

Identification of PV Fault Classes Using Intelligent Method KNN (K-Nearest Neighbours)

Godfrey Benjamin Zulu¹, Dr. C. Kara Mostefa Khelil², Godfrey Murairidzi Gatora³, Taane Zahreddine²

¹Mulungushi University, Department of electrical and electronics engineering

²Université Djilali Bounaama Khemis Miliana

³Arrupe Jesuit University, School of Engineering and ICT

DOI: <https://doi.org/10.51244/IJRSI.2024.1108093>

Received: 06 August 2024; Accepted: 13 August 2024; Published: 15 September 2024

ABSTRACT

Throughout many developing nations of our humble planet, renewable energy is a hot topic. Every country at this very moment is trying to move away from fossil fuels like petrol to complete renewable energy sources especially Photovoltaic systems.

The reliability and efficiency of renewable energy systems is now a frequent topic of discussion. Like all systems of production, renewable energy systems are subject to failures and defects in their normal operating functions with regards to the amount of power output. These systems break down and deteriorate during the period of their operation. This is why a system of diagnostic is required whose many objectives is to provide indicators with the given valuables like temperature, solar irradiation, voltage and current output to detect the faults and thus maintain the energy production at optimum.

The work in progress relates to the diagnostic of faults in the PV systems using artificial intelligent methods particularly the K-nearest Neighbour algorithm.

Keywords: Photovoltaic generator, faults, detection, diagnostic identification, KNN

GENERAL INTRODUCTION

Since the dawn of civilization, mankind has looked at the sky and wondered if it's possible to obtain energy from the stars especially our nearest star, the sun. So as we progressed as a civilization, we begun to use different energy sources to help with our daily energy consumption. Among these energy sources were fossil fuels like petrol, diesel, hydro carbons and natural gas.

For a time being we thrived on fossil fuels and hydro carbons until we discovered they were not the best of energy sources. They polluted and still continue to pollute our environment by the emission of carbon dioxide (CO_2) which causes the greenhouse effect. [1]

Since 1990s, Man has worked hard to discover various ways we to harness the solar energy from the sun and the most prominent way he has done this is through Photovoltaic systems. Today PV systems are everywhere and continue to be a subject of discussion in the aspect of renewable energy. [2]

Despite the PV systems having high efficiency rate, ideal for the environment and ease of use, they are often accompanied by system defaults which may not always be detected on time hence the use of Artificial intelligence algorithms like KNN, Naïve Bayes and probabilistic methods to classify different faults. In our work, we will focus more on KNN algorithm and how we can apply it to categorize different faults detected in PV systems at last give an example of a Mat-Lab simulation of the system. [3]

Our work includes five chapters that explain in detail the PV class systems, PV faults, different methods used to detect faults, the KNN algorithm and the conclusion from the experimental simulation using a PV set-up of the parameters of our choosing in comparison to already existing ones.

INTRODUCTION

Faults in any components (modules, connection lines, converters, inverters, etc.) of photovoltaic (PV) systems (stand-alone, grid-connected or hybrid PV systems) can seriously affect the efficiency, energy yield as well as the security and reliability of the entire PV plant, if not detected and corrected quickly. In addition, if some faults persist (e.g. arc fault, ground fault and line-to-line fault) they can lead to risk of fire. Fault detection and diagnosis (FDD) methods are indispensable for the system reliability, operation at high efficiency, and safety of the PV plant. In this paper, the types and causes of PV systems (PVS) failures are presented, then different methods proposed in literature for FDD of PVS are reviewed and discussed; particularly faults occurring in PV arrays (PVA). Special attention is paid to methods that can accurately detect, localize and classify possible faults occurring in a PVA. The advantages and limits of FDD methods in terms of feasibility, complexity, and cost-effectiveness and generalization capability for large-scale integration are highlighted. [4]

Improving the efficiency of photovoltaic (PV) systems has gained priority in current research due to the large volumes of PV panels installed. Moreover, the remarkable efforts made to investigate different methods of diagnosing PV failures have multiplied, giving additional impetus to research on the efficiency of PV systems. However, most of these methods are limited in the number of faults that can be identified; some are expensive and complex, and others require huge amounts of data to train. In this paper, a simple and robust multivariate statistical analysis method is proposed for the diagnosis and identification of faults in a PV system. [5]

Fault Classification

In this research, we are going to discuss the most common occurring faults in the photovoltaic installation. The faults chosen are classified according to their origin, intrinsic or extrinsic in the PV system.

To better understand the faults, we will elaborate more on the intrinsic and extrinsic faults by grouping them in a table by type of fault, its consequence and its degree of impact (weak, average, strong), also its phase of origin (Fabrication stage, installation, during process of use) .

Intrinsic Faults:

Table I.1: Intrinsic faults and their consequences [6]

FAULTS	CONSEQUENCE	CR IT	OC C
Inversion of output links	Incorrectly wired module, reduced performance	60 %	60 %
Bad orientation and inclination of the modules	Shady operating area which leads to reduced performance	40 %	60 %
Galvanic couple due to the mixing of the module junction	Corrosion	40 %	60 %
Bad module ventilation	Heating	40 %	40 %
Badly fixed module	Module displacement Reduced performance	40 %	40 %

Unwired module	Reduced performance	40 %	40 %
Cracking	Loss of its lightness nature Cell deterioration Decrease in its robust and performance	60 %	20 %
Rust by water infiltration	Cell deterioration, Loss of lightness	60 %	20 %
Poor insulation between modules and the inverter	Short-circuit Destruction of the module Fire	60 %	20 %
Humidity penetration	Hot spot Leakage current Reduced resistance of the short circuit current Corrosion Loss of adhesion and insulation	60 %	20 %
Difference in module performances	Decrease in the performance of light intensity	20 %	60 %
Cable gland plugs missing on the connection box	Water penetration Corrosion of the connections	20 %	60 %
Connection box put upside down	Water enters the box by plugs	20 %	60 %
Increase in the series resistance due to thermal cycle	Reduced performance	40 %	60 %
Deterioration of anti-reflective layer	Reduced performance	40 %	60 %
Inclination of the modules too low	Stagnation of water Earth deposit Mushroom overgrowth Sealing problem	40 %	60 %
Degradation of	Deterioration of joints	40	20

interconnections	Reduced performance Augmentation or resistance and heat	%	%
Inadequate mechanical support for the modules	Less mechanical effort to support the modules hence parts start separating	40%	20%
Bad mechanical resistance for the support modules	Deformation of the support system	40%	20%
Diffusion of phosphorous to the surface	Loss of encapsulation adhesion	40%	20%
Important leakage current	Heating problem	40%	20%
Overheating of the modules by the connection box	Detachment of the components and short-circuits of the system. Destruction of the insulation materials And reduced performance	40%	20%

Extrinsec Faults

Table I.2: extrinsic faults and their effects [6]

Fault	Consequence	Critic	Occur
Accumulation of soil, snow and other substances	Loss of power	60%	60%
Degradation of the modules by vandalism	Reduced performance, Lack of proper functioning of the system	60%	40%
Theft of modules	Functioning of the PV system affected	60%	40%
Deterioration of the sealing joints	Sealing losses Deterioration of cells	60%	20%
Deformation of the frame modules	Infiltration of water	60%	20%
Corrosion of the frame for modules	Deterioration of cells	60%	20%
Delamination	Poor performance Overheating	60%	20%
Lightening	Deterioration of the cells	60%	20%

Storms	Torn and broken module	60%	20%
Structure weaknesses to the wind	Torn and broken module	60%	20%
Lightning strikes on	Destruction of modules	60%	20%
Partial shade	Hot spot Deterioration of cells	60%	20%
Degradation of the encapsulation due to ultraviolet	Does not absorb radiation properly Reduced performance	60%	20%
Degradation due to light intensity	Destruction of diodes Reduced performance Voltage sage	40%	20%
Degradation due to heat	Reduced performance, Deterioration of joints and overheating	40%	20%
Nests of insects or birds on the modules	Reduced performance	40%	20%

Mismatch And Shading Faults

The mismatch and shading faults are the most frequent occurring faults in the PV systems. We are going to discuss these commonly occurring faults in depth.

Defination:

The mismatch fault is a fault caused by cell grouping which has the non-identical characteristic of I-V. Any change in the characteristic if I-V will cause tremendous amount of problems.

The shading fault problem is a specific type of mismatch fault because its presence signifies the reduction of solar radiation received by the solar cells. The change in the parameters affect two principal factors.

Firstly, the cells can have different physical properties caused by the fabrication tolerance, only the tolerance of power output of the cells are fixed by the fabricator and can vary from +/-3% and +/-5% depending on the fabricator.[7]

Secondly, the PV cell modules can be exposed to different working conditions caused by different faults. The parameters affected in this instance can be represented in the table below:

Table I.3: impact of different faults on the parameters of the PV cell [8]

NATURE OF FAULTS	PARAMETERS AFFECTED
Torn or broken modules	Variation of optimal current I
Shading caused by tree leaves, manure, sand, snow,	

pollution etc.	
Cell heating	Variation in temperature T
Degradation of the interconnections Cracking Corrosion of connections between cells	Variation in the series resistance R
Different module performances Deterioration of the cells Penetration of humidity	Variation of all the parameters of the cells.

Commonly Occurring Faults In Pv Site Systems And Installation

We have discussed different types of faults in the previous paragraphs and now it's time to talk about commonly occurring problems in PV installations. All the problems mentioned above are experienced in PV systems but not always and often. For example, it's not always it snows and corrosion doesn't happen in a single night, so while these maybe faults encountered, they may not be the most occurring of them all. Take an example of a car, a car experiences a lot of faults but a good driver knows where exactly to look for faults in an event that a car suddenly broke down, similarly a good engineer will start the diagnostic of PV faults with the following commonly occurring faults in the systems. [9]

Current Faults In Pv Systems (I)

The faults in the PV system can be described as temporally or permanent. The temporally faults are caused by shading and fouling of the solar cell modules. The permanent faults in the module are:

- The de-lamination.
- The bubbles and water drop on the surface of the cells.
- The yellowing of the cells due to radiation.
- Scratches and burnt cells.

The permanent faults are eliminated by replacing or repairing the destroyed or affected modules. The most serious and dangerous faults in a photovoltaic system are caused by short-circuits, between the lines, grounding and arc fault. Other factors which might minimize the power output during production is a point of maximum power (MPP), the power losses by joule effect in the cables and faulty equipment. The faults in the photovoltaic system can therefore be classified as the module faults, channels or grid according to the PV components involved. [9]

Hot Spot Faults

The hot spot faults occur in individual solar modules when they are shaded or broken by mechanical stress. These cells produce far less electrical current compared to the other cells not affected and can be polarized in the inverse direction which results in the production of heat by joule effect in the course of production.

This phenomenon affects the cells made from silicon crystalline and generally results in the fouling, shading of the damaged cells or diodes of the bypass damaged. The hot spot points release energy which increases the temperature of the surface and consequently the hot spot faults are diagnosed and analyzed by thermal and

infra-red. If the hot spot faults persist, it can damage the solar cells and the bypass diodes and provokes short circuit faults.

Degradation

The degradation of the solar modules facilitates the reduction in the power output as time passes. The degradation faults can be identified by reading the characteristic I-V of the module.

Partial Shade

The shading faults occur when certain parts of PV solar module receive less radiation compared to the rest of the module due to obstructions and shadows. Shading can be diagnosed by looking for unexpected current drops. A shadowing effect gives similar results to open circuit strings but are most often temporally. [10]

Open Circuit Faults

The open circuit faults are reference the faults of interconnection in the sub-systems of the PV generator or module. It will equally include the disconnections of the cells of the module, the chain of modules or the chain of the PV electrical grid.

The diagnostic in the PV grid can be done by inspecting the voltage and current indicators. The voltage of the PV grid remains constant, however, the fault results in the current drop. The open circuit faults can be caused by the damaged cells, defective diodes and wiring faults.

Short Circuit Faults

All like the open circuit faults, the short circuit faults can be produced in the different sub-systems of the PV installation. The modules having the short circuit in the chain of production experiences a significant voltage drop in the grid such that the current of the lines increase exponentially. The same effect is produced when the short circuit is produced between two branches of the system. An experimental study shows that the short circuit faults between the modules has harmful effects on the output voltage of the system as the short circuit of the strings. [11]

Ground Faults

The ground faults are considered to be the most commonly occurring faults in the PV systems. The faults result from the accidental electrical short circuit between the electrical conductor and the ground. This fault is principally caused by the wiring insulation. The grounding faults can cause serious harm and risks for the security of the workers in an event of the electrical arcs of DC currents generated at the point of failure on the system, the electrical shocks due to ground faults results in less voltage compared to the nominal voltage and risk fires.

Arc Faults

The involuntary passage of current in the air or in another dielectric is known by the name arc fault. The fault arcs can be produced by the discontinuity between two electrical conductors having different potential differences. Arc faults in photovoltaic system can risk serious dangers to the installation.

Line To Line Faults

The line-to-line fault designates the short circuit faults between the conductors of the PV system. The line-to-line faults can be caused by fault insulation of the wires and mechanical damage. [12]

Methods Of Detection And Diagnostic Of Pv Systems

Fault analysis and fault detection are important to the efficiency, safety and reliability of solar photovoltaic (PV) systems. Despite the fact that PV systems have no moving parts and usually require low maintenance, they are still subject to various fault conditions. Especially for PV arrays (dc side), it is difficult to shut down PV modules completely during faults, since they are always energized by sunlight in daytime. Furthermore, conventional series-parallel PV configurations increase voltage and current ratings, leading to higher risk of large fault currents or dc arcs.

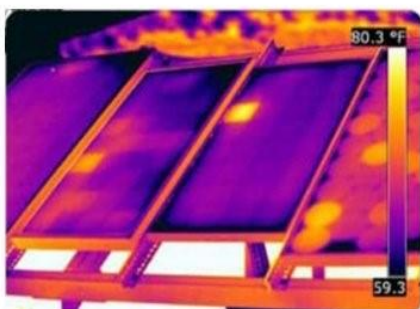
Once PV modules are electrically connected, any fault among them can affect the entire system performance. This means the PV system is only as robust as its weakest link (e.g., the faulted PV components). In a large PV array, it may become difficult to properly detect or identify a fault, which can remain hidden in the PV system until the whole system breaks down. In addition, conventional series-parallel PV configurations increase voltage and current ratings, leading to higher risk of large fault currents or dc arcs.

There are three methods used for diagnostic of PV systems in the industry:

- Method based on the analysis of current and volt (electrical method)
- Method based on the analysis of other parameters like the I-V curve (also called non electric method)
- Literature Method [13]
- Artificial Intelligence methods [14]

Non-Electrical Methods

There exist many non-electrical methods, destructive or non-destructive for the diagnostic of PV faults in the module. The main principal fault we can give much attention to is cell cracking. We can cite the methods as follows: mechanical bending tests, imagery by photo-luminescence, electro luminescence and the test of thermography. For the diagnostic of PV modules, the method of imagery (thermal camera) infra-red is widely used.



(a) cellule court-circuitée



(b) corrosion dans la boîte de jonction

Figure I.1: certain examples of the detection of PV faults using thermal camera.

There have been most successes in the localization and detection of PV faults using the thermal camera which are noted as: current leakage in the PV, increase in the resistance of the connection between the modules, abnormal heating of the cells, and conduction of the bypass diode. This method can equally be applied for the connections in the junction box and the functionality of the anti-reverse diode. [15]

Electrical Methods

Using electrical methods, the most considered parameters are:

- The current output of the GPV
- The voltage on the terminals of the GPV

- The insulation resistance between the positive and negative terminals of the GPV.

It is also possible to add additional parameters like the ambient temperature of the site and sunshine radiation in the electrical measures. The measures on the AC side are important because they are directly related to the energy which will be sold. It is necessary to take note of:

- The AC currents
- The AC voltage
- The frequency
- Impedance of the electrical grid as seen by the inverter

Out of these parameters written, it becomes much easier to deduce the following:

- ❖ DC instantaneous power
- ❖ AC instantaneous power
- ❖ Electrical energy produced on different periods (depending on the capacity of the storage system) on the side of both AC and DC.

We often add the following:

- The functioning duration of the inverter
- The date when put in service
- The CO₂ not released in the atmosphere (which is rather economical to the environment). [16]

Literature Methodes

The different methods proposed in the literature method type of detection and localization of PV faults are as follows:

Reflectometry Method

The reflectometry method is a diagnostic method used to send a signal in the system or on the diagnostic side. This signal is propagated according to the law of propagation in the medium in question and it encounters the discontinuity, and part of its energy is re-transmitted to the point of injection. Analysis of the signal allows us to deduce the information on the system or medium been considered.

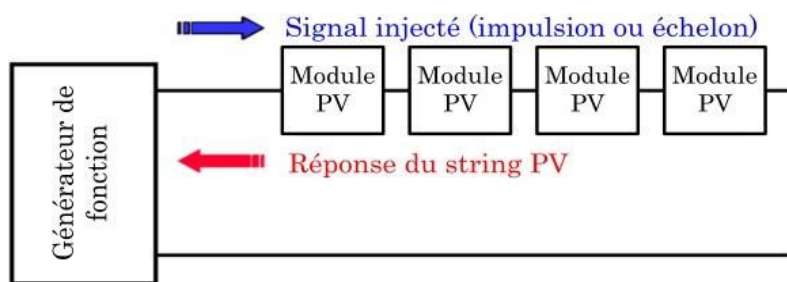


Figure I.2: principal reflectometry method for the detection and localization of PV faults in a string.

Analysis Of Power And Energy Produced

The power or energy measured is compared to the expected output and when there is an important deviation, we can be certain that there is surely a faulty.

The suggested analysis consists of generating the supplementary attributes on the power drop or energy produced such as: the duration, the amplitude, the frequency and the drop instances. These same attributes are equally predetermined for the different faults considered in our study.

In the course of their comparison, the fault whose value attributes are considered similar or close to those measured are considered as the faults responsible for the drop in the power output. [17]

Detection of PV faults by means of Artificial Intelligence methods

Introduction

Various factors, including maximum power point tracking error, environmental effects like shading and dust or snow buildup on the PV surface, wiring losses and aging, and malfunction in other PV components like the power conditioner unit and the inverter can all have an impact on how well a PV system operates. According to a monitoring study in [16], faults may cause a PV system to generate roughly 18.9 percent less power annually. In order to continually analyze the current, voltage, and output power characteristics of a PV system and find both existing and emerging defects, proper techniques have to be developed especially ones based on artificial intelligence like the following of the many examples.

Fuzzy Logic method Control Implementation

Among several renewable energy resources, Solar has great potential to solve the world's energy problems. With the rapid expansion and installation of PV system worldwide, fault detection and diagnosis has become the most significant issue in order to raise the system efficiency and reduce the maintenance cost as well as repair time.

The Fuzzy Control Implementation (FLC) is one of the modern artificial intelligence techniques used in fault diagnosis in the PV systems. The architecture of the implementation is based on the Max-Min arrangement procedure with a centroid type for the defuzzification. [24]

Artificial Neural Networks

Artificial neural networks, a pivotal technique of artificial intelligence, have been developed and applied in many fields including the fault diagnosis of PV systems, due to their strong self-learning ability, good generalization performance, and high fault tolerance. Artificial neural networks (ANN) are type of machine learning algorithms that are commonly used for PV fault diagnostic and detection. ANN's can be trained on large datasets of PV system performance data to recognize patterns associated with various types of faults. [19]

Deep Learning

Deep learning is a type of machine learning that uses neural networks with multiple layers to learn complex patterns in data. Deep learning has been shown to be effective in fault diagnosis and detection in PV systems particularly in cases where there are large amounts of data available. [20]

Genetic algorithms

Genetic algorithms are a type of optimization algorithms that can be used to optimize the parameters of a diagnostic systems. Genetic algorithms can be used to find the optimal set of parameters for a diagnostic system that can accurately diagnose faults in a PV system. [21]

Support Vector Machines (SVMs)

The support vector machines are another machine learning algorithms commonly used for PV fault diagnosis and detection. SVMs are particularly effective in cases where there are multiple types of faults that need to be distinguished from one another. [22]

Decision Trees

Decision Trees are a type of machine learning algorithm that can be used to create a model of the decision-making process used in fault diagnosis and detection. Decision trees can be used to identify the most likely cause of a fault based on the observed system performance. [24]

Operating Point Analysis

Other comparison between the power and energy produced to that which is expected, the comparison of the point of actual maximum power (current and voltage corresponding to the maximum power) to that which is expected can carry much information on the state of the PV system. The rational comparison between these currents and these voltages gives the two pairs of binary values (0 and 1). Depending on the comparison of these two pairs of values, the nature of the PV fault can be identified. The four families of problems are as follows:

- Faulty modules in a string
- Faulty string
- The family of non-discriminant faults: shading, MPPT error and old modules
- False alarms. [25]

These are just some of the AI methods used for fault detection and diagnosis.

The Bayesian Neural Networks

Bayesian neural network (BNN) combines ANN with Bayesian implication. Basically speaking, at BNN level, the treatment of both weights and outputs as variables and control over-fitting. The final goal of BNN is to quantify the uncertainties presented by the models, this approach employs the statistical methodology where the whole data has a probability distribution attached to it, In user interface design software, variables tend to take a specific value will turn the same result at every access to the dedicated variable. In comparative way, the Bayesian world can own similar entities as well-known as random variables that will present a various value at any moment you access it. In other terms, the historical data describe the prior information of the overall manner with each variable giving its own statistical properties which vary with time. [26]

Basically, Bayesian neural networks focus on marginalization comparing to other ANNs, they estimate by maximum a posteriori or predictive distribution. In addition, they depend on Markov Chain Monte Carlos, Variational Inference, and Normalizing Flows technics. Bayesian neural network are useful, in the area where data are rare, they have the capacity to obtain better results for a large number of labor as well as they can estimate the uncertainties in predictions. [27]

Analysis Of The Static Characteristic

Deformation of the current-voltage graph characteristic can be provoked by changing the working conditions (sunshine or temperature) or by the appearance of one or many faults in the PV system.

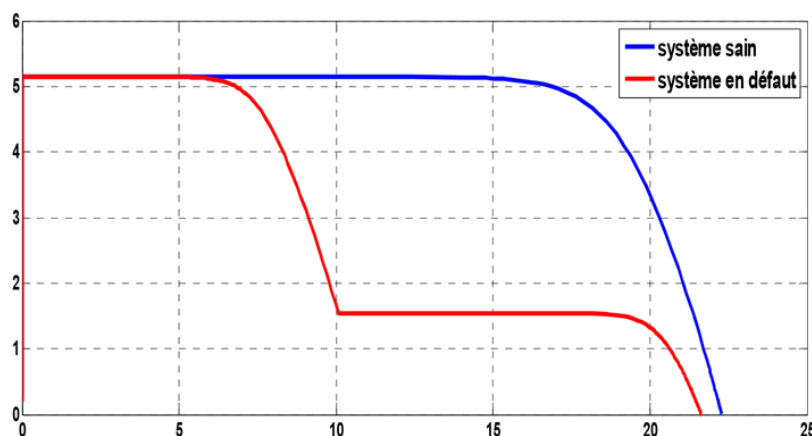


Figure I.3: appearance of the faulty graph characteristic of I-V.

Figure II.3 shows the faulty graph of I-V characteristic (shading of the module which consists of 36 solar cells at 50%) compared to the normal working conditions of the normal module. By exploiting the information on the I-V characteristic, the detection and localization of faults can be realized. [28]

Symptoms Of The Pv System

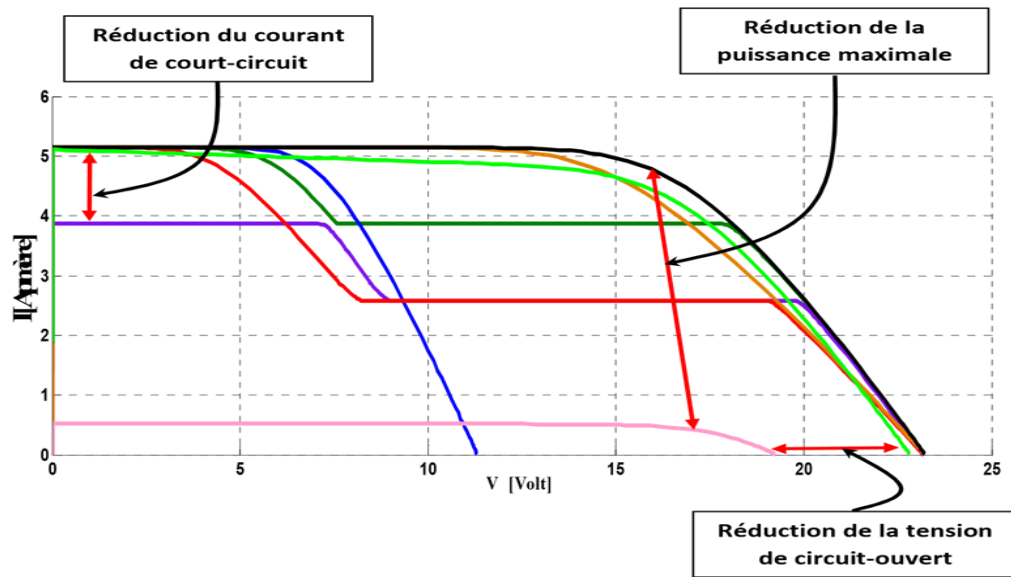


Figure I.4: symptoms of the I-V characteristic

The figure II.4 shows us how to identify faulty PV modules. From the graph we can extract information such as current, voltage and power. A normal functioning PV has a particular curve on the graph which we can compare to the faulty one as shown in the figure above. So in solar installations and solar farms, various soft wares such as MATLAB, PSIM, Scada and many more can be used to identify changes in the course of normal working conditions of the PV system. [29]

CONCLUSION

In a world where solar energy has become popular, it's imperative that engineers and technicians familiarize themselves with various problems encountered in PV installations in order to maximize the power output, diagnose and detect problems. It is clear solar power generation is the future of our lovely planet's energy and entails much attention is needed on the maintenance of PV systems.

Simulation And Programming Of Knn Algorithm Using Matlab

Introduction

When we human beings are sick, we tend to classify and identify the problem whether the causes are internal or external and how serious it actually is.

For instance, when we have a stomach ache, there are two possible causes, either internal or external. For internal causes, it may be ulcers due to the acid in the stomach. For external causes it may be the food we eat, either way we are able to identify and classify the problem and determine its causes and know whether it is serious or not. If it is serious we tend to go to the hospital, if not we just take some stomach drugs from a nearest pharmacy or wait for the problem to go on its own depending on how uncomfortable it is.

Now imagine the same scenario for machines, or more specifically our subject of discussion, PV systems. The area required for industrial installation of PV systems is huge and therefore engineers and technicians need to be alerted by the system itself about certain faults and how to identify them. This is the basis of our master's dissertation.

The Parameters Of Our Simulation

In order for us to better understand the program of our simulation, we need to know the parameters in question or rather the values we are trying to change.

The parameters in question are therefore temperature, sunlight radiation, voltage and current output.

Now we are going to be using values of temperature, irradiation, voltage and current output in our KNN algorithm to create a predictive program for our PV system.

The Values Of The Knn Program

As earlier mentioned before in chapter three, our KNN program depends on the values of k , so in order for the progressive learning of the program to continue, we are required to start with smaller values then advance to bigger values of k . in our case, we will begin with one till we reach one hundred.

The Knn Program Realised In Matlab Script

The basis of our work in to test a KNN PV program that detects and identifies faults based on the data of the system. Instead of using the Mat lab KNN tool box, we decided to code our own on MATLAB script in order to make the program more robust and so it can understand the different values been provided.

We have therefore used data science to classify and organize the data given for solar radiation, temperature, current and voltage in form of codes. Now the essence of our program is to give the code once a certain value is asked about its position in the data codes. The data is arranged in terms of training for five days, validation and test data for three days.

The Graphs Of The Simulation

Input data

The Training Phase for Temperature and solar irradiation

We first run the MATLAB script KNN program for temperature and Solar irradiation to better understand the relationship between the two variables, and the following are the graphs we obtained:

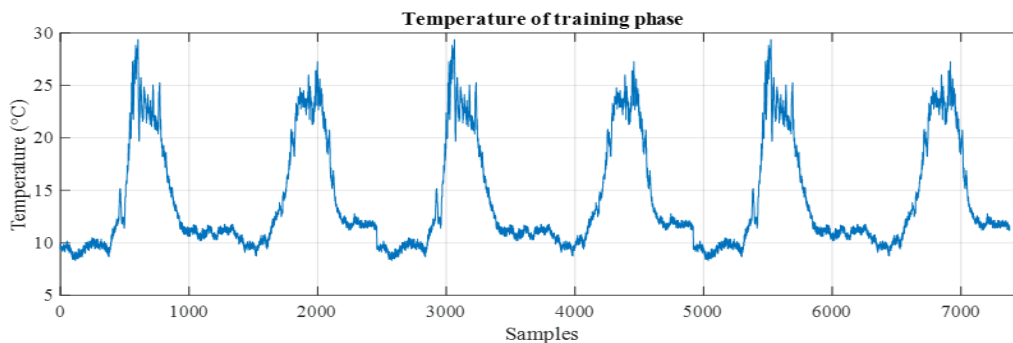


Figure II.1: the graph of temperature of training phase

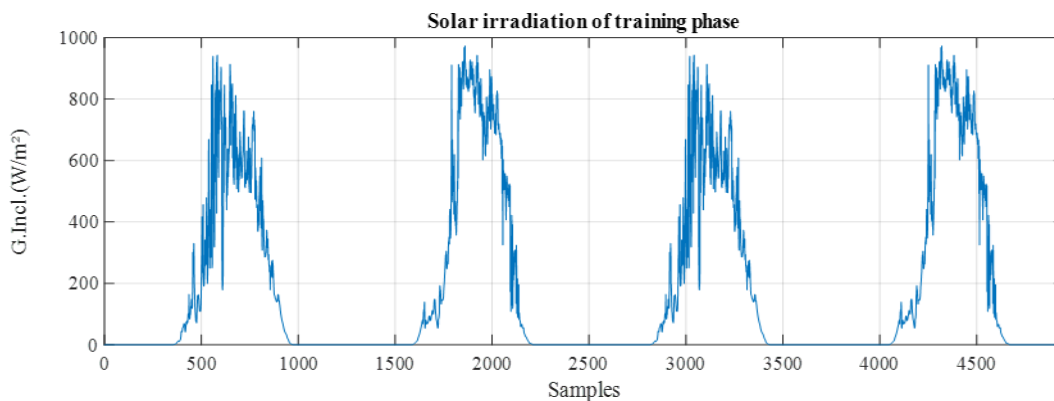


Figure II.2: the graph of solar irradiation of training phase

In solar renewable energy systems, the out values of voltage and current depend solely on two most important valuables, solar irradiation and temperature. From the first graph we see the graph of solar radiation recorded on certain days. The maximum or rather the optimum solar irradiation for the PV system is 1000w/m².

However due to the valuation of whether (summer, winter, autumn or spring) or time of the day (morning, afternoon and evening) the temperature fluctuates thereby affecting the current output as shown by figure IV.4

The Training Phase for Voltage and Current:



Figure II.3: the graph of training voltage

The current program was realized using the values of current in our sorted data and the following graphs were obtained:

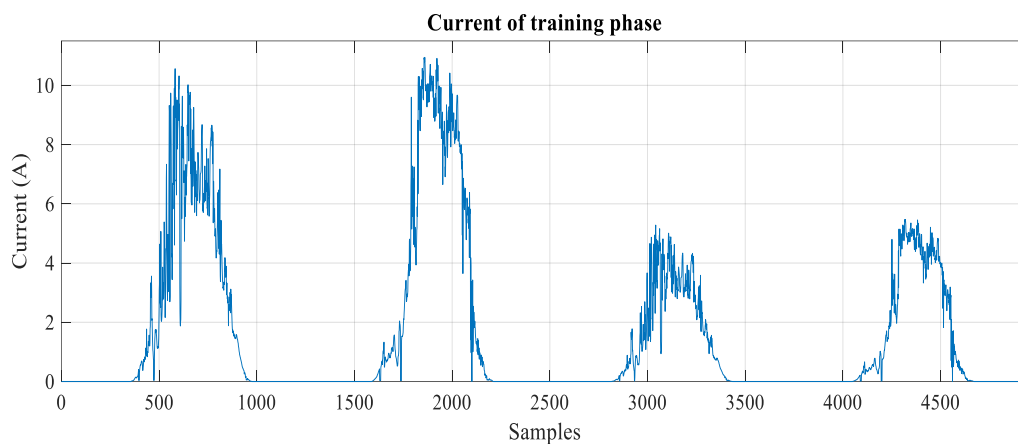


Figure II.4: the graph of training current

From the two graphs above, we can tell that the graphs are not ideal. This is because in reality, the temperature is not always at the optimum value which is 25 degrees Celsius and hence affects the graph of voltage. This is defined by one of the faults we earlier mentioned of partial or complete shading. Hence suffice it to say our system is running at a normal efficient of 65% efficient and for a photovoltaic system that's just about above average from the expected efficiency.

Output data of training phase:

Output data of training voltage classification

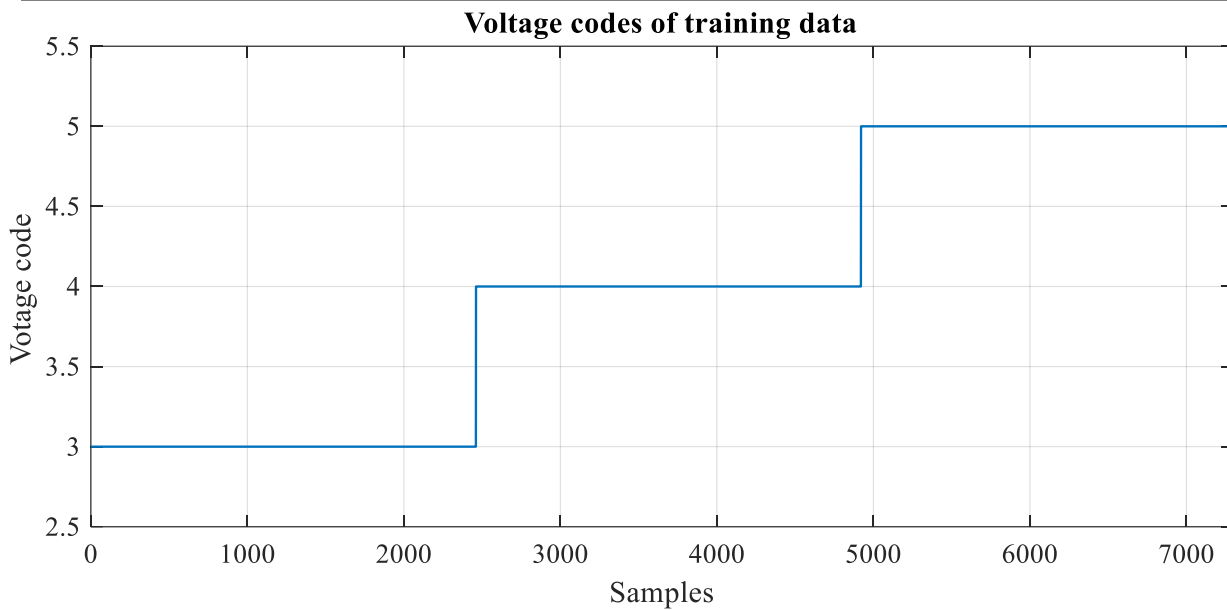


Figure II.5: the graph of training voltage classification

In the training phase, we had roughly 7388 samples in form of given values as shown by figure 4.5 on the x-axis. In PV systems, voltage output is directly related to temperature and hence each value of temperature input has a corresponding value of voltage as an output. This is shown by figure 4.5 and it equally shows the code values of the relationship between these two variables.

Output data of current classification for training phase:

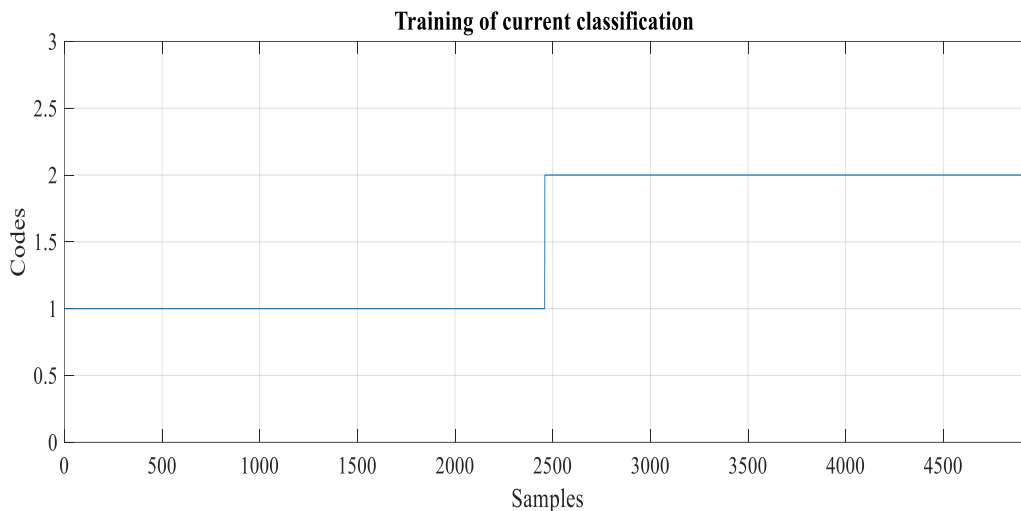


Figure II.6: the graph of training of current classification

In typical PV systems, it is the solar irradiation that directly affects the current, why? Because solar irradiation photons excite the electrons of the semi-conductor in PV systems and it is the movement of these electrons that we refer to as current hence the graph of figure IV.6. As in voltage and temperature, each value of solar irradiation has a corresponding value of current and the code of these values in our dataset is what is displayed by our figure. For example, the set of values of both solar irradiation and current between ``1-2000`` belongs to code 1.

The Validation Phase

The values tested in program simulation above were for the test data. As mentioned before, our program KNN works with training, validation and test valuables. We have so far done the simulation for training data and now we shall move to validation data.

Input data

The validation Phase for Temperature and solar irradiation

Upon running the program, following graphs were obtained:

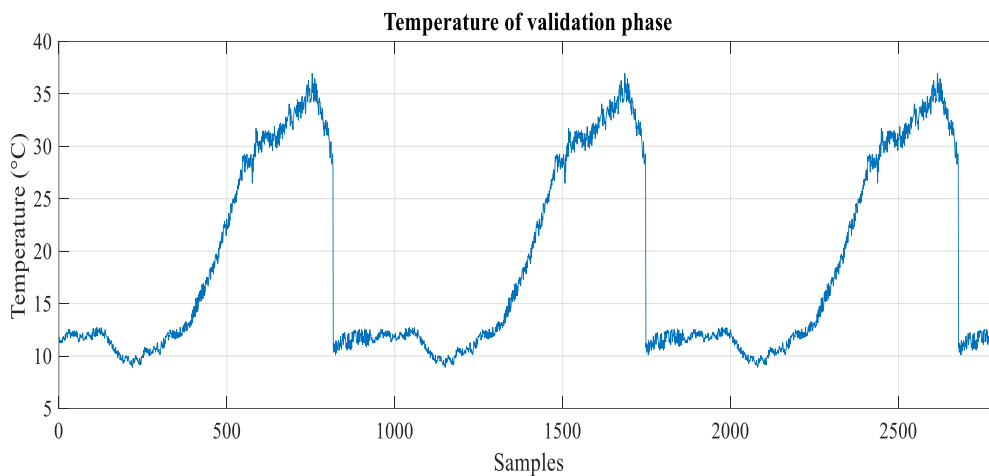


Figure II.7: the graph of temperature under the validation phase

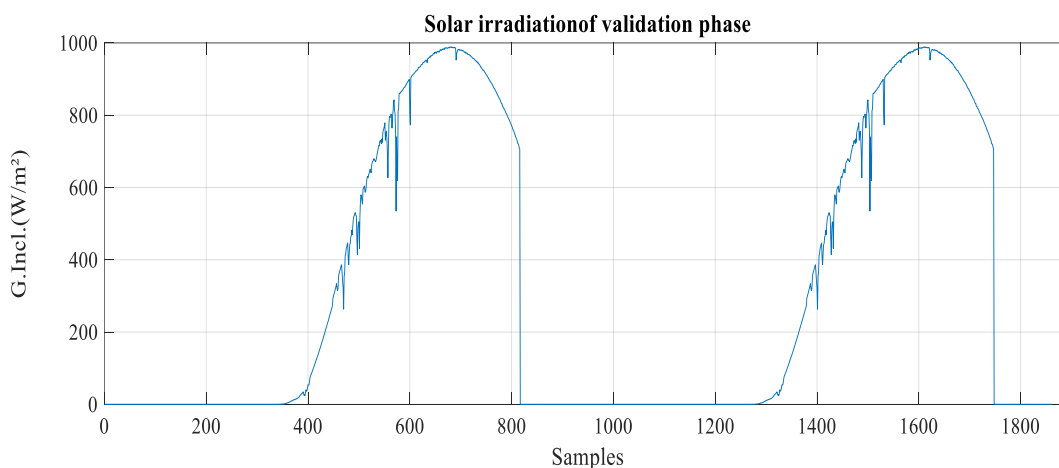


Figure II.8: the graph of solar irradiation under the validation phase

As shown by the two graphs, the noisy data has cleared and we are able to see how current is a direct function of the solar irradiation is. The graphs show how uniformly current is with regards to solar irradiation as discussed earlier.

The validation Phase for Voltage and current

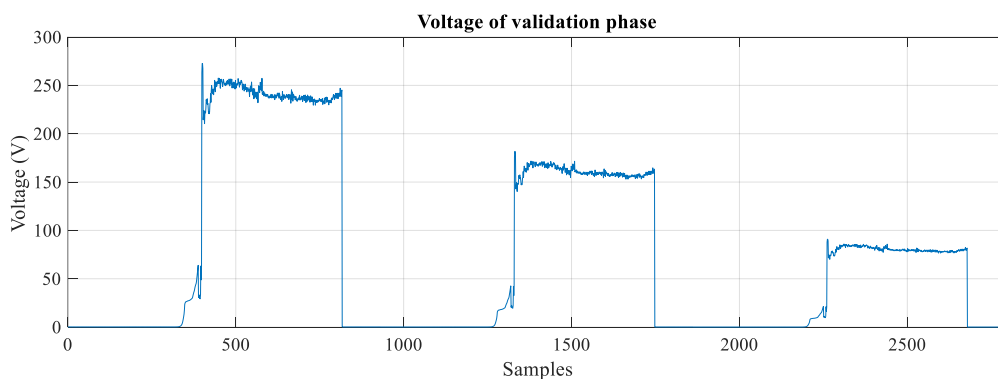


Figure II.9: the graph of voltage under validation phase

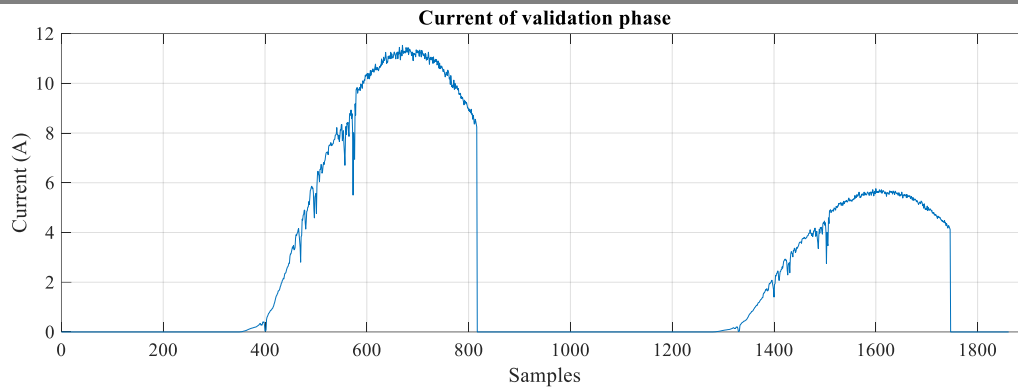


Figure II.10: the graph of current under validation phase

The KNN is a self-learning algorithm that takes time to learn and adapt. But like any learning and project, before testing whether the program or project is working, we need to be certain that it is. That certainty is validation, the need to know for sure that something will work.

In our fair case, we have the validation data which we just simulated to get the graphs of temperature and current, as observed by the two graphs, the voltage output is directly proportional to the temperature. Hence temperature is a crucial aspect of the PV systems and as such it is necessary to ensure that PV systems are running at optimum temperature (which is 25 degrees Celsius) and maximize the power output by giving the optimum voltage in return.

As shown by the two graphs, the noisy data has cleared and we are able to see how current is a direct function of the solar irradiation is. The graphs show how uniformly current is with regards to solar irradiation as discussed earlier.

Output data of validation phase:

Output data of Voltage and temperature:

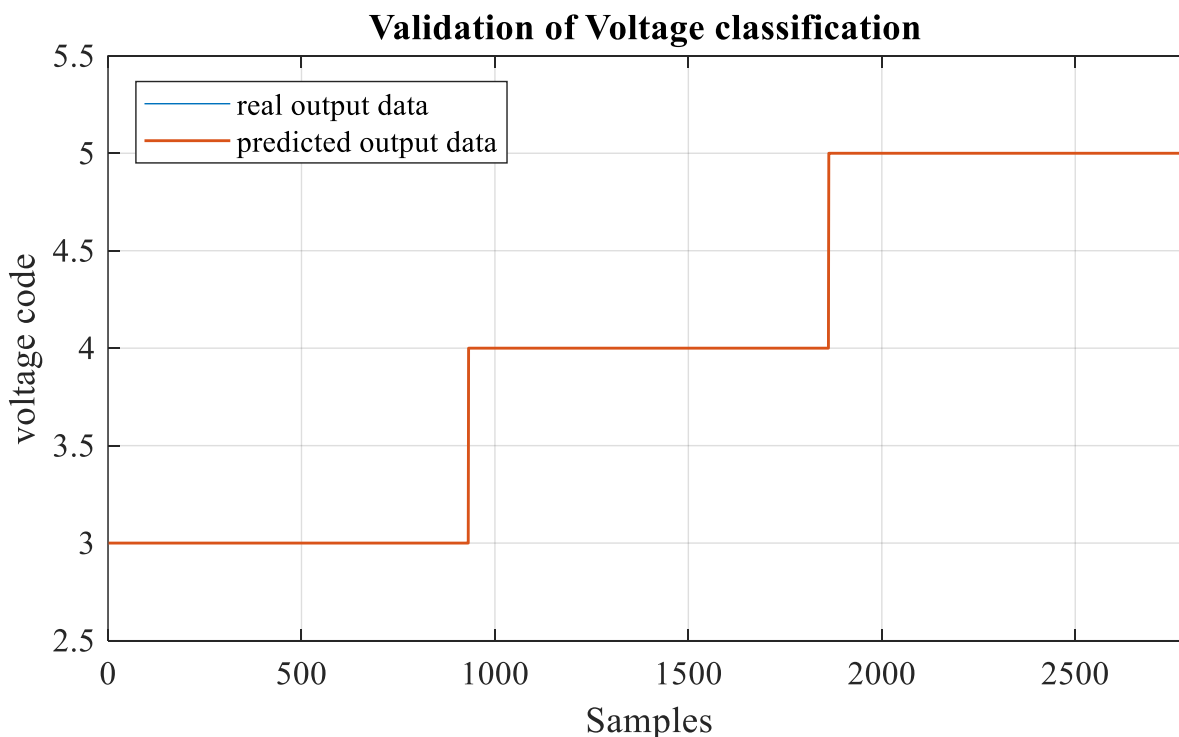


Figure II.11: the graph of validation of voltage classification

Output data of current classification for validation phase:

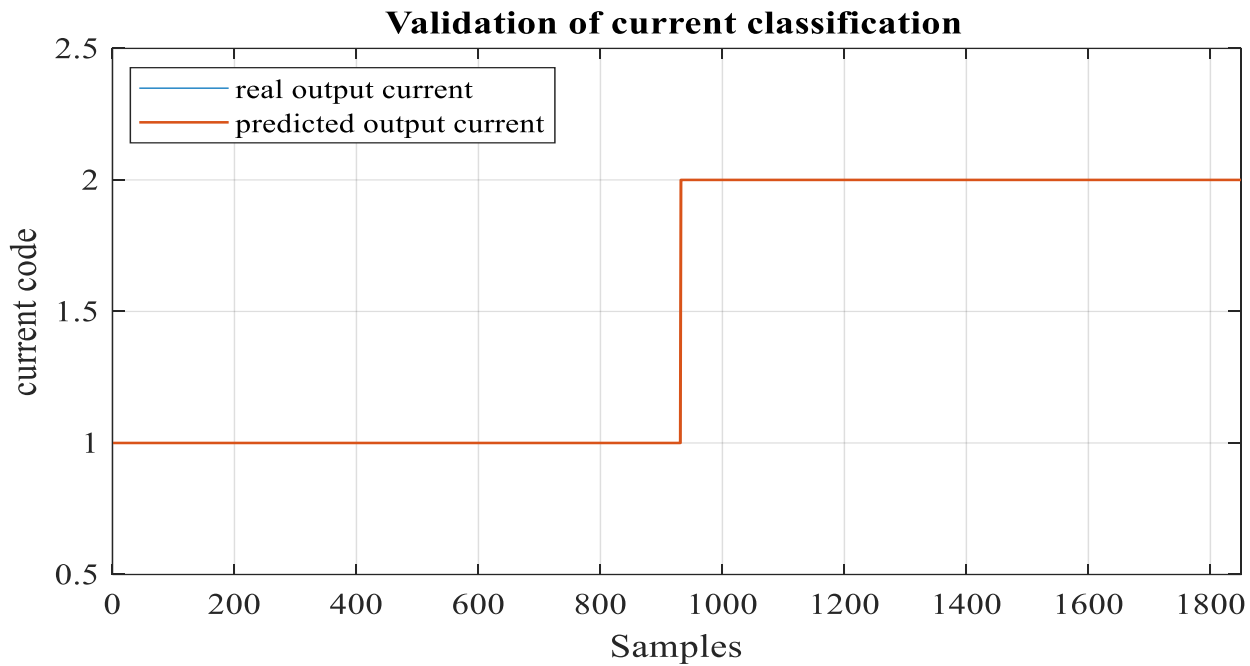


Figure II.12: the graph of validation of current classification

Figure IV.11 and figure IV.12 gives us the perfect results were the test data equals the predicted data. We can't say that KNN classifier always produces the results of 100% accuracy. This is because the simulation is completely dependent and based on the given dataset. In the best case scenario like ours, the training data is equal to the test data hence not much distinction is given by the real output data and the predicted output data.

The Testing Phase

Input data

The test Phase for Temperature and solar irradiation

The following are the graphs obtained from the validated data for testing (Temperature and solar irradiation):

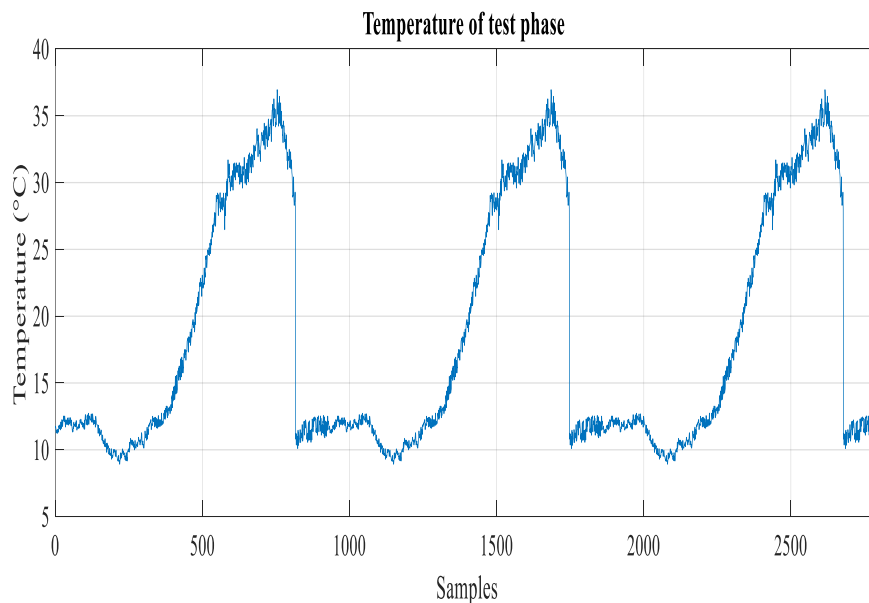


Figure II.13: the graph of temperature under testing phase

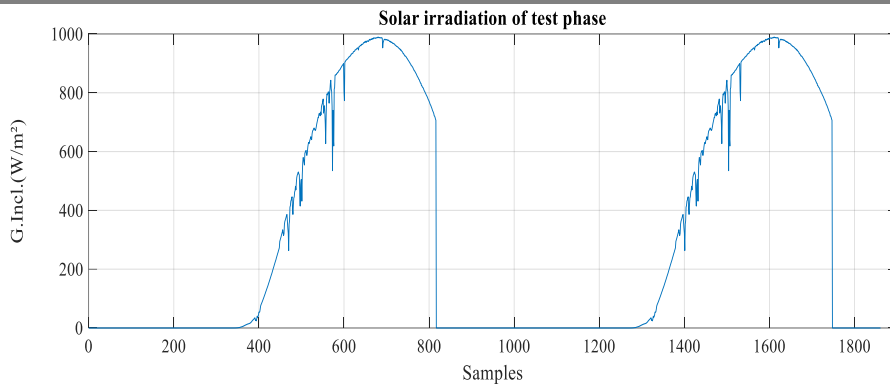


Figure II.14: the graph of solar irradiation in test data

As in any good sense of project making, after training and validating the project in question what follows is test. This is important because it answers the question of whether we have succeeded or not. In our humble case, we have truly succeeded in creating a working KNN script program because as shown by the two graphs of solar irradiation and current, these are the same graphs that can be obtained by measuring instruments of these physical quantities when observed by an oscilloscope. We can therefore be confident that our program works.

The testing Phase for Voltage and current

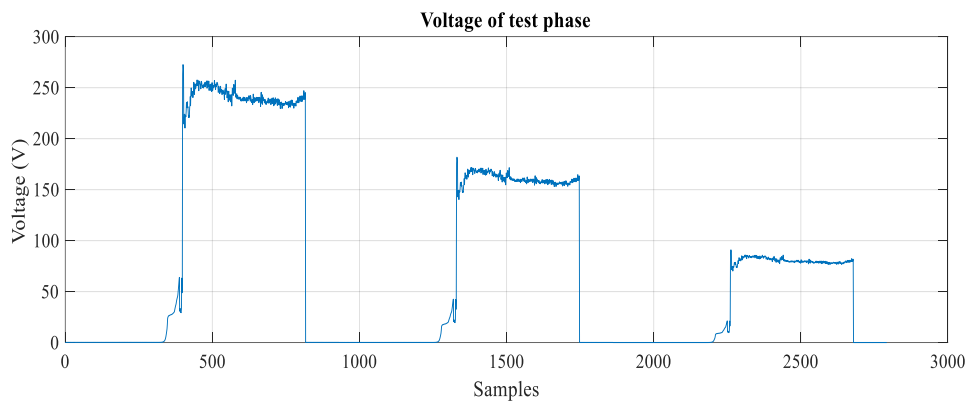


Figure II.15: the graph of voltage under test phase

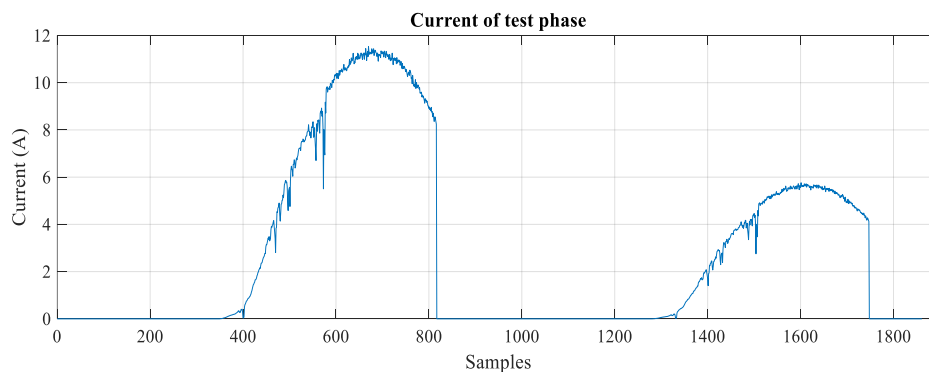


Figure II.16: the graph of current on the phase test data

The above testing phase graphs have exceptionally depicted the results of the two of the most important parameters in a PV system. The graphs show how actually voltage and temperature graphs looks like based on seven days obtained data of over 7838 values. These are the ideal graphs for PV systems exposed to the real world of perturbations such as shading faults, extreme weather, particle accumulation etc.

As in any good sense of project making, after training and validating the project in question what follows is test. This is important because it answers the question of whether we have succeeded or not. In our humble case, we have truly succeeded in creating a working KNN script program because as shown by the two graphs of solar irradiation and current, these are the same graphs that can be obtained by measuring instruments of these physical quantities when observed by an oscilloscope. We can therefore be confident that our program works.

Output data of test phase:

Output data of Voltage classification for test phase:

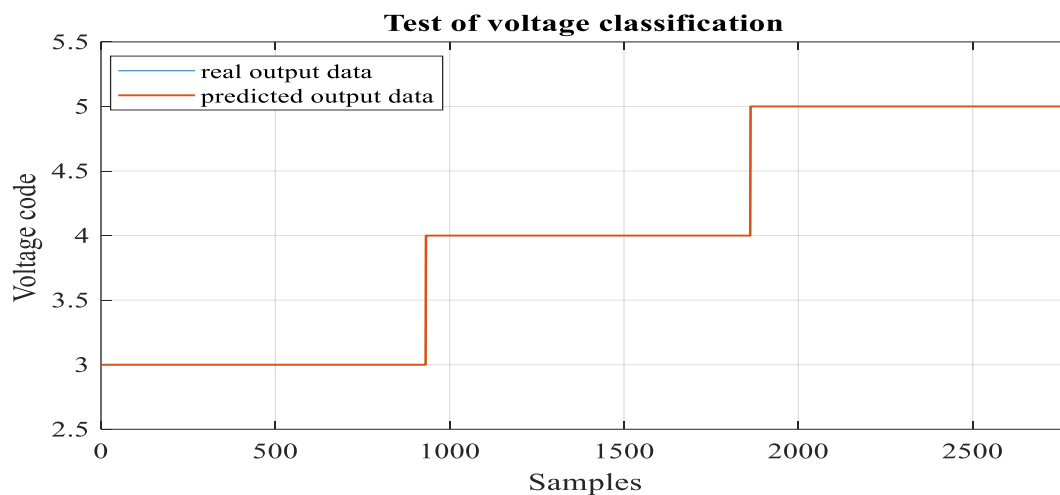


Figure II.17: the graph of test of voltage classification

In the figure IV.17, we notice the thin line between the voltage and the temperature. At the beginning of the week, the values obtained belonged to code 3 until they stabilized on code 5 before slowing dropping completely on code 3. So we see that on 2500 samples, the code became constant on three. This represents the time when the temperature is below optimum and therefore the voltage output is less.

Output data of current classification for test phase:

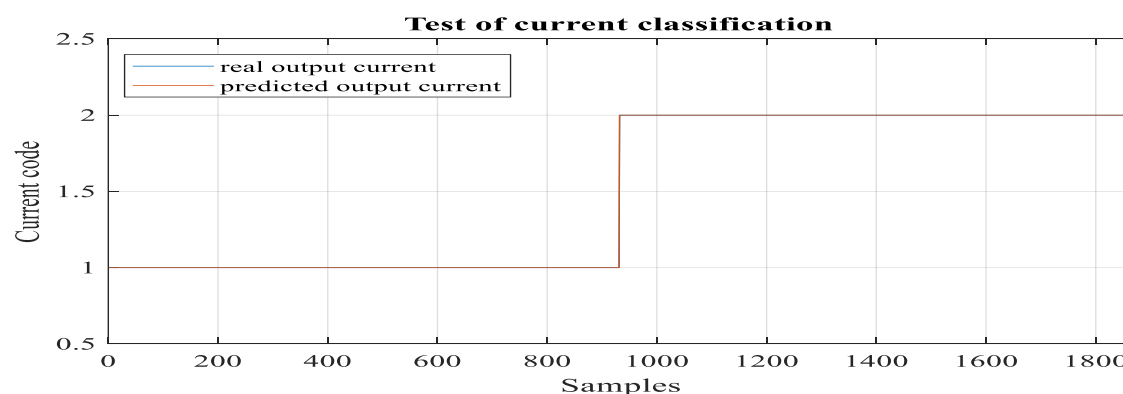


Figure II.18: the graph of test of current classification

In our final figure of Figure IV.18, we notice that our range of samples fluctuated between code 1 and code 2. This indicates a steady rise of current with respect to solar irradiation. When solar radiation becomes constant during the course of the day, the current follows the pattern and becomes constant as well. As earlier explained, it is the photons in the solar radiation that excites the electrons in the semi-conductor to create a potential difference which leads to the circulation of current. The in changing the solar irradiation we equally change the current and this is shown by our graph, the current follows the pattern of solar irradiation between code 1 and 2.

Output of KNN program:

Output KNN of Voltage classification for training phase:

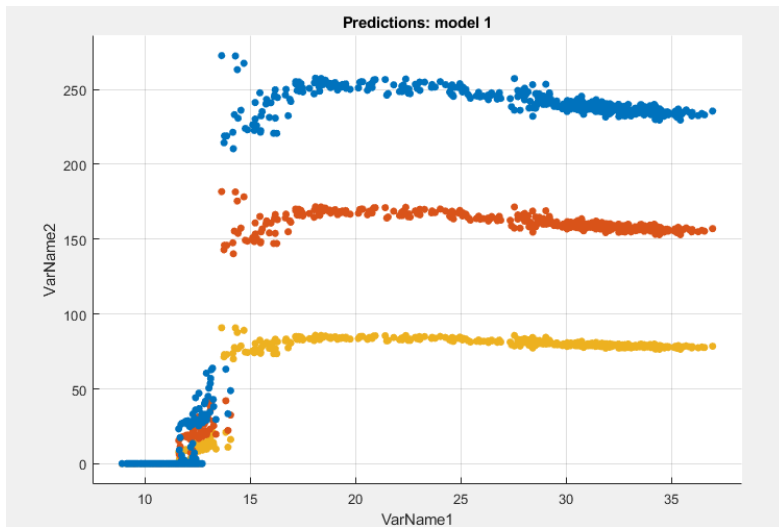


Figure II.19: Training KNN output for voltage classification with scatter plot

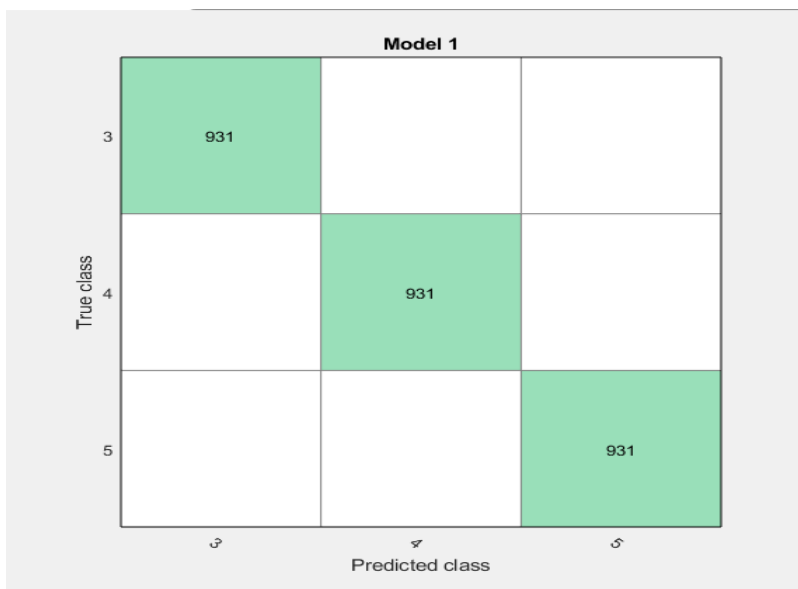


Figure II.20: Training KNN output for voltage classification with confusion matrix

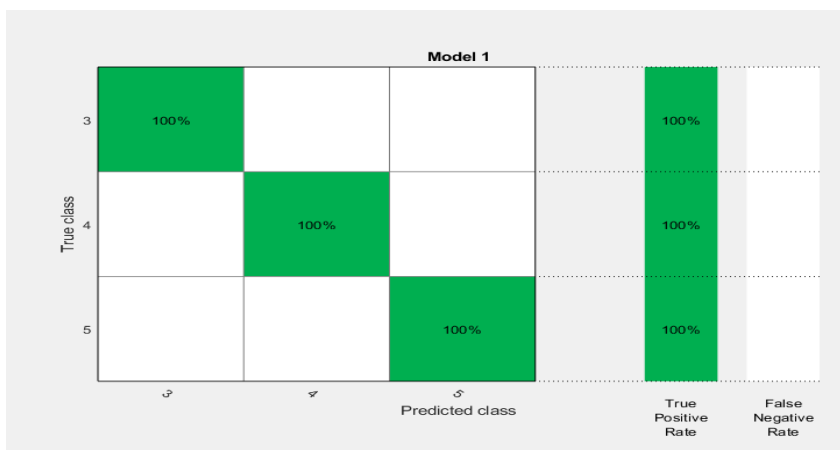


Figure II.21: Training KNN output for voltage classification with confusion matrix using true positive rate plot

When conducting any experiment, on charges, individual atoms or electrical conduction within a conductor, it is therefore necessary to create a model to better understand the system we are working with. In our case, we needed a second simulation of the G-scatter which allowed us to see the distinction between samples which are in form of data and the number of these samples which are in each code by means of a confusion matrix. In figure IV.19, we see samples of three different colors representing three codes and how they are spaced.

In figure IV.20, we see the first confusion matrix for the voltage classification. Now this matrix has both the predicted class and true class. In our theoretical work, it was mentioned that the value and worth of the KNN algorithm depends on the value of K. the greater the value the more accurate the results will be. Which is why in figure IV.21 we got the 100% accuracy when we used the value of 1 for K. why you might ask? Let's use an example of a phone, imagine you have a mobile smart phone, a computer, a photocopying machine and a fax machine. And you want to know whether the mobile smart phone is a computer or a phone. We have four samples in our case (a phone, computer, photocopying machine and fax machine). So our K will be equal to 4, the number of samples. So let's begin by choosing the value of k to be one, so $k=1$ meaning we compare the smartphone to itself, so is the smartphone a computer when compared to itself or not? The answer is it is 100% a mini portable computer. Think about it, it can type, communication, and basically work in the same manner a computer does. But let's choose the value of $k=4$. We include all the samples, the photocopying machine, fax machine, mobile smartphone and the computer itself. So when we compare the smartphone to these four samples, we discover that the smartphone is nearer to the functions of the computer but not the computer itself, so its true class is nearer to the computer than the fax and photocopying machine.

And that is exactly what we have with our matrices, they show the predicted class (the class we think is the code for our values) and the real class (the code where our values really belong to) and the number of samples in each class.

Output KNN of Voltage classification for validation phase:

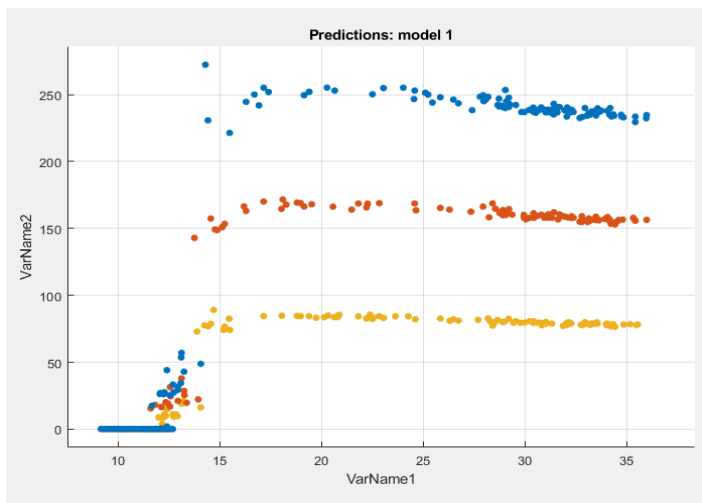


Figure II.22: Validation KNN output for voltage classification with scatter plot

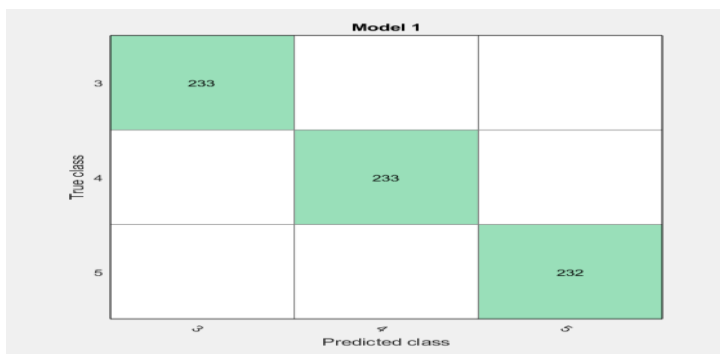


Figure II.23: Validation KNN output for voltage classification with confusion matrix

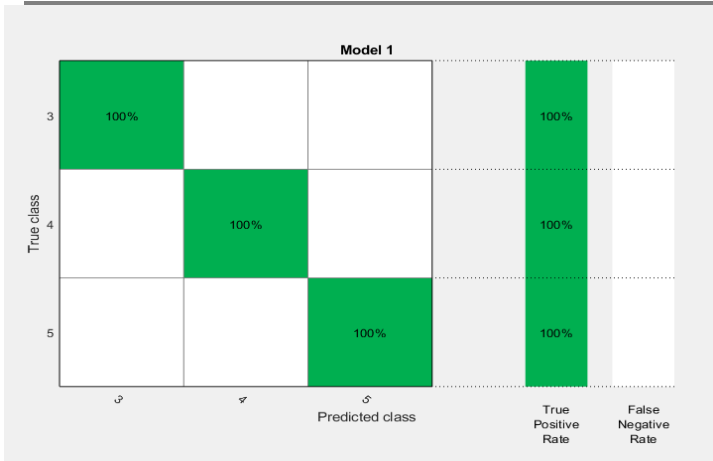


Figure II.24: Training KNN output for voltage classification with confusion matrix using true positive rate plot

In figure IV.22, we now see the scatter plot as we increased the value of k. we see that the samples are now much further to each other. The distance between each individual sample is the Euclidean distance we spoke of in chapter three. The closer the distance between samples, the more likely they are to belong to the same code as shown by the figure.

Output KNN of Voltage classification for test phase:

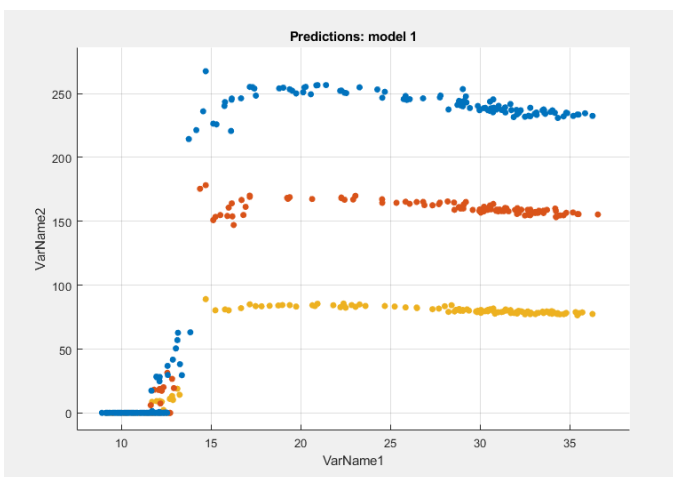


Figure II.25: Testing KNN output for voltage classification with scatter plot



Figure II.26: testing KNN output for voltage classification with confusion matrix

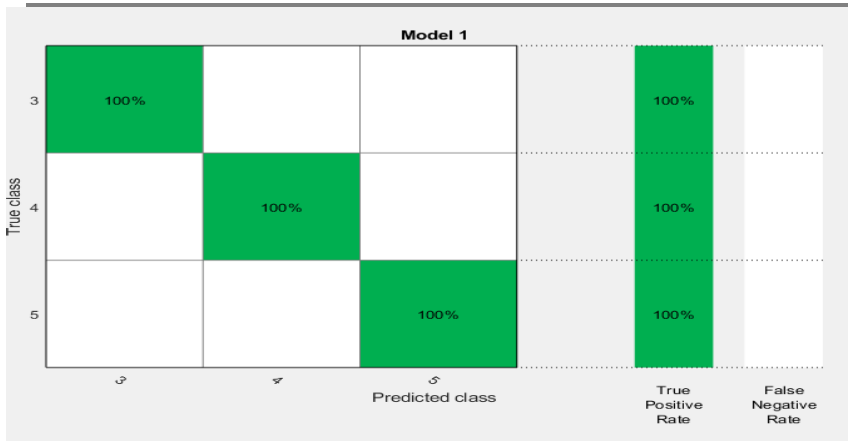


Figure II.27: Testing KNN output for voltage classification with confusion matrix using true positive rate plot

Now we consider the results of the voltage classification in the test phase. In figure IV.25 we see the separation of codes in forms of samples of different colors. Despite the fine distinction, we see that some samples are closer to another code than where their fellow samples are (samples of the same color code). For example, between 0 and 50 samples, we see that some bleu samples true class is that of the yellow samples than their bleu counterparts.

Figure IV.26 gives us the confusion matrix of the number of samples in each code and figure IV.27 illustrates the accuracy of the classification between the real and predicted classes.

Output KNN of current classification for training phase:

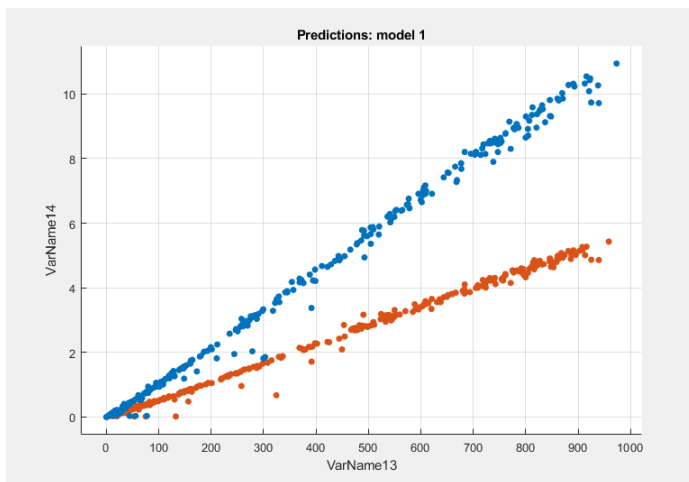


Figure II.28: the output of the current classification plot for the training data

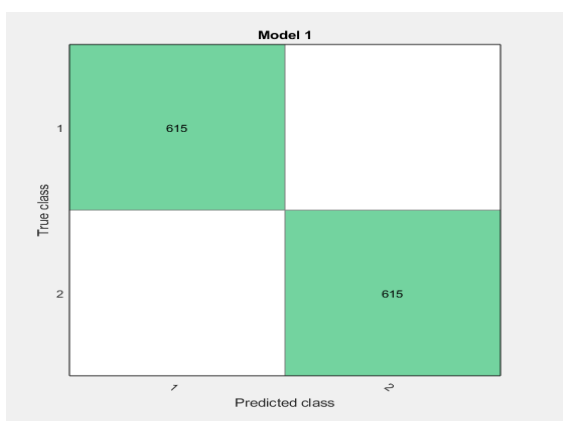


Figure II.29: the output confusion matrix for current training data

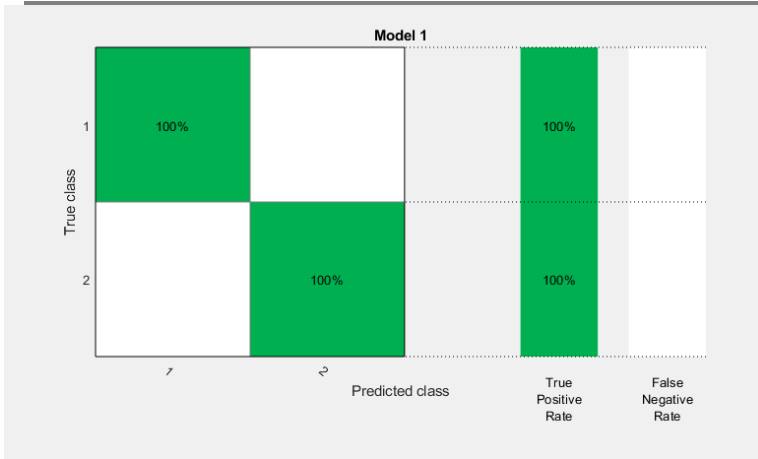


Figure II.30: the accuracy of the confusion data matrix for the current training phase

The parameters of interest for our work are of course voltage and current. In figure IV.28, IV.29 and IV.30, we notice that we are only associated with two codes, code 1 and 2. In figure IV.28 we see the separation of codes between these two samples and figure IV.29 displays the number of samples in each code by means of a confusion matrix and finally figure IV.30 the accuracy of our prediction.

II.8.5 Output KNN of current classification for validation phase:

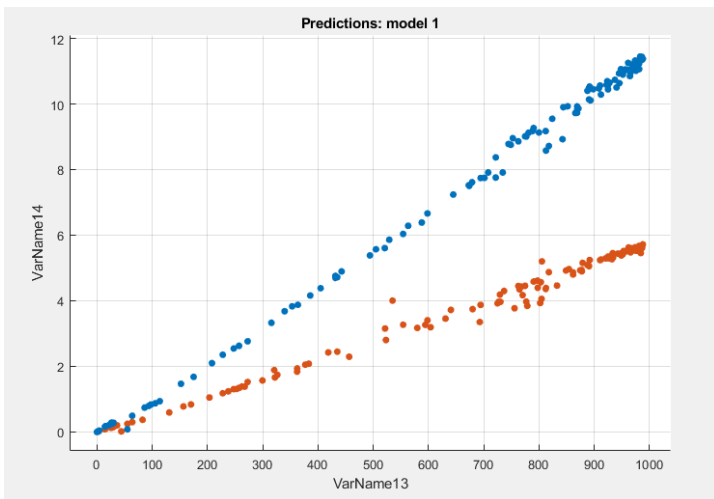


Figure II.31: the output of the current classification validation data

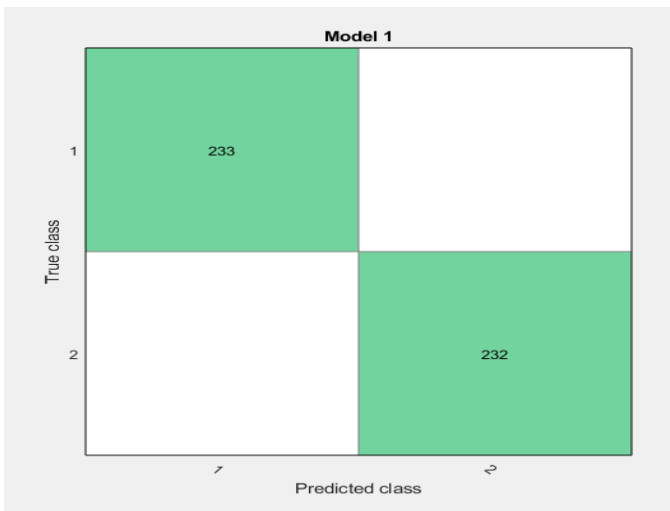


Figure II.32: the output of confusion matrix of the current classification validation data

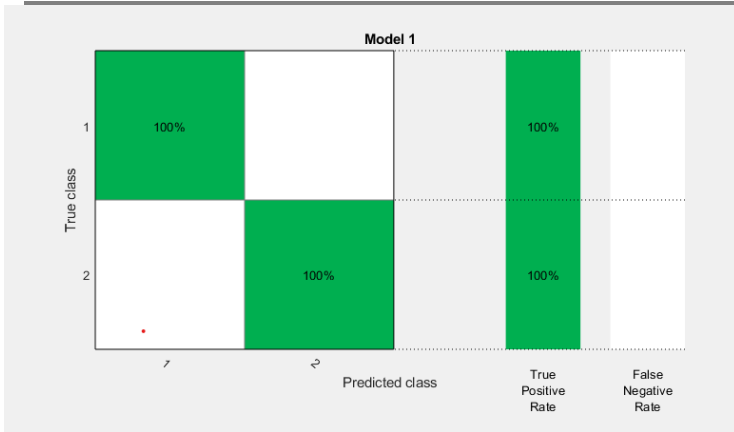


Figure II.33: the accuracy matrix for the current output of the validation data

The values of our simulation are based on real data obtained on different days which is why it is no surprising that the accuracy is high. As before only two codes (1 and 2) are of interest in the currents classification and we see the graph of samples of the two codes in IV.31 and the distribution of the number of samples in figure VI.32 and the accuracy of our validation phase.

The output Test of the current classification

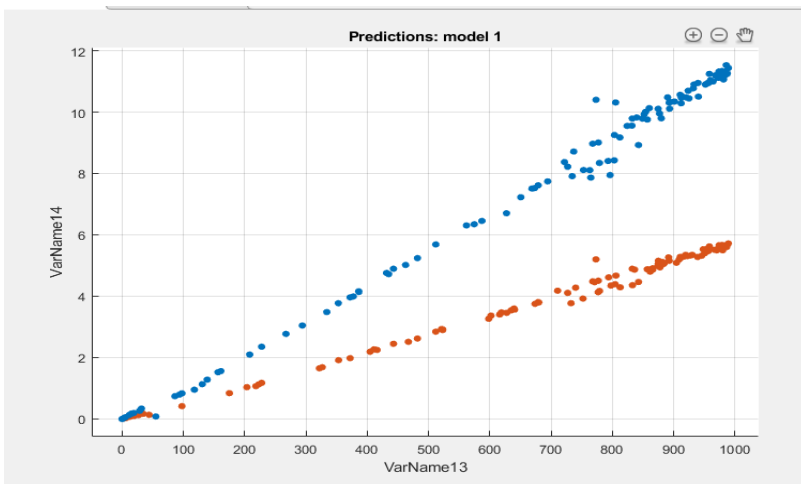


Figure II.34: the output test data of current classification

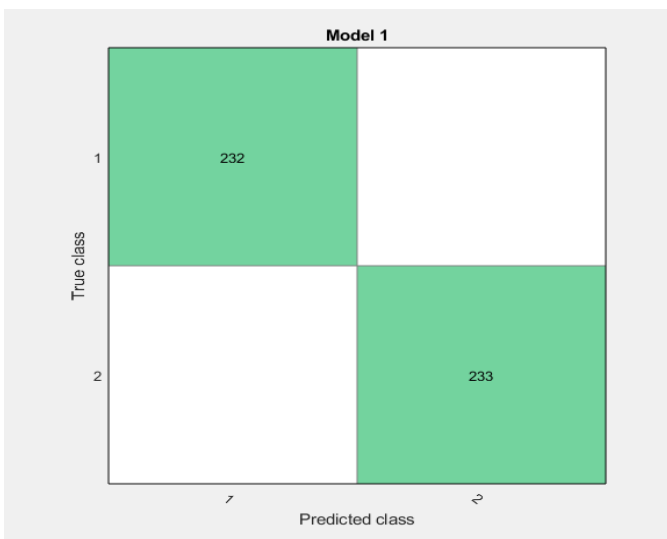


Figure II.35: the confusion matrix of the current classification of the test data

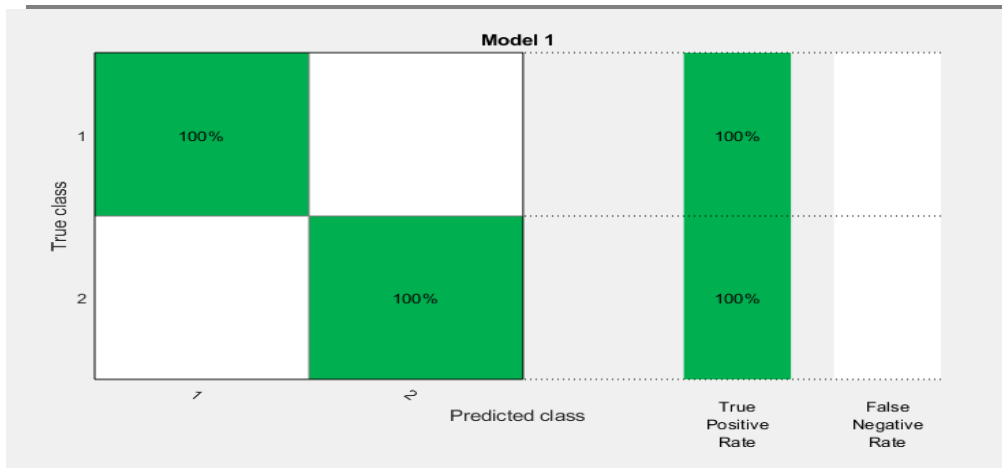


Figure II.36: the accuracy matrix of the current classification for the test data

Figure IV.34, IV.35, IV.36 shows the final graphs for the model of the output for the current classification. In our work for the output classification, we showed the graphs of samples, the number of samples (through the confusion matrix) and the accuracy. We were of course putting into consideration the distance and the k value to maximize the results and accuracy of our model.

GENERAL CONCLUSION

Renewable energy is one of the most interesting and exciting fields of modern science and engineering. As the world population increases, the demand for energy becomes more and more challenging. Renewable energy seems to be the partial answer to that challenge hence why it's imperative to consider the best method of detecting problems in solar panels considering they are the most widely used source of renewable energy.

In this Article, we briefly talked about the different types of Photovoltaic Faults and proceeded to elaborate more on the problems usually faced with engineers and technicians during maintenance of these systems. Which is why we further elaborated on how we can categorize and detect PV faults using data science and one of Artificial Intelligent (AI) algorithms KNN.

In the near future we hope to use more data science algorithms and compare them with each other to see which one gives the highest efficiency in detecting faults not just in PV systems but in all energy systems.

REFERENCES

1. An analysis of PV solar electrification on rural live hood transformation A case of Kisiju-Pwani in Mkuranga District, Tanzania page 18.
2. Ministère de l'Énergie et des Mines Algérie, Guide des Énergies Renouvelables, Édition 2007.
3. M.Boukli-Hacene Omar. Conception et réalisation d'un générateur photovoltaïque muni d'un convertisseur MPPT pour une meilleure gestion énergétique. Mémoire de magister. Université ABOU BAKR BELKAID de Tlemcen, 2011.
4. Long Bun, Détection et localisation de défauts dans un système photovoltaïque, Thèse de doctorat, Université de Grenoble, Novembre 2011.
5. S. Spataru, D. Sera, T. Kerekes, R. Teodorescu, Diagnostic method for Photovoltaic systems based on light I-V measurements, Solar Energy, Elsevier, Vol. 119, 2015, p. 29-43.
6. S.Saravanan, R. S. Kumar, A. Prakash, T. Chinnadurai, R. Tiwari, N. Prabakaran, et al., "Photovoltaic array reconfiguration to extract maximum power under partially shaded conditions", in Distributed Energy Resources in Microgrids. 2019, Elsevier. p. 225-241.
7. H.Bouzeria. Modélisation et commande d'une chaîne de conversion photovoltaïque. Thèse de Doctorat, Université de Batna 2 -Batna-, 2016.
8. Reference solar spectral irradiance: ASTM G-173, ASTM (American Society for Testing and Materials), IEEE Press, New York, 2000 .

9. Arani, M.S.; Hejazi, M.A. The comprehensive study of electrical faults in PV arrays. *J. Electr. Comput. Eng.* 2016.
10. M. K. Alam, F. H. Khan, J. Johnson, J. Flicker, PV arc-fault detection using spread spectrum time domain reflectometry (SSTDTR). in 2014 IEEE energy conversion congress and exposition (ECCE). 2014. IEEE.
11. D. Sera. Real-time modelling, diagnostics and optimised MPPT for residential PV systems. Doctoral Thesis, Institute of energy technology-Alborg university, Denmark, 2009.
12. Firth, S.K.; Lomas, K.J.; Rees, S.J. A simple model of PV system performance and its use in fault detection. *Sol. Energy* 2010, 84, 624–635.
13. V. Tamrakar, S. Gupta, Y. Sawle. Single-diode and two-diode PV cell modeling using Matlab for studying characteristics of solar cell under varying conditions. *Electrical & Computer Engineering: An International Journal*. Vol 4, p. 67-79, 2015;
14. W. Rezgui. Système intégré pour la supervision et le diagnostic des défauts dans les systèmes de production d'énergies: les installations photovoltaïque. Thèse de Doctorat, Université de Batna 2 - Batna-, 2015.
15. S. R. Madeti, S. N. Singh. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Solar Energy*. Vol 158, p. 172-185, 2017;
16. R. Platon, J. Martel, N. Woodruff, T. Chau. Online fault detection in PV systems. *IEEE Transactions on Sustainable Energy*. Vol 6, p. 1200-1207, 2015;
17. Cherifa K, Khelil M., Amrouche B, A. S. Benyoucef, Kamel K, Aissa C, New Intelligent Fault Diagnosis (IFD) approach for grid-connected photovoltaic systems, p 30
18. T. Pei, X. Hao. A Fault Detection Method for Photovoltaic Systems Based on Voltage and Current Observation and Evaluation. *Energies*. Vol 12, p. 1712, 2019;
19. Fault detection and diagnosis of photovoltaic system using fuzzy logic control January 2019 E3S Web of Conferences 107(5):02001 DOI:10.1051/e3sconf/201910702001 License CC BY Authors: Zaki K, Cairo University, Hong Lu Zhu, North China Electric Power University
20. Mellit A, Benghanem M, Hadj Arab A, Guessoum A, Modeling of sizing the photovoltaic system parameters using artificial neural network. In: *Proceedings of IEEE conference on control application*, p 3, 2003.
21. Brett L. *Machine Learning with R* second edition p64-78
22. Harsh B., Surbhi B *Application of Genetic Algorithms in Machine learning*, Uttar Pradesh, India.
23. Vikramaditya J, *Tutorial on Support Vector Machine (SVM)*, School of EECS, Washington State University, Pullman 99164.
24. Jerry Z, *Machine Learning: Decision Trees CS540* University of Wisconsin-Madison
25. T. Pei, X. Hao. A Fault Detection Method for Photovoltaic Systems Based on Voltage and Current Observation and Evaluation. *Energies*. Vol 12, p. 1712, 2019;
26. Christopher M B 1997 Bayesian Neural Networks *Journal of Brazilian Computer Society* 10 1590.
27. G. M. Tina, F. Cosentino, C. Ventura, Monitoring and diagnostics of photovoltaic power plants. In *Renewable Energy in the Service of Mankind Vol II*. 2016, Springer. p 503-516.
28. Mahmoud D, *Fault Detection and Performance Analysis of Photovoltaic Installations* by A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy p23-30.