University kasdi Merbah Ouargla
Faculty of New Technologies of Information and Communication
Department of Electronics and Telecommunications

Master's Degree Dissertation
In Automatics And Systems

Dissertation Title

Fault detection and prognosis for multilevel inverter

By Ouargli Mohamed Riadh, Boukerche Imad Eddine

SUPERVISOR: Dr. Kafi Mohamed Redouane
Kasdi Merbah University
Automatics And Systems Master's Degree (2023)

TITLE: fault detection in multicellular converter
AUTHOR: Ouargli Mohamed Riadh - Boukerche Imad Eddine (Kasdi Merbah University)
SUPERVISOR: Dr. kafi Mohamed Redouane
JURY:
Dr. Rouabeh Boubakeur
Dr. BOUHAFS Ali
NUMBER OF PAGES: 1, 56

Abstract

Multilevel inverters have become increasingly popular due to their numerous advantages, such as providing high-voltage generation and improved power quality. However, due to their complexity and the presence of numerous components, they are more likely to suffer faults and failures. The purpose of this dissertation is to develop efficient fault detection and prognosis techniques for multilevel inverters employed by Non-Intrusive Load Monitoring (NILM) applications. This study aims to identify and predict failure by analyzing the output voltage of inverters, ensuring reliable and uninterrupted power supply. The proposed methodologies make use of advanced signal processing techniques, machine learning algorithms, and statistical analysis to accurately detect and diagnose faults while also predicting the remaining useful life of faulty components.

Keywords : Fault diagnosis based machine learning, multi cellular power converter, photovoltaic system,control,NILM,prognosis,diagnosis

ملخص

تلعب المحولات ذات المستويات المتعددة دورًا حاسمًا في أنظمة الطاقة الحديثة، حيث تمكّن من توليد جهد عالي وتحسين جودة الطاقة. ومع ذلك، فإن تعقيد هذه المحولات ووجود العديد من المكونات يزيد من احتمال حدوث أعطال وفشل. يركز هذا البحث على تطوير تقنيات فعالة لكشف الأعطال والتنبؤ بها مصممة خصيصًا للمحولات ذات المستويات المتعددة المستخدمة في تطبيقات مراقبة الأحمال غير التداخلية. من خلال تحليل الجهد الناتج عن المحولات، يهدف هذا البحث إلى تحديد وتوقع الأعطال، مما يضمن توفير الطاقة بشكل موثوق ومستمر. تعتمد المنهجيات المقترحة على تقنيات متقدمة لمعالجة الإشارات وخوارزميات تعلم الآلة وتحليل الإحصاءات للكشف عن الأعطال وتشخيصها بدقة، مع توقع العمر المفيد المتبقي للمكونات المعيبة. تؤكد النتائج التجريبية فعالية الأساليب المقترحة، مما يبرز إمكانية تعزيز أداء وموثوقية المحولات ذات المستويات المتعددة في أنظمة .

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

Résumé

Les onduleurs multilevel jouent un rôle crucial dans les systèmes électriques modernes, permettant la génération de hautes tensions et une meilleure qualité de puissance. Cependant, leur complexité et la présence de nombreux composants augmentent le risque de pannes et de défaillances. Cette thèse se concentre sur le développement de techniques efficaces de détection de défauts et de pronostic spécifiquement adaptées aux onduleurs multilevel utilisés dans les applications de Non-Intrusive Load Monitoring (NILM). En analysant la tension de sortie des onduleurs, cette recherche vise à identifier et prédire en temps réel les défauts, garantissant ainsi une alimentation électrique fiable et ininterrompue. Les méthodologies proposées exploitent des techniques avancées de traitement du signal, des algorithmes d'apprentissage automatique et des analyses statistiques pour détecter et diagnostiquer avec précision les défauts, tout en prédisant la durée de vie utile restante des composants défectueux. Les résultats expérimentaux valident l'efficacité des méthodes proposées, mettant en évidence leur potentiel pour améliorer les performances et la fiabilité des onduleurs multilevel dans les systèmes NILM.

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

Acknowledgements

I'd like to thank my advisor, Dr. Kafi, for his sincerity and encouragement, which I'll never forget. Dr. Kafi has been an inspiration to me as I struggled through this Masters degree program. He epitomizes leadership and serves as the ultimate role model. This dissertation would not have been possible without Dr. Rouabah's guidance from the beginning of the research process, which allowed me to develop an understanding of the subject. I am grateful for the extraordinary experiences He arranged for me as well as the opportunities He provided for me to grow professionally. It is a honor to learn from Dr. Kafi and Dr.Rouabah.

I would also like to express my gratitude to my family for their unconditional support throughout this experience. Their constant encouragement, love and support have been essential to me. Their confidence in my abilities motivated me to persevere even in difficult times. I am grateful to have such a wonderful family by my side.

Contents

List of Figures

List of Tables

Declaration of Authorship

We, Ouargli Mohamed Riadh - Boukerche Imad Eddine, declare that this dissertation titled, Fault detection and prognosis for multilevel inverter and the work presented in it are our own.

Introduction

Multilevel inverters play a crucial role in modern power systems [32] , enabling efficient and high-quality power conversion for various applications. However, the complexity and criticality of these inverters demand robust diagnostic and prognostic techniques to ensure their reliable and optimal operation. This thesis focuses on the development and implementation of diagnosis and prognosis methods for multilevel inverters, incorporating the capabilities of NILM (Non-Intrusive Load Monitoring) technology.

The integration of NILM within the diagnosis and prognosis framework offers a non-intrusive and cost-effective approach to monitor and analyze the performance of multilevel inverters. By leveraging the existing power signal data from smart meters or other strategically placed sensors, NILM enables the disaggregation of energy consumption information, allowing for detailed analysis of individual appliances or loads connected to the multilevel inverter.

The primary objective of this dissertation is to investigate and propose novel techniques that utilize NILM for both diagnosis and prognosis of multilevel inverters. Diagnosis involves identifying and localizing faults, failures, or abnormalities in the inverter system, enabling timely maintenance and troubleshooting. On the other hand, prognosis focuses on predicting the future behavior of the multilevel inverter, such as component degradation, performance degradation, or potential failures, enabling proactive maintenance and system optimization.

By combining the capabilities of NILM with advanced signal processing and machine learning algorithms, this research aims to develop effective fault detection, fault location, and fault classification methods for multilevel inverters. Furthermore, the dissertaation seeks to explore the potential of NILM in predicting the remaining useful life of critical components, estimating their degradation levels, and recommending appropriate maintenance actions. [33]

The outcomes of this study will contribute to enhancing the reliability, efficiency, and longevity of multilevel inverters in power systems. The proposed diagnosis and prognosis techniques, integrated with NILM technology, will provide valuable insights into the performance and health condition of the inverters, enabling timely maintenance interventions and facilitating the transition towards condition-based maintenance strategies.

Overall, this dissertation aims to bridge the gap between multilevel inverter diagnostics, prognostics, and NILM technology, paving the way for more effective monitoring, analysis, and optimization of these essential components in modern power systems.

Chapter 1

Multicell Power Converter

1.1 Introduction

1.2 Photovoltaic power system:

1.2.1 Introduction:

Renewable energy sources are particularly appealing due to their ease of access and pollution-free environment for power generation. Recently, photovoltaic energy has been used to generate power from solar energy.

(PV) systems have seen rapid growth. It has the advantages of low repair costs, the absence of moving or rotating parts, and no global impact. However, the conversion efficiency of solar energy to electricity is lower, at around 18-23. Furthermore, the power generated by a PV system is affected by solar insolation and ambient temperature. When the solar insolation is uniform, the output of a PV system shows a single operating point where the power generated is maximum[5] .

1.2.2 Difinition:

Solar PV systems are power systems that use the photovoltaic effect to convert sunlight into electricity. When semiconducting materials are exposed to light, they generate voltage and current. This effect is typically implemented in realworld applications using solar cells, which are individual devices whose electrical characteristics change when exposed to light. These cells are typically made of polycrystalline or monocrystalline silicon and can be linked in series or parallel to achieve the desired voltage and current. A solar module or solar panel is a collection of solar cells packed into a metal frame, and this is the form in which solar PVs are commercially available for use. [5]

The solar panel in a solar PV system receives sunlight and transforms the incident photons into electrical energy. A solar inverter changes the direct current (DC) energy generated by the panel into alternating current (AC). A solar tracking system to enhance overall performance and a battery bank to store the generated power are additional optional extras. The balance of system (BOS) refers to all components of a PV system other than the solar panels. Solar PV systems come in a variety of sizes, ranging from

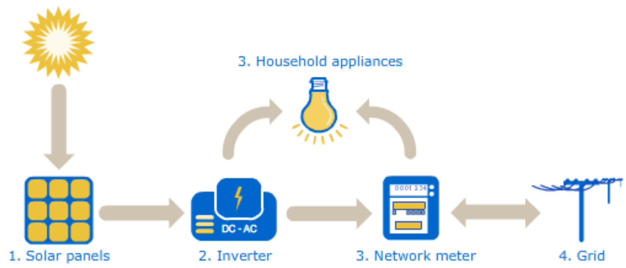

FIGURE 1.1: photovoltaic systeme

small roof-mounted installations with a few panels and kilowatts of capacity to large scale installations with several megawatts of capacity by utilizing large arrays of solar panels. The increasing complexity of a solar PV cell and the constituent devices of a functional solar PV system is depicted in Fig. 1.

Currently, photovoltaic systems have seen tremendous growth and have established themselves as a ripe technology for electricity generation as a result of the global shift to clean energy and reduction in carbon emissions. Within 0.7 to 2 years, a roof-mounted photovoltaic system recovers the energy used in manufacturing it, and over the course of its 30-year service life, it generates 95 clean energy [1]. Due to a developed manufacturing sector and advancements in the semiconductor industry, this growth is accompanied by a swift decline in the cost of solar panels themselves. In the recent past, the panels typically make up less than half of the overall system cost.

1.2.3 Multilevels converters:

A MULTILEVEL converter is a power electrical device which synthesizes various levels of dc voltages as inputs to create the desired voltage output. The multilevel converter can offer a way to integrate fuel cells, wind turbines, solar cells, and other dispersed energy sources into an existing three phase power grid [40]. Heavy duty hybridelectric vehicles (HEVs), such as tractor trailers, transfer trucks, or military vehicles, are another area of application interest. The development of massive electric drive trains for these vehicles will lead to improved vehicle performance (acceleration and braking), higher fuel efficiency, and lower emissions. A cascaded H-bridges multilevel inverter can be used to power the traction motor in parallel-configured HEVs from a group of batteries, ultracapacitors, or fuel cells In high-voltage dc/ac transmissions [8] , motor drives [4] , energy storage systems [20], power quality enhancement equipment [28], modular solidstate transformers [37], special power supplies [42], high-power amplifiers [15], electric vehicles [29], and high-power chargers [30], multilevel converters have emerged as the preferred technology option. In terms of semi-conductor voltage ratings, power quality, passive filter size, electromagnetic interference noise, and system redundancy, multilevel converters are superior to their two-level equivalents. as seen in the figure (2) Three popular multilevel converters are the cascaded-bridge (CBC), flying capacitor, and diode-clamped converters [26]. Because of its modularity, scalability, lack of extra

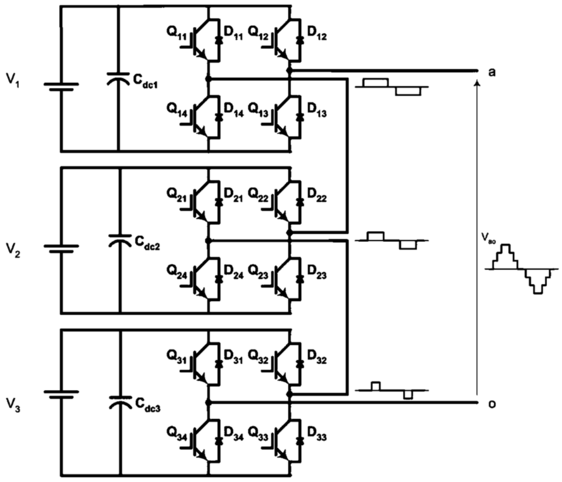

FIGURE 1.2: Seven-level-cascaded-multilevel-inverter

diodes, and lack of high-voltage capacitors, CBCs stand out among them. Furthermore, by using CBCs to replace individual active switches in two-level converters, we are able to create modular multilevel converters (MMCs) [22]. As a result, because MMC arms are made up of CBCs, MMCs inherit advantages from CBCs. In the past 20 years, MMC technology has advanced at an unprecedented rate and breadth [23]. Early in the 1990s, the CM converter was released [21]. As seen in Fig. 2, this architecture is based on the series connection of cells , each cell's structure is built around a separate voltage source. A large and intricate multi-secondary input transformer is necessary when there is only one dc voltage source available,. As a result, the converter's price and size have gone up. Since this design uses series power conversion cells, it is simple to scale the voltage and power levels and achieve the highest possible output voltage levels. The sum of each cell's output voltage, or the total output voltage, is: An additional benefit of this architecture is that, in the event of an internal failure, the problematic cell may be quickly isolated using an external switch and changed out for a functioning cell without switching the converter off [3].

1.3 related works:

1.3.1 classical converters:

Work 1:[17] a comparaison betwen half-bridge, full-bridge, and push-pull configurations as well as other conventional inverter topologies. evaluated variables like output voltage quality, productivity, and harmonic distortion. Their analysis,

Work 2: [41] this work concentrated on pulse width modulation (PWM) strategies to raise the effectiveness of conventional inverters. They researched various modulation techniques and assessed their effects on inverter performance, including sinusoidal PWM and space vector PWM.

Work 3: [43] In order to increase the reliability and robustness of standard inverters,this work, investigated fault detection and protection mechanisms. They suggested a number of methods, such as circuitry for protection, fault-diagnosis algorithms, and current- and voltage-sensing techniques.

Work 4: [36] this work concentrated on control methods for conventional inverters connected to the grid in renewable energy systems. They examined various control algorithms, including proportional-integral (PI) control, predictive control, and droop control, and evaluated how well they controlled power flow and preserved grid stability.

Work 5:[38] In order to scale up standard inverters for industrial applications, this work looked into their modular design. They recommended a modular architecture that permits adaptable power rating configurations and simple expansion. The effect of electromagnetic interference (EMI) and conducted emissions on the functionality of modular inverters,

1.3.1.1 discussion:

Classical inverters have been widely used in power electronics systems to convert DC to AC. However, as discussed in the previous section, these inverters have some limitations. In recent years, there has been a surge of interest in multilevel inverters as an alternative solution. Multilevel inverters have several advantages over standard inverters, including improved voltage quality, reduced harmonic distortion, and increased power handling capabilities. However, the use of multilevel inverters brings with it new challenges and limitations. This paragraph investigates the reasons for the shift toward multilevel inverters, addresses the previous limitations of standard inverters, and highlights the emerging challenges associated with multilevel inverters.

One of the primary reasons for switching to multilevel inverters is their ability to produce high-quality output voltages. Normal inverters, particularly those with a single-phase half-bridge or full-bridge configuration, frequently exhibit voltage distortion and harmonic content in the output waveform . These distortions can cause inefficient operation as well as unwanted effects on connected loads. Multilevel inverters, on the other hand, generate stepped output voltage waveforms by using a series of power semiconductor switches and capacitor voltage sources. Multilevel inverters can significantly reduce harmonic distortion and improve voltage quality by utilizing multiple voltage levels, resulting in improved performance and reliability of connected systems .

Multilevel inverters, as opposed to standard inverters, offer higher power handling capabilities in addition to better voltage quality. The lower voltage levels at which standard inverters typically operate prevent high-power systems from using them. With a greater variety of voltage levels, multilevel inverters can efficiently handle higher voltage ratings and larger power outputs. They are well suited for a variety of applications where high power requirements are common, including electric vehicle charging stations, industrial motor drives, and renewable energy systems .

1.3.2 multilevell converters:

Work 1: [7] this work compared different multilevel inverter topologies, such as the diode clamped, flying capacitor, and cascaded H-bridge inverters, by analyzing performance parameters like output voltage quality, harmonic distortion, and switching losses.

wor2: [35] this work focused on multilevel inverter control strategies for renewable energy systems. To increase the system's efficiency and power quality, they investigated a number of modulation techniques, including phase-shifted carrier-based PWM and selective harmonic elimination. They did not, however, fully analyze how unbalanced grid conditions and grid voltage fluctuations affect multilevel inverter performance. Future studies should look into these variables to create reliable control schemes that can adapt to changing grid conditions.

Work 3: [6] this work compared different modulation strategies for multilevel inverters, such as space vector modulation, carrier-based PWM, and hybrid modulation strategies. They assessed variables like computational complexity, switching losses, and harmonic distortion. The dynamic response and transient performance of the modulation techniques under changing loads,

Work 4: [39] this work studied the use of multilevel inverters in electric vehicle (EV) charging systems. They looked at how multilevel inverters performed in terms of power quality, charging efficiency, and grid integration. Their study,

Work 5: [13] this work studied the efficiency of multilevel inverters and proposed optimization techniques to increase their overall efficiency. They looked at things like conduction losses, switching losses, and power losses.

1.3.2.1 discussion:

The works on multilevel inverters that have been reviewed have made significant contributions to our understanding of their performance, control strategies, modulation techniques, applications, and efficiency analysis. however, Each study has limitations, but the most notable limitation in previous works on multilevel inverters is the lack of comprehensive analysis and consideration of fault diagnosis techniques.

Fault diagnosis is critical in ensuring the reliability and robustness of multilevel inverters by detecting and identifying any abnormalities or malfunctions in their components or operation. However, the majority of the studies reviewed missed going into detail about this. Future research in the field of multilevel inverters should focus on developing advanced fault diagnosis techniques tailored specifically for multilevel inverters, taking into account their unique characteristics and complex circuit topologies. These techniques should allow for the timely detection, accurate identification, and effective mitigation of faults in order to prevent system failures and minimize downtime. By incorporating fault diagnosis into the design and operation of multilevel inverters, their overall reliability and performance can be significantly improved, resulting in improved system efficiency and lower maintenance costs.

In addition to the limited consideration of fault diagnosis techniques, previous works on multilevel inverters have also overlooked the development of prognosis methods. Prognosis entails predicting the remaining useful life and performance degradation of components or systems based on their current condition and operating parameters. By implementing effective prognosis methods, it is possible to anticipate and mitigate potential failures in multilevel inverters before they occur, improving system reliability and minimizing unexpected downtime.

However, the use of prognosis techniques that are especially suited for multilevel inverters has not been fully explored in the existing research in this field. Future research should concentrate on creating effective prognosis techniques that account for the particular features and working conditions of multilevel inverters. These techniques ought to make it possible to identify degradation patterns, estimate remaining useful life, and predict component failures with accuracy. The analysis of multilevel inverters can take a comprehensive approach to condition monitoring and maintenance, leading to more effective and dependable operation of power electronics systems. This is done by via integrating fault diagnosis and prognosis into the analysis of multilevel inverters.

1.4 Contribution:

Our contribution to this work is to develop sophisticated fault diagnosis and prognosis techniques in order to address the shortcomings of earlier research on multilevel inverters. We acknowledge the value of these techniques in improving the multilevel inverters' dependability, performance, and periodic upkeep. By putting forth new techniques for fault diagnosis and prognosis designed specifically for multilevel inverters, we hope to fill the gap in the literature that currently exists.

First, in order to create specialized fault diagnosis methods, we will research and examine the special qualities and intricate circuit topologies of multilevel inverters. In addition to monitoring important operating parameters, this will entail identifying and detecting potential faults in components like semiconductor devices, capacitors, and inductors. We will try to accomplish precise and timely fault detection, localization, and classification, enabling quick mitigation actions, by utilizing advanced signal processing, pattern recognition, and machine learning algorithms.

Second, we will focus on the creation of prognosis methods for multilevel inverters in order to predict component failures and determine how long they will function. Predictive modeling methods will be used, along with the analysis of degradation patterns and the relationships between operating conditions and component performance. We aim to provide accurate estimates of component health, enabling early maintenance and replacement strategies to be implemented, by combining real-time monitoring data, historical performance data, and advanced data analytics.

Our significant contribution to this work is the creation of a prototype system capable of diagnosing and prognosing multilevel inverters. We recognize that practical implementation is critical for the validation and effectiveness of fault diagnosis and prognosis techniques in real-world scenarios. As a result, we hope to design and build a functional prototype that combines advanced sensing, data acquisition, signal processing, and machine learning capabilities.

Our work aims to improve the field of multilevel inverters by focusing on previously unconsidered crucial elements of fault diagnosis and prognosis. We look to improve the general performance, reliability, and maintenance of multilevel inverters in various power electronics applications by offering reliable and effective methods for identifying faults, forecasting failures, and calculating remaining useful life. Ultimately, the goal of our research is to increase the knowledge of multilevel inverters and the ways in which they can be used, creating new opportunities for their integration into today's power systems.

1.5 conclusion:

The focus of this chapter was multicell power converters, specifically in the context of photovoltaic power systems. The introduction provided an overview of the subject, emphasizing the significance of photovoltaic systems and defining key concepts. The discussion then moved on to multilevel converters, with an explanation of their significance and application in photovoltaic systems. The chapter also examined related works, including both classical and multilevel converters. Finally, the chapter concluded by outlining the current study's contribution. Overall, this chapter lays the groundwork for future research and analysis of multicell power converters in photovoltaic systems.

Chapter 2

Multilevel Inverters with Their Applications in PV Systems and NILM overview

2.1 Introduction

In this chapter, we will take a look at multicell inverters and how they can be used in photovoltaic (PV) systems. With the growing demand for renewable energy sources and the integration of PV systems into the power grid, efficient power conversion and control techniques are in high demand. Multicell inverters have emerged as a promising solution, with benefits including higher voltage levels, improved power quality, and increased system efficiency.

We'll start by talking about the various topologies of multicell inverters, such as flying capacitor, cascaded H-bridge, and diode-clamped configurations. Each topology has its own set of advantages and disadvantages, which will be discussed in depth.

We will also delve into the operation principles and modulation strategies associated with these topologies, providing a thorough understanding of their functionality.

The chapter will then go over the benefits of using multicell inverters in PV systems. We will focus on their ability to generate higher voltage levels, which results in lower current levels and power losses, ultimately improving system efficiency.

In addition, we will discuss how multicell inverters contribute to improved power quality by reducing harmonics and improving voltage waveform quality, both of which are critical for seamless grid integration.

Effective control and modulation techniques are critical in the operation of multicell inverters in PV systems. We will investigate advanced control strategies such as selective harmonic elimination (SHE), space vector modulation (SVM), and model predictive control (MPC). These techniques enable precise voltage level regulation, harmonic reduction, and optimal power conversion, resulting in optimal performance of multicell inverters in PV systems.

While multicell inverters have many advantages, they also have some drawbacks as well. We will go over the increased complexity, increased component count, and cost implications of multicell inverters.

We will also discuss the need for efficient control algorithms to handle varying operating conditions and ensure consistent system performance.

Finally, we will discuss future research directions and potential solutions to the problems caused by multicell inverters in PV systems. We can unlock the full potential of multicell inverters and improve their integration with PV systems by researching advanced control techniques, optimizing system design, and improving reliability.

Overall, the goal of this chapter is to provide a thorough overview of multicell inverters and their applications in PV systems. We will gain valuable insights into the potential of multicell inverters for efficient and reliable PV system integration by studying their topologies, benefits, control techniques, experimental validation, and challenges.

2.1.1 Multilevel Inverters:

Multilevel inverters became known as a promising power electronics solution, offering numerous advantages over conventional two-level inverters. These advanced power electronic systems can produce high-quality output voltage waveforms with low harmonic distortion. Multilevel inverters are becoming increasingly popular in a variety of applications, including renewable energy systems, electric vehicles, and high-power industrial drives. The working principles, topologies, modulation strategies, control techniques, and recent advancements in multilevel inverters will all be covered in this comprehensive discussion. The basic idea behind multilevel inverters is to create the desired output voltage waveform by combining multiple levels of DC voltages. Unlike two-level inverters, which can only produce two voltage levels, multilevel inverters can produce several discrete voltage levels. This feature allows them to approximate sinusoidal voltage waveforms, reducing harmonic content and improving output voltage quality. One of the primary advantages of multilevel inverters over conventional inverters is their ability to operate at higher voltage levels. Multilevel inverters can generate higher output voltage levels by utilizing multiple levels of DC voltage, resulting in less voltage stress on power semiconductor devices. This feature improves component reliability and increases component lifespan.

FIGURE 2.1: multilevel inverter

2.1.2 Topologies of the multicell inverters:

Multilevel inverters exist in a variety of topologies, each with their own benefits, trade-offs, and uses. The cascaded H-bridge, diode-clamped (neutral-pointclamped), and flying capacitor inverters are three common topologies.

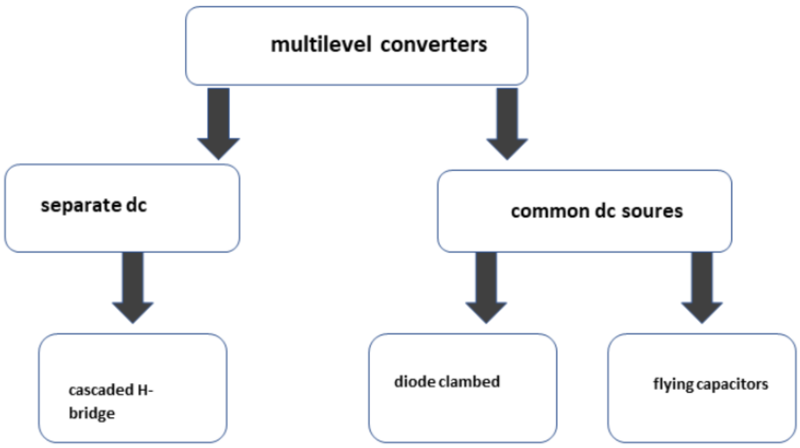

FIGURE 2.2: multilevel converters topologie

2.1.2.1 cascaded H-bridge topology:

The cascaded H-bridge multilevel inverter topology is a well-known configuration that has received a lot of attention in the field of multilevel power electronics.

This topology is suitable for high-power applications due to its excellent modularity, scalability, and fault tolerance. The cascaded H-bridge multilevel inverter works on the basis of multiple H-bridge cells connected in series to generate the desired output voltage waveform. Each cascaded H-bridge multilevel inverter cell is made up of four power semiconductor switches (typically insulated gate bipolar transistors - IGBTs or MOSFETs) and two DC voltage sources (typically capacitors or batteries).

Each Hbridge cell's switches are controlled in a complementary manner, allowing current to flow in either direction. The voltage across each H-bridge cell can be controlled by varying the switching states of the switches, resulting in the generation of multiple voltage levels.

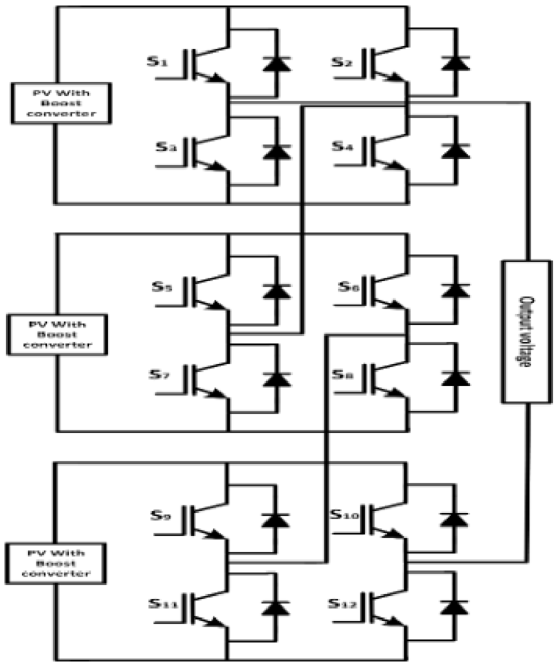

FIGURE 2.3: Seven level Cascaded H-bridge based multilevel inverter connected with three PV modules

To obtain the overall output waveform, the output voltage of each H-bridge cell is stacked together. The number of available voltage levels is determined by the number of H-bridge cells. A three-level cascaded H-bridge inverter, for example, would be made up of three H-bridge cells that could generate three voltage levels: positive, zero, and negative. There are several advantages to using a cascaded H-bridge multilevel inverter.

For starters, its modular design allows for simple expansion and scaling by adding or removing H-bridge cells. This feature qualifies it for high-power applications that necessitate higher voltage levels. Furthermore, the topology's fault tolerance allows the system to continue operating even if one or more Hbridge cells fail. Different modulation techniques can be used to control the output voltage waveform of the cascaded H-bridge multilevel inverter. PWM (pulse width modulation) is a popular technique. The amplitude and frequency of the output voltage waveform can be altered by varying the duty cycle of the switching signals applied to the H-bridge cells. Several PWM strategies, including carrierbased PWM and space vector modulation (SVM), have been investigated for cascaded H-bridge multilevel inverters in order to optimize harmonic performance and minimize switching losses [31].

Furthermore, advanced control strategies to improve the performance of cascaded H-bridge multilevel inverters have been proposed. Selective harmonic elimination (SHE) is a popular control technique that solves a set of nonlinear equations to eliminate specific harmonics in the output voltage waveform. Another strategy is model predictive control (MPC), which uses a predictive algorithm to optimize the switching states of the H-bridge cells based on a predefined cost function. These advanced control techniques improve cascaded H-bridge multilevel inverter efficiency, reliability, and harmonic performance [18].

It is important to note that the cascaded H-bridge multilevel inverter topology has certain disadvantages. The main difficulty is the increased complexity and component count when compared to conventional two-level inverters. The requirement for multiple H-bridge cells and associated control circuitry can increase the cost and complexity of the system. Furthermore, maintaining optimal performance requires ensuring proper voltage balancing across the Hbridge cells and mitigating the effects of capacitor voltage ripple.

To address these issues, researchers have proposed a variety of solutions. To maintain voltage balance among the H-bridge cells, active balancing techniques such as capacitor voltage balancing control algorithms have been developed. Furthermore, methods for reducing capacitor voltage ripple, such as interleaved operation and hybrid modulation strategies, have been investigated to mitigate the negative effects of voltage ripple [14].

2.1.2.2 diode-clamped topology:

The neutral-point-clamped (NPC) inverter, also known as the diode-clamped multilevel inverter, is a popular topology that offers several advantages in terms of voltage control, harmonic performance, and reduced switching losses [27]. Clamping diodes and capacitors are used in diode-clamped inverters to control voltage levels and limit voltage stress on power semiconductor devices . This section will go through the diode-clamped multilevel inverter topology's ,operation principles, modulation techniques, control strategies, and recent advancements.

The clamping diodes and capacitors connected between the DC input source and the output phase legs operate the diode-clamped multilevel inverter . The diodes are connected to the DC source's neutral point, allowing the voltage across each phase leg to be clamped within a specific range. Capacitors provide the necessary energy storage to keep voltage levels stable during switching transitions . The diode-clamped inverter can synthesize the desired output voltage waveform by appropriately controlling the switching states of the power semiconductor devices.

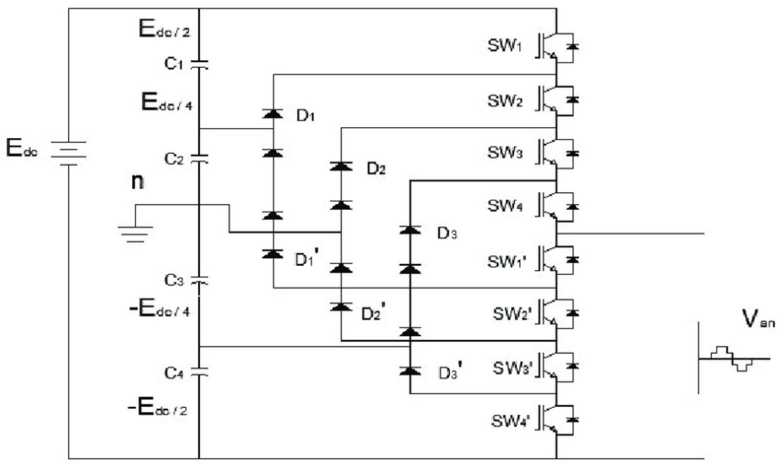

FIGURE 2.4: Five-Level Clamped Topology

One of the main advantages of the diode-clamped inverter over other multilevel topologies is the reduced voltage stress on the power switches . Because clamping diodes limit the voltage across the switches, lower voltage-rated devices can be used . This feature improves overall system reliability while also extending the life of power semiconductor devices.

Various modulation techniques for diode-clamped multilevel inverters have been developed to regulate the output voltage waveform and minimize harmonic distortion. Techniques for pulse width modulation (PWM), such as carrier-based PWM and space vector modulation (SVM), are widely used [45]. To generate switching pulses for power devices in carrier-based PWM, the reference waveform is compared with triangular carrier signals. In contrast, space vector modulation operates in the space vector domain and makes better use of the DC voltage levels .

Control strategies are critical for optimizing performance in diode-clamped multilevel inverters. Selective harmonic elimination (SHE) is a popular control technique that solves a system of nonlinear equations to eliminate specific harmonics in the output voltage waveform . The SHE technique calculates the switching angles required to satisfy the harmonic elimination conditions and produce the desired voltage waveform. Model predictive control (MPC) is another approach that uses a predictive model of the system to optimize control actions over a finite control horizon [1]. MPC has shown promising results in improving diode-clamped multilevel inverter dynamic response, harmonic performance, and fault tolerance.

2.1.2.3 flying capacitor multilevel inverter topology:

The flying capacitor multilevel inverter is a well-known topology that has been extensively researched and used in a variety of applications. To generate the desired voltage levels, this topology employs flying capacitors connected between the DC sources and

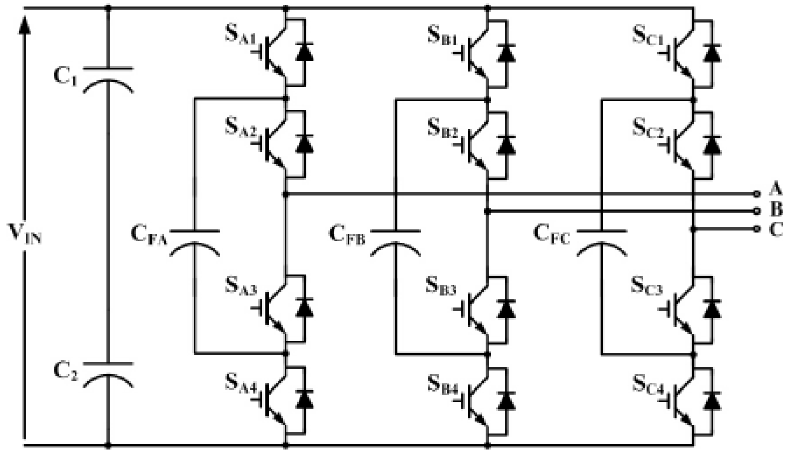

FIGURE 2.5: Three-level-flying-capacitor-multilevel-inverter-topology

the output phase legs. The flying capacitor multilevel inverter provides several benefits by charging and discharging the flying capacitors in a coordinated manner, including high-quality output voltage, reduced voltage stress on the switches, and improved reliability

The development of advanced modulation techniques is another area of research in the flying capacitor multilevel inverter. PWM techniques are commonly used to control the switching of power semiconductor devices and synthesize the desired output voltage waveform. To improve the harmonic performance and efficiency of the flying capacitor multilevel inverter, researchers investigated various PWM strategies such as carrier-based PWM and space vector modulation (SVM). To reduce the number of switching transitions and minimize harmonic distortion, Li et al. proposed a hybrid PWM strategy that combines the benefits of carrier-based PWM and SVM .

The flying capacitor multilevel inverter topology has been used in a variety of applications, such as renewable energy systems and motor drives. The flying capacitor multilevel inverter is critical in renewable energy systems such as solar photovoltaic (PV) systems for converting the DC power generated by the PV panels into high-quality AC power for grid integration. Researchers studied the performance and control of the flying capacitor multilevel inverter in PV systems, focusing on issues such as maximum power point tracking (MPPT) algorithms, grid synchronization, and power quality improvement [56].

2.1.3 benefits of using multicell inverters in PV systems:

When multicell inverters are integrated into PV systems, they improve performance, reliability, and overall efficiency. Here are some key benefits of using multicell inverters in PV systems, backed up by references:

2.1.3.1 Improved Power Quality:

Multicell inverters produce high-quality output voltage waveforms with low harmonic distortion. This is especially important in PV systems because it helps to reduce grid disturbances, improve power quality, and meet grid connection standards.

2.1.3.2 Higher Efficiency:

Since multicell inverters can use higher voltage levels, the current flowing through the power electronic components is reduced. As a result, conduction losses are reduced and the overall efficiency of the PV system is improved [12].

2.1.3.3 Voltage Boosting Capability:

Multicell inverters can boost output voltage levels, which is useful in PV systems, particularly when operating in low solar irradiance or shaded conditions. The voltage boost capability ensures optimal power generation and increases the PV system's energy yield .

2.1.3.4 Multicell inverters provide the flexibility to implement advanced maximum power point tracking (MPPT) algorithms

. These algorithms optimize the PV system's power output by continuously tracking the maximum power point of the solar panels under changing environmental conditions. As a result, energy harvesting increases and overall system performance improves [24].

2.1.3.5 Modularity and scalability:

Multicell inverters can be designed modularly, allowing for easy system expansion or reconfiguration. This modularity enables scalability, allowing the PV system to be adapted to changing energy demands or future expansion requirements .

2.1.3.6 Fault Tolerance and Redundancy:

Multicell inverters have built-in fault tolerance. If one cell or module fails or fails, the system can continue to function using the remaining cells. This redundancy improves the PV system's reliability and availability while reducing downtime and maintenance costs .

2.1.3.7 Reduced Voltage Stress:

Because multicell inverters use multiple cells or modules, voltage stress is distributed across the components, reducing stress on individual power electronic devices. As a result, the system components are more reliable and have a longer lifespan .

2.1.3.8 Grid-Friendly Operation:

Multicell inverters can be designed to meet grid requirements and integrate seamlessly with the utility grid. They can include grid support functions such as reactive power control, low-voltage ride-through capabilities, and voltage regulation, allowing the PV system to operate in a stable and reliable manner within the grid infrastructure .

2.2 overview about NILM:

NILM stands for Non-Intrusive Load Monitoring. It is a technique for analyzing and disaggregating the energy usage of individual appliances within a household or building without the requirement for extra sensors or meters on each device. [34]

The purpose of NILM is to provide detailed information about how much energy each device or appliance in a building consumes without the requirement for separate monitoring of each unit. It entails using data from a single or a few strategically positioned sensors, such as smart meters, to monitor the entire energy use of the building. NILM methods can predict individual appliance energy usage by evaluating power signal data and applying signal processing and machine learning algorithms.

NILM provides a wide range of applications and benefits. It can assist users in better understanding their energy consumption patterns, identifying energy inefficiencies, and making informed decisions to reduce their energy consumption. It can also be utilized in smart grids for demand response programs, energy conservation, and load balancing. NILM can also be used to monitor appliances at the appliance level in residential, commercial, and industrial environments. [2]

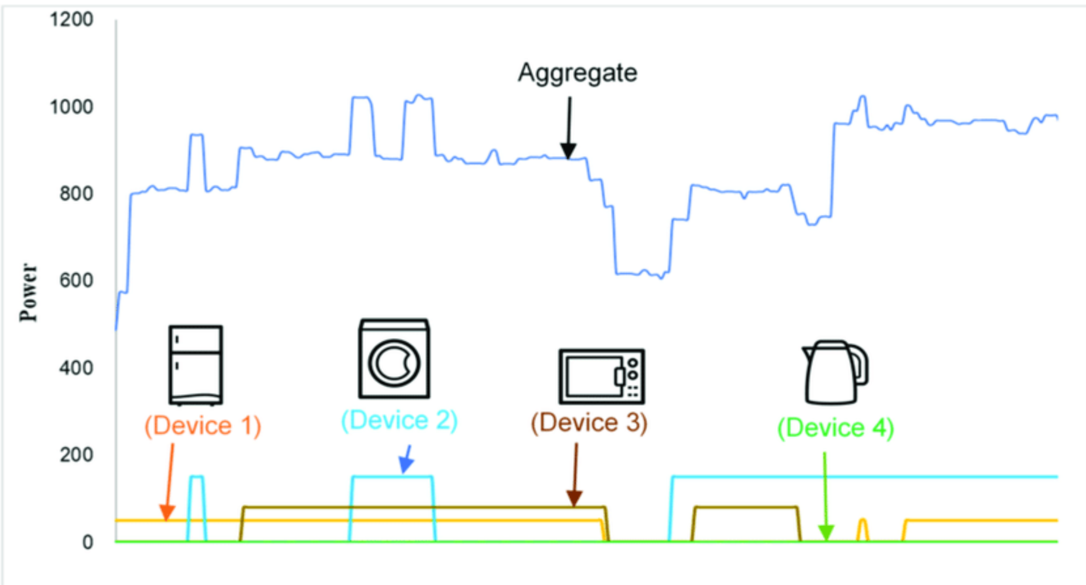

FIGURE 2.6: NILM example

2.2.1 Importance of NILM in the diagnosis of machines :

Importance of NILM in diagnosis of machines Non-Intrusive Load Monitoring (NILM) is also useful in diagnosing machinery and technology. Although NILM is well recognized for its use in energy management and load monitoring, their techniques and principles can also be applied to industrial diagnosis and fault detection ,Here are some of the reasons why NILM is crucial in machine diagnosis:

2.2.1.1 Non-Intrusive Monitoring:

One of the primary benefits of NILM is that it allows for monitoring and analysis without the use of extra sensors or invasive methods. NILM can provide insights into the operating behavior of machines and equipment by studying their electrical load signatures and identifying any abnormalities that may indicate problems or anomalies.

2.2.1.2 Early Fault Detection:

By constantly observing the electrical load profiles of equipment, NILM can detect early symptoms of problems or irregularities in their functioning. This allows for timely intervention and preventative maintenance, which helps to minimize costly breakdowns, production losses, and safety issues.

2.2.1.3 Comprehensive Monitoring:

NILM can record and analyze the electrical fingerprints of numerous subsystems and parts within a machine. This holistic monitoring approach detects faults at multiple levels, from individual components to the total system, offering a full view of machine health.

2.2.1.4 Pattern Recognition and Anomaly Detection:

NILM approaches use pattern recognition algorithms to learn normal operating patterns and spot deviations from these patterns. NILM can detect anomalous load signatures associated with malfunctioning circumstances such as excessive power consumption, irregular duty cycles, or odd operational states by comparing real-time load data with historical data.

2.2.1.5 Data-Driven Insights:

NILM generates huge quantities of data relating to machine functioning and energy use. By examining this data, machine operators and maintenance workers can acquire useful insights into machine performance, discover root causes of defects, and make informed decisions about maintenance, repair, or replacement.

2.2.1.6 Cost and Time Efficiency:

NILM offers a cost-effective and time-efficient alternative to machine diagnosis as compared to standard diagnostic approaches that include physical inspections or invasive measurements. It eliminates the need for extra sensors or specialized tools, and the analysis may be done remotely, allowing for speedy and proactive defect identification.

overall, Non-Intrusive Load Monitoring (NILM) provides important advantages for machine diagnosis. Due to its non-intrusive design, it is possible to continuously monitor electrical load signatures, allowing for the early identification of errors and irregularities in machine operation. By examining load data, NILM can offer thorough insights into machine health, revealing potential problems at various system levels. The ability of NILM to recognize patterns and detect anomalies makes it easier to spot departures from regular operating patterns, which helps with the early detection of defects. The data-driven approach of NILM helps companies optimize maintenance plans and decrease downtime by offering useful insights for maintenance decisions. Overall, NILM is a useful instrument for raising machine reliability, lowering expenses, and increasing overall operational effectiveness.

2.3 coclusion:

As a summary, integrating multicell inverters into PV systems provides significant benefits such as improved power quality, higher efficiency, voltage boosting capability, enhanced MPPT performance, modularity, fault tolerance, reduced voltage stress, and grid-friendly operation. Because of these benefits, multicell inverters are an excellent choice for improving the performance and reliability of PV systems, thereby contributing to the growth of clean and sustainable energy generation.

Chapter 3

Machine Learning Fault Detection and prognosis Methodologies

In recent years, machine learning has become a potent tool in various fields, revolutionizing how we approach challenging issues. One example of this is in the field of fault detection and prognosis, where machine learning techniques have been widely used to increase the effectiveness and precision of fault diagnosis and prediction. Machine learning algorithms can automatically detect deviations identify faults, and even predict potential failures before they happen by using the enormous amounts of data that are available from sensors, tools, and systems. This chapter examines how machine learning is applied to fault detection and prognosis, focusing on its importance in ensuring the reliability, safety, and best performance of critical systems.

3.1 fundamentals of Fault Detection and Prognosis

3.1.1 Definition and Importance of Fault Detection

Fault detection is the process of identifying deviations or irregularities in the behavior or performance of a system or its components that indicate the presence of faults or failures. It is essential in many industries, including manufacturing, aerospace, automotive, and energy, since it enables prompt interventions to prevent costly malfunctions, decrease downtime, and assure the safety and dependability of critical systems [9]. Fault detection is a critical component of condition monitoring systems, which aim to detect and diagnose errors before they cause serious damage . Fault detection algorithms can detect small changes in patterns, outliers, or abnormalities that signal the presence of faults or degradation by continuously monitoring operational data and sensor measurements from equipment and systems [16].

Effective defect detection is critical in industries where equipment downtime or failure can have catastrophic consequences. In the aerospace industry, for example, early detection of defects in aircraft engines can avoid catastrophic occurrences and ensure passenger safety [44]. Similarly, in the manufacturing industry, fault detection assists in identifying process irregularities or equipment faults, resulting in enhanced product quality, less scrap, and increased productivity [25]. Fault detection systems can automatically learn the typical behavior of a system or its components and spot deviations from it by employing machine learning techniques. Machine learning algorithms are capable of handling vast amounts of data, capturing complicated patterns, and adapting to changing situations, allowing for precise and dependable defect detection .

3.1.2 Early Fault Detection and its Significance:

Early fault detection is the proactive identification and diagnosis of defects in systems or equipment at an early stage, prior to them escalating into major failures or causing considerable damage. It entails tracking and examining different signals, parameters, or indicators to find minute variations or deviations that could point to underlying flaws .[19]

Due to the following factors, early fault identification is extremely important in many industries:

3.1.2.1 Reduced Downtime:

Early identification enables prompt maintenance and intervention, which can stop unforeseen failures and lower unplanned downtime. By identifying flaws early on, systems can be repaired or replaced before they entirely malfunction, limiting the impact on business operations .

3.1.2.2 Saving costs:

Catching faults early can save a lot of money on repairs and avoid potential harm to other parts or systems. Early detection of problems can reduce the extent of damage, avert expensive repairs or replacements, and increase the overall cost-effectiveness of maintenance tasks .

3.1.2.3 Enhanced Safety and Reliability:

By identifying possible risks or issues in advance of accidents or serious failures, early fault detection improves safety. The dependability of systems and components can be increased by proactively correcting problems, assuring their maximum performance and lowering the possibility of catastrophic events .

3.1.2.4 Condition-Based Maintenance:

The application of condition-based maintenance solutions is made possible by early fault identification. Maintenance operations can be planned based on the actual health status of equipment and parts rather than on specified maintenance intervals by continuously monitoring their condition. This method optimizes resource allocation while reducing unnecessary maintenance tasks .

3.1.3 Definition and Importance of prognosis

Prognosis is a method of forecasting or estimating the future behavior, condition, or performance of a system, component, or process. Prognosis is essential in decision-making, planning, and maintenance strategies in many sectors, including medical, engineering, and business. It entails using existing knowledge, historical data, and prediction models to forecast probable events and make educated decisions. [11]

3.1.4 Importance of Prognosis:

3.1.4.1 Predictive Maintenance:

Prognosis is critical for applying predictive maintenance solutions. Proactive maintenance actions can be scheduled by projecting the future behavior or degradation of essential components or systems in order to prevent failures, minimize downtime, and optimize maintenance costs.

3.1.4.2 System dependability and Availability:

Prognosis aids in the dependability and availability of systems or equipment. Potential failures can be foreseen and addressed in advance by monitoring and predicting component health and performance degradation, decreasing unexpected system downtime and improving overall system reliability.

3.1.4.3 Cost Optimization:

Prognosis allows for cost optimization by eliminating needless maintenance operations or component replacements. By precisely projecting the remaining usable life of assets or evaluating their degradation levels, maintenance resources can be deployed more efficiently, resulting in cost savings and greater asset utilization.

3.1.4.4 Decision-Making:

Prognostic information assists in making informed decisions about repair, replacement, or upgrade options. Understanding the future behavior of a system or component allows stakeholders to make decisions based on cost-benefit calculations, risk assessments, and overall system performance.

3.1.4.5 Process Optimization:

Prognosis aids in process optimization by identifying potential bottlenecks, inefficiencies, or performance restrictions. Adjustments can be performed in real time to enhance efficiency, productivity, and resource usage by forecasting the behavior of crucial process variables.

3.1.4.6 Risk Management:

Prognosis plays an important role in risk management by identifying and assessing potential hazards associated with system or component failures. Understanding the probability and effects of failures allows for the implementation of suitable risk mitigation techniques, assuring safety and limiting potential dangers.

Overall, prognosis provides useful information on how systems, components, or processes will behave in the future. It enables decision-makers to optimize maintenance, improve dependability, and improve system performance by employing predictive models, historical data, and advanced analytics, ultimately resulting to cost savings, better efficiency, and improved overall operations.

3.2 Machine Learning :

3.2.1 Machine Learning Overview:

Machine learning is a branch of artificial intelligence (AI) that focuses on the creation of algorithms and models that allow computer systems to learn and improve without being explicitly programmed. It involves extracting patterns, insights, and knowledge from data in order to create predictions or judgments. Machine learning approaches have gotten a lot of attention and have been used successfully in a variety of fields, including image identification, natural language processing, recommendation systems, and predictive analytics [11].

3.2.2 Types of machine learning:

Machine learning can be categorize into three categories: supervised learning, unsupervised learning, and reinforcement learning.

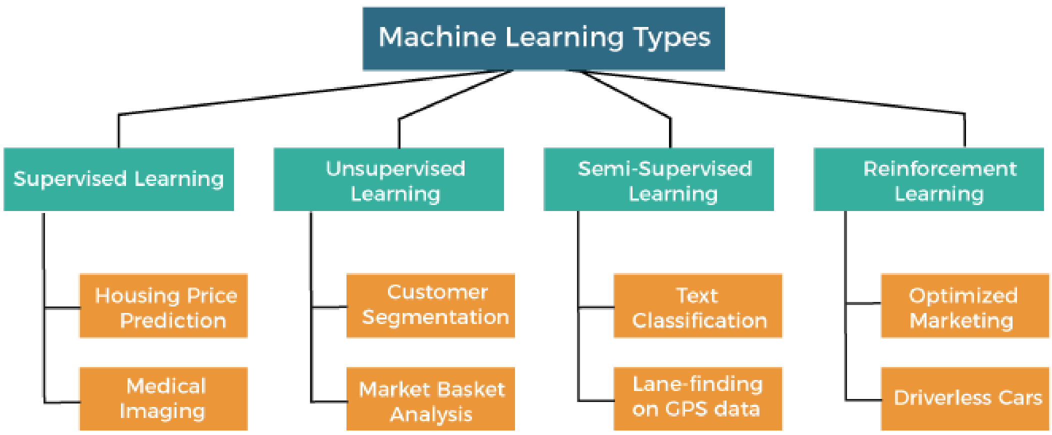

FIGURE 3.1: types-of-machine-learning

3.2.2.1 Supervised learning :

Supervised learning entails training a model on a labeled dataset, with each data occurrence assigned a target or output label. The model learns to generalize from the examples presented and may make predictions on new, previously unseen data. Linear regression, decision trees, support vector machines (SVM), and neural networks are examples of popular supervised learning techniques [10].

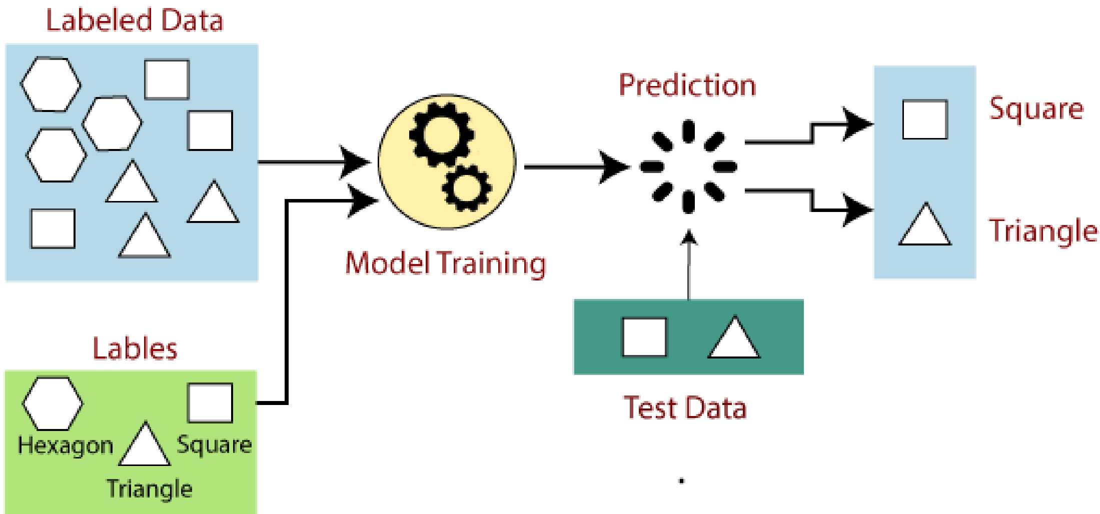

FIGURE 3.2: supervised-machine-learning

3.2.2.2 Unsupervised learning:

on the other hand, works with unlabeled data, with the model exploring hidden patterns or structures within the data in the absence of any predefined labels. Unsupervised learning often employs clustering algorithms such as k-means and hierarchical clustering, as well as dimensionality reduction techniques such as principal component analysis (PCA) and t-SNE [77].

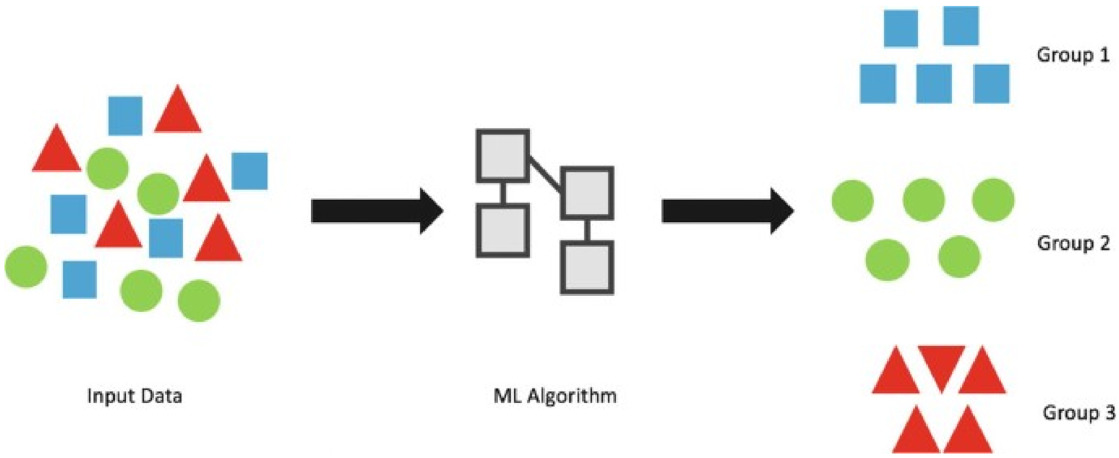

FIGURE 3.3: unsupervised machine learning

3.2.2.3 Reinforcement learning :

concerned with training an agent to interact with an environment and learn optimal actions through trial and error. The agent receives input in the form of rewards or penalties based on its behaviors, and it strives to maximize its cumulative rewards over time. Reinforcement learning has found applications in robotics, gaming, and autonomous systems [78].

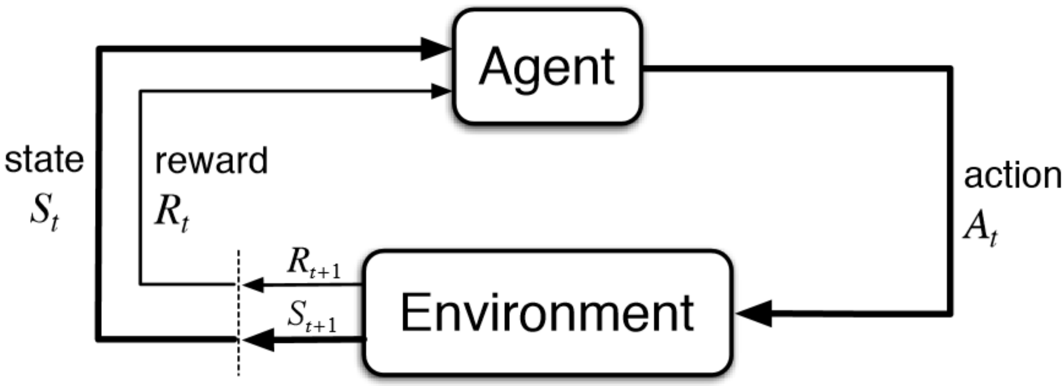

FIGURE 3.4: h

3.2.3 Machine Learning Classification:

Classification is a key process in machine learning in which input is classified into preset classes or categories. It is essential in various kinds of applications, such as fault identification and prognosis . Depending on the nature of the problem and the number of classes involved, many types of classification methods are utilized in machine learning.

3.2.4 Classification Types:

3.2.4.1 Binary Classification:

A binary classification is a sort of classification in which data must be classified into one of two categories. It is frequently employed when a situation can be reduced to a yes/no or true/false answer. Numerous techniques for binary classification tasks have been developed, including logistic regression ,support vector machines (SVM), and decision trees. In addition to fraud detection , binary classification has found uses in spam email filtering .

3.2.4.2 Multiclass Classification:

Multiclass classification is the process of categorizing data into more than two groups. It is utilized when the task necessitates sorting data into numerous unique categories. For multiclass classification tasks, techniques such as random forests , k-nearest neighbors , and neural networks can be used. Multiclass classification has been used in a variety of domains, including image recognition.

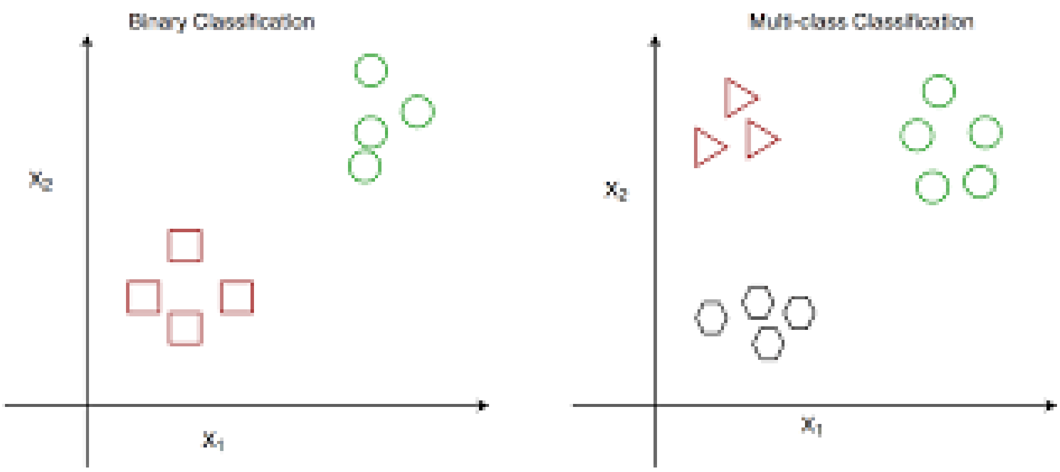

FIGURE 3.5: binary and multi-class classification

3.2.5 Dataset:

Machine learning algorithms rely on data to learn patterns and make predictions. The choice and quality of the dataset are critical in the context of defect detection and prognosis. This section covers the role of learning data in both supervised and unsupervised learning systems.

3.2.6 Learning Data in Supervised and Unsupervised Learning:

In supervised learning, the machine learning model is trained using a dataset that contains input data and labeled outputs. The input data is the features or measurements acquired from the system under observation, whereas the labeled outputs are the known fault or anomaly labels associated with each input data point . The training data enables the model to learn the relationship between the input features and the corresponding fault labels, allowing it to generalize and predict on previously unknown data. In unsupervised learning, the learning data does not include labeled outputs. Instead, the machine learning model aims to identify hidden patterns or anomalies within the dataset without any prior knowledge of the fault labels. The absence of labeled data presents both challenges and opportunities. Unsupervised learning methods, such as clustering , dimensionality reduction , and outlier detection , can uncover interesting structures or deviations in the data that may indicate the presence of faults. However, the lack of labeled data makes it difficult to evaluate the performance of the model objectively.

3.2.6.1 Training Data:

Training Data refers to the subset of the dataset used to train the machine learning model. It typically comprises a significant portion of the available data and is used to optimize the model's parameters or weights through an iterative learning process. The training data should be representative of the system's operational conditions and cover a diverse range of fault scenarios to ensure the model's robustness and generalization capabilities.

3.2.6.2 Labeled Data:

Labeled Data is an important part of supervised learning. It is developed by experts or topic specialists who manually annotate the dataset's defects or oddities. During the training process, these labels serve as the fundamental truth, guiding the model to understand the mapping between input data and the relevant fault classes or severity levels. The availability of labeled data makes it much easier to construct and test supervised learning models for defect detection and diagnosis.

3.2.6.3 Datasets for training, validation, and testing

No AI model can be trained and evaluated on the same data.Because the model is being evaluated on what it already knows, the model's estimate will be biased. This is like to giving students the same questions on an exam that they have already answered in class. We don't know if the student memorized the answers or really understood the issue this way.The same rules apply to machine learning models. Their data volume percentages are as follows:

3.2.6.4 Training data:

The training set is used to train the machine learning model. It is the largest section of the dataset and is often utilized for model parameter optimization. A typical split is 70-80 size and complexity of the dataset.

3.2.6.5 Validation data:

The validation set is used to fine-tune the model throughout the training process. It aids in analyzing various model configurations, hyperparameter tuning, and preventing over-fitting. The validation set is required for model selection and performance comparison. A typical split is roughly 10-15whole dataset.

3.2.6.6 Testing data:

The testing set is used to evaluate the trained model's ultimate performance. It gives an unbiased assessment of the model's capacity to generalize on previously encountered data. To minimize overfitting, the testing set should not be used during model building or parameter adjustment. Typically, 10-20the overall dataset is set aside for testing purposes.

3.2.7 Machine Learning Data Quantitative Requirements:

Having the proper amount and quality of data is critical when using machine learning algorithms to achieve accurate and dependable outcomes. The quantitative needs of data in machine learning include a number of factors that must be considered. Here are some important aspects to consider when assessing the quantitative data requirements for machine learning:

3.2.7.1 Sufficient Data Volume:

In order to train successfully, machine learning algorithms often require a large amount of data. The dataset size can vary based on the intricacy of the problem at hand and the algorithm used. Larger datasets, on average, produce stronger models by capturing a wider range of patterns and variances in the data.

3.2.7.2 adequate Sample Size:

Having an acceptable sample size, in addition to the overall volume of data, is critical for avoiding sampling bias and generating representative results. The sample size should be large enough to capture the population's variety and diversity. Power analysis, for example, can be used to determine the required sample size depending on parameters such as the desired level of confidence and the estimated effect size.

3.2.7.3 balanced Class Distribution:

An imbalanced class distribution happens when the number of instances in distinct classes is unbalanced. Unbalanced data can have a negative impact on model performance in machine learning tasks involving classification or anomaly detection, resulting in biased predictions. To overcome this, it is critical to ensure a fair representation of each class or to correct class imbalances using techniques such as oversampling, undersampling, or synthetic data generation.

3.2.7.4 Data Continuity and Consistency:

In some circumstances, it is critical to assure data collecting process continuity and consistency. This means that the data should be acquired consistently over time or from multiple sources in order to prevent biases introduced by differences in data collecting methods or surroundings.

3.2.7.5 Data Quality:

Data quality is critical for generating dependable and accurate outcomes. The data should be devoid of errors, outliers, missing numbers, and other issues that could induce biases or influence the machine learning algorithm's performance. To achieve excellent data quality, preprocessing techniques such as data cleaning, normalization, and outlier elimination may be required.

3.2.8 Metrics in classification problems:

Metrics are used in machine learning operations to quantify model quality and compare different methods, and their selection and analysis is an important part of a data scientist's job. We'll look at many quality criteria in classification problems and discuss what important when choosing a metric and what may go wrong. Accuracy, precision, recall and specificity:

The confusion matrix is a crucial notion that must be introduced in order to define these metrics in terms of classification mistakes.If we have two classes and an algorithm that predicts whether an object belongs to one of the classes, the classification error matrix will look like this:

Here, represents the algorithm's reaction to the object, and represents the true label of the class on this object. Thus, there are two sorts of classification errors:

false negative (FN) false positive (FP).

	Y=1	Y=0
Y=1	True positive(TP)	False positive(FP)
Y=0	True negative(TN)	False negative(FN)

3.2.8.1 Accuracy:

Accuracy calculates the ratio of correctly classified examples to the total number of instances to determine the overall correctness of the model's predictions. However, if the dataset is imbalanced, meaning the classes have unequal representation, accuracy can be misleading.

$$\frac{TP+TN}{TP+TN+FP+FN}$$

3.2.8.2 Precision :

Precision is the percentage of accurately anticipated positive cases (true positives) out of all positive instances predicted (true positives + false positives). It assesses the model's ability to avoid false positives and indicates the precision of positive predictions.

$$\frac{TP}{TP+FP}$$

3.2.8.3 Recall:

Recall is a percentage of accurately anticipated positive instances (true positives) out of all actual positive instances (true positives + false negatives). It represents the model's capacity to detect positive cases while avoiding false negatives.

$$\frac{TP}{TP+FN}$$

3.2.9 Training Data Preparation:

Training data preparation is an important phase in the machine learning pipeline that includes collecting, cleaning, preparing, and transforming data before it can be used to train a model. The applicability and quality of the training data have a direct impact on the performance and reliability of the final machine learning model.

3.2.9.1 Data Collection and Acquisition:

Data collection and acquisition include locating relevant data sources and utilizing appropriate ways to collect the data required for machine learning tasks. This procedure includes human data gathering, automatic data retrieval from databases or APIs, and data integration from many sources. To ensure the quality and reliability of the collected data, data quality evaluation and filtering must be performed, which includes reviewing the data for completeness, accuracy, consistency, and relevance. Effective data collection and acquisition provide the groundwork for later phases in the machine learning pipeline, making high-quality data available for analysis, model training, and decision-making processes.

3.2.9.2 Data Cleaning:

Data cleaning is a key phase in the training data preparation process that involves finding and dealing with inconsistencies, errors, missing values, outliers, and noise in the dataset. It ensures the data's quality and trustworthiness, resulting in more accurate and dependable machine learning models. Typically, data cleansing is divided into multiple sections:

3.2.9.3 Missing Data:

Missing data refers to the absence of values in a dataset. It is critical to deal with missing data in order to avoid biased or incorrect outcomes. Sections dealing with missing data may include:

3.2.9.4 Identifying missing values:

- Missing value deletion or imputation approaches such as mean, median, or regression imputation

- Advanced imputation methods such as k-nearest neighbors (KNN) or multiple imputation

3.2.9.5 Dealing with Outliers:

Outliers are data points that differ significantly from the rest of the dataset. They can have a negative impact on model performance and lead to skewed predictions. Sections on dealing with outliers may include: - Identifying outliers using statistical approaches or visualization techniques

- Handling outliers with procedures such as pruning, Winsorizing, or substituting with statistical measures such as mean, median, or mode

- Advanced outlier detection methods, such as clustering-based or density-based approaches

3.2.9.6 Check for duplicate:

the same sample data may appear many times in a dataset. This can be developed by combining data from many sources, resulting in comparable data. They should be removed since they may cause the model to overfit certain patterns and make inaccurate predictions

3.3 The k Nearest Neighbors (kNN):

The k Nearest Neighbors (kNN) algorithm is a non-parametric, instance-based machine learning approach that is often used for classification and regression applications. It is a simple yet powerful algorithm that makes predictions based on the similarity of input data points to their nearest neighbors in the training set.

The notion underlying kNN is based on the assumption that comparable instances have similar outputs. In other words, if a new data point is comparable to its k nearest neighbors in the training set, it is likely to belong to the same class or have a similar value as those neighbors. The kNN algorithm applies this notion to create predictions for unknown data.

The k in kNN indicates the number of nearest neighbors evaluated while making predictions. When a new data point needs to be categorised, the method computes the distances between it and all of the instances in the training set. Based on these distances, it then chooses the k closest neighbors. The new data point's class or value is selected by a majority vote or by averaging the classes or values of its k nearest neighbors.

In kNN, the value of k is a significant parameter. A lower k number may result in more flexible decision boundaries, but it may also make the model more susceptible to noise or outliers. A bigger value of k, on the other hand, can result in smoother decision boundaries but may lose some local details.

kNN is a versatile technique that may be used for classification as well as regression. It has been widely used in a variety of disciplines, including image recognition, recommender systems, anomaly detection, and bioinformatics. kNN is frequently used as a baseline technique to measure the performance of more complicated models due to its simplicity and intuitive nature.

3.3.1 Theoretical component of the k-NN algorithm :

k-Nearest Neighbors (kNN) is a flexible technique that uses distance metrics to determine the similarity of data points. The selection of distance measure is critical to the performance of the kNN algorithm. In this section, we will go through some of the most widely used distance measures in kNN:

3.3.1.1 Euclidean Distance:

This metric computes the straight-line distance between two data points in a multi-dimensional space. It is the most commonly used distance metric in kNN and works effectively when the data attributes are continuous and of equal value.

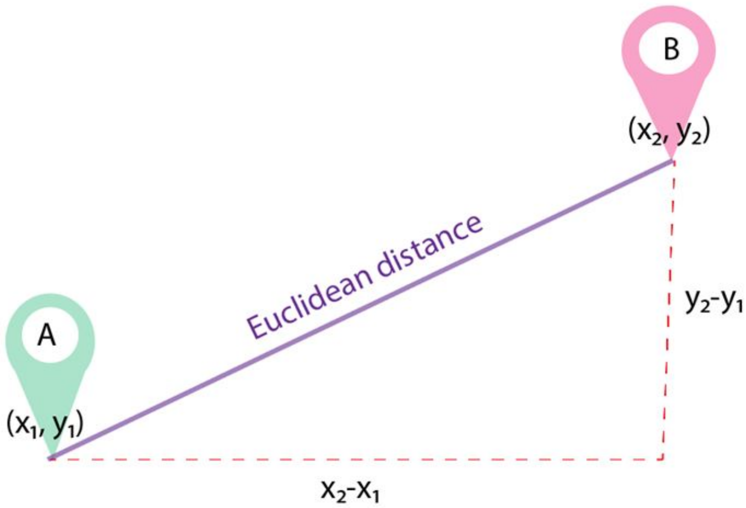

FIGURE 3.6: Euclidean Distance

3.3.1.2 Manhattan Distance:

Also known as city-block distance or L1 norm, it is the sum of the absolute differences between the coordinates of two places. It is beneficial when dealing with traits that are not continuous or when the relevance of distinct attributes varies.

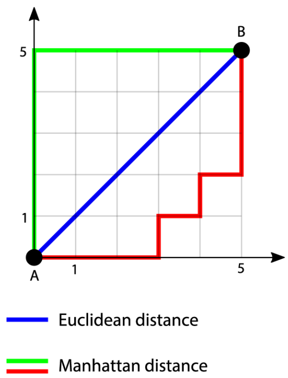

FIGURE 3.7: comparison between Euclidean Distance and Manhattan distance

Cosine Similarity: Unlike the previous metrics, cosine similarity evaluates the angle between two vectors instead of their geographic distance. It is frequently used in text analysis and recommendation systems, where vector magnitude is less important than vector orientation.

A suitable distance metric must be chosen based on the problem at hand and the characteristics of the data. Scale, attribute type, and domain expertise must all be considered. To ensure that all attributes contribute equally to the distance calculation, the features may need to be scaled or normalized. Experimentation and validation with the appropriate evaluation metrics are required to select the best distance metric for a specific kNN application.

As a conclusion, the k Nearest Neighbors (kNN) algorithm is a straightforward yet effective machine learning method for classification and regression tasks. It works under the premise that instances with similar characteristics are more likely to share labels or values. It can predict the future or categorize new data based on the consensus of the neighbors' labels or values by identifying the k closest neighbors to a given data point.

The kNN algorithm has a number of benefits, including simplicity, ease of use, and the capacity to handle both categorical and numerical data. Additionally, it is a non-parametric method, which means it does not make any assumptions about the distribution of the underlying data. However, kNN has some drawbacks. Especially for large datasets, it can be computationally expensive, and if the feature space is high-dimensional or the data is unbalanced, performance may suffer.

Data preparation is essential for kNN to perform at its best. The data must be scaled and normalized, missing values must be handled, and the right distance metrics must be chosen. Cross-validation and other methods can be used to find the ideal k value, which is also crucial.

Although kNN is popular in many fields, including image classification, recommender systems, and anomaly detection, it is not appropriate in every situation. Before using kNN or considering alternative algorithms, it is crucial to take into account the precise specifications and characteristics of the current problem.

In summary, the kNN algorithm provides an approach that is simple and natural to use for pattern recognition and prediction tasks. Effective use of this algorithm in real-world applications depends on an understanding of its advantages, constraints, and recommended data preparation methods.

3.4 Decision Trees alghoritm:

A decision tree is a non-parametric supervised learning algorithm that can be used for classification as well as regression tasks. It has a tree structure that is hierarchical and consists of a root node, branches, internal nodes, and leaf nodes.

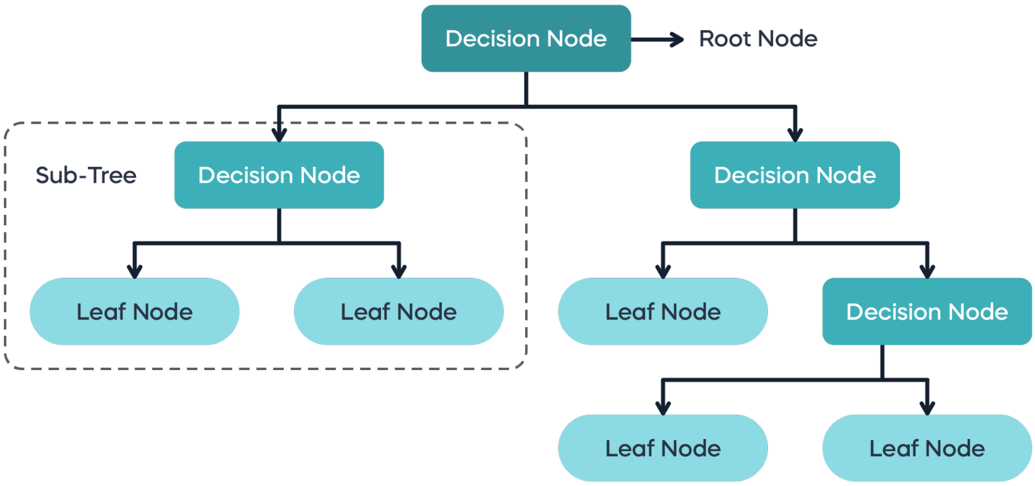

FIGURE 3.8: Decision Trees

A decision tree, as shown in the diagram above, begins with a root node that has no incoming branches. The root node's outgoing branches then feed into the internal nodes, also known as decision nodes. Both node types evaluate the available features to form homogeneous subsets, which are denoted by leaf nodes or terminal nodes. The leaf nodes represent all of the dataset's possible outcomes.

3.4.1 Decision Tree Types

Hunt's algorithm, which was developed in the 1960s in Psychology to model human learning, serves as the foundation for many popular decision tree algorithms, including the following:

- ID3: Ross Quinlan is credited with creating ID3, which stands for "Iterative Dichotomiser 3." To evaluate candidate splits, this algorithm uses entropy and information gain as metrics.

- C4.5: This algorithm is a later iteration of ID3, which was also created by Quinlan. It can evaluate split points within decision trees using information gain or gain ratios.

3.4.2 Advantages and disadvantages of Decision Trees

While decision trees have a wide range of applications, other algorithms generally outperform decision tree algorithms. Decision trees, on the other hand, are especially useful for data mining as well as knowledge discovering tasks. Let's take a closer look at the key benefits and drawbacks of using decision trees:

3.4.2.1 Advantages:

-Decision trees are easy to understand and consume due to their Boolean logic and visual representations. A decision tree's hierarchical nature also makes it easy to see which attributes are most important, which isn't always clear with other algorithms, such as neural networks.

- Requires only minimal data preparation: Decision trees have several characteristics that make them more flexible than other classifiers. It can handle a variety of data types, including discrete and continuous values, and values that are continuous can be converted to categorical values using thresholds. It can also handle missing values, which can be challenging for other classifiers such as Nave Bayes.

- More flexible: Decision trees are easily used for classification and regression tasks, making them more adaptable than other algorithms. It's also insensitive to underlying relationships between attributes; for example, if two variables are highly correlated, the algorithm will only split on one of them.

3.4.2.2 Disadvantages:

- Prone to overfitting: Complex decision trees are prone to overfitting and do not generalize well to new data. This scenario can be avoided by using pre-pruning or post-pruning techniques. When there is insufficient data, pre-pruning halts tree growth, whereas post-pruning removes subtrees with insufficient data after tree construction.

- High variance estimators: Small variations in data can result in a very different decision tree. Bagging, or averaging of estimates, can be used to reduce the variance of decision trees. However, this approach has limitations because it can result in highly correlated predictors.

3.5 conclusion:

provides an overview of fault detection and prognosis methodologies based on machine learning. It begins by going over the fundamentals of fault detection and prognosis, emphasizing their definitions, significance, and the importance of early detection. Following that, the chapter delves into machine learning, providing an overview of its principles and various types. It delves into machine learning classification, including different types of classification, datasets, and the quantitative requirements of learning data. The chapter also discusses metrics in classification problems and training data preparation. Two specific machine learning algorithms are introduced and explained, k Nearest Neighbors (kNN) and Decision Trees, highlighting their theoretical components, types, and the advantages and disadvantages of Decision Trees. Overall, this chapter provides an overview of machine learning-based fault detection and prognosis methodologies.

Chapter 4

Implementation And Results

4.1 work methodology:

4.1.1 Simulation in MATLAB:

The first step is to create a simulation environment in MATLAB to model the behavior of the multi-cell power converter system. This entails designing and implementing the necessary circuitry while taking into account the photovoltaic power system and the selected multilevel inverter topology (such as the cascaded H-bridge or diode-clamped topology). Configure simulation parameters like cell count, switching frequency, and control strategies.

4.1.2 Extraction of Voltage Output Features:

Run simulations for both faulty and healthy cases of the multi-cell power converter. Capture the simulated system's voltage output signals. Then, from these voltage output signals, extract relevant features that can distinguish between faulty and healthy operations. Harmonic content, amplitude variations, and waveform irregularities are examples of such features.

4.1.3 Faulty and Healthy Case Generation:

Create a dataset with a variety of faulty and healthy cases. Introduce specific fault conditions into the simulation model for each case to generate faulty voltage output signals. A short-circuit in a cell, an open-circuit fault, or a faulty switching device are all examples of faults. Similarly, for healthy cases, simulate the system under normal, healthy conditions to generate voltage output signals.

4.1.4 Feature Selection:

From the preprocessed voltage output data, select the most relevant and discriminative features. This step reduces dimensionality by focusing on the key characteristics that contribute to distinguishing between faulty and healthy cases. The feature selection process can be guided by techniques such as correlation analysis, feature importance ranking, or domain knowledge.

4.1.5 Machine Learning Model Development:

Create machine learning models for fault detection and prognosis using the KNN and decision tree algorithms. Divide the preprocessed data into two sets: training and testing. Use the training set to train the models, with the extracted voltage output features as input and the corresponding fault or health status as the target variable. To improve the performance of the models, adjust their hyperparameters, such as the number of neighbors in KNN or the maximum depth in the decision tree.

4.1.6 Model Evaluation:

Using the testing dataset, evaluate the performance of the developed machine learning models. Assess the effectiveness of the models in detecting and classifying faults by measuring key performance metrics such as accuracy, precision, recall, and F1-score. Analyze the strengths and weaknesses of the KNN and decision tree models' performance.

4.1.7 Data Preprocessing:

Prepare the collected voltage output data for machine learning algorithms by preprocessing it. This may entail normalizing the data, handling missing values, and removing any outliers or noise that could impair the algorithms' performance.

4.2 Diagnosis

4.2.1 Data collection and processing:

The construction of the multi-cellular power converter is shown in the following picture, which is made up of four cells and two capacitors that are relevant to the study:

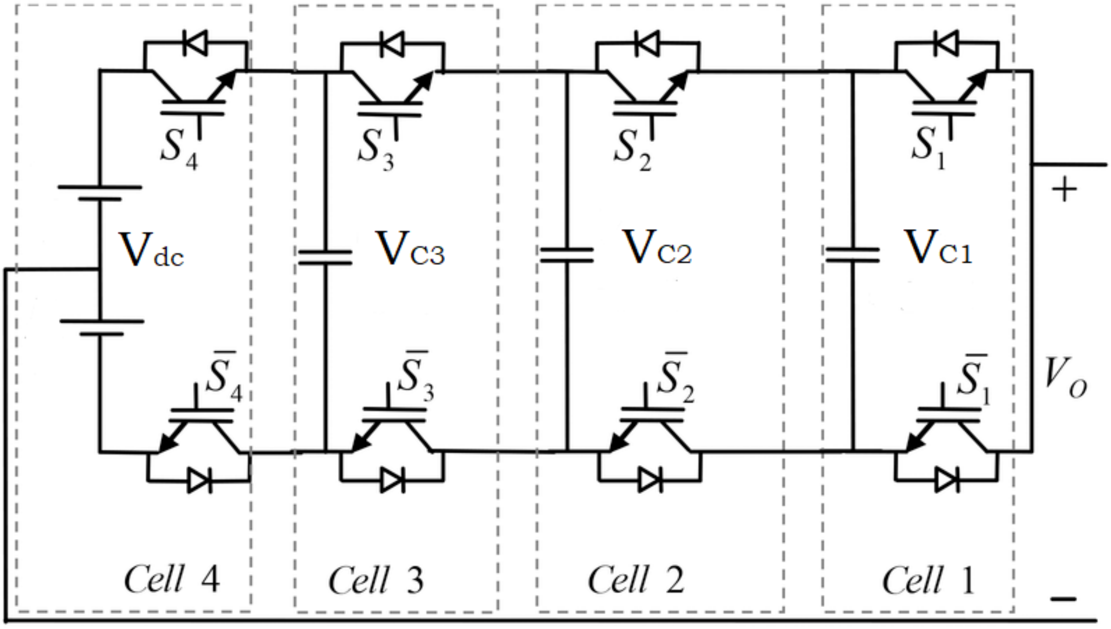

FIGURE 4.1: 4-Cell-FCM-converter-configuration-19

4.2.2 Duty cycle modulation command control

Duty cycle modulation command control will be used to gather the data throughout the operation of this converter. Because the system has not yet attained a stable state or stable behavior to deal with its data, the transitory regime was not used during data collection. If the transitory regime had been used in this study, the results would be questionable. In the course of this system's operation, there will be two potential states: a healthy mode and a malfunctioning mode.

Healthy mode, (absence of failure).

Faulty C1, (failure in the first capacitor).

Faulty C2, (failure in the second capacitor).

Faulty C3, (failure in the Third capacitor).

Faulty C1C2, (failure in both capacitors).

Faulty C1C3, (failure in both capacitors).

Faulty C2C3, (failure in both capacitors).

Faulty C1C2C3, (failure in all capacitors).

4.2.2.1 Duty cycle modulation command control

The signal voltages forms in operation with Duty cycle modulation command control without transitory mode are illustrated in Where we took a period from the main voltage of T = 0.02 sec to apply our study to it, because the voltage is a sinusoidal signal with a frequency of f = 50 Hz, meaning that the voltage is repeated every 0.02 sec. the Figure 4.2 bellow:

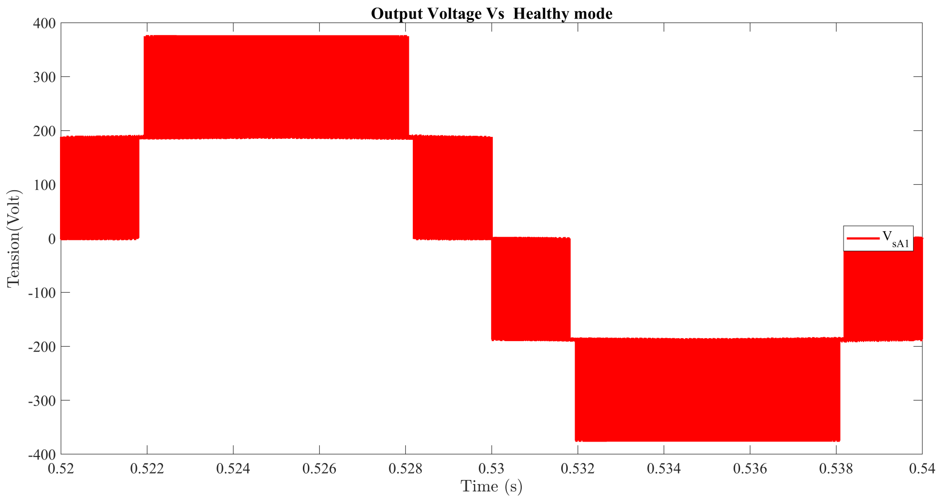

FIGURE 4.2: Healthy case

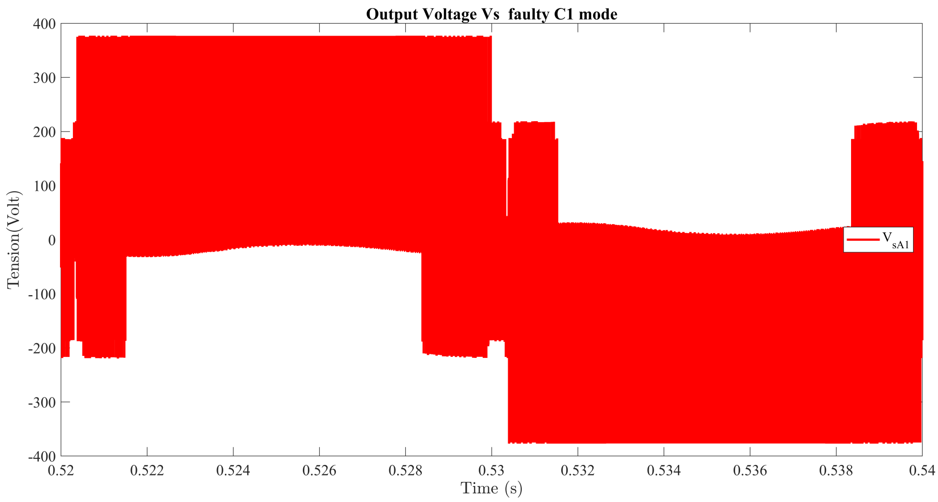

FIGURE 4.3: Faulty C1 case

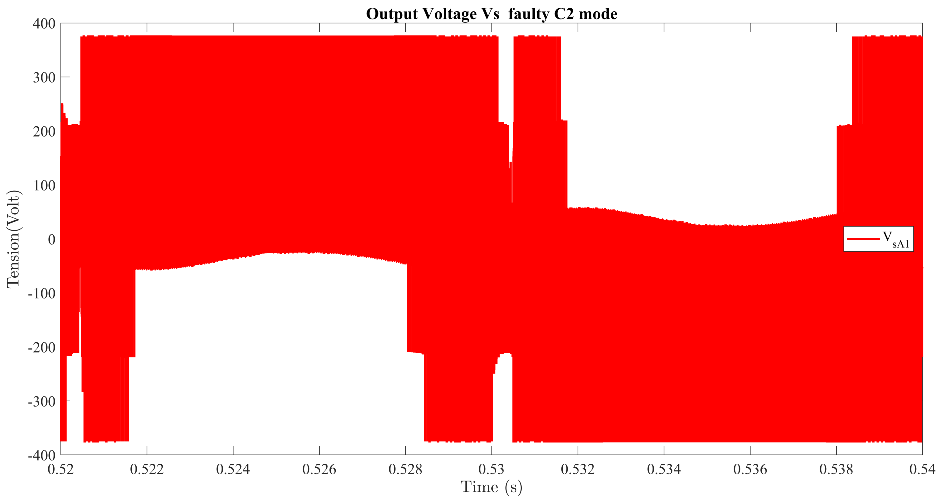

FIGURE 4.4: Faulty C2 case

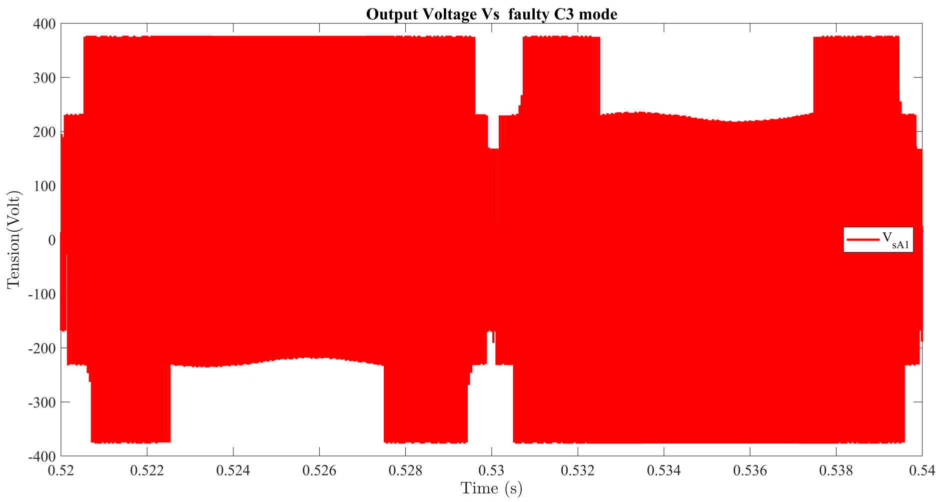

FIGURE 4.5: Faulty C3 case

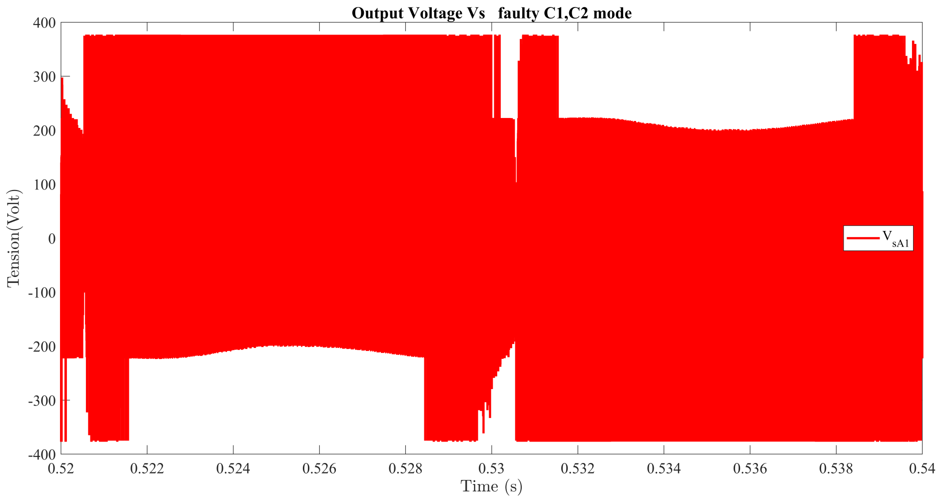

FIGURE 4.6: faulty C1,C2 case

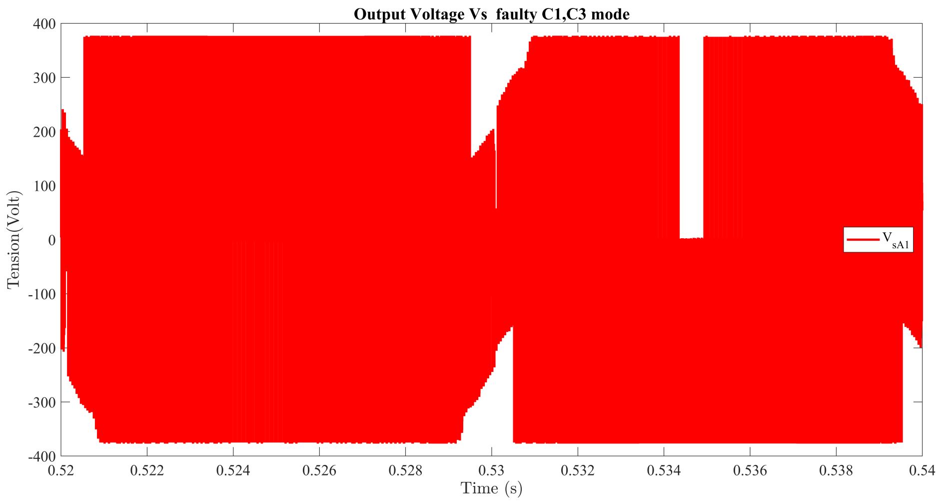

FIGURE 4.7: faulty C1,C3 case

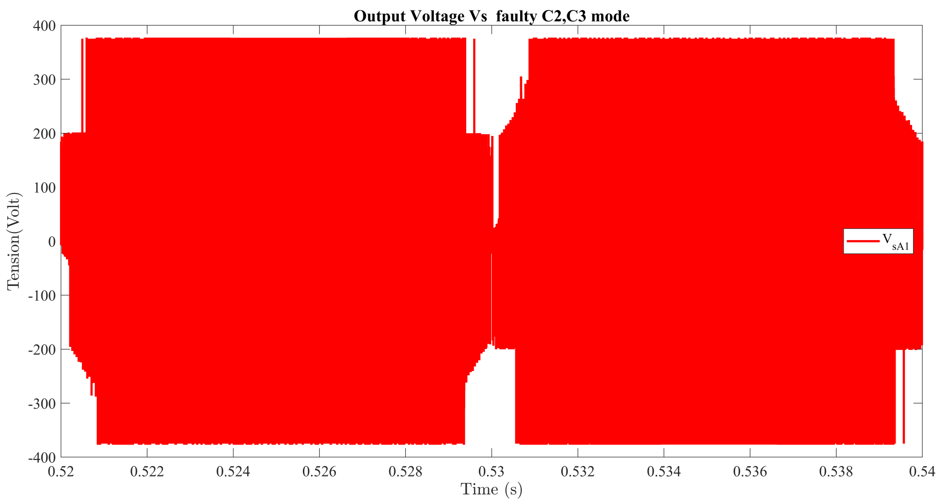

FIGURE 4.8: faulty C2,C3 case

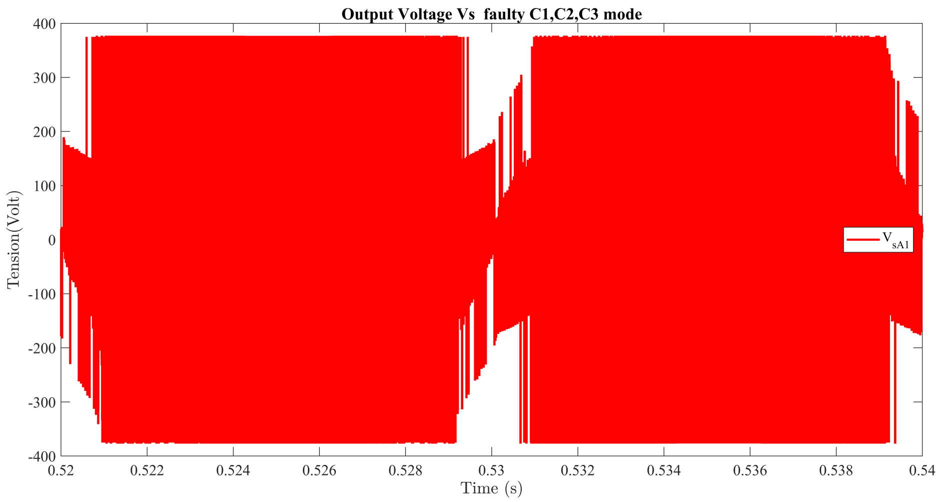

FIGURE 4.9: Faulty C1,C2,C3 case

4.2.2.2 Comment:

As a result of the signal forms, we see a drop and increase in voltage levels, as well as a change in wave-forms, which is caused by the failure of one of the capacitors each time... When a capacitor fails, it creates an open circuit in the circuit, which has a detrimental impact on the voltage levels produced by the Multi-cell power converter.

4.2.3 Data preparation:

After collecting the data, we computed the standard deviation as an additional feature in the different conditions during multi-cellular power converter operation. It is a measure of how much the signal fluctuates from the mean, as the use of this measure is important for the accuracy of the machine learning model against real-world data.

4.2.4 Feature selection:

Drawing the feature space is a necessary step in the prediction model construction process because it allows researchers to examine how data change behaves and is dispersed during system operation. This provides a preliminary evaluation of both the suitability of the features employed and the accuracy of the model's predictions. So, here the features that are thus available are output voltage (Vs), and standard deviation (STD). As shown in the following figure 4.10:

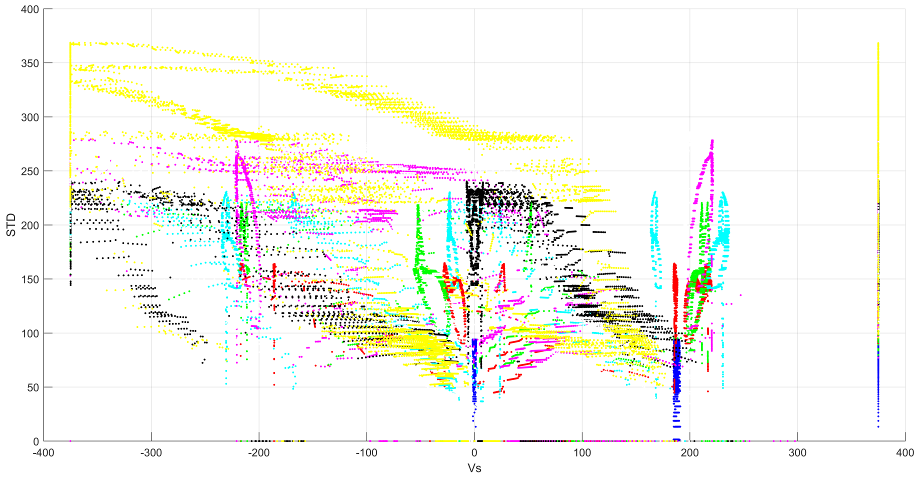

FIGURE 4.10: Feature space in Two dimensions (Vs,STD).

Feature space in two dimensions indicates in the different classes 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.

4.2.5 fault detection application method

4.2.5.1 The k-nearest neighbors (KNN)

The k-Nearest Neighbors (KNN) algorithm is a popular and simple classification and regression algorithm. It works on the principle of finding the closest k data points in the training set to a given test point and using their classification labels to make predictions. The algorithm will be trained using the provided cases, complete with all its features, and its dependencies will be tested for accuracy.

4.2.5.2 The decision tree

The decision tree algorithm is a popular and powerful machine learning algorithm that can be used for both classification and regression tasks. It builds a tree-like model of decisions and their possible consequences based on the features (attributes) of the input data.

After gathering the data and performing the treatments, there are various processes in developing the classification model:

1. A unique symbol must be used to identify each case.
2. Merge all data into a single table using randomization.
3. Divide the whole amount of data into 30% for testing and 70% for training.

each case will be given its own symbol or name to distinguish it. So, it will be 1 for the healthy mode, 2 for faulty C1 mode, 3 for faulty C2 mode, 4 for faulty C3 mode, 5 for faulty C1 and C2 mode, 6 for faulty C1 and C3 mode, 7 for faulty C2 and C3 mode, and 8 for faulty C1 and C2 and C3 mode. To train the algorithm to recognize them, all of this data (cases) will be put together in one table, randomized, and divided into 70%. The model categorization will next be tested using the remaining 30% of the data. It may be determined if the model is incorrect in determining the classification of the case because the data on which the model will be evaluated has a classification. As a result, it is possible to assess the model's reliability and accuracy. .

4.2.6 Applying K-nearest neighbor and decision tree :

Once all necessary actions have been done, the data is ready for model creation. There are several algorithms, including neural networks, KNN, RNN, and SVM. Due to its superior classification accuracy, efficiency, and simplicity, the K-nearest neighbor technique and the decision tree approach were chosen for this study. The first stage in the process is to segment the data. To accomplish this, first classify each occurrence into its own category, then collect the data into a single table and randomly divide it into two sets of data: one for training and one for testing. With the remaining 30% of the data being tested, as much data as possible. To make sure the model is accurate and error-tolerant, the test model is trained.

4.2.6.1 Building the classification model and efficiency:

After building the prediction model in K-nearest neighbor technique and decision tree, their performance should be evaluated using a number of settings, as shown in the following Table 4.4::

	Accuracy	recall	precision	specificity	F1-score
k-nearest neighbors (K=3)	94.60%	94.61%	94.59%	94.59%	99.22%
The decision tree	96.37%	96.37%	96.37%	96.37%	99.48%

Table 4.1: Table metrics of evaluation in the two method

4.2.7 Comparison between the two method

After trying the algorithms and evaluating their performance, we find that the decision tree provides a transparent decision-making structure that is easy to understand and interpret. The resulting tree can be analyzed to gain insight into the decision-making process.

4.2.8 Conclusion

Two multi-cell transformer failure classification models were built in this study using the Non-Invasive Load Monitoring (**NILM**) approach. We discovered that the decision tree model yields the best results because it can identify and locate faults in multi-cell power transformers at an early stage. The main lessons were that duty cycle adjustment commands were managed smoothly, and we were able to detect and classify errors when controlling duty cycle adjustment using the decision tree machine learning algorithm with high accuracy. This was made possible by the system's operation, which displays the output signal in response to all potential faults and is independent of the signal from the capacitors inside a multi-cell converter. These results are important because they enable the identification and classification of faults in similar power inverters, thereby minimizing downtime, minimizing damage, while preventing system shutdowns for an extended period of time and consequent catastrophic damage.

4.3 Prognosis

4.3.1 Data collection and processing:

The control of the service cycle modification order will be used to collect data and our study will be applied to the occurrence of failure at the level of the capacitor C1 only, but we added changes during the control as we included variables such as the fault slope and the fault amplitude, which in turn helps us to collect data to estimate the expected failure time to occur on Intensive level C1. We fix the value of the amplitude error at 100 and then change the value of the slope of the amplitude error according to the time to expect the failure to occur. The temporary system was not used during data collection. If the interim regimen was used in this study, the results would be questionable. In the course of operating this system, there will be two principles for study: Healthy mode, (failure in the first capacitor C1). • The remaining time for failure to occur is 15 days.
• The remaining time for failure to occur is 30 days.

4.3.2 Duty cycle modulation command control

The signal voltages forms in operation with Duty cycle modulation command control without transitory mode are illustrated in Where we took a period from the main voltage of T = 9.5 sec into T = 10 sec to apply our study to it, until the difference is in the voltage output values. the Figure 4.2 bellow :

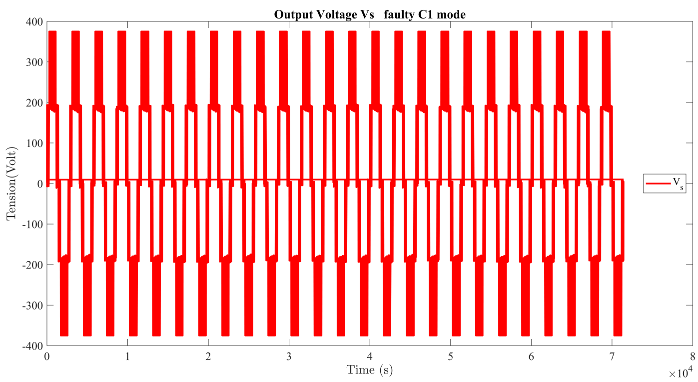

FIGURE 4.11: Fault model until 15 days

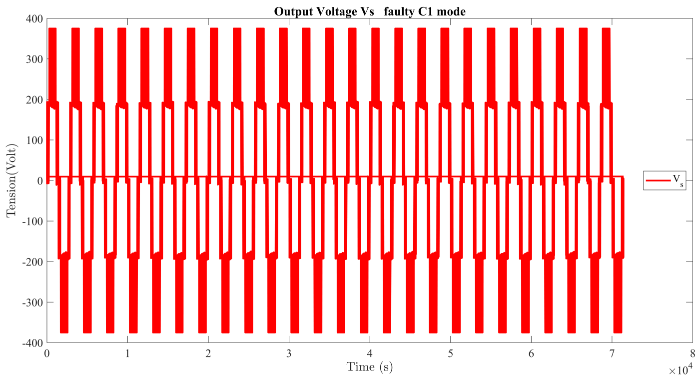

FIGURE 4.12: Fault model until 30 days

4.3.2.1 Comment:

Signal distortion causes minute changes and alterations that are invisible to the unaided eye to be observed. This is explained by how the degree of deterioration has been changing over time.

4.3.3 Data preparation:

After collecting the data, we computed the median and mean value as additional features at different error amplitude slope values during multi-cell transducer operation. It may even be beneficial to use both the mean and median as separate features to capture different aspects of the data distribution.

4.3.4 Feature selection:

The selection of features depends on their ability to capture relevant information related to the faults and their impact on the inverter's performance. In the case of multilevel inverters, Output Voltage (Vs), Mean, and Median are commonly chosen as features for the following reasons:

Output Voltage (Vs): The output voltage of the multilevel inverter is a critical parameter that directly reflects its operational status. Changes in output voltage can indicate the presence of faults or abnormalities in the inverter's components or control system. Monitoring the output voltage allows you to identify deviations from expected behavior and detect potential faults.

Mean: The mean value of various electrical parameters, such as current or voltage, indicates the system's average behavior. It is possible to detect any deviations or shifts from normal operating conditions by analyzing the mean values over time, which may indicate the presence of faults. Changes in the mean value can indicate component degradation, unbalance, or other problems affecting inverter performance.

Median: The median is a statistical measure that identifies the middle value in a dataset. Compared to the mean, it is less sensitive to outliers and offers a reliable evaluation of the data's central tendency. It is possible to find anomalies or flaws that the mean might not have caught by taking into account the median value of the pertinent parameters. This enhances the AI model's robustness and broadens its scope for defect detection.

In addition to the chosen features of Output Voltage (Vs), Mean, and Median, it is important to note that the current parameter is rarely used as a stand-alone feature for fault diagnosis in multilevel inverters. This is because the current is directly proportional to the inverter system's charge and load conditions.

In conclusion, selecting Output Voltage (Vs), Mean, and Median as characteristics for the AI model in the diagnosis of failures in multilayer inverters enables the capture of crucial data regarding the behavior and functioning of the inverter. These characteristics give the model information about the output voltage, typical behavior, and statistical patterns of pertinent parameters, allowing it to correctly identify and forecast failures.

4.3.5 prognosis application method

4.3.5.1 The k-nearest neighbors (KNN)

A well-liked and straightforward classification and regression approach is the k-Nearest Neighbors (KNN) algorithm. It operates on the idea of locating the k data points in the training set that are most similar to a particular test point and utilizing their classification labels to predict the future. The algorithm will be trained using the supplied examples, including all of its characteristics, and the correctness of its dependencies will be examined.

There are several steps in constructing the classification model after collecting the data and conducting the treatments:

1. Each instance must be identified by a special symbol. 2. Use randomization to combine all data into a single table. 3. Split the total quantity of data in half, with 30% used for training and 70% for testing.

Each instance will have a unique symbol or name to help identify it. It will thus be 1 for RUL=15 days and 2 for RUL=30 days. All of this data (cases) will be combined into a single table, randomly divided into 70%, and used to train the algorithm to distinguish them. The remaining 30% of the data will then be used to evaluate the model categorisation. Because the data on which the model will be judged includes a classification, it may be determined if the model is inaccurate in determining the classification of the case. As a consequence, the model's dependability and accuracy may be evaluated.

4.3.6 Applying K-nearest neighbor

When all required steps have been taken, the data is prepared for model building. The K-nearest neighbor technique were selected for this investigation due of their higher classification accuracy, effectiveness, and simplicity. Data segmentation is the initial step in the procedure. To achieve this, categorize each instance into its own category first. Then, compile the data into a single table and randomly split it into two sets of information: one for training and one for testing. As much data as feasible, with the remaining 30% of the data being tested. The test model is trained to ensure that the model is accurate and error-tolerant.

4.3.6.1 Building the classification model and efficiency:

After building the prediction model in K-nearest neighbor technique, their performance should be evaluated using a number of settings, as shown in the following Table 4.4:

	Accuracy	recall	precision	specificity	F1-score
k-nearest neighbors (K=3)	98.87%	98.87%	98.87%	98.87%	98.87%

TABLE 4.2: Table metrics of evaluation in K-nearest neighbor method

4.3.7 Comparison between the two method

After trying the algorithms and evaluating their performance, we find that the decision tree provides a transparent decision-making structure that is easy to understand and interpret. The resulting tree can be analyzed to gain insight into the decision-making process.

4.3.8 Conclusion

In this work, models were created to categorize the multicellular transducer failure time using a non-invasive load monitoring (NILM) strategy. In this work, a model was created utilizing the closest neighbor machine learning technique to forecast the failure time of multicellular inverters. Because it can detect multi-cell transducer failure at an early stage, we discovered that it produces the greatest results.

General Conclusion

In conclusion, the focus of this research was on the creation and implementation of diagnosis and prognosis methods for multilevel inverters using only the output voltage and Non-Intrusive Load Monitoring (NILM) with the K-Nearest Neighbors (KNN) machine learning algorithm. While eliminating the need for extra sensors or intrusive monitoring, the research sought to meet the requirement for efficient failure detection, localisation, and prediction approaches in multilevel inverters.

The feasibility and effectiveness of diagnosing and prognosticating faults in multilevel inverters were effectively proved by this research through the use of NILM and the analysis of output voltage data. Accurate fault localisation and detection were made possible by utilizing the KNN algorithm, which is renowned for its effectiveness and simplicity in classification jobs. This allowed for prompt intervention and decreased system downtime. The dissertation made a contribution to the area by highlighting the potential of employing output voltage data and NILM as a non-intrusive and economical method for diagnosing multilevel inverter problems. It stressed the significance of utilizing machine learning methods to evaluate and understand the voltage data for defect detection, such as KNN.

The research also showed how the established methodology could predict the future. It was feasible to predict performance decline and calculate the remaining useful life of multilevel inverters by examining patterns and trends in the output voltage data. This prognostic data enables proactive maintenance planning, asset management optimization, and the reduction of unexpected failures. Overall, the dissertation provided insights into the potential of NILM and the KNN algorithm for defect detection and performance prediction, highlighting the significance of diagnosis and prognosis in multilevel inverters. The study helps to improve multilevel inverter systems' availability, availability, and efficiency while lowering maintenance costs and raising overall system performance.

In order to improve the diagnostic and prognostic capabilities of multilevel inverters, more study can be done to investigate the integration of other data sources and sensor information. For useful applications in industry and power systems, it will also be helpful to look at the scalability and real-time implementation of the proposed methodology.

Finally, the dissertation sheds light on the significance of diagnosis and prognosis in multilevel inverters by utilizing NILM with output voltage data and the KNN machine learning technique. It lays the groundwork for future advances in fault detection, localization, and performance prediction, ultimately contributing to the creation of more reliable and efficient multilevel inverter systems.

Bibliography

[1] Abdul Afram and Farrokh Janabi-Sharifi. "Theory and applications of HVAC control systems–A review of model predictive control (MPC)". In: *Building and Environment* 72 (2014), pp. 343–355.

[2] Georgios-Fotios Angelis et al. "NILM applications: Literature review of learning approaches, recent developments and challenges". In: *Energy and Buildings* (2022), p. 111951.

[3] Ebrahim Babaei et al. "Reduction of dc voltage sources and switches in asymmetrical multilevel converters using a novel topology". In: *Electric Power Systems Research* 77.8 (2007), pp. 1073–1085.

[4] E Cengelci et al. "A new medium-voltage PWM inverter topology for adjustable-speed drives". In: *IEEE Transactions on Industry Applications* 35.3 (1999), pp. 628–637.

[5] Wael Charfi et al. "Performance evaluation of a solar photovoltaic system". In: *Energy Reports* 4 (2018), pp. 400–406.

[6] Davide De Simone, Luigi Piegari, and Salvatore D'Areo. "Comparative analysis of modulation techniques for modular multilevel converters in traction drives". In: *2018 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*. IEEE. 2018, pp. 593–600.

[7] Atieh Delavari, Innocent Kamwa, and Amin Zabihinejad. "A comparative study of different multilevel converter topologies for high power photovoltaic applications". In: *2016 7th Power Electronics and Drive Systems Technologies Conference (PEDSTC)*. IEEE. 2016, pp. 415–420.

[8] Nikolas Flourentzou, Vassilios G Agelidis, and Georgios D Demetriades. "VSC-based HVDC power transmission systems: An overview". In: *IEEE Transactions on power electronics* 24.3 (2009), pp. 592–602.

[9] Janos Gertler. *Fault detection and diagnosis in engineering systems*. CRC press, 1998.

[10] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

[11] Michael I Jordan and Tom M Mitchell. "Machine learning: Trends, perspectives, and prospects". In: *Science* 349.6245 (2015), pp. 255–260.

[12] Samir Kouro et al. "Recent advances and industrial applications of multilevel converters". In: *IEEE Transactions on industrial electronics* 57.8 (2010), pp. 2553–2580.

[13] Devineni Gireesh Kumar et al. "Performance Analysis of an Optimized Asymmetric Multilevel Inverter on Grid Connected SPV System". In: *Energies* 15.20 (2022), p. 7665.

[14] Jih-Sheng Lai and Fang Zheng Peng. "Multilevel converters-a new breed of power converters". In: *IEEE Transactions on industry applications* 32.3 (1996), pp. 509–517.

[15] Enrique Ledezma et al. "Development of a modular configurable multi-megawatt power amplifier". In: *IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2013, pp. 631–636.

[16] Guannan Li et al. "Review on fault detection and diagnosis feature engineering in building heating, ventilation, air conditioning and refrigeration systems". In: *IEEE Access* 9 (2020), pp. 2153–2187.

[17] S Llorente et al. "A comparative study of resonant inverter topologies used in induction cookers". In: *APEC. Seventeenth Annual IEEE Applied Power Electronics Conference and Exposition (Cat. No. 02CH37335)*. Vol. 2. IEEE. 2002, pp. 1168–1174.

[18] Shuai Lu and Keith A Corzine. "Advanced control and analysis of cascaded multilevel converters based on PQ compensation". In: *IEEE Transactions on Power Electronics* 22.4 (2007), pp. 1242–1252.

[19] Bo Luo et al. "Early fault detection of machine tools based on deep learning and dynamic identification". In: *IEEE Transactions on Industrial Electronics* 66.1 (2018), pp. 509–518.

[20] Laxman Maharjan et al. "State-of-charge (SOC)-balancing control of a battery energy storage system based on a cascade PWM converter". In: *IEEE Transactions on Power Electronics* 24.6 (2009), pp. 1628–1636.

[21] Mario Marchesoni, Maurizio Mazzucchelli, and Sandro Tenconi. "A nonconventional power converter for plasma stabilization". In: *IEEE Transactions on Power Electronics* 5.2 (1990), pp. 212–219.

[22] R Marquardt. "A new modular voltage source inverter topology". In: *Conf. Rec. EPE 2003*. 2003.

[23] Rainer Marquardt. "Modular multilevel converters: State of the art and future progress". In: *IEEE Power Electronics Magazine* 5.4 (2018), pp. 24–31.

[24] Shivinder Mehta and Vinod Puri. "A Review of different Multi-Level Inverter topologies for Grid Integration of Solar Photovoltaic System". In: *Renewable Energy Focus* (2022).

[25] Liang Ming and Jinsong Zhao. "Review on chemical process fault detection and diagnosis". In: *2017 6th international symposium on advanced control of industrial processes (AdCONIP)*. IEEE. 2017, pp. 457–462.

[26] Akira Nabae, Isao Takahashi, and Hirofumi Akagi. "A new neutral-point-clamped PWM inverter". In: *IEEE Transactions on industry applications* 5 (1981), pp. 518–523.

[27] Ali Nami et al. "A new configuration for multilevel converters with diode clamped topology". In: *2007 International Power Engineering Conference (IPEC 2007)*. IEEE. 2007, pp. 661–665.

[28] Fang Zheng Peng and Jih-Sheng Lai. "Dynamic performance and control of a static var generator using cascade multilevel inverters". In: *IAS'96. Conference Record of the 1996 IEEE Industry Applications Conference Thirty-First IAS Annual Meeting*. Vol. 2. IEEE. 1996, pp. 1009–1015.

[29] Mahran Quraan, Taejung Yeo, and Pietro Tricoli. "Design and control of modular multilevel converters for battery electric vehicles". In: *IEEE Transactions on Power Electronics* 31.1 (2015), pp. 507–517.

[30] Sebastian Rivera et al. "Electric vehicle charging station using a neutral point clamped converter with bipolar DC bus". In: *IEEE transactions on Industrial Electronics* 62.4 (2014), pp. 1999–2009.

[31] Jose Rodriguez, Jih-Sheng Lai, and Fang Zheng Peng. "Multilevel inverters: a survey of topologies, controls, and applications". In: *IEEE Transactions on industrial electronics* 49.4 (2002), pp. 724–738.

[32] 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.

[33] Antonio Ruano et al. "NILM techniques for intelligent home energy management and ambient assisted living: A review". In: *Energies* 12.11 (2019), p. 2203.

[34] Antonio Ruano et al. "NILM techniques for intelligent home energy management and ambient assisted living: A review". In: *Energies* 12.11 (2019), p. 2203.

[35] Mamatha Sandhu and Tilak Thakur. "Multilevel inverters: literature survey–topologies, control techniques & applications of renewable energy sources-grid integration". In: *Int. Journal of Engineering Research and Applications* 4.3 (2014), pp. 644–652.

[36] Guoqiao Shen et al. "An improved control strategy for grid-connected voltage source inverters with an LCL filter". In: *IEEE Transactions on Power Electronics* 23.4 (2008), pp. 1899–1906.

[37] Jianjiang Shi et al. "Research on voltage and power balance control for cascaded modular solid-state transformer". In: *IEEE Transactions on Power Electronics* 26.4 (2011), pp. 1154–1166.

[38] Gui-Jia Su, John W McKeever, and Kelly S Samons. “Modular PM motor drives for automotive traction applications”. In: *IECON’01. 27th Annual Conference of the IEEE Industrial Electronics Society (Cat. No. 37243)*. Vol. 1. IEEE. 2001, pp. 119–124.

[39] Leon M Tolbert, Fang Z Peng, and Thomas G Habetler. “Multilevel inverters for electric vehicle applications”. In: *Power Electronics in Transportation (Cat. No. 98TH8349)*. IEEE. 1998, pp. 79–84.

[40] Leon M Tolbert, Fang Zheng Peng, and Thomas G Habetler. “A multilevel converter-based universal power conditioner”. In: *IEEE Transactions on Industry Applications* 36.2 (2000), pp. 596–603.

[41] Xiang Wu et al. “Optimized common-mode voltage reduction PWM for three-phase voltage-source inverters”. In: *IEEE Transactions on Power Electronics* 31.4 (2015), pp. 2959–2969.

[42] Lu Yue, Inhwan Lee, and Xiu Yao. “Tokamak vertical stability coil power supply based on modular multilevel converter”. In: *2016 IEEE International Power Modulator and High Voltage Conference (IPMHVC)*. IEEE. 2016, pp. 447–452.

[43] Seyed Fariborz Zarei, Hossein Mokhtari, and Frede Blaabjerg. “Fault detection and protection strategy for islanded inverter-based microgrids”. In: *IEEE Journal of Emerging and Selected Topics in Power Electronics* 9.1 (2019), pp. 472–484.

[44] Zhen Zhao, Yi-gang Sun, and Jun Zhang. “Fault detection and diagnosis for sensor in an aero-engine system”. In: *2016 Chinese Control and Decision Conference (CCDC)*. IEEE. 2016, pp. 2977–2982.

[45] Keliang Zhou and Danwei Wang. “Relationship between space-vector modulation and three-phase carrier-based PWM: a comprehensive analysis [three-phase inverters]”. In: *IEEE transactions on industrial electronics* 49.1 (2002), pp. 186–196.