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

University Kasdi Merbah– OUARGLA

Faculty of New Technologies of Information and Communication

Department of Computer Science and Information Technology

Dissertation

Field: Mathematics and Computer Science
Sector: Computer Science
Specialty: Network Administration and Security

Presented by:

BAHLOUL Nousseiba

MOKHTARI Nouciba

Title:

Fault detection in photovoltaic power converter

Jury members:

Dr KAFI M.Redouane	Supervisor	UKM Ouargla
Dr KHALDI Amine	Examinator	UKMOuargla
Dr BOUKHAMLA Akram	Examinator	UKMOuargla

Academic Year: 2021 – 2022

Acknowledgments

رَبِّ أَوْزِعْنِي أَنْ أَشْكُرَ نِعْمَتَكَ الَّتِي أَنْعَمْتَ عَلَيَّ وَعَلَىٰ وَالِدَيَّ وَأَنْ أَعْمَلَ صَالِحًا تَرْضَاهُ وَأَدْخِلْنِي بِرَحْمَتِكَ فِي عِبَادِكَ الصَّالِحِينَ –النمل19-

Praise be to God, so much praise for his innumerable blessings, and from him he has succeeded us to complete this work, where it was bounty to us, was great.

In gratitude we extend our sincere thanks, sincere praise and sincere appreciation to all of us help us accomplish this work and especially mention Professor **K. Mohamed Redouane** for his acceptance supervising this modest study with all effort and time with care, guidance and support.

We are pleased to extend our sincere thanks and appreciation to Professor **Hicham Henna** for helping us with the Dissertation.

We also do not forget the professors of the Department of Computer Science for their advice and guidance.

Finally, we are pleased to extend our sincere thanks and appreciation to all those who helped us little or much, hoping for them from God Almighty reward, reward and praise, and praise be to God, Lord of the worlds.

Dedication

From the bottom of my heart, and with love and sincerity, I would like to dedicate this thesis to everyone who was with me in my difficult times before the happy ones, in my weakness before my strength, and in every stage of my life.

To the first woman of my life ... To the one who welcomes me with her smile and bid me farewell with her prayers ... To the one who taught me and gave me love and affection ... To whomever, her supplication was the secret of my success ... To the one who taught me that loving goodness to others, beautiful giving, Good heart, optimism in every matter

My Mother, my love and my paradise may God protect her.

To the one who taught me the giving without waiting ...To the one who holds his name with great pride ... To the one who taught me modesty, values and morals ... To the one who taught me pride in religion ... To the one who taught me to trust in Allah

My father and the crown of my head may God protect him.

My Sisters and My Brothers ...Words are racing and competing to organize the contract of thanks and praise that all of you deserve ... I wish you all success and reimbursement in your lives ... And also to everyone whom I love, to each member of my extended family, my teachers, and all my friends ... I love you all, thank you very much

Finally, my graduation dream came true, Thanks to Allah.

Nousseiba BAHLOUL

I would like to dedicate this thesis to everyone who was with me in my difficult times before the happy ones, in my weakness before my strength, and in every stage of my life...

To the first woman of my life...To the fragrance and the wind of Heaven...To the moon that illuminated my days and my path... To the one who welcomes me with her smile and bid me farewell with her prayers... To the one who taught me to trust in Allah, To the one who taught me and gave me love and affection, which was and will remain my source of strength for as long as I live...

My Mother, and my role model, my love.

*To my dear father ... Whatever I try to give you, I will not fulfill your right so accept my love and appreciation and a kiss that I print it on your forehead ... To the best supporter, bond, and the secret of my happiness...**My darling Father**, all gratitude and respect to you...*

To my brothers those who have had a great impact on many obstacles and difficulties: Amel, Radia, Kaouthar, Khadidja, Fatima.

To my loved ones and friends, I thank them all:Rayhana,Icherak, Wiam

To anyone who has helped me do this work, near or far.

Nouciba MOKHTARI

Abstract

Fault detection is a sub-field of control engineering that is concerned with monitoring a system, and identifying expected fault.

The main objective of this thesis is to propose a method for detecting the expected errors in the system converter for photovoltaic energy, And this process is controlled by the expected ratio in the same converter system between its normal state and its abnormal state using machine learning algorithms and determining the detection of expected faults with their location. The KNN, NB and SVM algorithms was used to help us in the study.

Keywords: fault detection, machine learning, converter, system.

Résumé

La détection des défauts est un sous-domaine de l'ingénierie de contrôle qui concerne la surveillance d'un système et l'identification des défauts attendues.

L'objectif principal de cette thèse est de proposer une méthode pour détecter les défauts attendus dans le convertisseur système pour l'énergie photovoltaïque, Et ce processus est contrôlé par le rapport attendu dans le même système de convertisseur entre son état normal et son état anormal en utilisant des algorithmes d'apprentissage automatique et en déterminant la détection des défauts attendus avec leur emplacement. L'algorithme KNN, NB et SVM a été utilisé pour nous aider dans l'étude.

Mots-clés : détection des défauts, apprentissage automatique, convertisseur, système.

ملخص

يعد اكتشاف الأعطال مجالًا فرعيًا لهندسة التحكم يهتم بمراقبة النظام وتحديد الخطأ المتوقع.

الهدف الرئيسي من هذه الرسالة هو اقتراح طريقة للكشف عن الأخطاء المتوقعة في محول النظام للطاقة الكهروضوئية ، ويتم التحكم في هذه العملية من خلال النسبة المتوقعة في نظام المحول نفسه بين حالته الطبيعية وحالته غير الطبيعية باستخدام خوارزميات التعلم الآلي وتحديد كشف الأعطال المتوقعة بموقعها. تم استخدام خوارزمية KNN، NB و SVM لمساعدتنا في الدراسة.

الكلمات المفتاحية: كشف الأخطاء ، التعلم الآلي ، المحول ، النظام.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
FD	Fault Detection
FDI	Fault Detection And Isolation
FL	Fuzzy Logic
FTC	Fault Tolerant Control
HOF	Hard-Over-Failure
JDBC	Java Data Base Connectivity
KNN	K-Nearest Neighbor
LIP	Lock-In-Place
LOE	Loss Of Effectiveness
ML	Machine Learning
NB	Naïve Bayes
NN	Neural Networks
ODBC	Open Data Base Connectivity
PV	PhotoVolatic
SVM	Support Vector Machine

General Introduction

In order to improve efficiency and safety, methods for detecting faults have become increasingly important for many operations, and this applies to systems related to safety and systems important in our lives. The presence of faults is one of the most important problems that you face on a group of systems, machines, motors, and even sensors. It also plays an important role in high-cost operations. Early detection of faults can also help avoid many abnormal events of the systems

PVconverter systems are one of the systems on which fault detection can be applied, so the proposed method of avoiding total or partial damage to the system and its costs and reducing its symptoms is to detect faults on a set of data in the normal (correct state) and abnormal state (with a set of fault) of the system. a variety of faults that may encounter PV converter systems and compare them.

The primary goal of our thesis is highlighted by applying machine learning algorithms to faults detect.

This thesis is divided into three chapters:

The first chapter provides a general description of faults in various systems, their definition, types, consequences, various methods of detection and the importance of initial detection of faults.

In the second chapter, we will give a simple description on the PV converter highlighting machine learning, its divisions, algorithms, and fault detection based on machine learning techniques.

In the third chapter, we will provide an overview of the work environment and data analysis with the application of machine learning techniques and the results obtained.

Chapter1: Fault Detection

Overview

1.1 Introduction

Faults detection has been the subject of many studies since ancient times. And it was developed over the years and using different techniques and applied in various fields. A fault is defined as an unacceptable deviation of one or more features in the system from its usual behavior.

In this chapter, we will give an overview of fault detection: Defining faults, their causes, consequences and their various types. Fault detection definition and their importance and their various methods. And some real-world fault detection applications.

1.2 Faults

A fault is an unpermitted change or deviation of at least one parameter or characteristic of the system from the acceptable, usual and standard condition [1]. This is a change in the system that produces an, inadmissible reduction in system quality and performance [27].

•Causes

Faults occur due to wear, implementation errors, design errors, damages, use errors or human errors. All these faults yield deviations of input and output properties of the system from the normal state which further results in a degradation or even a loss of the system task [28].

•Consequences

Many consequences result from these faults, which could be catastrophic, for example: economic losses, worse performances, waste of raw materials, environmental damages, human damages, lower quality, lower production, energy waste... [2]

1.3 Type of faults

1.3.1 Depending on the faulty component

plant component, sensor, actuator [2].

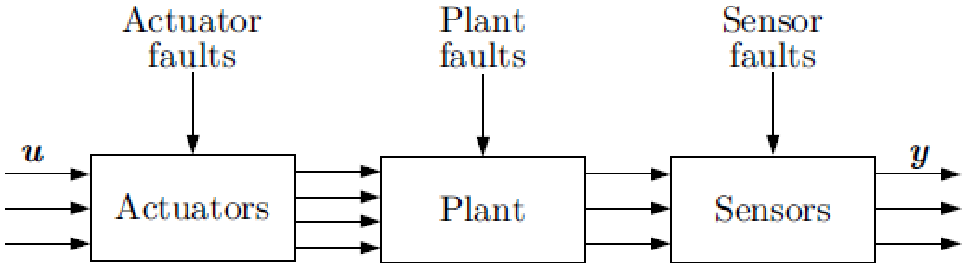


Figure 1: Fault models based on faulty component

1.3.1.1 Components faults

The plant component faults are commonly represented as cases where some condition changes in the system rendering the nominal dynamic equation of the system invalid. Component faults are also dependent on the system being monitored [3].

Component faults may have slight to severe consequences. For example, an unexpected failure of the gearbox in a helicopter may cause significant economic loss.

Some examples include power source (e.g., battery, solar arrays) failures in satellites; friction faults due to lubricant deterioration.

Nonetheless, these kinds of faults commonly occur due to corrosion of system components. Thus, it is extremely crucial to diagnose these faults at initial stages of component degradation in order to avoid catastrophic consequences. Initial diagnosis of incipient component faults allows performing timely, on-demand upkeep operations on the degraded component, which may also involve component replacement.

1.3.1.2 Sensor faults

The sensors are the ones that transmit information about the behavior of the system and its internal states, they are the interface of the system's output to the external world. Therefore, sensor faults may cause substantial performance decline of all decision-making systems or processes that depend on data integrity for making decisions. Thus, the existence of faults in sensors may decline state estimates and consequently result in inefficient and/or inexact control [3].

For example, feedback control systems, safety control systems, quality control systems, and particularly health monitoring and fault diagnosis systems. In a feedback control system, sensors are used either to directly measure system states or to generate state estimates for the feedback control law.

Common sensor faults/failures include: (a) bias; (b) drift; (c) performance degradation (or loss of accuracy); (d) sensor freezing; and (e) calibration error.

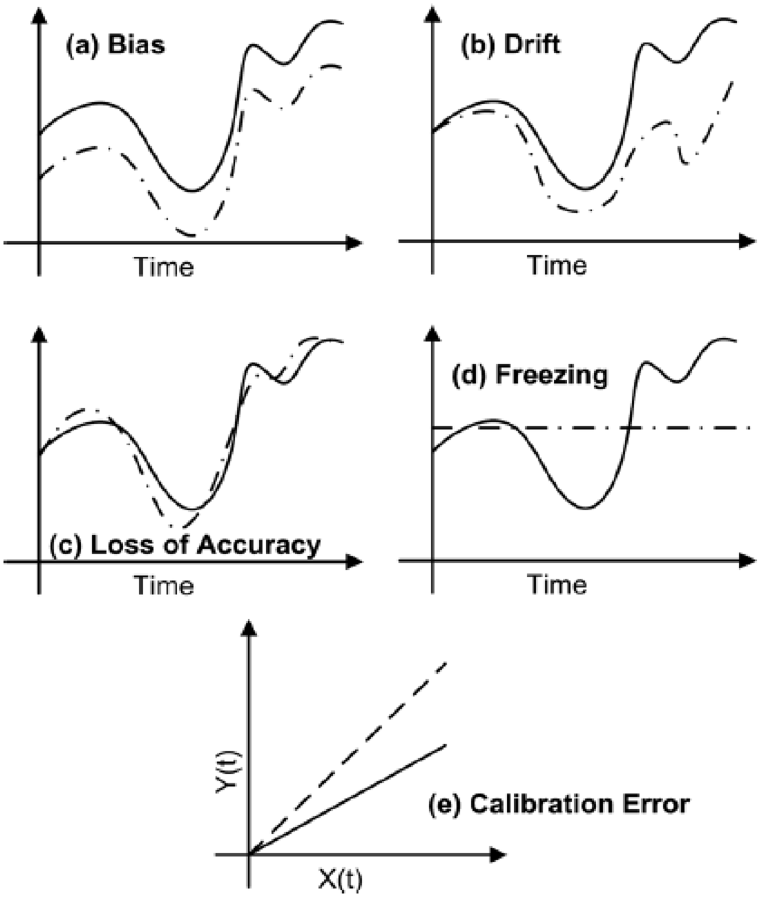


Figure 2: Depicts the effect of the above faults on system measurements

1.3.1.3 Actuator faults

An actuator fault is a form of failure influence the system inputs. because to unusual operation or material obsolescence, System actuator faults may occur. Actuator failures may alteration the system action, Which leads to deterioration or instability [4].

Common types of faults have been identified for specific types of actuators. e.g. common set of actuator faults especially in servomotors include lock-in-place (LIP) or freezing; float; hard-over-failure (HOF) and loss of effectiveness (LOE). Figure 1.3 illustrates the effect of these faults on the operating signal [3].

In the case of freezing faults, the actuator freezes at a Specific situation and does not respond to next commands. Float fault occurs when the actuator floats with zero moment and does not participate to the control authority. HOF is characterized by the actuator moving to lower or top position limit Whatever the command. LOE is characterized by reducing the actuator gain regarding its nominal value.

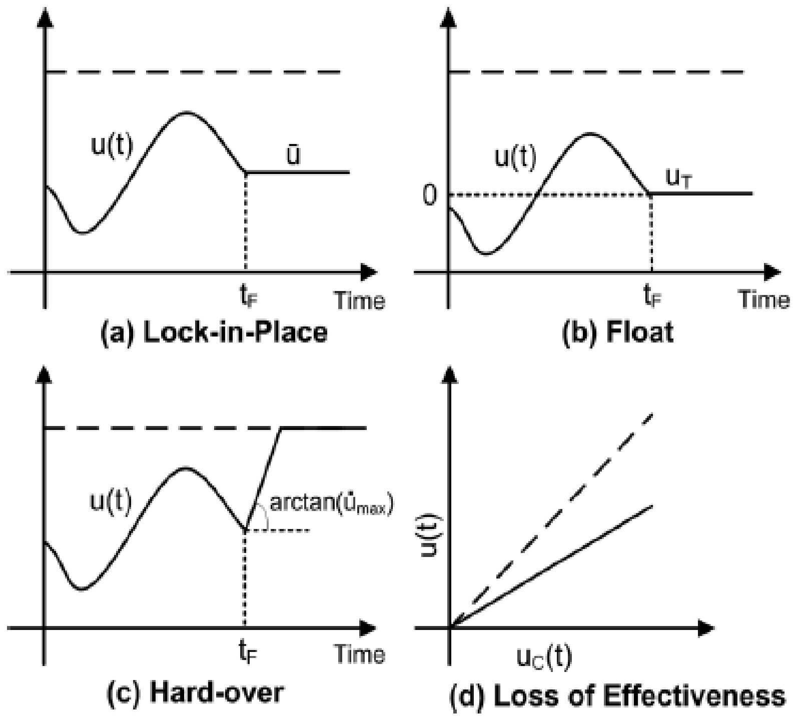


Figure 3: Common types of actuator faults

1.3.2 Depending on the faulty form

1.3.2.1 Incipient or evolutionary fault:

Affects slowly and gradually (like drift).

1.3.2.2 Abrupt fault:

It happens suddenly (like step).

1.3.2.3 Intermittent fault:

With interruptions (like pulses) [2].

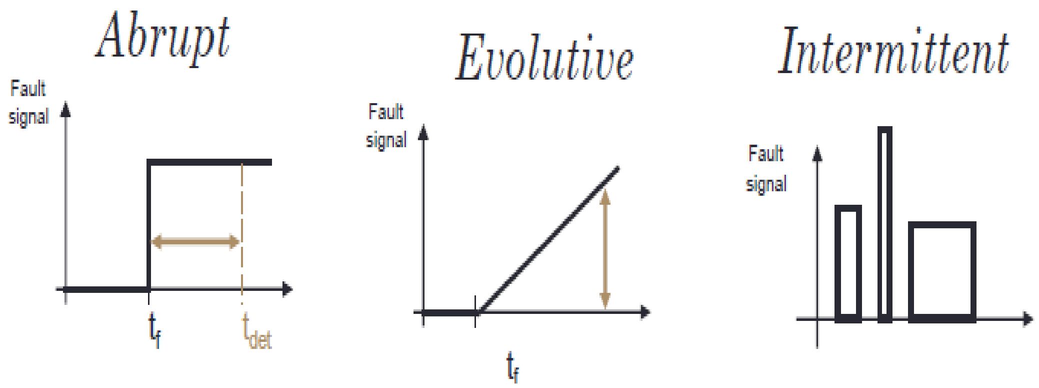


Figure 4: Fault models based on faulty form

1.3.3 Based on the effect of fault on system behavior

1.3.3.1 Multiplicative fault

Changes at output rely of the magnitude of the fault and of inputs. Changes of the parameters of the process (such as changes in resistances).

1.3.3.2 Additive fault

Changes at output only rely of size of the fault, this happens because offsets in actuators and sensors.

1.4 Fault tolerant control (FTC)

It is a control system with fault-tolerant capability. Is intended to continue the system operation as long as possible in the event of one or Many faults, provided both efficiency and security remain acceptable, and to give operators or automatic monitoring systems (Systems of fault detection FD) enough time to repair the damage or take alternative measures to avoid catastrophe [6].

FTC technologies are divided into two parts: Active and Passive: Active FTC uses detection techniques to find the faults (Monitor system performance and detect faults) and readjust the control structure to compensate for the effect of faults in the system, and passive is responsible for minimizing the faults effects of system or stabilize the system in case of a fault.

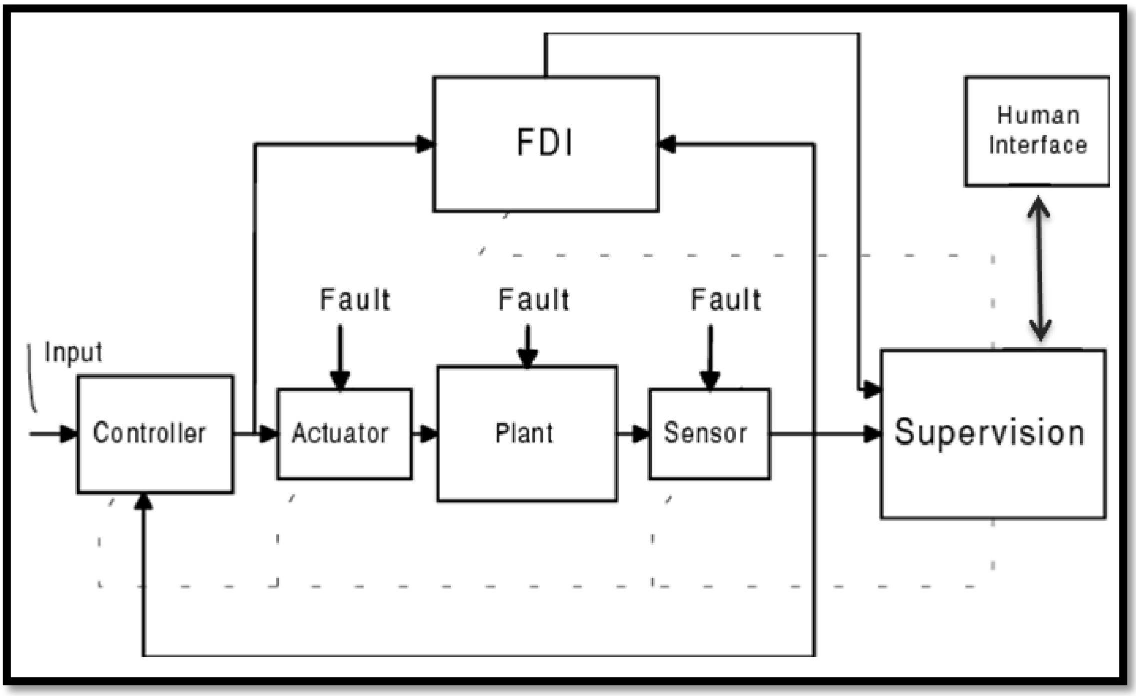


Figure 5: Scheme of fault-tolerant control system

The figur.5.represents Scheme of fault-tolerant control system with supervision subsystem. It contains 4 main components: the plant itself (including actuators and sensors), the fault detection and isolation (FDI) unit, the feedback (or feed-forward) controller, and the supervision system.

The solid line is signal flow, and the dashed line is adaption (restructure, tuning, reconfiguration or scheduling).

A component of the system is faulty (actuator, sensor or other component), the FDI unit provides the supervision system with information about the beginning of any fault based on the system inputs and outputs (Just as solid lines appear).

The supervision system will reconfigure components of the system to isolate the faults and adapt or tune the controller to accommodate the fault effects (Just as dashed lines appear).

In order to fault-tolerant control, the location and the magnitude of the fault have to be to know (find it). There are several names are used to distinguish the diagnostic steps according to the depth of study: [27]

• **Fault detection:** In this step, an error is detected or not, later we will talk about it in detail.

• **Fault isolation:** In this step, we determine the location of the fault (in which component the fault occurred).

• **Fault identification and fault estimation:** In this step, identify the fault and estimate its greatness (the severity and type of fault).

1.5 Faults Detection

The detection of fault is a delicate work because it is subject to unexpected conditions of the system and environmental disturbances, it is necessary to determine whether there is a defect or not.

Therefore, the problem of detecting faults is to measure data during operation of the system and generate residues to determine whether the system's operation is normal or failed after comparing with the main system (System normal measurements)**.**

Fault detection is divided or consists into detection of faults in actuators, sensors, and plant(Components faults) by using dependencies between different measurable signals [1].

The quality of the fault detection system is tested by detection time (detection late or no), false alarms (and the time between them) and missed detection [27].

1.6 Importance, benefits, and advantages of fault detection in systems

Fault detection and diagnostics and machine monitoring, are very difficult and important topics. Through proper machine monitoring and fault detection plans, the safety and reliability of a system can be achieved [33].

Defective systems or damaged devices can cause economic, or even human, losses in the event of fatal accidents.

For example, a report contained in Reference (J. Zarei and J. Poshtan, 2011) says that 20$ billion is the losses incurred by the petrochemical industries in the United States alone within one year, and the larger the petrochemical plants, the greater the costs of loss and maintenance, along with the economic loss [34].

Early and initial faults detection is important in the saving costs for maintaining systems and machines, If the malfunction is discovered before its occurrence, in whole or in part and as soon as possible [33].

Then, with early warning, the system can be maintained and kept more reliable (repair or replacement the system to most convenient time and as soon as possible). With minimal loss of time and low productivity today [34].

1.7 The problems of fault detection

Despite the efforts of researchers and engineers (such as Polycarpou and Trunov, Zhang, Polycarpou and Vemuri others ...) in many fields (such as artificial intelligence, mathematics, statistics, and others...) in the problem of fault detection and with the presence of disturbances, measurement noise and modeling errors,the problems of detecting faults such as false alarm, Missing detection and detection time are still existing and largely difficult to solve[8].

Detection time: Delay in detecting a malfunction is one of the important problems in detecting faults because this is a delay that may result in more difficult problems.

False alarm: Is the system detecting a malfunction and issuing an alarm, despite its absence , low false alarm rate is essential for fault detection systems to be reliable.

Missed detection: It is a fault that is not discovered despite its presence, and this matter is very dangerous to the integrity of the system and its components [2].

1.8 Fault Detection Methods

There exist several taxonomies of the field. like as control engineering approach, statistical, mathematical, AI approach [1]. Currently, it is possible to identify 3type of detection methods:

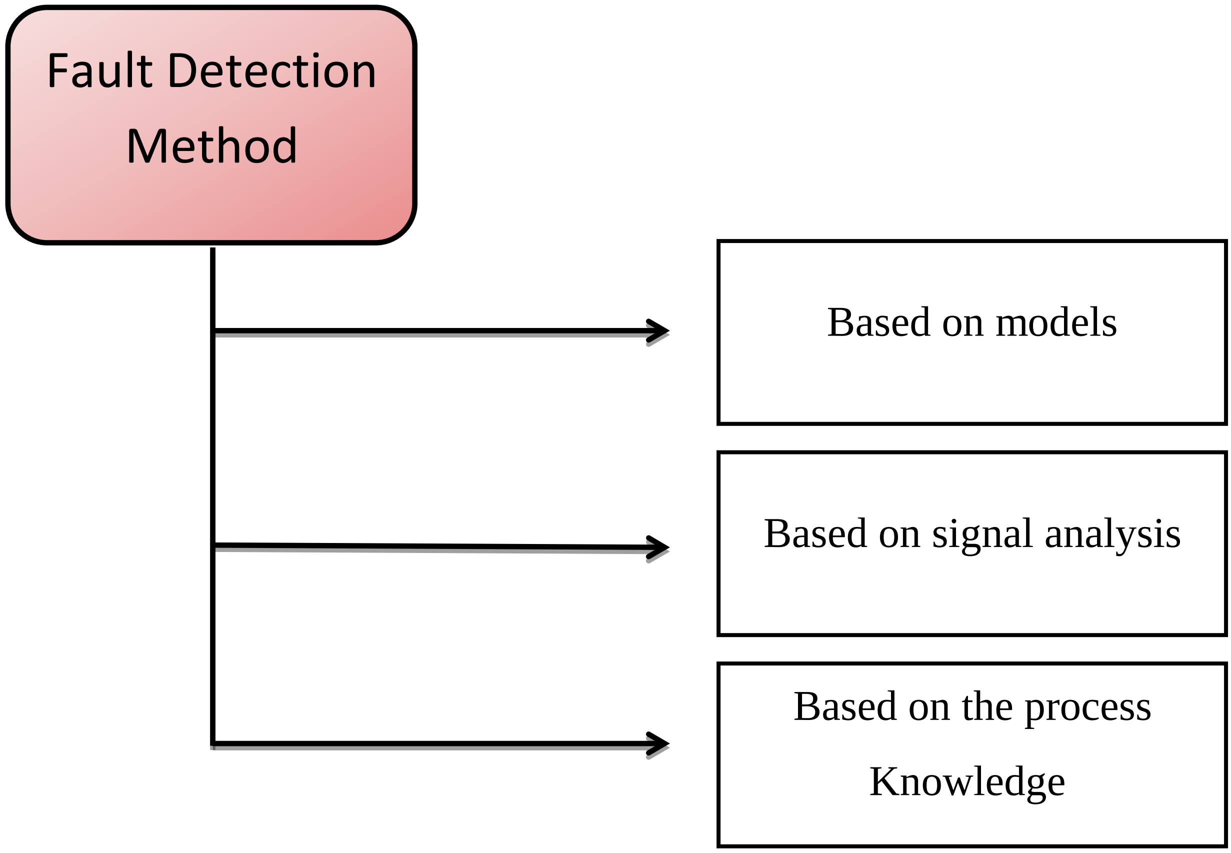


Figure 6: Fault detection method

1.8.1 Based on models

Model-based fault detection based on the experimental experience on a lot of equipment and systems such as motors, hydraulic servo steering, pipelines, machine tools (cutting, grinding, and drilling), heat exchangers, feeding motors, robots, circulation pumps, vehicle suspension and electric and pneumatic motors. By reviewing the effects of faults on the mathematical process.

We measure process input (actuator input) and process output (sensor signal), and sometimes some parameters in between.

Often, system parameters or variables are affected by faults such as mechanical stiffness or friction parameters, inductive or electrical resistance, heat exchange coefficients and fluid resistance parameters. Then they appear as additive or multiplicative faults [30].

1.8.2 Based on the signal analysis

The signal analysis technique can be applied in fault detection by comparing the mathematical models of the measured signal with measurements of normal behavior by training artificial intelligence techniques like as ANNs [1].

An example of fault detection is by analyzing the time-field reflection signal, where a signal is sent through an electric line and comparing the reflected signal mathematically with the original signal in order to identify faults.

For example, by means of wavelet analysis to reveal the abnormal behavior of the signal, which has been studied in the following literature ZoltánGermán-Salló, Gabriela Strnada, 2018 [29].

1.8.3 Based on the process knowledge

Knowledge-based fault detection techniques can use of expert resolution in the field in making judgments, as researchers model the relationship between the phenomenon of fault and cause [32].

Among the methods for fault detection are knowledge-based methods such as causal analysis, fuzzy logic, expert systems, and pattern recognition.

These techniques are based on qualitative modeling, obtained by means of fault-symptom examples, causal modeling of the system, expert knowledge or detailed description of the system.

Causal analysis techniques are primarily used for fault detection, while expert systems are used to imitate human experts' logic when detecting faults and which have a wide range of applications for diagnostic tasks [31].

1.9 Conclusion

In this chapter, we provided a background to the topics related to our research. The chapter consists of an overview of the faults and types related to it, then we identified the fault detection and problems and ways to treat it and the most important applications to use.

Chapter2: Machine learning and fault detection in photovoltaic converter

2.1 Introduction

We have seen machine learning as a buzzword over the past few years, and this may be due to the large amount of data production by applications, the increase in computational power in the past few years, and the development of better algorithms.

In this chapter, we will provide a brief description of the PV converter and an overview of machine learning and its different types. Famous algorithms and applications. We will also talk about fault detection using machine learning.

2.2 Photovoltaic converter:

It is a linear internal current source of limited power, which has the characteristics of constant current depending on the operating point. One or more converter that convert the output direct current (constant) into alternating current at a given volt [40].

2.3 Previous types of faults in photovoltaic cells

- **a.** Inverse bypass diode.
- **b.** Bypass diode breakdown.
- **c.** Voltage sag.
- **d.** Open-circuited converter IGBT(Insulated Gate Bipolar Transistor).
- **e.** Short-circuited bypass diode.

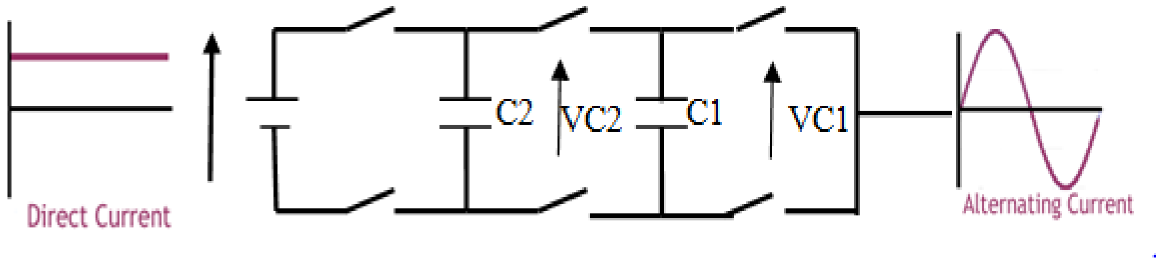


Figure 7: PV Converter Scheme

2.4 Machine learning

Machine learning is an application of artificial intelligence (AI) .Machine learning focuses on developing computer software and adopting machine learning algorithms as a mathematical model based on a data sample[10]. known as "training data", in order to make predictions or decisions without being explicitly programmed.

2.5 Type of Machine learning

Machine learning can be classified into 3 types of algorithms:

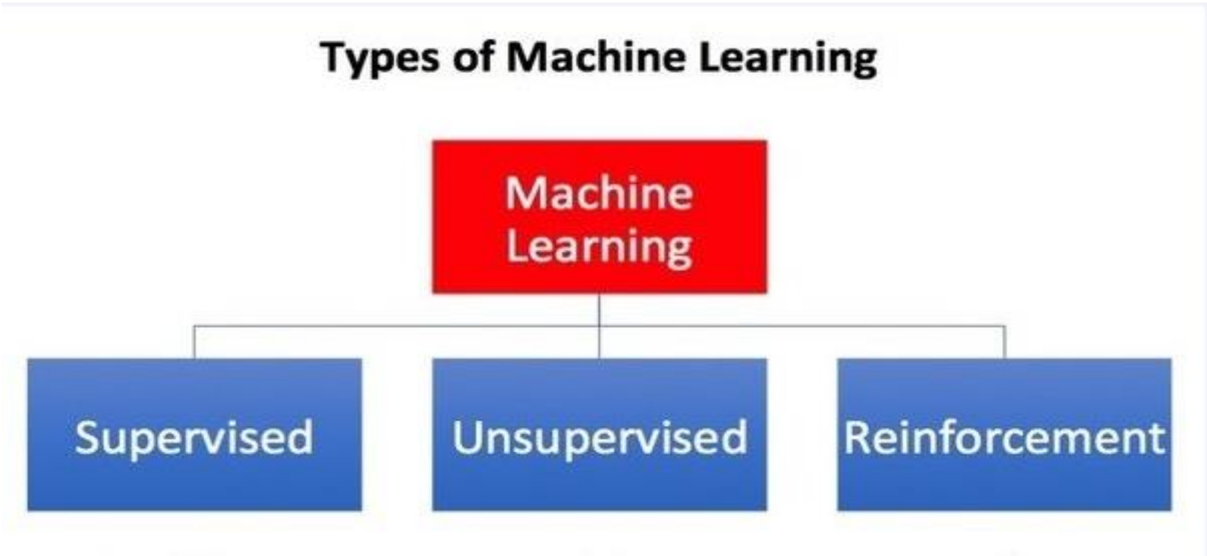


Figure 8: Types of Machine Learning

2.5.1 Supervised learning

Supervised learning relies on classified examples, a data set classified to conclude a learning algorithm and uses data as a basis for accurately predicting or predicting results through machine learning algorithms [12].

2.5.1.1 Type of Supervised learning

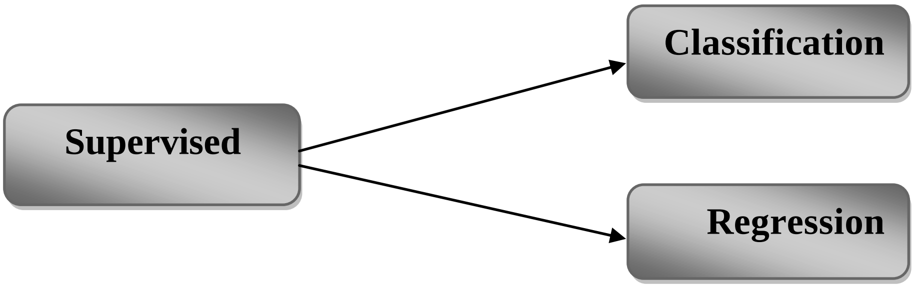


Figure 9: Supervised learning

- Classification:

Is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data [13]. The process starts with predicting the class of given data points. The classes are often referred to as target, categories or label.

- Regression:

Regression is used to predict continuous values and predictions can be made by learning the relationship between data properties and results that have been combined in uninterrupted value and using regression in finance, investment and other disciplines [14].

2.5.1.2 Supervised learning algorithms:

We'll look at the methods and algorithms most commonly used in machine learning processes.

A. Neural networks:

Neural networks process training data by interconnecting the human brain through layers of nodes, each node consists of input and output. Neural networks learn this mapping function through supervised learning, adjusting based on the loss function through the process of gradient descent. When the cost function is at or near zero, we can be confident in the model's accuracy to yield the correct answer .Neural networks are applied to many real-life problems today, including speech and image recognition, finance, and medical diagnosis [15].

B. Naive Bayes:

It is a classification algorithm based on Bayes theory and uses the theory to classify objects. Naive Bayes classifiers assume strong independence between data points. Common uses of naive Bayes classifiers include spam filters and text analysis, and it is multiclass algorithm [12].

C. Linear regression:

The regression algorithm creates a regression model between variables and obtains the relationship between dependent variables and variables through the learning process (training), regression analysis can be used for prediction or classification models, Linear regression analysis is the most widely used of all statistical techniques [16].

D. Logistic regression:

Logistic regression is a model used to predict the likelihood of an event by matching data on a logistics curve .It is a classification algorithm in machine learning that uses one or more independent variables to determine an outcome. The outcome is measured with a dichotomous variable meaning it will have only two possible outcomes. Logistic regression is mainly used to solve binary classification problems (medicine and social sciences.....)[17].

E. Support vector machine (SVM):

A support vector machine is a model developed by Vladimir Vapnik, the algorithm is often used in classification and regression, and the goal of the algorithm is to find the best way to divide data into two groups by placing a super-level between them hyperplane [12].

F. K-nearest neighbor:

It is an algorithm that can be used for both classification and regression problems. is a simple algorithm that stores all the available cases and classifies the new data or case based on a similarity measure. It is mostly used to classifies a data point based on how its neighbors are classified, and it is multiclass algorithm [17].

G. Random forest:

The Algorithm are often used in classification and regression, a data building that is applied to machine learning that develops large numbers of random trees to analyze combinations of variables. which are then merged together to reduce variance and create more accurate data predictions [18]. This type of algorithm helps improve the ways in which techniques analyze complex data.

2.5.2 Unsupervised learning

Unsupervised learning is a type of machine learning that allows machine training to use unclassified or disaggregated information and to allow the algorithm to act. The organ's mission here is to collect information that is not classified according to similarities, patterns and differences without any prior training of data. Also known as self-organization [19].

2.5.2.1 Type of unsupervised learning:

- Clustering:

Clustering is a data mining technique which groups unlabeled data based on their similarities or differences [19]. Clustering algorithms are used to process raw, unclassified data objects into groups represented by structures or patterns in the information. Clustering algorithms can be categorized into a few types, specifically exclusive, overlapping, hierarchical, and probabilistic.

2.5.3 Reinforcement learning:

Reinforcement Learning is defined as a Machine Learning method that is concerned with how software agents should take actions in an environment. Reinforcement Learning is a part of the deep learning method that helps you to maximize some portion of the cumulative reward. This neural network learning method helps you to learn how to attain a complex objective or maximize a specific dimension over many steps [20]. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

2.6 Machine learning applications

Machine learning apps are everywhere, and many people use them every day. We cite a few examples of these applications.

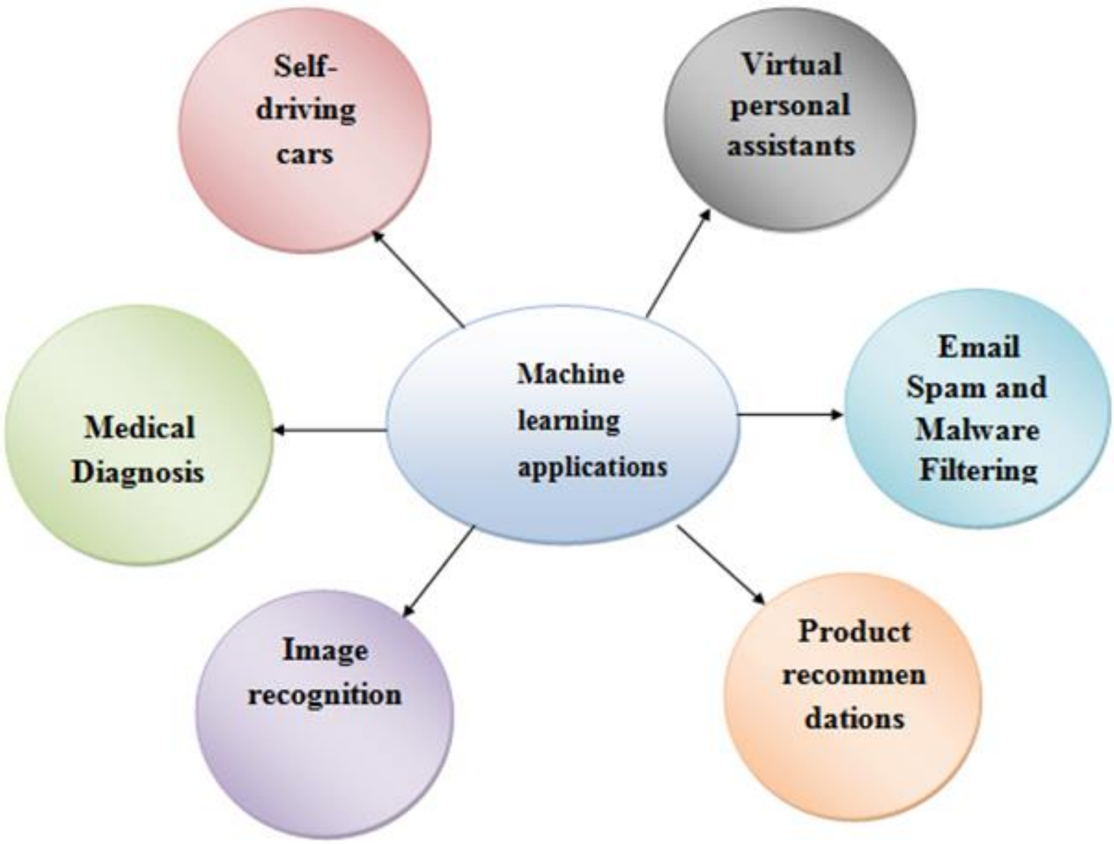


Figure 10: Machine Learning applications

2.6.1 Virtual personal assistants

We have various virtual personal assistants such as Google assistant, Alexa, Cortana, Siri, they help find information when questioned aloud. Just activate them and ask a question. To respond, the personal assistant searches for information, recalls related requests, or sends a command to other resources (such as phone applications) to collect information. Virtual assistants even receive instructions to perform certain tasks including setting the alarm for the next day, reminding appointments and deadlines [21].

Machine learning is an important part of these personal assistants as they collect information from daily interactions with the user.

2.6.2 Email Spam and Malware Filtering

To ensure that these spam filters are constantly updated, they are powered by machine learning. When spam filtering is done from a rule base, it fails to track down the latest techniques adopted by spammers. Some machine learning algorithms such as Multi-Layer Perception, Decision tree, and Naïve Bayes classifier are used for email spam filtering and malware detection.

2.6.3 Product recommendations

Current approaches supporting consumers in their buying decision are, amongst others, provided through web-based product recommendation systems [22]. Machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest. As similar, when we use Netflix, we find some recommendations for entertainment series, movies, and this is also done with the help of machine learning.

2.6.4 Self-driving cars

One of the most exciting applications of machine learning is self-driving cars. The Google self-driving car without steering wheel or pedals is tested on public roads in early 2015 [23].Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

2.6.5 Medical Diagnosis

Machine learning is used for diseases diagnoses. there has been a dramatic increase in the use of computation-intensive methods to analyze biomedical signals ,in which a computer program “learns” important features of a dataset to enable the user to make predictions about other data that were not part of the original training set. It helps in finding brain tumors and other brain-related diseases easily [24].

2.6.6 Image recognition

Image recognition, in the context of machine vision, is the ability of software to identify objects, places, persons, writing and actions in images.

Face book provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Face book friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's face detection and recognition algorithm [25].

2.7 Terminology machine learning

2.7.1 Definition of under-fitting

We say a machine learning algorithm is under-fitting if it does not fit the data well enough. This usually happens when we have less data to build an accurate model, or when we try to build a linear classification model with a non-linear data.

2.7.2 Definition of over-fitting

We say a machine learning algorithm is over-fitting if when the model is trained with so much of data, the algorithm learns from inaccurate data and noise, and then does not classify the model correctly.

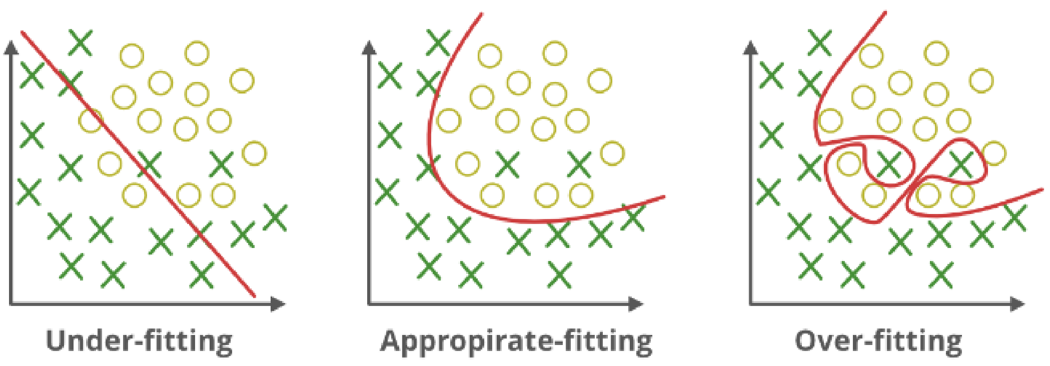


Figure 11: Under-fitting and over-fitting

2.7 Fault detection based on machine learning

The study of systems in various fields such as industry, economics and others... produces a huge amount of data, which is difficult to analyze individually. In this context, the use of machine learning (ML) techniques, which is one of the important branches of artificial intelligence science, appears to deal with the fault detection challenges in these systems [41].

Over the past 30 years, and with the increase in studies and literature in the field of fault detection and diagnosis, many artificial intelligence (AI) techniques have been developed in fault detection systems, including Artificial Neural Networks (ANNs), Naive Bayes, Fuzzy Logic (FL) and Support Vector Machines (SVM) [11].

The practical applications and application fields of artificial neural networks (ANNs) are increasing with very high rates, and one of these areas which has achieved promising results is fault detection [42], because it is a programming technique capable of solving nonlinear problems easily [43]. However, they can be considered as black boxes because they offer little explanation and detail in terms of detecting faults [11].

On the contrary, it is possible to perform fault detection in fuzzy logic (FL) systems, and to interpret and analyze their results with a good basis. However, fuzzy logic presents some difficulties with rule definitions and input processing [11].

The theoretical foundation of Naive Bayes classifier is Bayesian decision theory, and Naive Bayes (NB) have gained more applications in the field of classification and fault detection, and it has given good results.

Recently, support vector machines (SVM) have gained more applications in the field of classification and fault detection due to their high success rate, their ability to better generalize and their ability to classify linear and nonlinear, but with this the SVM is less efficient in the case of the presence of many data and chapters [41].

2.8 Application of fault detection using machine learning

In the early 1990s, most of the research and development in fault detection was on nuclear power plants, processing plants, the automobile industry ,aircraft and national defense. Today bug detection is established in many industries.

2.8.1 Machines and Engines

Signal analysis technique for internal combustion engine using ANN is given and engine cylinder fault detection.

2.8.2 Electrical Motors

Fault detection of induction motors and electrical motors fault detection.

2.8.3 Pumps

Fault detection for centrifugal pumps combining neural networks and neuro-fuzzy.

2.8.4 Manufacturing

Fault detection in automated electronics manufacturing systems, fault detection is important to achieving maximum productivity.

2.8.5 Bearings and Machinery

Detection of faults in rotating and hydraulic mechanical systems using neural networks.

2.8.6 Aircraft

Detection of faults in a field of flight critical Sensitive engine control systems.

2.8.7 Automotive Systems

Fault detection is existing engine electronic control unit of engine or gearbox.

2.8.8 Chemical Processes

Due to deterioration of the equipment and components of the chemical process, the performance of the process deteriorates. Such as, fractionation of fatty refinery or acids, sewage plant.

2.8.9 Steam turbines

In scientific research (Gang Zhao, Dongxiang Jiang, JinghuiDiao, LijunQian, 2004) a study has been conducted Fault detection for steam turbine is described, and also wind turbines have been studied in many researches [35].

2.8.10 Railway

Railway faults have been studied in many literature and this is due to their importance, because a minor fault in the railways may lead to catastrophic consequences, like studying of (MoussaHamadache, SaikatDutta, Osama Olaby, RamakrishnanAmbur, Edward Stewart and Roger Dixon, 2019) [26].

2.9 Conclusion

In this chapter, we provide a background on machine learning and its fields. The chapter consists of fault detection by machine learning in a photovoltaic converter.

This study led us to apply some useful machine learning algorithms for our guidance study in the next chapter.

Chapter3: Implementation and results

3.1 Introduction

In this chapter, we will provide a general description of the tools used in our work, we interpret and design PV system data we worked on. In addition to implementing the KNN, SVM, and NB algorithms, and comparing their results.

The experiment aims to evaluate the performance and result of each algorithm and to select the best classification technique for this type of problem.

3.2Work environment

3.2.1 Hardware

In order to carry out this project, we used a set of materials the main characteristics of which are as follows:

Manufacturer: Toshiba

Rating: 4.7 Windows performance index

Processor: Intel (R) Core— i3-3217U @ 1.80 GHZ.

RAM: 4.00 GB

System type: 64 bite operating system.

3.2.2 Software

3.2.2.1MATLAB programming language

MATLAB (Short form of "Matrix Laboratory") is a computer program or programming language that provides the user with a convenient environment for performing many types of calculations and technical computing. It integrates visualization, computation and programming in easy to use environment [36].

It is used to solve many technical problems which require matrix and vector formulations, besides it is usually used to solve differential equations and it is an effective way and can be considered as easy and quick. Typical uses of MATLAB are as follows:

• Computation and Math.

• Data analysis, visualization and exploration.

• Algorithm modeling and simulation.

• Scientific computations.

• Image processing applications.

• Application development..

- **MATLAB Window**

When MATLAB is run, below window pops up. MATLAB environment is divided into 4 windows:

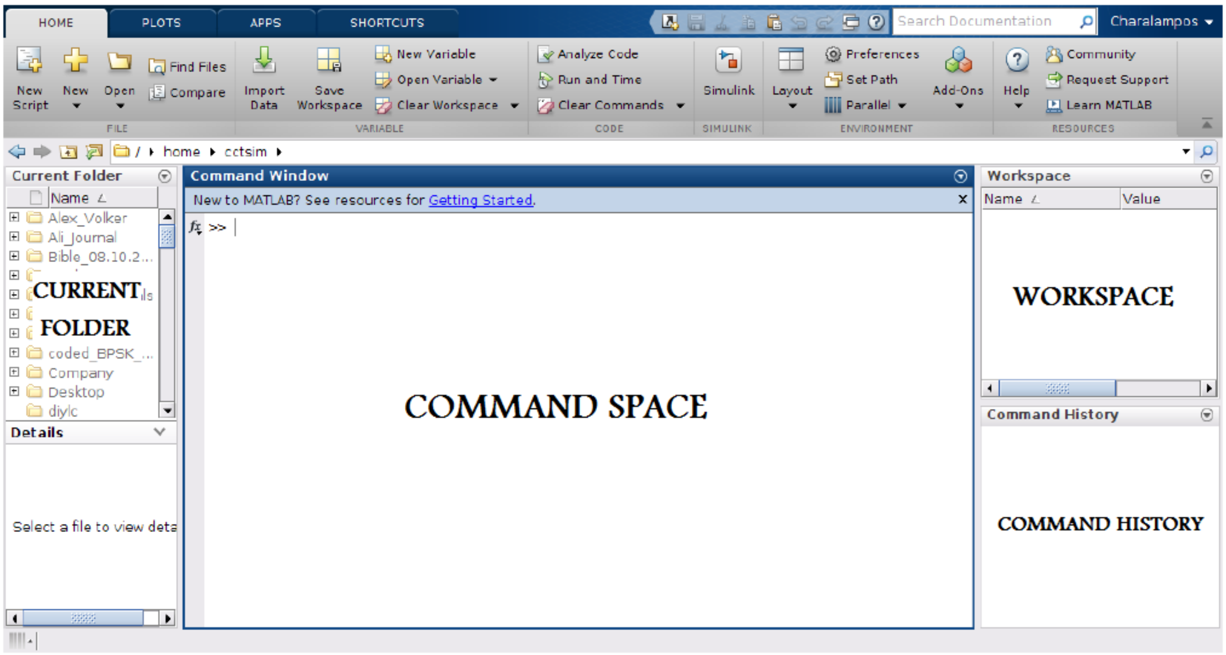


Figure 12: MATLAB Window Components

Command window: This is the main window. It contains command prompt (>>). The place where user type all the command.

Workspace: It lists all the variables program or user has generated in the current session, in addition totype and size of the variable.

Command history: It displays list of previously typed commands.

Current directory: It shows files and folders in the current directory [37].

Table 1: The Advantage and Disadvantage of MATLAB Language

Advantage	Disadvantage
- It is easy to use, e.g. it is not required to define variable types (unless needed). - Variables (results of simulations) are stored in the workspace for inspection and debugging, at any time. - MATLAB is platform independent and hence it can be installed on different Operating Systems like Vista, Windows, Macintosh and Linux. - It is more convenient and faster because it has huge built-in library of functions for many tasks such control systems, neural networks, signal processing, communications, image processing etc. - Excellent visualization (plots) capabilities.	- It is very expensive for non-students (It is not free and hence users need to obtain licensed version), but not 100%. - Code execution can be slow if programmed carelessly. - It requires fast computer with sufficient amount of memory [38].

3.2.2.2 Matlab with machine learning applications

We know that in order to learn machine learning we have many programming languages, and Matlab is one of these languages, when it comes to machine learning, Matlab proves to be very helpful:

- It's a perfect platform for analysis and data visualization with different tools and functions.
- Automatically generate code for the analysis of data from sensors.
- Matlab helps in areas such as image processing, computer vision, signal processing etc.
- With MATLAB we can have automatic access to functions and specialized applications for regression, classification and clustering [39].

3.3 System offered and Dataset

We suggest a way to detect detection method for Capacitors faults located in PV energy converter system exist in solar energy systems (in our case) based on SVM, KNN and NB algorithms. These faults are sure to pose a threat to the stable operation of the photovoltaic (PV) system.

The studied data made up of 4 classes:

H = Healthy data.

FC1 = Data captured in the event of a fault at capacitor1 level.

FC2 = Data captured in the event of a fault at capacitor2 level.

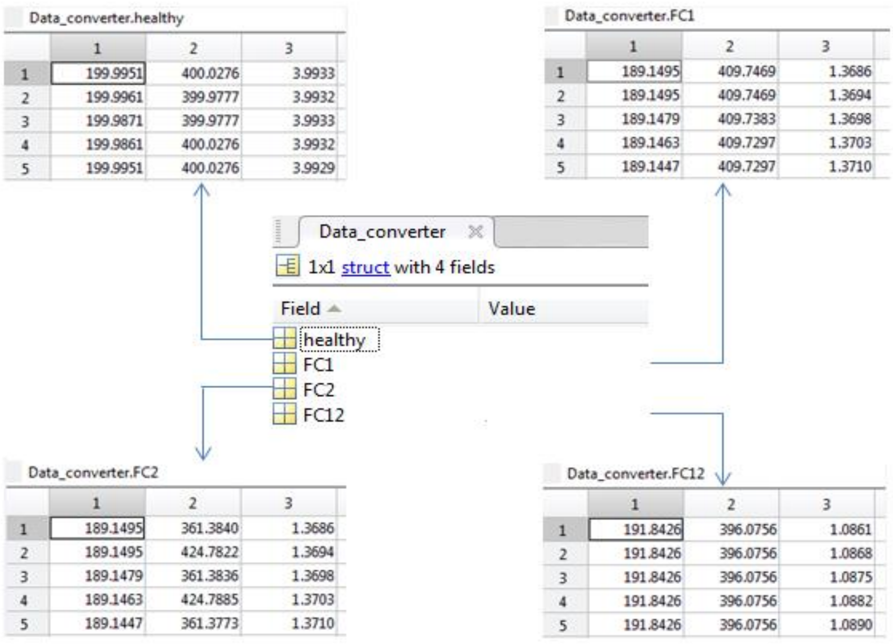
FC12 = Data captured in the event of a combined fault at capacitor 1 and2.

- These are the classes that must be separated to detect faults and their location.
- Each of these classes contains 3 variables:

To construct our feature space, we have used three features extracted from system parameters:

- Capacitor 1 Voltage,
- Capacitor 2 Voltage,
- Load current.

- Each feature is a time series of 50,000 values out of 600,000 real data values (because the data is large and repetitive, only a portion of it was used). These values are delivered by high fidelity PV simulator.



Data_converter.healthy

	1	2	3
1	199.9951	400.0276	3.9933
2	199.9961	399.9777	3.9932
3	199.9871	399.9777	3.9933
4	199.9861	400.0276	3.9932
5	199.9951	400.0276	3.9929

Data_converter.FC1

	1	2	3
1	189.1495	409.7469	1.3686
2	189.1495	409.7469	1.3694
3	189.1479	409.7383	1.3698
4	189.1463	409.7297	1.3703
5	189.1447	409.7297	1.3710

Data_converter.FC2

	1	2	3
1	189.1495	361.3840	1.3686
2	189.1495	424.7822	1.3694
3	189.1479	361.3836	1.3698
4	189.1463	424.7885	1.3703
5	189.1447	361.3773	1.3710

Data_converter.FC12

	1	2	3
1	191.8426	396.0756	1.0861
2	191.8426	396.0756	1.0868
3	191.8426	396.0756	1.0875
4	191.8426	396.0756	1.0882
5	191.8426	396.0756	1.0890

Figure 13: Part of dataset

3.4 Data analysis

The data were analyzed, in order to show the normal and abnormal condition in case of failure in capacitor 1, failure in capacitor 2 and failure in both capacitors 1 and 2 at the same time.

In all simulation scenarios, the fault was injected at sample time 200000 (see Fig. 14 to Fig. 17).

Failure C1:

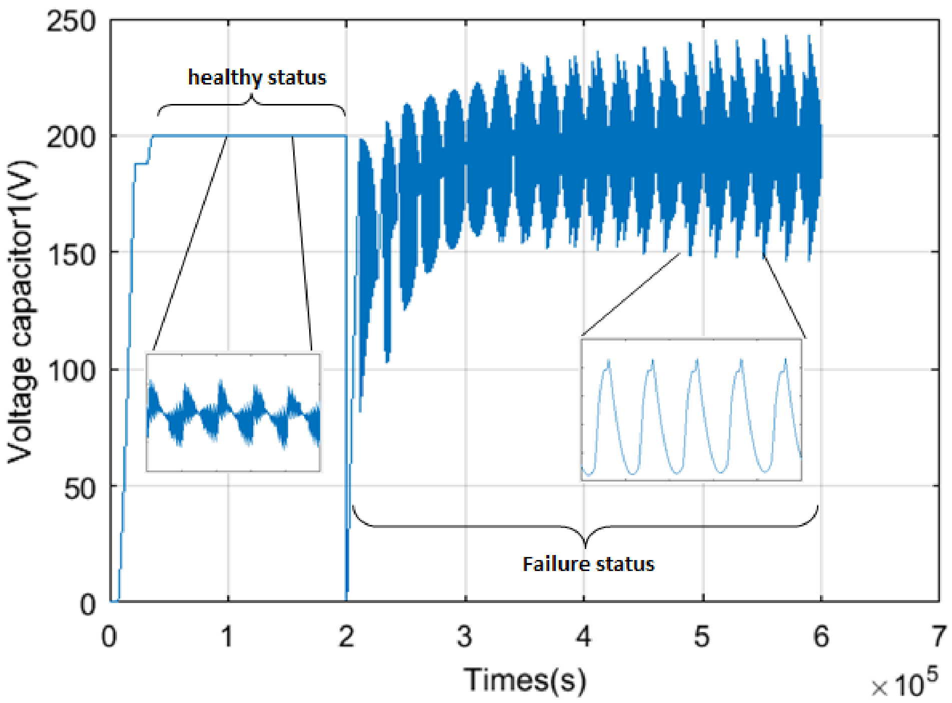


Figure 14: Voltage capacitor 1 in case of capacitor 1 fault.

Failure C2:

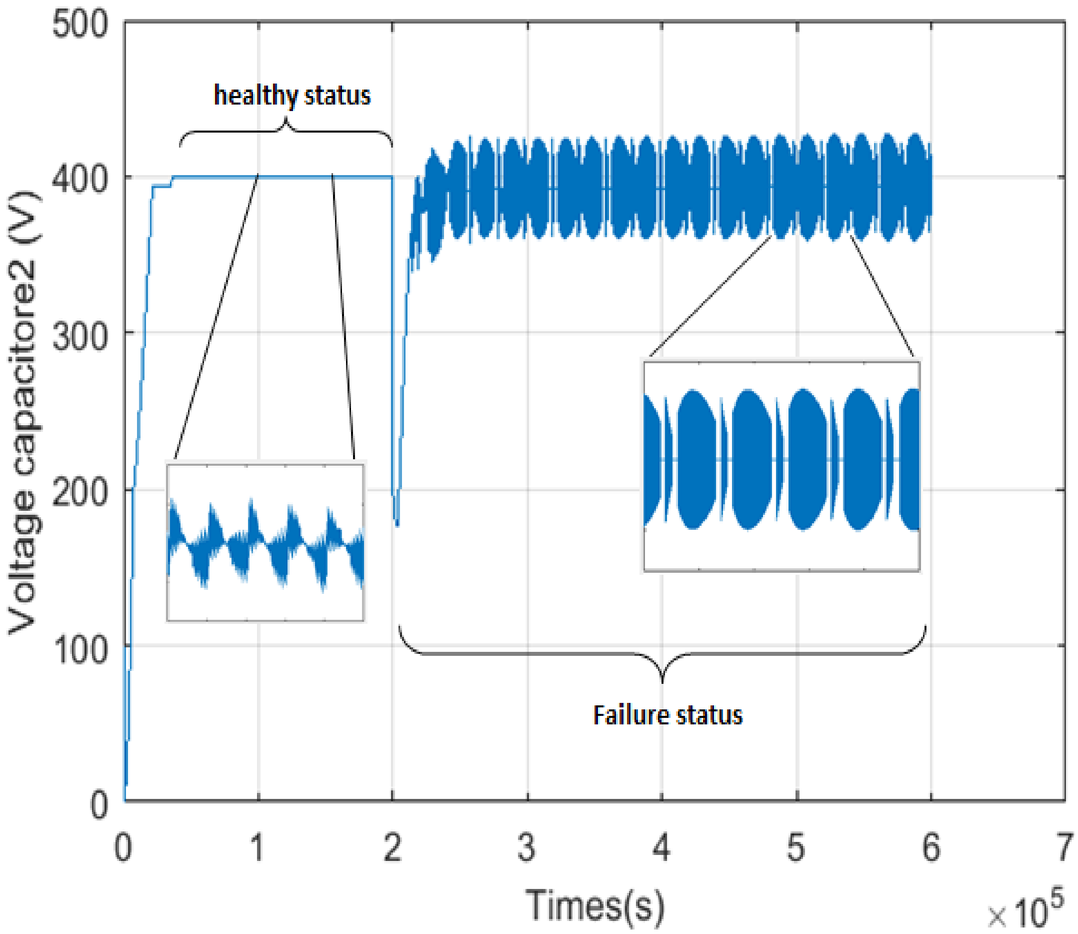


Figure 15: Voltage capacitor 2 in case of capacitor 2 fault.

Failure C1 and C2 both:

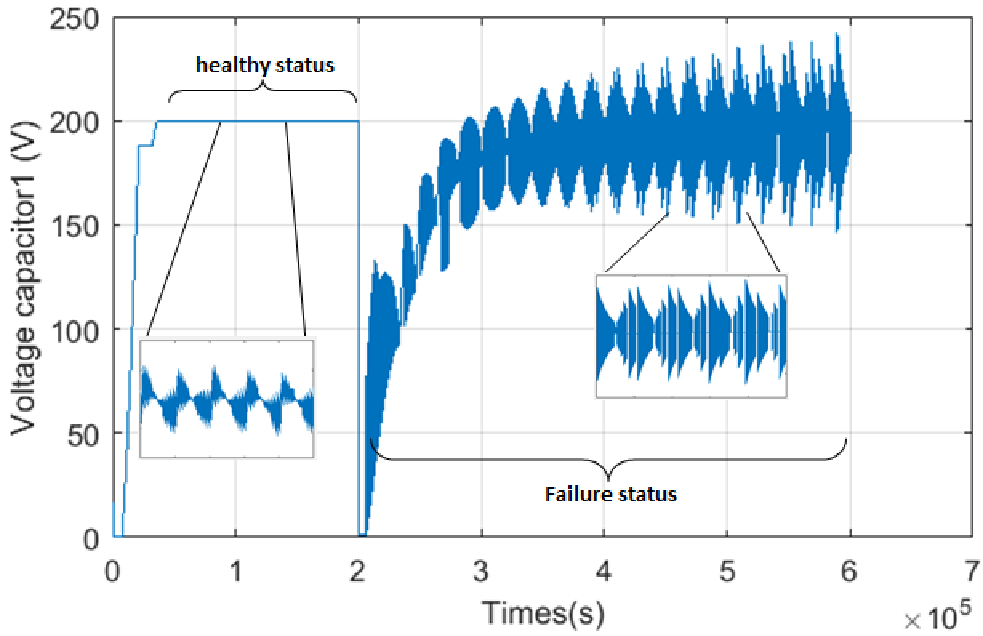


Figure 16: Voltage capacitor 1 in case of capacitor 1 and capacitor 2 fault

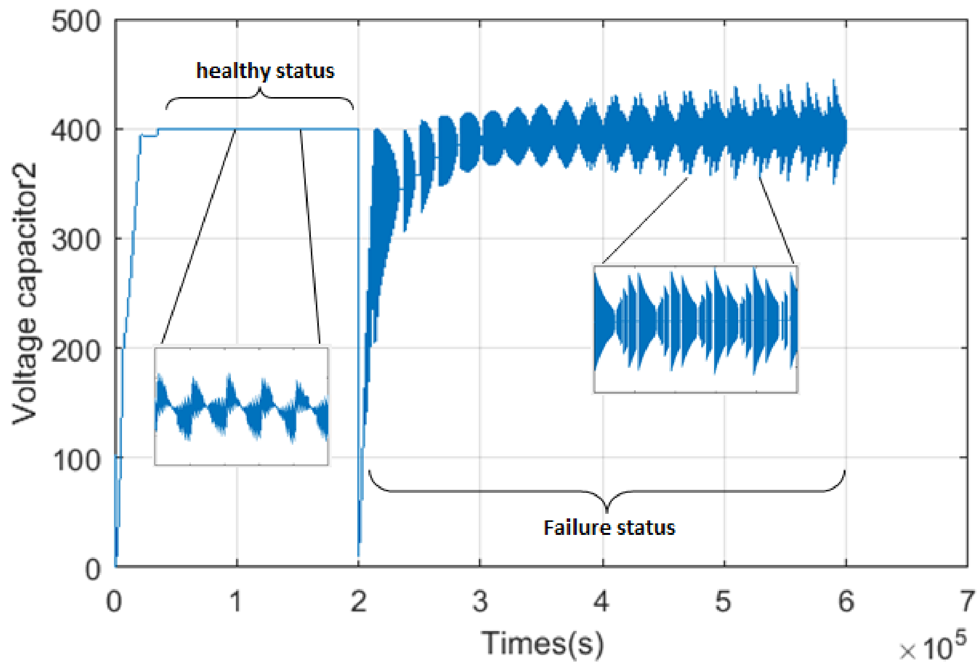


Figure 17: Voltage capacitor 2 in case of capacitor 1 and capacitor 2 fault

3.5 Data pre processing

In order to prepare data for training, we have performed preprocessing that includes:

- Extraction of physical values (voltage and current) and omitting time scale,
- Avoiding transients regimes at the start of converter and begin of faults,
- Split of data to 4 classes: healthy mode and 3 fault modes,
- Labeling the data as we will use supervised learning algorithms, we have chosen the following labels: "0" for healthy mode, "1" for C1 fault mode, "2" for C2 fault mode, and "3" for C1/C2 fault mode.

```
    % prepare data for all classes
data(i,:) = [data_in.healthy((i+3)/4,1:3),0];
data(i+1,:) = [data_in.FC1((i+3)/4,1:3),1];
data(i+2,:) = [data_in.FC2((i+3)/4,1:3),2];
data(i+3,:) = [data_in.FC12((i+3)/4,1:3),3];
```

Figure 18: Prepare data for all classes

After data preprocessing, we have obtained five datasets, one global dataset for multiclass classifiers and four elementary datasets for binary classifiers. The global dataset contains 200000 rows and 3 columns (column for each feature).In order to classify, we have randomly divided the data into two parts (train_set and test_set):

Train:

Training data represents 70% of all data.

Test:

Test data represents 30% of all data.

On case of SVM: The SVM algorithm works based on binary classification:

- In our case there are 4 classes, so we have developed a solution in order to separate 4 classes.

- Each time we took one of the four classes (for example, healthy) as one class (Label 0) and the rest of the classes as one class (Label 1) as shown in the figure 19.

- Then we applied a SVM algorithms detailed in Table. 2.

- In each test case, based on the maximum score, the One-Vs-All strategy helps decide for each test point the class that it belongs to.

```
    % prepare data for healthy class
data_h(i,:) = [data_in.healthy((i+3)/4,1:3),0];
data_h(i+1,:) = [data_in.FC1((i+3)/4,1:3),1];
data_h(i+2,:) = [data_in.FC2((i+3)/4,1:3),1];
data_h(i+3,:) = [data_in.FC12((i+3)/4,1:3),1];
    % prepare data for FC1 class
data_c1(i,:) = [data_in.healthy((i+3)/4,1:3),1];
data_c1(i+1,:) = [data_in.FC1((i+3)/4,1:3),0];
data_c1(i+2,:) = [data_in.FC2((i+3)/4,1:3),1];
data_c1(i+3,:) = [data_in.FC12((i+3)/4,1:3),1];
%      % prepare data for FC2 class
data_c2(i,:) = [data_in.healthy((i+3)/4,1:3),1];
data_c2(i+1,:) = [data_in.FC1((i+3)/4,1:3),1];
data_c2(i+2,:) = [data_in.FC2((i+3)/4,1:3),0];
data_c2(i+3,:) = [data_in.FC12((i+3)/4,1:3),1];
%      % prepare data for FC12 class
 data_c12(i,:) = [data_in.healthy((i+3)/4,1:3),1];
 data_c12(i+1,:) = [data_in.FC1((i+3)/4,1:3),1];
 data_c12(i+2,:) = [data_in.FC2((i+3)/4,1:3),1];
 data_c12(i+3,:) = [data_in.FC12((i+3)/4,1:3),0];
```

Figure 19 : Prepare data for each classes alone

- **Representation of data as feature space:**

Data representation in Matlab code:

```
scatter3(data_in.healthy(:,1),data_in.healthy(:,2),data_in.healthy(:,3));hold on;
scatter3(data_in.FC1(:,1),data_in.FC1(:,2),data_in.FC1(:,3));
scatter3(data_in.FC2(:,1),data_in.FC2(:,2),data_in.FC2(:,3));
scatter3(data_in.FC12(:,1),data_in.FC12(:,2),data_in.FC12(:,3));
```

Figure 20: Data representation in Matlab code

Result of representing data as a feature space:

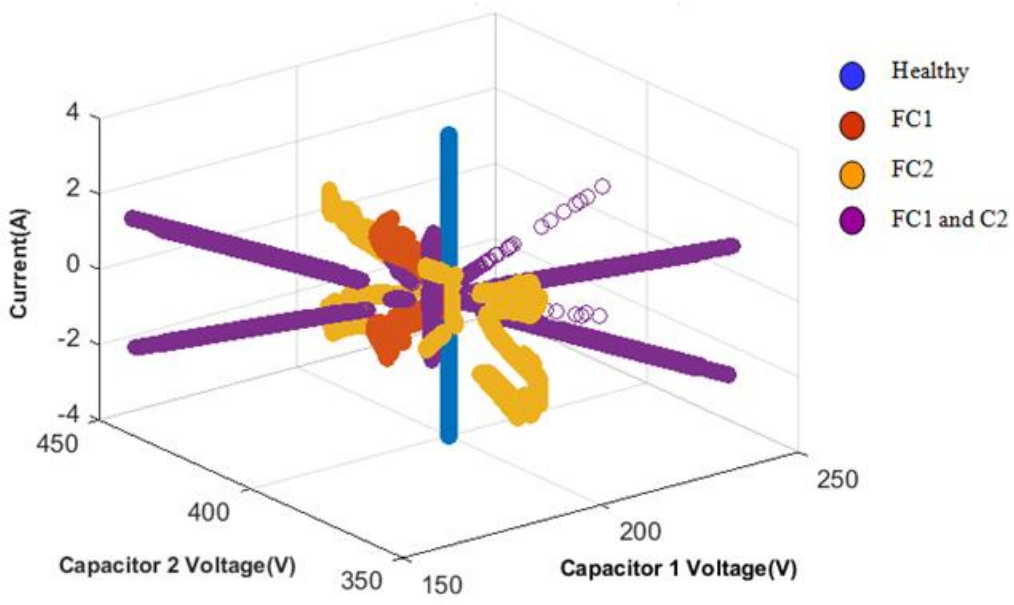


Figure 21: Representation data as a feature space

3.6 Comparison between SVM, KNN and NB Algorithm

SVM Classification:

We applied SVM technology to 2 types: polynomial, RBF

In polynomial kernel SVM we change the degree of the polynomial starting from 3 or more and note the accuracy of the classification.

At polynomial degree 5 the best classification accuracy, less than that, the accuracy was low and this indicates under-fitting, and after degree 5 it decreases again and this is evidence of the occurrence of over-fitting.

And RBF gave the best result compared to the polynomial.

KNN Classification:

We change the of the NN value which represents the number of neighbors in the KNN algorithm, show is as table 2. We got the best result (best accuracy) in the case of the number of neighbors 2.

Table 2: Comparison bettween SVM, KNN and NB on detection accuracy

Method	Parameterization		Accuracy (%)	Time
KNN	of neighbors	1	98.8183	1.64s
		2	98.82	1.84s
		3	98.81	1.84s
		4	98.8150	1.84 s
		5	98.80	1.66 s
SVM	kernel	polynomial-3	30.15	330.76 m
		polynomial-5	47.51	461.03 m
		polynomial-6	28.60	321.35 m
		RBF	91.44	10.41 m
NB			95.51	1.80 s

3.7 Representation of results

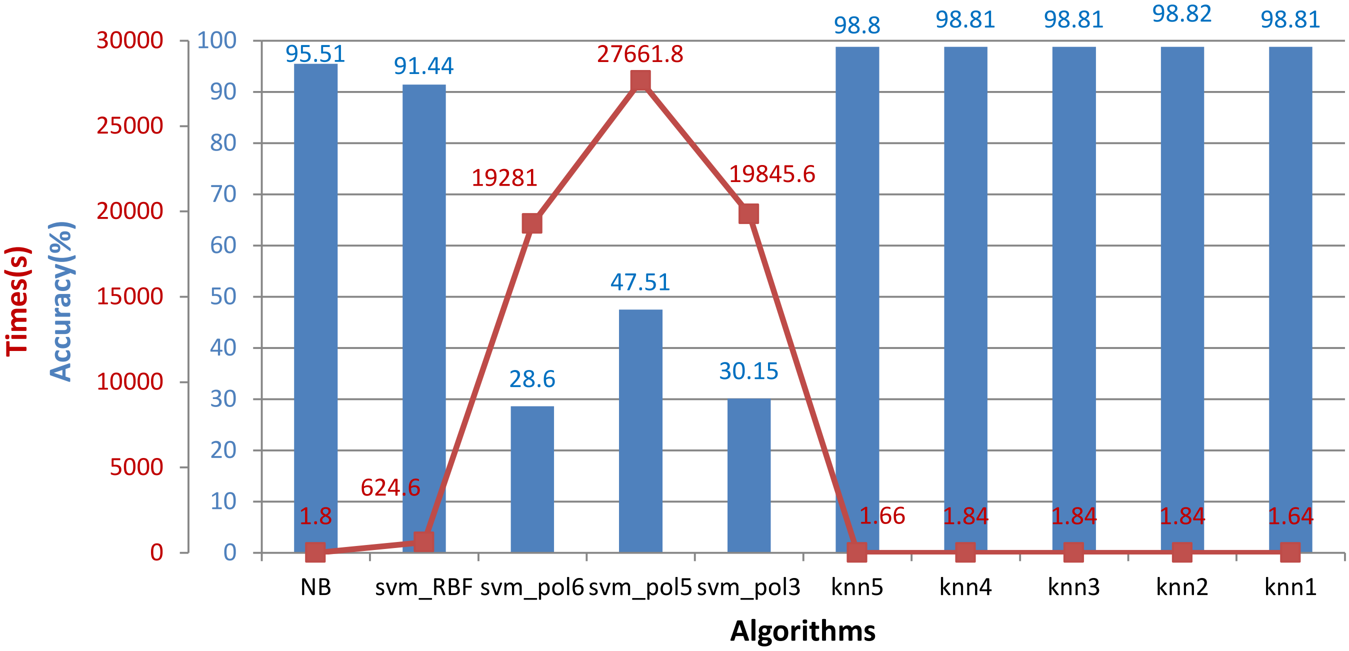


Figure 22: Accuracy comparison for SVM, KNN and NB

- After comparing three algorithm (KNN, SVM and NB), the result of KNN (number of neighbors = 2) was the best result in terms of accuracy and detection time.

3.8 Conclusion

In this chapter, we introduce the tools and various machine learning techniques used to classify data with explanations, then we compare the results obtained.

Finally, we show the results obtained and compare them and choose the best solution to the problem.

General Conclusion

In this work, we conducted a study on discovering errors in the field of photovoltaic converter with a set of data, and the data was divided from 600,000 thousand values to 50,000 values to facilitate the classification process.

For this purpose, models were proposed to compare the SVM, KNN, and NB algorithms in classification and how to detect faults. The results showed that the "KNN" algorithm had a high detection rate compared to other algorithms (SVM, NB).

The KNN technique was chosen as the best technique in this type of detection.

The work carried out in the framework of this dissertation aims at accurate fault detection using techniques machine learning..

In the first chapter, we presented an overview of fault detection with its fields, and in the second chapter, we presented a comprehensive overview of machine learning and its broad uses, with knowledge of error detection using machine learning algorithms.

In the third and final chapter, we presented data analysis, algorithms used, and obtained results.

References

[1] Dubravko Miljković, Hrvatska elektroprivreda, Zagreb, Croatia " Fault Detection Methods: A Literature Survey" Conference Paper · May 2011.

[2] MaríaJesús " FAULT DETECTION AND ISOLATION:AN OVERVIEW ", Dpto. Ingeniería de Sistemas y Automática, Escuela de IngenieríasIndustriales, Universidad de Valladolid.

[3] Sobhani-Tehrani, Khorasani, Fqult "Diqgnosis of Nonlinear Systems Using Hybrid Approqch".

[4] Blanke et al., 2003https://www.igi-global.com/"What is Actuator Fault" March 18, 2021, 15:02.

[5] M.Blanke" Fault-tolerant control Systems" Department of control, Aalborg University, Fredrik Bajers.

[6] Gao Fei, Zhang Hong-yue "Analysis of Observer-Based Fault Tolerant Control Systems with Markovian Parameters " 2006 - 2007.

[7] Gary G. Yen, 2005 " The Dilemma of Online Fault Detection and Diagnosis", in The Electrical Engineering Handbook.

[8] Applied Soft Computing, 2015 https://www.sciencedirect.com/topics/engineering/missed-detection"Missed detection ". March 27, 2021, 09:32.

[9] Farshad Harirchi, Necmiye Ozay, "Guaranteed Model-Based Fault Detection in Cyber-Physical Systems: A Model Invalidation Approach".

[10] " What is Machine Learning ? A definition" , https://expertsystem.com/machine-learning-definition/, March 16, 2021, 20:25.

[11] Vinicius A.D. Silvam Robson, " Fault detection in induction motors based on artificial intelligence", University of Campinas.

[12] OULMI Mehdi- KALOUNE Salim: Classification d'objets avec le Deep Learning,p 4,UniversityAkli Mohand Oulhadj de Bouira,2017/2018.

[13] " Classification in Machine Learning edureka ", https://www.edureka.co/blog/classification-in-machine-learning/#:~:text=Artificial%20Neural%20Networks.,Classification%20Algorithms,recognition%2C%20document%20classification%2C%20etc , March 17, 2021, 16:30

[14] " Regression in Machine Learning | medium ", https://medium.datadriveninvestor.com/regression-in-machine-learning-296caae933ec , March17, 2021, 11:15.

[15] https://www.ibm.com/cloud/learn/neural-networks , March 17, 2021, 16:30.

[16] Khushbu Kumari, Suniti Yadav, Linear regression analysis study,2018.

[17] " K-Nearest Neighbor(KNN) Algorithm for Machine Learning – Javatpoint " , https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning, March24, 2021, 15:10.

[18] Mariana Belgiu, Lucian Drăguţ, Random forest in remote sensing: A review of applications and future directions,2016.

[19] https://www.ibm.com/cloud/learn/unsupervised-learning, March 28, 2021, 20:14.

[20] https://searchenterpriseai.techtarget.com/definition/reinforcement-learning, March 28, 2021, 20:30.

[21] IATARIEN Abderraouf, Classification d'âge par Deep Learning, University Kasdi Merbah Ouargla, 2019/2020.

[22] Von Reischach F, Guinard d, Michahelles F, Fleisch E, A mobile product recommendation system interacting with tagged products, 2009.

[23] Schellekens M, Self-driving cars and the chilling effect of liability law, Computer law & Security Review, 2015.

[24] Kanneth R Foster, Robert Koprowski, Joseph D Skufca,Machine learning, medical diagnosis, and biomedical engineering research-commentary, 2014.

[25] https://www.javatpoint.com/applications-of-machine-learningApril 30, 2021, 22:14.

[26] MoussaHamadache ,SaikatDutta, Osama Olaby , RamakrishnanAmbur , Edward Stewart and Roger Dixon " On the Fault Detection and Diagnosis of Railway Switch and Crossing Systems: An Overview".

[27] Worden, K., &Dulieu-Barton, J. M. (2004). *An Overview of Intelligent Fault Detection in Systems and Structures. Structural Health Monitoring: An International Journal, 3(1)*

[28] M.Blanke, M.Kinnert, J.Lunze, M.Staroswiecki, "Diagnosis and Fault-Tolerant Control". (200*6)*

[29] ZoltánGermán-Salló, Gabriela Strnada, "Signal processing methods in fault detection in manufacturing systems"

[30] R.Isermann, " Integration Of Fault Detection And Diagnosis Methods", Institute of Automatic Control, Laboratory for Control Engineering and Process Automation, Technical, University Darmstadt, Landgraf-Georg- StraBe 4, D-64283 Darmstadt, F.R. Germany.

[31] Leo H. Chiang, Evan L. Russell, Richard D. Braatz, " Knowledge-based Methods"

[32] J. Phys, "A Survey of Knowledge-Based Intelligent Fault Diagnosis Techniques", Journal of Physics: Conference Series Conf. Ser. 1187 032006.

[33] Mo-Yuen Chow, & Goode, P. V. (n.d.). "The advantages and challenges of machine fault detection using artificial neural network and fuzzy logic technologies", Proceedings of 36th Midwest Symposium on Circuits and Systems. doi:10.1109/mwscas.1993.342948.

[34] J. Zarei,J. Poshtan, (2011). *"Sensor Fault Detection and Diagnosis of a Process Using Unknown Input Observer", Mathematical and Computational Applications, 16(1), 31–42.* doi:10.3390/mca16010031

[35] Gang Zhao, Dongxiang Jiang, Jinghui Diao, Lijun Qian,"APPLICATION OF WAVELET TIME-FREQUENCY ANALYSIS ON FAULT DIAGNOSIS FOR STEAM TURBINE", Octeber 2004.

[36] Waleed K Ahmed, "Advantages and Disadvantages of Using MATLAB/ode45 for Solving Differential Equations in Engineering Applications", Article February 2013.

[37] https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-MATLAB.html , May 25, 2021, 12:00.

[38] Dr. Harris Tsimenidis, "Introduction to MATLAB", Newcastle University School of Engineering, September 2018.

[39] https://fr.mathworks.com/discovery/machine-learning.html, May 26, 2021, 13:00, May 25, 2021, 13:42.

[40] TuomasM, JuhaP,Lari N, Anssi M Photovoltaic Generator as an Input Source for Power Electronic Converter, June 2013.

[41] Ndeye Lo, Jean-Marie Flaus, Olivier Adrot " Review of Machine Learning Approaches In Fault Diagnosis applied to IoT System " https://hal.archives-ouvertes.fr/hal-02344344/document, April 16, 2021, 11:34.

[42] "fault diagnosis using artificial neural networks (ANNs)" https://link.springer.com/chapter/10.1007/978-94-015-8300-8_5, April 13, 2021, 21:34.

[43] Majid Jamil, Sanjeev Kumar Sharma " Fault detection and classification in electrical power transmission system using artificial neural network".