x)} \tag{2.7}$$

where: x_i is the original value. $min(x)$ is the minimum value in data set . $max(x)$ is the maximum value in data set .

What is the course of action?

- Download our data.
- Initialize k by choosing the optimal number of neighbors.

For each of the data samples:

. Using the data, calculate the distance between the querying example and the actual example.

. Add the index of the sample to the ordered collection, just like its distance.

- Sort an ordered list of distances and indices in ascending order from least to greatest.
- Choose the first k data points from the sorted collection.
- Take the labels of the chosen k entries.
- If we have a regression problem, we'll return to the average of the previously selected k labels.
- If we have a classification problem, we'll return to the most frequently occurring value of the previously selected labels k .

2.7 Spectral analysis

One of the most powerful techniques for processing experiments is spectral analysis. In particular, it is used for data analysis, detection of characteristic frequencies [41], noise suppression[72], etc.

The spectrum of the data set y(x) is some function of another coordinate (or coordinates, if we are talking about a multidimensional spectrum [44]) $F(\omega)$, Obtainable through the use of a specific algorithm .

Fourier transform, power spectrum, and wavelet transform are examples of spectral analysis approaches.

2.7.1 Fourier Transform

The Fourier transform mathematically presents the signal y(x) as an infinite summation of sinusoids of the format $F(\omega)\ sin(\omega x)$ .The function $F(\omega)$ is called the Fourier transform, or the Fourier spectrum of the signal. Its argument represents the frequency of the appropriate component of the signal. The inverse Fourier transform transforms the spectrum $F(\omega)$ into the original $y(x)$ signal. By definition,

$$F(\omega) = \int_{-\infty}^{\infty} \mathrm{y(x)} * e^{-i\omega x}\,\mathrm{d}x \tag{2.8}$$

Even if the signal is real, the Fourier transform is a complex value [11], as defined by definition in formula 2.8.The Fourier transform is particularly important in many mathematical applications, and a very efficient algorithm for it has been constructed. Called the FFT (Fast Fourier Transform) algorithm. it's so popular due to its super-economical, which in almost all mathematical packages is organized as a subroutine.

The FFT algorithm has a rather strong limitation, which is not critical in practice. The idea is that the direct Fourier transform argument, i.e. the sample size $y(xi)$, must contain exactly $2n$ elements (n is any integer). Accordingly, the result of the FFT algorithm is a vector with $1+2^{(n-1)}$ elements [6]. If the number of data does not match the power of 2, then to run the FFT algorithm, it is enough to pad the missing elements with zeros.

Considering the most typical situation for a physical experiment in calculating the Fourier spectrum of a real signal. we utilize the discretization of the following deterministic signal as model data (Figure 2.11(a)) To further comprehend the Fourier transform:

$y(x) = 1\sin(2\pi 0.05x) + 0.5\sin(2\pi 0.1x) + 0.1\sin(2\pi 0.5x)$

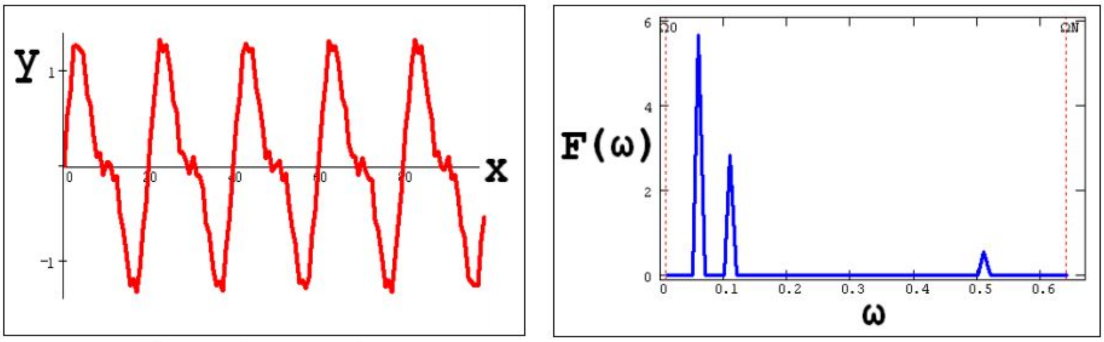


FIGURE 2.11: (a)-Model data (b)-Selective Fourier Spectrum

On Figure 2.11 (b) shows the results of the FFT algorithm in the form of the Fourier spectrum modulus $|F(\omega)|$, since, again, the spectrum itself is complex. It is very useful to compare the obtained amplitudes and the location of the spectral peaks in Figure 2.11 (b) with the definition of sinusoids [22] in the formula 2.11. It is significant that if we subject the obtained absolute value of the Fourier spectrum Figure 2.11 (b) to the inverse transformation

Fourier, which is also provided by the FFT algorithm, then the profile of the original signal will be reconstructed correctly, but will be shifted by a certain distance along the x-axis Figure 2.12. This is because taking the absolute value of the complex spectrum destroys information about the relative phase of the data samples. Otherwise, the signal $y(x)$ is restored with great accuracy, which is typical for a smooth signal change [8]. If, however, the complex Fourier spectrum is used as the input data of the inverse Fourier transform, then the match will be complete [69].

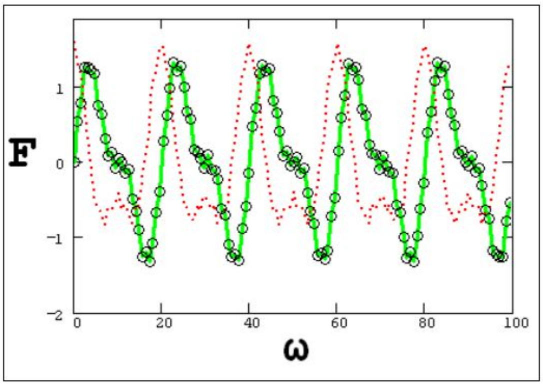


FIGURE 2.12: Inverse Fourier transform of the complex (circles) and real (dots) Fourier spectrum

2.7.1.1 Conclusion

The k-nearest neighbors approach is a basic supervised machine learning algorithm that may be used to handle classification and regression issues. It is simple to implement and understand, but has a significant drawback - a significant slowdown when the amount of data grows.

The k-Nearest Neighbor algorithm classifies based on the distance to a certain number (K) of training samples. This family of algorithms is called instance-based learning, since there are no parameters to study. the model assumes that the distance is sufficient for inference, otherwise it makes no assumptions about the underlying data or its distribution.

Chapter 3

Implementation And Results

3.1 Data collection and processing:

The following figure represents the structure of the multicellular power converter, consisting of three cells, meaning that it has two capacitors, which are concerned with the study:

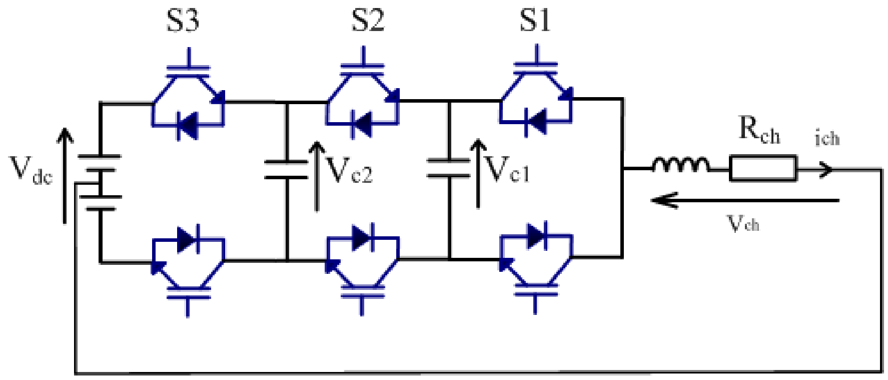


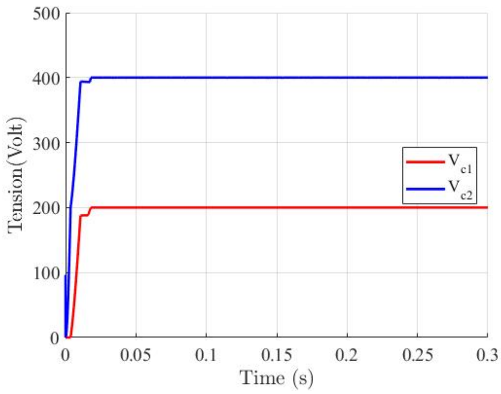
FIGURE 3.1: electrical circuit of multicellular power converter DC-AC

The data will be collected during the operation of this converter using two methods of control; sliding mode control and exact linearization control, When data collection, the transitory regime was excluded, and this is because the system at this stage has not yet reached a stable state or stable behavior in order to deal with its data, because if the transitory regime is adopted in this study, its results will be suspect. There will be two possible situations, a healthy mode and faulty mode, the all possible cases during the functioning of this system are:

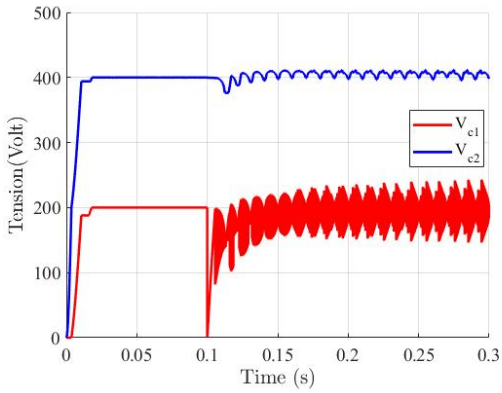
- Healthy mode, (absence of failure).
- Faulty C1, (failure in the first capacitor).
- Faulty C2, (failure in the second capacitor).

- Faulty C1C2, (failure in both capacitors).

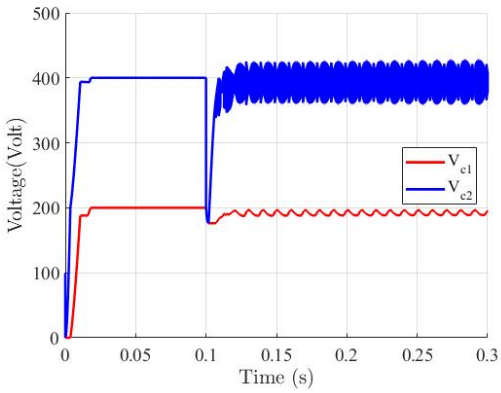
The voltages and current signals forms in sliding mode control with transitory mode are illustrated in Figure3.2 and 3.3 :



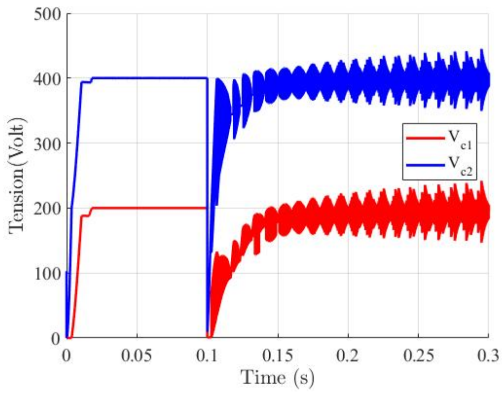
(A) Healthy case



(B) Faulty C1 cas

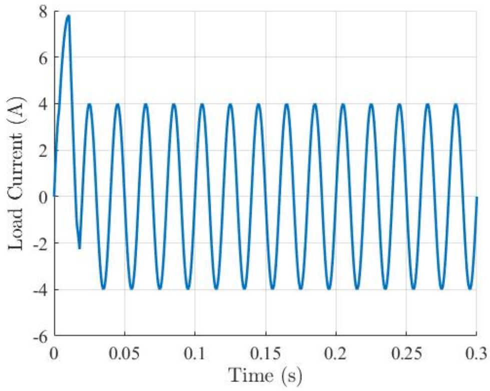


(C) Faulty C2 case

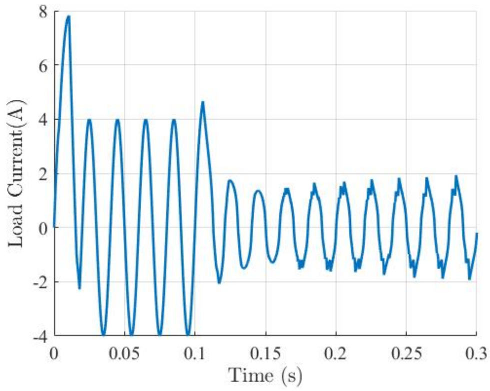


(D) Faulty C1C2 case

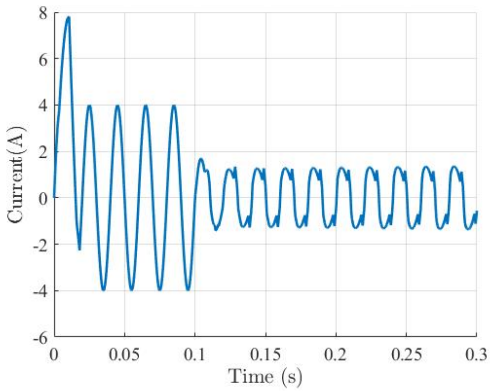
FIGURE 3.2: C1 and C2 voltages in the four cases in sliding mode control with the transitory mode (time domain).



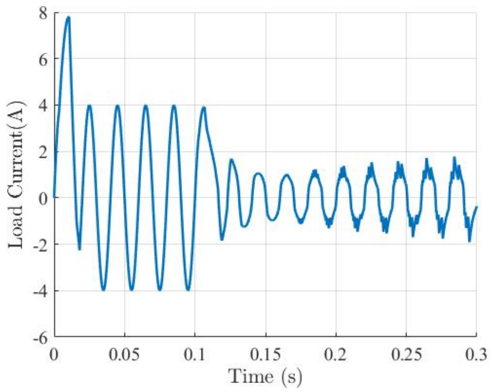
(A) Healthy case



(B) Faulty C1 cas



(C) Faulty C2 case

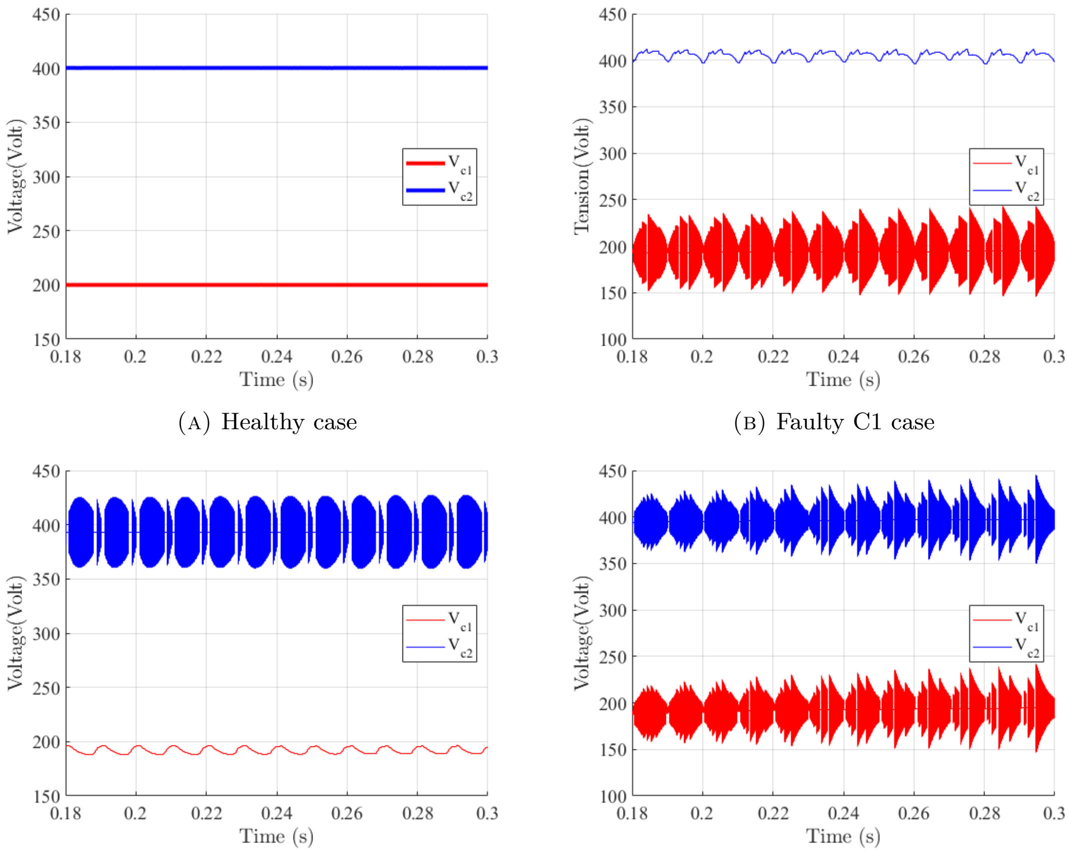


(D) Faulty C1C2 case

FIGURE 3.3: Load current in the four cases in sliding mode control with the transitory mode (time domain).

3.1.1 Sliding mode control

The signal voltages forms in operation with sliding mode control without transitory mode are illustrated in the Figure 3.4 bellow:



(A) Healthy case

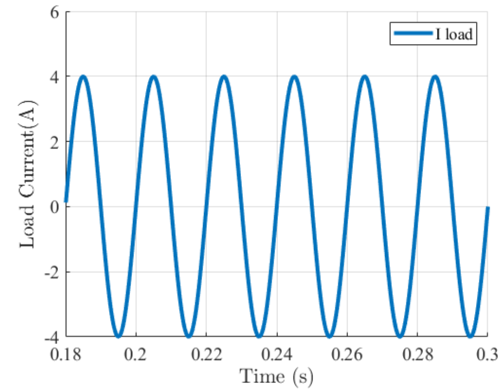
(B) Faulty C1 case

(C) Faulty C2 case

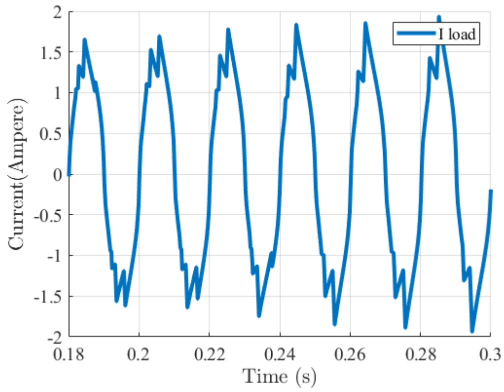
(D) Faulty C1C2 case

FIGURE 3.4: C1 and C2 voltages in the four cases in sliding mode control (time domain).

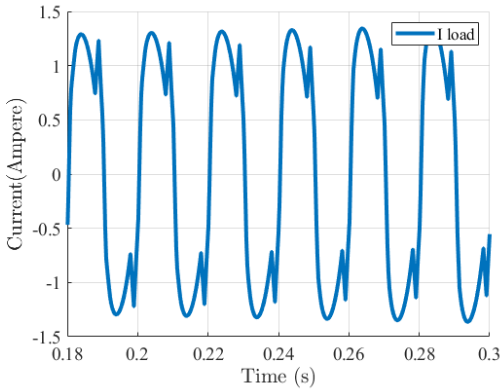
As a result of the signal forms, voltages signals are gradually shift to the permanent mode, and the current established in a sinusoidal form. The appearance of harmonics in the remaining forms is also noted, and the faulty situations are classified according to the failure of each capacitor. As a result, in the voltage of the first capacitor (Vc1) signal, Figure 3.4b shows the appearance of high level of harmonics, compared to the voltage of the second capacitor (Vc2) signal, it appears to preserve its natural shape, but because the two capacitors are connected through the circuit this affects the Vc2 signal, it shows the presence of small noises compared to Vc1 signal. As a result of the analysis, the failure in this situation is in the first capacitor(C1), which is logical. This last will not operate normally in the event of a failure case in it, and its behavior will be oscillating in the permanent mode. to the latest results, depending on the extracted figures, it can be observed that the remaining cases Figure3.4c and 3.4d based on the first conclusion, Figure 3.4c indicates a fault in the second capacitor(C2) because large harmonics are observed in the signal of Vc2 and the signal of Vc1 affected by this failure. Figure 3.4d indicates the occurrence of a failure in both capacitors, and this is evident through the emergence of large and clear harmonics in both voltages. As indicated in the Figure 3.5, the influence of capacitors faults on the load current appears, with the shape of the pulses changing from smooth to sharp, affecting the rotation of the machines, creating vibrations and a high temperature of the device.



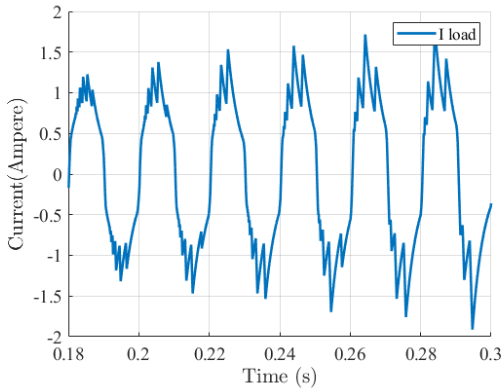
(A) Healthy case



(B) Faulty C1 cas



(C) Faulty C2 case



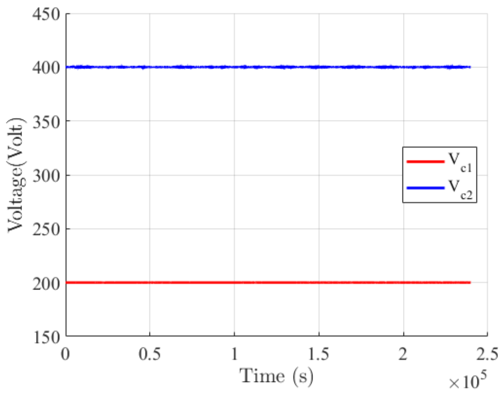
(D) Faulty C1C2 case

FIGURE 3.5: Load current signals in the four cases in sliding mode control (time domain).

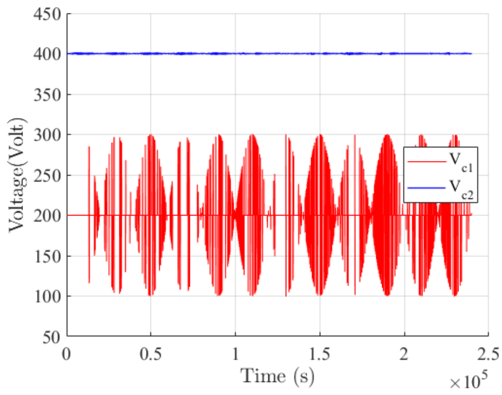
The current signals in the healthy mode, which is entirely healthy and has a sinusoidal form, are represented in figure 3.5a. When there is a problem on C1. A sudden anomaly will emerge at the peak of each pulse, as seen in Figure 3.5b , causing the peak's form to become non-smooth and spiky. In the event of a failure in C2, the fault will also appear at the peak of each pulse with a slight difference in the harmonic shape as shown in Figure 3.5c . However, when there is a fault in both capacitors, the harmonics in the current signal will be greater and more sharp, as the figure shows. It can be said that in sliding mode control the current signal maintains its sinusoidal form in all cases, with failure harmonics appearing when a fault occurs.

3.1.2 Exact linearization mode

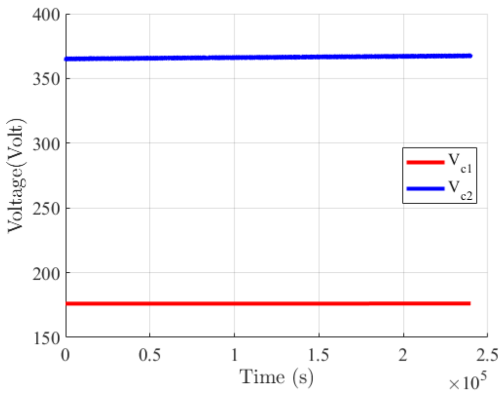
Here the reading of the results of the signal forms differs to the first mode, the Figure3.6 bellow are the signal forms of voltages Vc1 and Vc2 in the four cases:



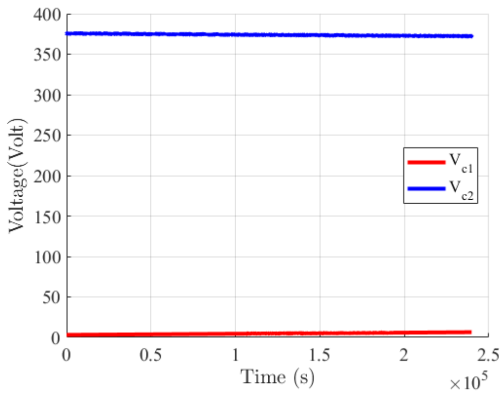
(A) Healthy case



(B) Faulty C1 cas



(C) Faulty C2 case

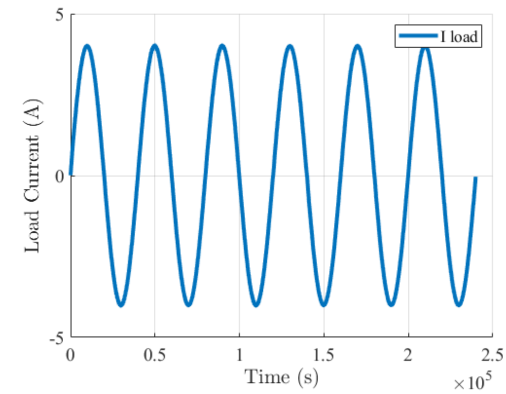


(D) Faulty C1C2 case

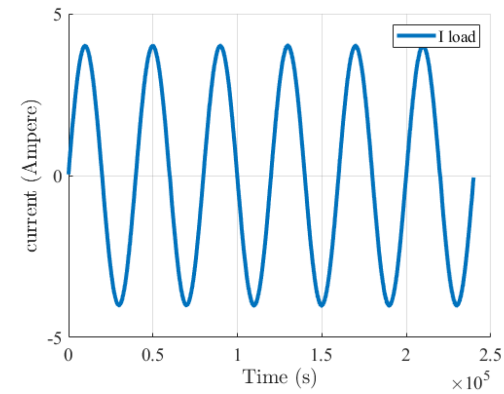
FIGURE 3.6: C1 and C2 voltages in the four cases in exact linearization control (time domain).

Looking at the signals forms obtained in the exact linearization control Figure 3.6, the signals of the two voltages Vc1 and Vc2 are in their normal form healthy mode Figure 3.6a in the absence of faults, except that a slight noise appears in Vc2. The signals of Vc1 and Vc2 in the malfunctioning of C1 case are illustrated in Figure 3.6b. The C1 faults are plainly obvious in the signal of Vc1 by the appearance of huge harmonics in it, and the second was affected with. as illustrated, the appearance of noise in the signal of Vc2. The forms significance of the voltages signals differs in the case of the C2 faults; the shape of the signals voltages Figure 3.6c does not indicate whether there is a fault or not, meaning that this control is not sensitive to faults. Both signals are affected in amplitude, where the value of Vc1 and Vc2 in the healthy case was 200 volt and 400 volt, respectively, and where, after a malfunction in the C2, both voltages drop to 176 volt and 365 volt, respectively, and without condoning on the appearance of some small noises in the signal of Vc2.

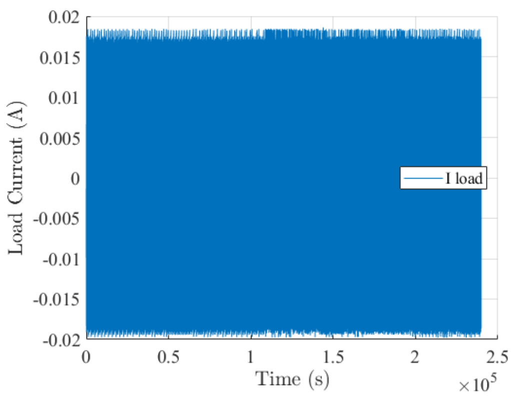
In the event of a malfunction in both capacitors, according to the Figure 3.6d that shows the voltages signals Vc1 and Vc2. The existence of a tiny noise in Vc2 draws attention here, But the voltage Vc1 is free of noise, which indicates that it is not affected by the faults of the capacitor. However, there was a drop in the amplitude of the Vc2, which was previously high, that it was before faults 400 volt but then it became 374.6 volt. Vc2 was also affected by the fault, but more significantly, as there was a considerable fall in the amplitude of the voltage Vc1, as the signal amplitude decreased from 200 volt to 3.9 volt, and this was greatly influenced by the fault. as it is noted that all voltages are taken in the permanent system only. With regard to current, the following Figure 3.7 show the load current in all cases:



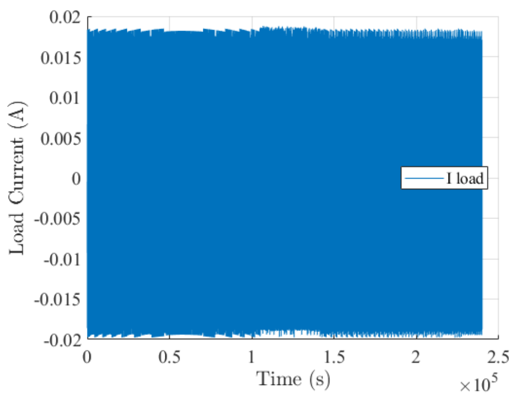
(A) Healthy case



(B) Faulty C1 case



(C) Faulty C2 case



(D) Faulty C1C2 case

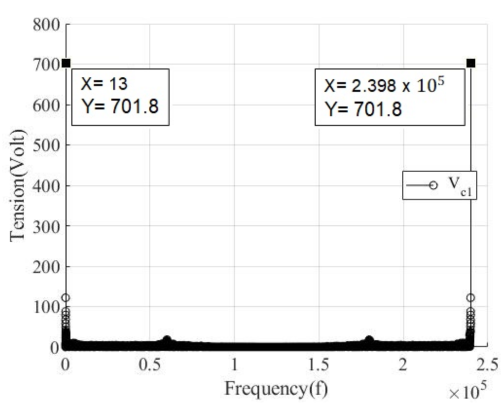
FIGURE 3.7: Load current signals in the four cases in exact linearization control (time domain).

Despite the presence of a defect in the first capacitor, the current signals in Figure 3.7a and Figure 3.7b, which are the healthy mode and the faulty C1 mode, does not show any harmonics or distortion of the signal shape .indicating that the current arrives in the health conditions even when there is a failure in the first capacitor. In Figures 3.7c and 3.7d, which is the case of faulty C2 and the case of faulty C1C2, the current signals appears filled with harmonics to the extent that the shape of the current signal is distorted, which indicates that in the event of a failure in the second capacitor or both together, the current does not reach the health condition of the load. The objective of this part is to preview the data and to make a partial comparison between the sliding mode control and the exact linearization control, for this system.

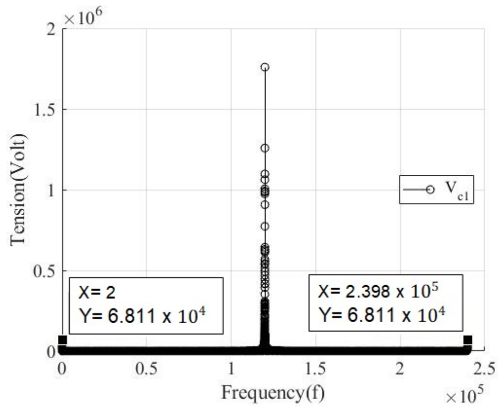
Based on the resulting figures and observations, it appears that the sliding mode control data makes sense; In the absence of a failure, the current and the voltages are in the normal state. In the event of a defect in one or both capacitors, the harmonics appear in the signal of the voltage of the capacitor concerned with the defect, in addition to the influence in the load current.

3.2 Data preparation

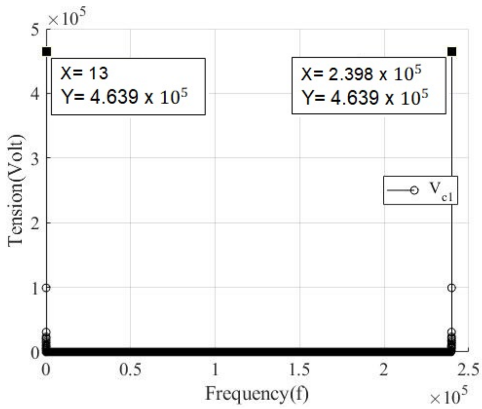
The data spectrum analysis is the subject of this research. As a different look, the data will be transformed from the time domain to the frequency domain using the Fast Fourier transform, to track the spectral distribution. The Fast Fourier transform (FFT) is a typical signal processing transformation method for converting data to the frequency domain [64]. The fast Fourier transform allows you to transform data and drop it into frequency component of sinus and cosinus. So the real reason behind the need the FFT is speed or efficiency. The samples obtained for the signals of capacitor voltages and load current in the different cases during the converter's operation after transforming data to the frequency domain are shown in the following Figures 3.8 and 3.9 in sliding mode control, Figure3.10 and 3.11 in exact linearization control:



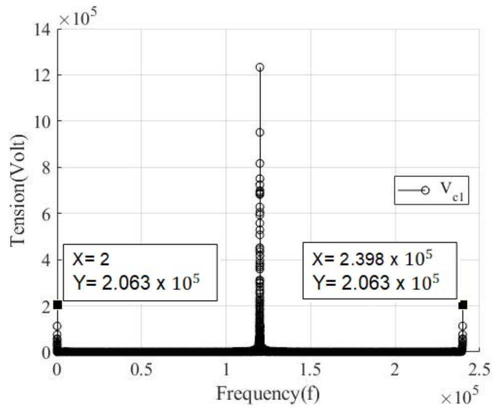
(A) Healthy case



(B) Faulty C1 case

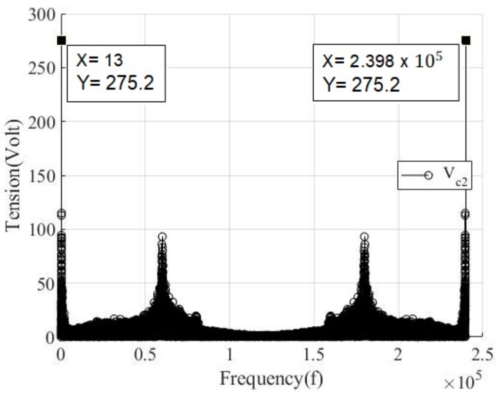


(C) Faulty C2 case

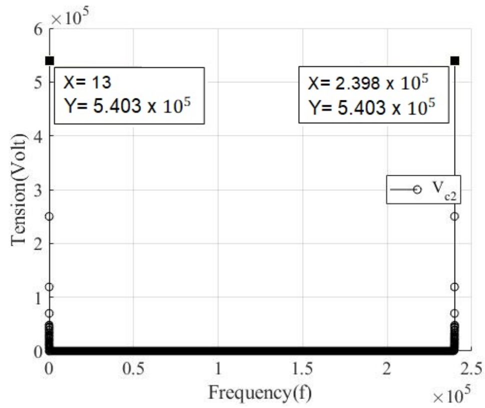


(D) Faulty C1C2 case

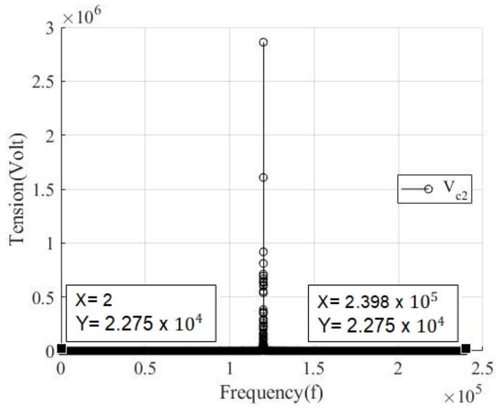
FIGURE 3.8: C1 voltages in the four cases in sliding mode control (frequency domain).



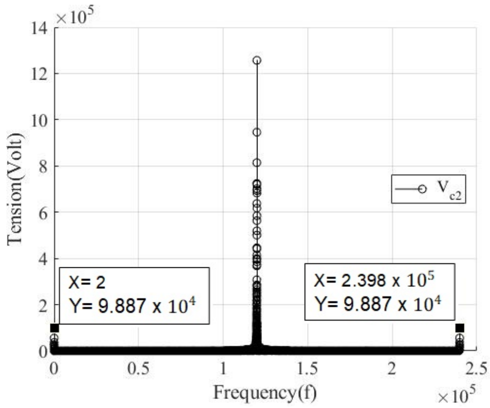
(A) Healthy case



(B) Faulty C1 cas

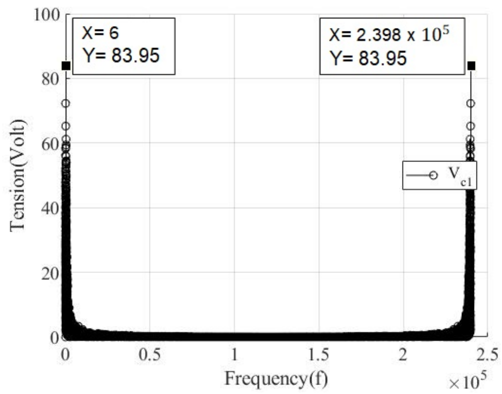


(C) Faulty C2 case

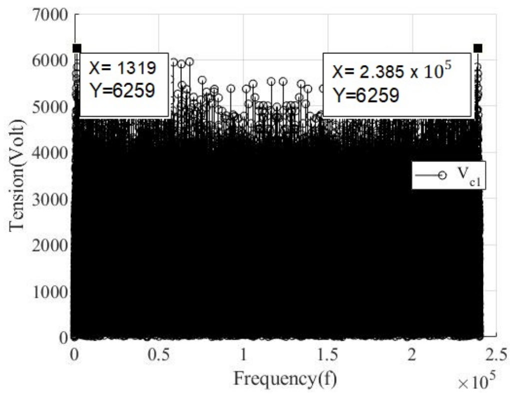


(D) Faulty C1C2 case

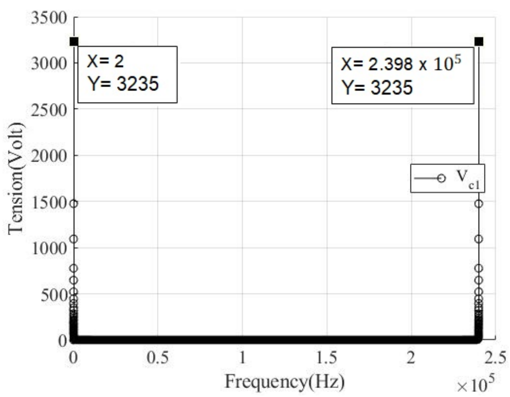
Figure 3.9: C2 voltages in the four cases in sliding mode control (frequency domain).



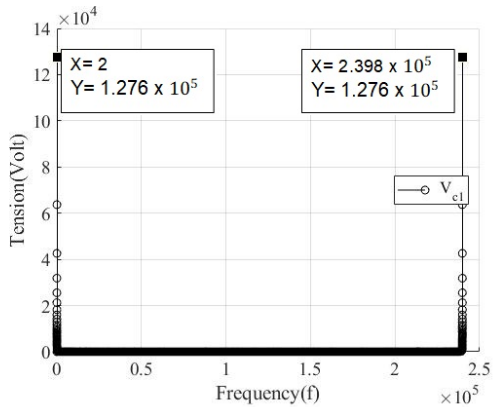
(A) Healthy case



(B) Faulty C1 case

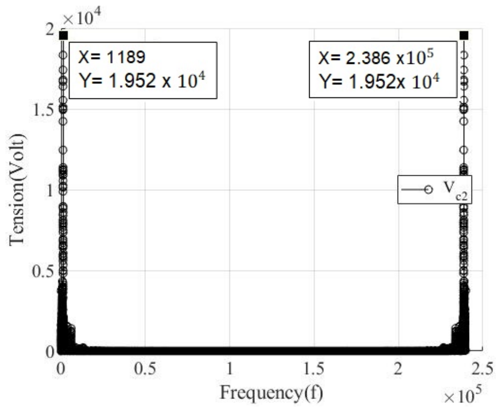


(C) Faulty C2 case

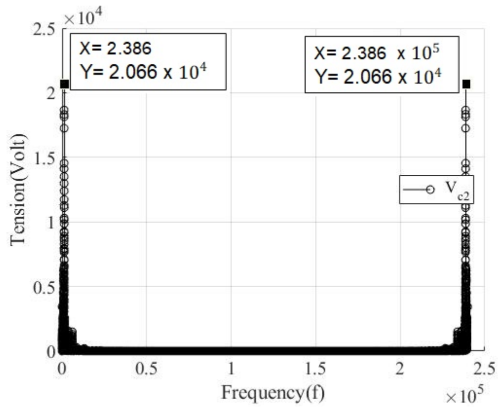


(D) Faulty C1C2 case

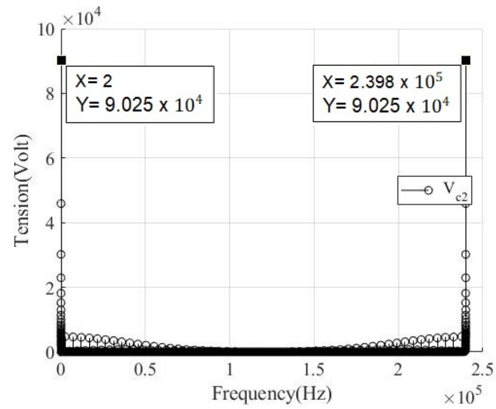
FIGURE 3.10: C1 voltages in the four cases in exact linearization control (frequency domain).



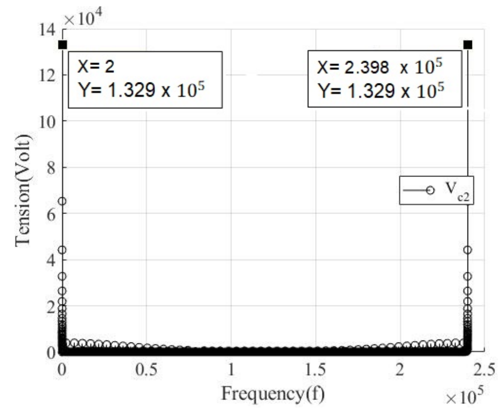
(A) Healthy case



(B) Faulty C1 cas



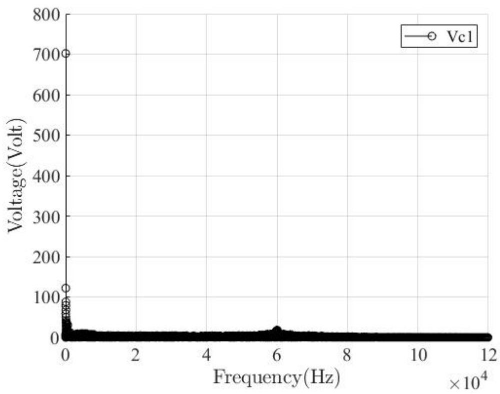
(C) Faulty C2 case



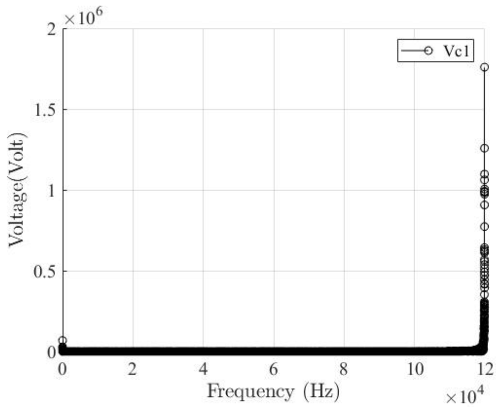
(D) Faulty C1C2 case

Figure 3.11: C2 voltages in the four cases in exact linearization control (frequency domain).

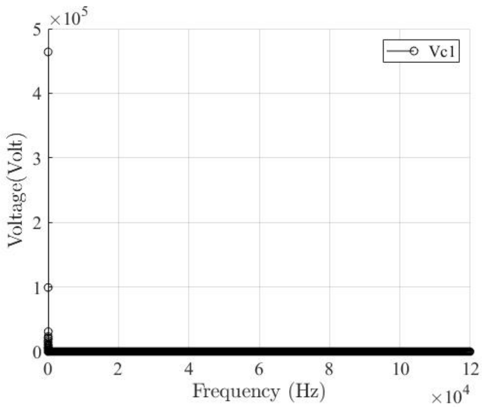
After converting the data to the frequency domain, the mean value must be subtracted from the data. The data is in the form of samples of various amplitudes, but a close examination of the forms indicates that the produced signals are repeated sections over particular time periods, as a result, studying a section of these signals over a period of time and generalizing the data will suffice to get the desired results.



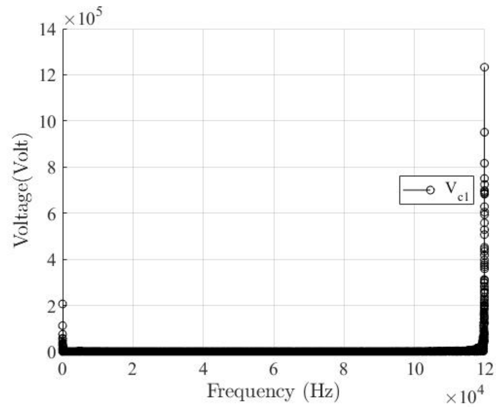
(A) Healthy case



(B) Faulty C1 cas

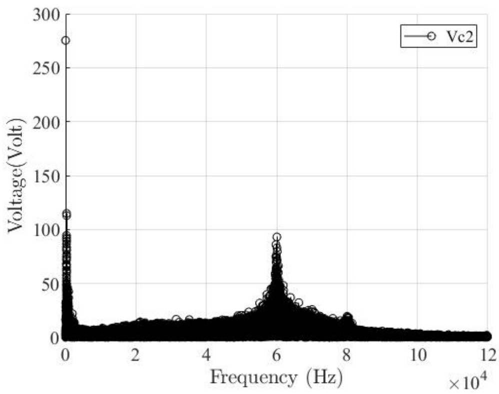


(C) Faulty C2 case

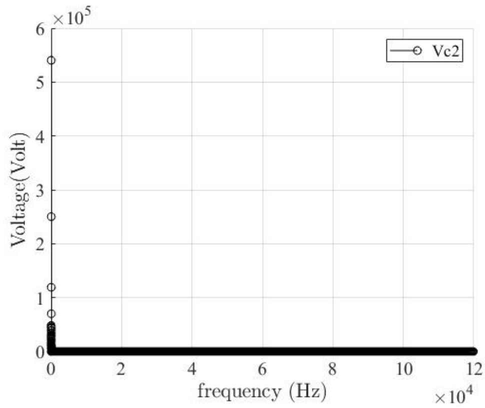


(D) Faulty C1C2 case

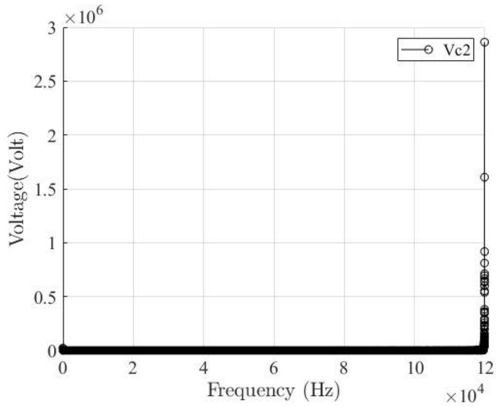
FIGURE 3.12: C1 voltages in the four cases in sliding mode control (frequency domain).



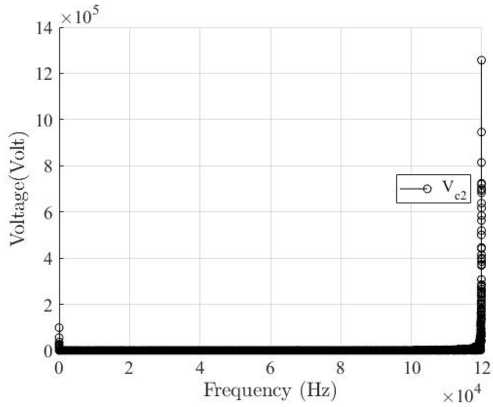
(A) Healthy case



(B) Faulty C1 cas

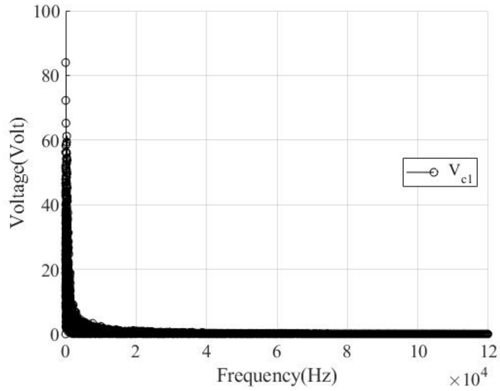


(C) Faulty C2 case

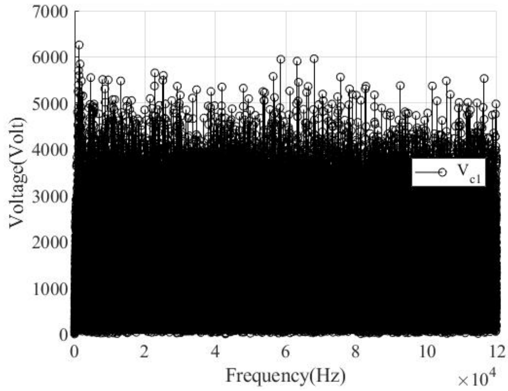


(D) Faulty C1C2 case

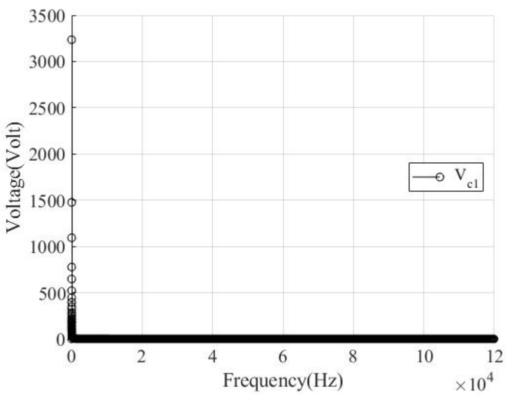
FIGURE 3.13: C2 voltages in the four cases in sliding mode control (frequency domain).



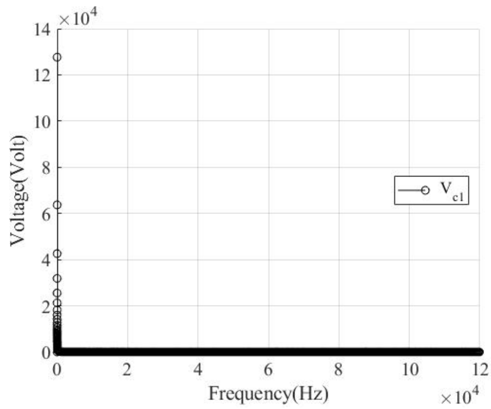
(A) Healthy case



(B) Faulty C1 case

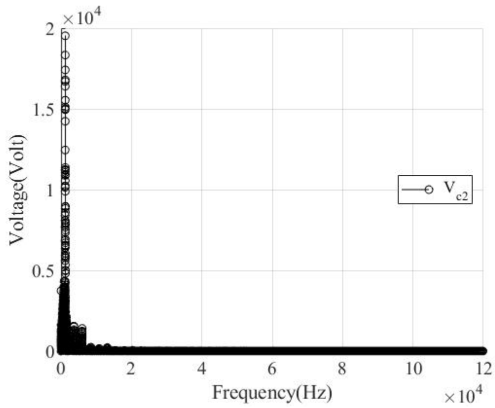


(C) Faulty C2 case

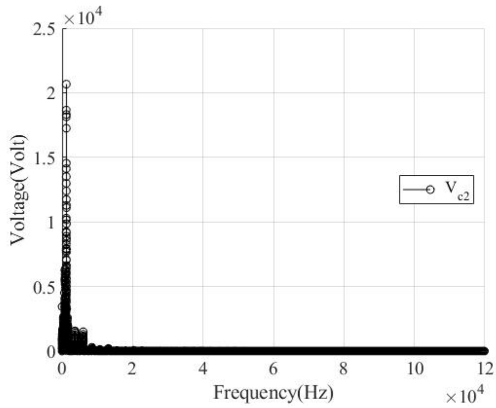


(D) Faulty C1C2 case

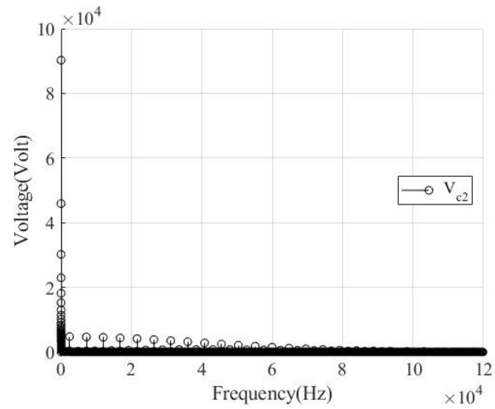
FIGURE 3.14: C1 voltages in the four cases in exact linearization control (frequency domain).



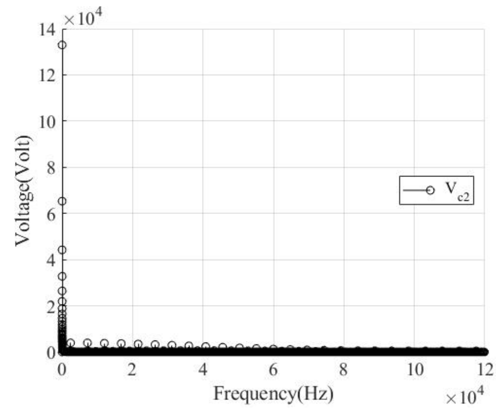
(A) Healthy case



(B) Faulty C1 cas



(C) Faulty C2 case



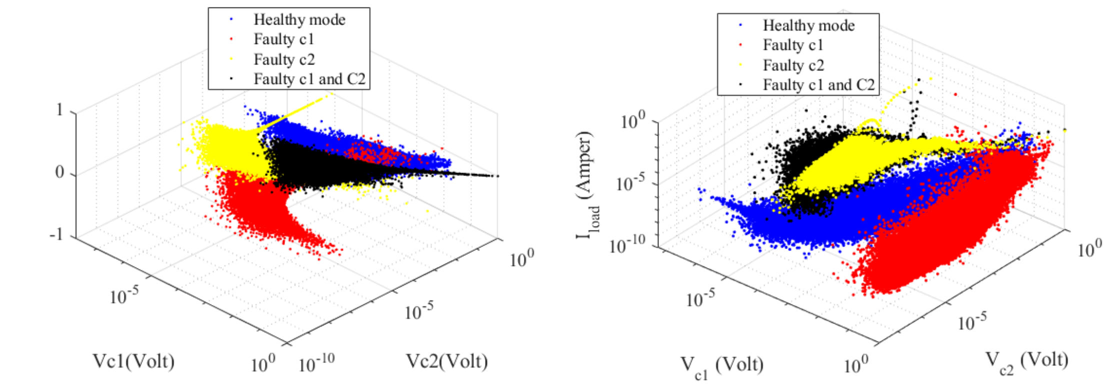
(D) Faulty C1C2 case

FIGURE 3.15: C2 voltages in the four cases in exact linearization control (frequency domain).

As a last treatment for the data, it is preferable to restrict it by adapting it between two values, 0 as the smallest value and 1 as the largest value. This type of conditioning is called a Min-Max normalization . Its purpose is to bring the data together, elimination of blank spaces and normalizing point flow sizes.

3.3 Feature selection

In the stages of building the prediction model, the feature space should be drawn, the purpose of which is to take a look at the behavior of data change during the operation of the system and how it is distributed. This gives a first interpretation about the reliability of the model in its predictions and the appropriateness of the features used. So, here the available features are voltage of first capacitor (Vc1), voltage of second capacitor (Vc2) and the load current (Id) for drawing the feature space is in three dimensions. As shown in the following figures in sliding mode control Figure (3.16a) and exact linearization Figure(3.16b):

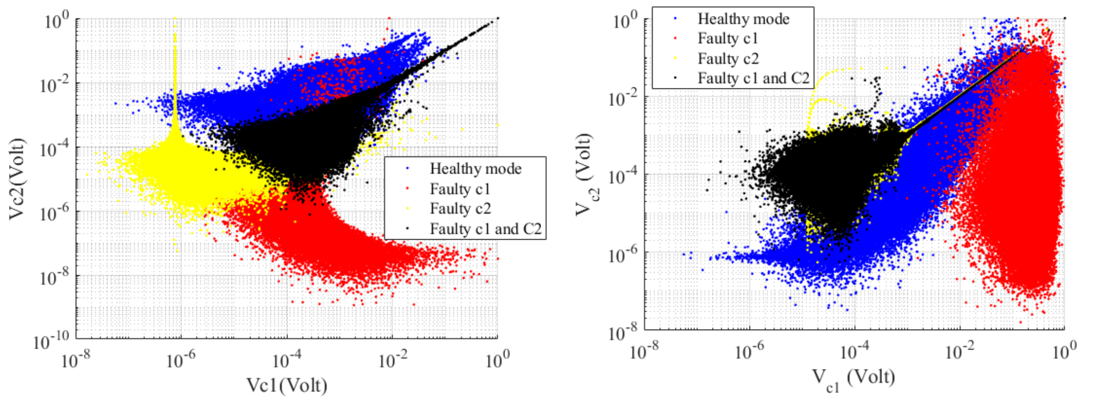


(A) Sliding mode control

(B) Exact linearization control

FIGURE 3.16: Feature space in three dimensions(load current).

In the obtained feature space, the points flow in each case with some overlap between them The fixed features must be determined under a variety of situations where the inaccuracy can be detected, distinguished, and the field of change recognized. However, because the current varies with the load, it cannot be depended upon as a constant feature. So, the voltages of the first and second capacitors will be the deciding factors in drawying feature spaces but in two dimensions and in the case of constructing classification model for all kinds of load used.



(A) Sliding mode control

(B) Exact linearization control

FIGURE 3.17: Feature space in two dimensions(Vc1,Vc2).

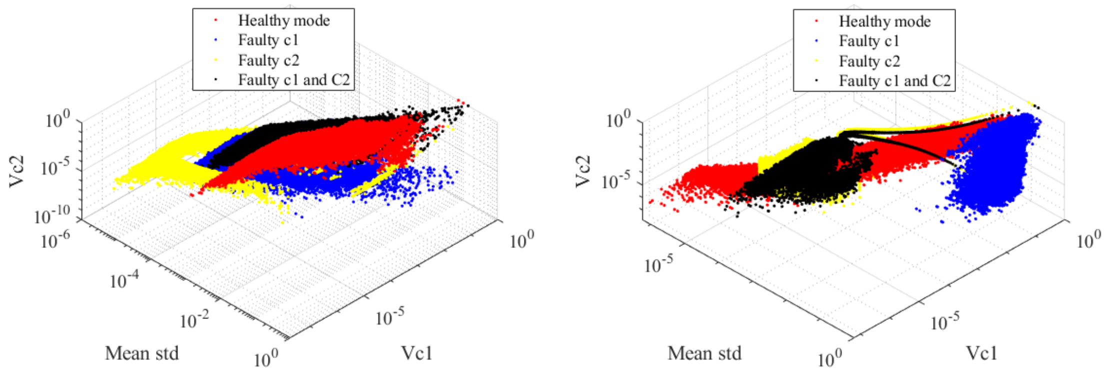
In sliding mode control figure 3.17a the feature space is only in two dimensions and this couldn't be clear enough to demonstrate the distribution of cases, necessitating the addition of a third dimension to clarify the view. Where in exact linearization control figure 3.17b, as it appears that the feature space here does not show the distribution of each case clearly, as the distribution of the faulty C2 mode is identical to the faulty C1C2.

Determining the features depends on the significance of the latter for each of the studied cases[21].in other words, the extent to which this feature changes in each of the cases. tha available feature are Vc1, Vc2 and the load current, so here it is required to create new complex features from the available features according to the state of the feature space distribution,From the surface view of the feature space distribution, we see that each state takes a path in a specific direction. This means that each case deviates from its clusters in a different direction, from the other cases As a result, in each case computing the variance of voltages will be crucial to note how variate in each cases in order to determine the feature that distinguish each case to other, then after calculating the variance, the results are in the table 3.1 as follows:

	voltages	Healthy mode	Faulty C1 mode	Faulty C2 mode	Faulty C1C2 mode
Sliding mode control	vc1	$1.33 \cdot 10^{-5}$	$5.46 \cdot 10^{-5}$	$8.88 \cdot 10^{-5}$	$6.98 \cdot 10^{-5}$
	vc2	$3.64 \cdot 10^{-4}$	$1.20 \cdot 10^{-5}$	$1.83 \cdot 10^{-5}$	$6.68 \cdot 10^{-5}$
Exact linearization control	vc1	$6.58 \cdot 10^{-4}$	0.01	$1.32 \cdot 10^{-5}$	$1.36 \cdot 10^{-5}$
	vc2	$2.07 \cdot 10^{-4}$	$1.67 \cdot 10^{-4}$	$1.40 \cdot 10^{-5}$	$1.36 \cdot 10^{-5}$

Table 3.1: Table of voltages variances in the the two modes of control

As a result, given the contrast values of the two voltages shown in the table, they differ from one case to the next, and all the flow of points of cases in feature space deviates in a definite direction. As a result, the standard deviation is the feature that demonstrates a difference between all possible cases because it reflects the deviation of the flow points from their mean. So, in this situation the standard deviation turns out to be the third most acceptable dimension to separate the feature spaces. So, it will represent the direction of voltages divergence on the mean. So, the third dimension will be the mean of the two standards deviations of the two voltages, the feature space in three dimensions (Vc1,Vc2 and the mean of the two standard deviation) figure 3.18



(A) Sliding mode control

(B) Exact linearization

FIGURE 3.18: Feature space in three dimensions(Vc1,Vc2 and the standard deviation of the two voltages).

3.4 fault detection method application

The K-nearest neighbor algorithm will be used to perform the classification model after drawing the feature space and observing the distribution of each case in this space. This will be done after training the algorithm on the given cases with all of their features and testing the accuracy of its prediction and dependence on it. There are several steps to building the classification model after collecting the data and treatments:

1. Each case must have its own symbol to distinguish it.
2. Concatenate all data Randomly into one table.
3. Split all data into 70% for training and 30% for testing.

each case will be given its own symbol or name to distinguish it. So, it will be 0 for the healthy mode, 1 for faulty C1 mode, 2 for faulty C2 mode, and 3 for faulty C1 and C2 mode. All of these data (cases) will be grouped together in one table, then randomized and split it into 70% to train the model to recognize them. The remaining 30% of the data will subsequently be utilized to test the model classification. Because the data on which the model will be tested has a classification, it can be determined if the model is wrong in determining the case's classification. As a result, the accuracy of this model and the extent of its reliability may be determined.

3.5 Applying K-nearest neighbor

After all the necessary steps have been done, the data is ready to build the model. There are many algorithms KNN, RNN, SVM, neural network … The K-nearest neighbor algorithm was chosen in this study due to its classification efficiency, simplicity and hight accuracy[19]. The first step is to divide the data; this is done by first assigning each case to its own class, then collecting the data in one table and randomly dividing it into two data the first is for training and the second is for testing. As much data as possible, with 30% of the remaining data being tested. The goal of training the testing model is to determine the model's correctness and the extent to which it can make mistakes when relied upon.

3.6 Comaparison between the two controls

The objectives of this section is to compare the two modes of control, sliding mode and exact linearization, in addition to building a classification model, as well as in the objective of obtaining the optimal model we have to variate the number of neighbor selection , The following figure 3.19 are the characteristics and performance of the classification models after they were built in both control modes using the KNN algorithm with variation of five values of k :

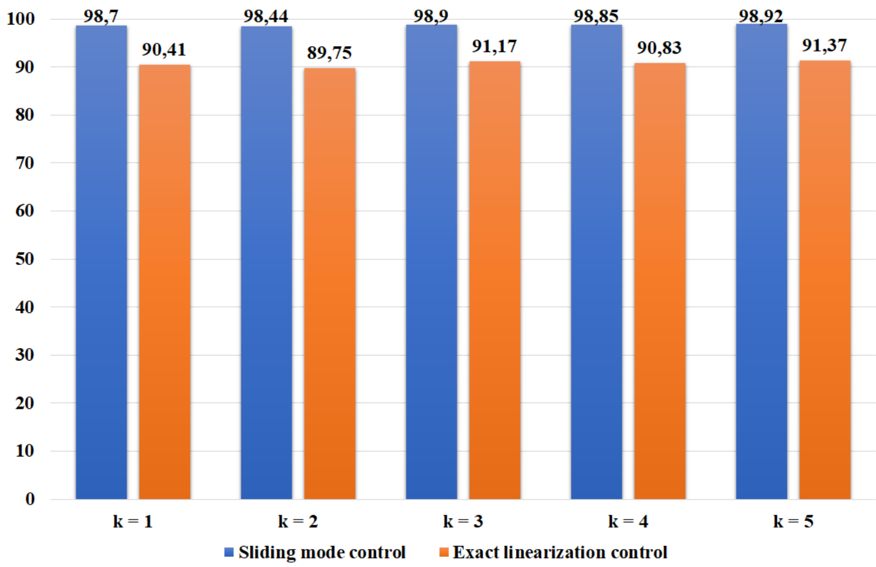


FIGURE 3.19: Comparison between the sliding mode control and exact linearization control the five values of k

the histogram shows that the sliding mode control have the better accuracy compared the exact linearization control. so the selection of the model depends on the best accuracy as a result

According to the results of this study, the sliding mode is the best choice for this system.in terms of:

3.6.0.1 system control and robustness

Looking at the results, it appears that the sliding mode control makes the system to behave naturally in reaction to failures, which aids in the diagnosis of voltages and current signals by give an overview of their behavior. Furthermore, when a failure occurs despite the appearance of harmonics in the load current signal, it maintains the general shape, which indicates the durability of this mode of control. In contrast to the exact linearization control , the obtained signals do not reflect the real state of the system, the system cannot be diagnosed through the signals of voltages. In addition, in the event of a failure in the second capacitor, or in the event of a failure in both, exact linearization is ineffective, as a total distortion in the load current occurs, making it useless.

3.6.0.2 Building the classification model and efficiency

After building the prediction model, and choosing the optimal k. its performance must be evaluated through some settings, as shown in the following table 3.2:

K = 3	recall	specificity	precision	F1-score
Sliding mode control	96.8%	99.63%	98.91%	97.84%
Exact linearization control	77.47%	96.87%	21.16%	83.76%

TABLE 3.2: Table metrics of evaluation in the two controls

3.7 Conclusion

This study presented the creation of a classification model for a multicellular power converter failure, allowing it to operate in safe power for devices and early fault detection. in addition to determining the best mode of operation for the system. the main points gained was that the best control for this system is sliding mode control and this is compared with the exact linearization control, this is due to the system's smooth operation, which shows the harmonics caused by probable faults while keeping the current's general shape, aside from a model for detecting and classifying faults with high accuracy in sliding mode control using KNN machine learning algorithm. The significance of these findings lies in the ability to diagnose and identify a similar power converter system, as well as the ability to use the sliding mode, Which can be applied to similar reel examples of these systems, preventing the system from operating under failure conditions, assure the continuity of operation, avoid shut down of the system and catastrophic damages. In general, sliding mode control showed more robustness against capacitors failure compared with exact linearization control. A further direction we propose to introduce fault prognosis and fault tolerant control of multicellular converter.

Conclusion

This work was an opportunity to treat a major problem in the context of renewable energy specifically in multicellular converters. The study was thus carried out following a professional project approach which guarantees the achievement of objectives in terms of detecting capacitor faults in multicellular powerconverters with the KNN machine learning algorithm as well as determining the best control approach between exact linearization and sliding mode control. The realization of this project has constituted an excellent opportunity to study the main faults and their effect on the voltage and the output current of the multicell converter.

This work has been very beneficial for the field of renewable energy. Indeed, it made it possible to locate capacitor faults to handle high-performance equipment due to the multicellular power converter system's smooth operation and a stable load current generated by it.

Bibliography

[1] Mohammad Zawad Ali et al. "Machine learning-based fault diagnosis for single- and multi-faults in induction motors using measured stator currents and vibration signals". In: *IEEE Transactions on Industry Applications* 55.3 (2019), pp. 2378–2391.

[2] Erin L Allwein, Robert E Schapire, and Yoram Singer. "Reducing multiclass to binary: A unifying approach for margin classifiers". In: *Journal of machine learning research* 1.Dec (2000), pp. 113–141.

[3] Daniel Aloise et al. "NP-hardness of Euclidean sum-of-squares clustering". In: *Machine learning* 75.2 (2009), pp. 245–248.

[4] Stefan Baschel et al. "Impact of component reliability on large scale photovoltaic systems' performance". In: *Energies* 11.6 (2018), p. 1579.

[5] CH Basha and C Rani. "Different conventional and soft computing MPPT techniques for solar PV systems with high step-up boost converters: a comprehensive analysis". In: *Energies* 13.2 (2020), p. 371.

[6] Glenn D Bergland. "A guided tour of the fast Fourier transform". In: *IEEE spectrum* 6.7 (1969), pp. 41–52.

[7] Michael J Bianco et al. "Machine learning in acoustics: Theory and applications". In: *The Journal of the Acoustical Society of America* 146.5 (2019), pp. 3590–3628.

[8] JW Brault and OR White. "The analysis and restoration of astronomical data via the fast Fourier transform". In: *Astronomy and Astrophysics* 13 (1971), p. 169.

[9] B Chandra and Manish Gupta. "An efficient statistical feature selection approach for classification of gene expression data". In: *Journal of biomedical informatics* 44.4 (2011), pp. 529–535.

[10] Bahzad Charbuty and Adnan Abdulazeez. "Classification based on decision tree algorithm for machine learning". In: *Journal of Applied Science and Technology Trends* 2.01 (2021), pp. 20–28.

[11] Kui Fu Chen and Shu Li Mei. "Composite interpolated fast Fourier transform with the Hanning window". In: *IEEE Transactions on instrumentation and measurement* 59.6 (2010), pp. 1571–1579.

[12] Younghoon Cho. "Dual-buck residential photovoltaic inverter with a high-accuracy repetitive current controller". In: *Renewable Energy* 101 (2017), pp. 168–181.

[13] Alexandre De Bernardinis. "Synthesis on power electronics for large fuel cells: From power conditioning to potentiodynamic analysis technique". In: *Energy conversion and management* 84 (2014), pp. 174–185.

[14] M Elsied et al. "Real-time implementation of four-phase interleaved DC–DC boost converter for electric vehicle power system". In: *Electric Power Systems Research* 141 (2016), pp. 210–220.

[15] Hans Ertl, Johann W Kolar, and Franz C Zach. "A novel multicell DC-AC converter for applications in renewable energy systems". In: *IEEE Transactions on industrial Electronics* 49.5 (2002), pp. 1048–1057.

[16] Dimitrios Fragkoulis. "Détection et localisation des défauts provenant des actionneurs et des capteurs: application sur un système non linéaire". PhD thesis. Toulouse 3, 2008.

[17] Cyrille Gautier et al. "Control for the currents balancing of a multicell interleaved converter with ICT". In: *2013 15th European Conference on Power Electronics and Applications (EPE)*. IEEE. 2013, pp. 1–9.

[18] Damien Guilbert et al. "Power switch failures tolerance and remedial strategies of a 4-leg floating interleaved DC/DC boost converter for photovoltaic/fuel cell applications". In: *Renewable Energy* 90 (2016), pp. 14–27.

[19] Gongde Guo et al. "KNN model-based approach in classification". In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2003, pp. 986–996.

[20] Muhammad Ali Imron and Budi Prasetyo. "Improving algorithm accuracy k-nearest neighbor using z-score normalization and particle swarm optimization to predict customer churn". In: *Journal of Soft Computing Exploration* 1.1 (2020), pp. 56–62.

[21] Alan Jović, Karla Brkić, and Nikola Bogunović. "A review of feature selection methods with applications". In: *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*. Ieee. 2015, pp. 1200–1205.

[22] Steven M Kay and Stanley Lawrence Marple. "Spectrum analysis—a modern perspective". In: *Proceedings of the IEEE* 69.11 (1981), pp. 1380–1419.

[23] Lyubomir Kerachev, Yves Lembeye, and Jean-Christophe Crébier. "Analysis and prototyping of distributed series connected DC/AC converters integrated in PV panels". In: *2017 19th European Conference on Power Electronics and Applications (EPE'17 ECCE Europe)*. IEEE. 2017, P–1.

[24] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. "A survey of feature selection and feature extraction techniques in machine learning". In: *2014 science and information conference*. IEEE. 2014, pp. 372–378.

[25] Mathieu Le Bolloch, Marc Cousineau, and Thierry Meynard. “Current-sharing control technique for interleaving VRMs using intercell transformers”. In: *2009 13th European Conference on Power Electronics and Applications*. IEEE. 2009, pp. 1–10.

[26] TAME LEFEUVRE and TA MEYNARD. “Fast line voltage conditioners using new PWM AC chopper opologies”. In: *European Power Electronics Conference*. 2001, pp. 27–29.

[27] Q Li et al. “Nonlinear robust control of proton exchange membrane fuel cell by state feedback exact linearization”. In: *Journal of Power Sources* 194.1 (2009), pp. 338–348.

[28] Shengbo Eben Li et al. “Multiple-model switching control of vehicle longitudinal dynamics for platoon-level automation”. In: *IEEE Transactions on Vehicular Technology* 65.6 (2016), pp. 4480–4492.

[29] Shuhui Li et al. “Integrating photovoltaic and power converter characteristics for energy extraction study of solar PV systems”. In: *Renewable Energy* 36.12 (2011), pp. 3238–3245.

[30] Yifeng Li, Chih-Yu Chen, and Wyeth W Wasserman. “Deep feature selection: theory and application to identify enhancers and promoters”. In: *Journal of Computational Biology* 23.5 (2016), pp. 322–336.

[31] Henrique R Mamede et al. “A multicell Dual-Active Bridge converter for increasing the reliability of power supply in a DC microgrid”. In: *2015 IEEE First International Conference on DC Microgrids (ICDCM)*. IEEE. 2015, pp. 274–279.

[32] Jérôme Mavier. “Convertisseurs génériques à tolérance de panne-Applications pour le domaine aéronautique”. PhD thesis. Institut National Polytechnique de Toulouse-INPT, 2007.

[33] TA Meynard and H Foch. “Imbricated cells multi-level voltage-source inverters for high voltage applications”. In: *EPE journal* 3.2 (1993), pp. 99–106.

[34] TA Meynard and H Foch. “Multi-level choppers for high voltage applications”. In: *EPE journal* 2.1 (1992), pp. 45–50.

[35] Thierry Meynard. *Analysis and design of multicell DC/DC converters using vectorized models*. John Wiley & Sons, 2015.

[36] Thierry Meynard et al. “Parallel multicell converters for high current: Design of intercell transformers”. In: *2010 IEEE International Conference on Industrial Technology*. IEEE. 2010, pp. 1359–1364.

[37] KN Nwaigwe, P Mutabilwa, and E Dintwa. “An overview of solar power (PV systems) integration into electricity grids”. In: *Materials Science for Energy Technologies* 2.3 (2019), pp. 629–633.

[38] Tokuo Ohnishi. “Three phase PWM converter/inverter by means of instantaneous active and reactive power control”. In: *Proc. IEEE IECON*. Vol. 91. 1. 1991, pp. 819–824.

[39] Martin Ondera and Mikulas Huba. "Web-based tools for exact linearization control design". In: *2006 14th Mediterranean Conference on Control and Automation*. IEEE. 2006, pp. 1–6.

[40] Yifan Peng et al. "DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs". In: *Ophthalmology* 126.4 (2019), pp. 565–575.

[41] Aurélien Pichon et al. "Spectral analysis of heart rate variability: interchangeability between autoregressive analysis and fast Fourier transform". In: *Journal of electrocardiology* 39.1 (2006), pp. 31–37.

[42] J Srinivas Rao, P Srinivasa Varma, and T Suresh Kumar. "Novel switching design structure for three phase 21-level multilevel inverter fed BLDC drive application". In: *International Journal of Power Electronics and Drive System (IJPEDS)* 9.3 (2018), pp. 1202–1213.

[43] Muhammad H Rashid. *Power electronics: circuits, devices, and applications*. Pearson Education India, 2009.

[44] Andre Rauh and Gonzalo R Arce. "Optimized spectrum permutation for the multidimensional sparse FFT". In: *IEEE Transactions on Signal Processing* 65.1 (2016), pp. 162–172.

[45] Jonathan G Richens, Ciarán M Lee, and Saurabh Johri. "Improving the accuracy of medical diagnosis with causal machine learning". In: *Nature communications* 11.1 (2020), pp. 1–9.

[46] Boubakeur Rouabah. "Commande d'un convertisseur multicellulaire pour une application de filtrage actif". PhD thesis. 2018.

[47] Boubakeur Rouabah. "Contribution à l'amélioration des performances d'un filtre actif parallèle de puissance par l'utilisation d'un convertisseur multicellulaire". PhD thesis. 2021.

[48] Boubakeur Rouabah, Houari Toubakh, and Moamar Sayed-Mouchaweh. "Fault tolerant control of multicellular converter used in shunt active power filter". In: *Electric Power Systems Research* 188 (2020), p. 106533.

[49] Boubakeur Rouabah et al. "Adaptive and exact linearization control of multicellular power converter based on shunt active power filter". In: *Journal of Control, Automation and Electrical Systems* 30.6 (2019), pp. 1019–1029.

[50] Boubakeur Rouabah et al. "Adaptive data-driven fault-tolerant control strategy for optimal power extraction in presence of broken rotor bars in wind turbine". In: *ISA transactions* (2022).

[51] Boubakeur Rouabah et al. "More Efficient Wind Energy Conversion System Using Shunt Active Power Filter". In: *Electric Power Components and Systems* 49.4-5 (2021), pp. 321–332.

[52] B Rouabaha and L Rahmani. “Direct LYAPUNOV control of multi-cells inverter applied to an active power filter with unbalanced canditions”. In: *The International Conference on Electronics & Oil: From Theory to Applications*. 2013.

[53] Boubakeur Roubah, Houari Toubakh, and M Sayed. “Advanced fault-tolerant control strategy of wind turbine based on squirrel cage induction generator with rotor bar defects”. In: *Annual Conference of the PHM Society*. Vol. 11. 1. 2019.

[54] Sébastien Sanchez, Frédéric Richardeau, and Damien Risaletto. “Design and fault-operation analysis of a modular cyclic cascade inter-cell transformer (ICT) for parallel multicell converters”. In: *Mathematics and Computers in Simulation* 131 (2017), pp. 190–199.

[55] Viktor Sebestyén. “Renewable and Sustainable Energy Reviews: Environmental impact networks of renewable energy power plants”. In: *Renewable and Sustainable Energy Reviews* 151 (2021), p. 111626.

[56] Kazuma Sekiguchi, Wataru Eikyu, and Kenichiro Nonaka. “Feedback Control for a Drone with a Suspended Load via Hierarchical Linearization”. In: *Journal of Robotics and Mechatronics* 33.2 (2021), pp. 274–282.

[57] Pratap Chandra Sen, Mahimarnab Hajra, and Mitadru Ghosh. “Supervised classification algorithms in machine learning: A survey and review”. In: *Emerging technology in modelling and graphics*. Springer, 2020, pp. 99–111.

[58] Ebrahim Shahzad et al. “Sensor Fault-Tolerant Control of Microgrid Using Robust Sliding-Mode Observer”. In: *Sensors* 22.7 (2022), p. 2524.

[59] Vasudha Sharma and Vijay Kumar Tayal. “Hardware implementation of sun tracking solar panel using 8051 micro-controller”. In: *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE. 2017, pp. 483–486.

[60] Marif Daula Siddique and Adil Sarwar. “Performance analysis of carrier based PWM technique for three level diode clamped multilevel inverter with different reference signals”. In: *2016 IEEE 7th Power India International Conference (PIICON)*. IEEE. 2016, pp. 1–6.

[61] Bahman Taheri et al. “A new controller for DC-DC converters based on sliding mode control techniques”. In: *Journal of Control, Automation and Electrical Systems* 30.1 (2019), pp. 63–74.

[62] Khaled Tamizi. “Control of multicellular power converters for microgrids and renewable energies applications”. PhD thesis. Université Paris Saclay (COmUE), 2018.

[63] Houari Toubakh and Moamar Sayed-Mouchaweh. “Hybrid dynamic classifier for drift-like fault diagnosis in a class of hybrid dynamic systems: Application to wind turbine converters”. In: *Neurocomputing* 171 (2016), pp. 1496–1516.

[64] Bengt Ulriksson. “Conversion of frequency-domain data to the time domain”. In: *Proceedings of the IEEE* 74.1 (1986), pp. 74–77.

[65] Indragandhi Vairavasundaram et al. “A review on small power rating PV inverter topologies and smart PV inverters”. In: *Electronics* 10.11 (2021), p. 1296.

[66] Helmut L Votzi and Hans Ertl. “Cell balancing of a multi-cell battery storage system for renewable energy dc micro-grids”. In: *Int. Conf. on Power Conversion (PCIM)*. 2012, pp. 1480–1486.

[67] Kilian Q Weinberger, John Blitzer, and Lawrence Saul. “Distance metric learning for large margin nearest neighbor classification”. In: *Advances in neural information processing systems* 18 (2005).

[68] Tyler Westenbroek et al. “Feedback linearization for unknown systems via reinforcement learning”. In: *arXiv preprint arXiv:1910.13272* (2019).

[69] Jason WH Wong, Caterina Durante, and Hugh M Cartwright. “Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets”. In: *Analytical chemistry* 77.17 (2005), pp. 5655–5661.

[70] Yunda Yan et al. “Non-linear-disturbance-observer-enhanced MPC for motion control systems with multiple disturbances”. In: *IET Control Theory & Applications* 14.1 (2020), pp. 63–72.

[71] Emmanuel Yofo and Philippe Djondiné. “Parallel Active Filtering based on a three–phase multicell converter”. In: *SCIREA Journal of Electrical Engineering* 4 (2019), pp. 48–62.

[72] Toshio Yoshizawa, Shigeki Hirobayashi, and Tadanobu Misawa. “Noise reduction for periodic signals using high-resolution frequency analysis”. In: *EURASIP Journal on Audio, Speech, and Music Processing* 2011.1 (2011), pp. 1–19.

[73] Neng Zhang, Danny Sutanto, and Kashem M Muttaqi. “A review of topologies of three-port DC–DC converters for the integration of renewable energy and energy storage system”. In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 388–401.