An-Najah National University

Faculty of Graduate Studies

Predicting I-V curve for photovoltaic modules using Random Forests Technique

By

Areej Ahmad Alia

Supervisor

Dr. Tamer Khatib

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Clean Energy Conservation Engineering, Faculty of Graduate Studies, An Najah National University, Nablus-Palestine.

2020

Predicting I-V curve for photovoltaic modules using Random Forests Technique

By

Areej Ahmad Alia

This Thesis was defended successfully on 28 /7/2020 and approved by:

<u>Defense Committee Members</u>	<u>Signature</u>
– Dr. Tamer Khatib /Supervisor	....
– Dr. Fouad Zaro /External Examiner	....
– Dr. Emad Al-Natsheh /Internal Examiner	....

Dedication

To those stars that have given me the path to progress and success ... my beloved father ... my beloved mother...

To my dear son Wesam and my wonderful husband who supported me in every step I take in my scientific career....

To my dear brothers and sisters ... who always support me...

To those hidden stars that appear to give us the path when we need it.... To everyone who loved me and wanted me progress and success...

I dedicate this work to you...

Acknowledgment

First of all, I thank "God", who has helped me in my scientific career and to carry out this practical research.

I would like to thank my supervisor, Dr. Tamer Al-Khatib, for the valuable supervision and continuous support throughout the period of work on this research.

Also, I would like to thank all my professors in the Department of Energy Engineering and energy conservation at An-Najah University, and in particular I thank Dr. Muhammad Al-Sayed for all the information he provided to me throughout the period of my studies at the university.

Last but not least, I thank my boss, Mr. Bassam Al-Qasem and Mrs. Salwa Al-Khatib for their support to me throughout my university studies.

الإقرار

أنا الموقعة أدناه ، مقدمة الرسالة التي تحمل العنوان:

Predicting I-V curve for photovoltaic modules using Random Forests Technique

أقر بأن ما شملته هذه الرسالة إنّما هو نتاج جهدي الخاص، باستثناء ما تمّت الإشارة إليه حيثما ورد، وأنّ هذه الرسالة ككل، أو أيّ جزء منها لم يقدّم من قبل لنيل أيّ درجة أو لقب علميّ لدى أيّ مؤسسة تعليمية أو بحثية أخرى.

Declaration

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's Name: Areej Ahmad Alia **اسم الطالبة:**

Signature: **التوقيع:**

Date: 28 /7/2020 **التاريخ:**

List of Abbreviations

GHG	greenhouse gas
PV	photovoltaic
RFs	Random Forests
NOCT	Nominal operating cell temperature
FF	Fill factor
MPPT	Maximum power point
LM	Levenberg–Marquardt Method
NRM	Newton–Raphson method
AI	Artificial Intelligence
ANN	Artificial Neural Network
MBE	Mean bias error
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
P_{max}	Rated Maximum Power
V_{oc}	Open Circuit voltage
I_{sc}	Short Circuit Current
V_{mp}	Maximum Voltage
I_{mp}	Maximum Current
k	Kelven
W/m^2	Watt per meter square
C°	Degree Celsius

Table of Contents

List of Tables

List of Figures

Predicting I-V curve for photovoltaic modules using Random Forests Technique

By
Areej Ahmad Alia
Supervisor
Dr. Tamer Khatib

Abstract

The study of the special curves of solar PV modules are of great importance in developing cells and increasing their capacity, hence the idea of this thesis that was intended to predict the current-voltage curve for solar PV module by developing a new developed model that relies on random forest technique in training and testing data using MATLAB program.

The random forest technique is a machine learning method, where this technique relies on decision trees (classification trees, regression trees). The regression trees were released in the new proposed model to predict the output variable (PV module output current), depending on a set of inputs represented by five parameters (ambient temperature, solar radiation, PV DC voltage, short circuit current, and open circuit voltage). This data sets were obtained by conducting several experiments on a (STF - 120P6) poly-crystalline PV module 14.0% module efficiency. Seven experiments were done on the (STF - 120P6) PV module in different values of cell temperatures and solar radiations to measure the current and the voltages by using the (I-V CURVE TRACER DEVICE).

Through training and testing these data, high accuracy results were obtained for the proposed model, where the metric error values (RMSE, MAPE, and MBE), which are equal (0.04251%, 4.315097%, -0.3959%), respectively. A

value of (MAPE) was used to evaluate this model and compare it with previous models that adopted different methods to predict the I-V curve of the solar PV module. These methods can be classified into offline methods and online methods. Online methods depend on real devices to extract the I-V curve as (capacitor, resistor, inductor, and switches), Offline methods represented in Artificial Intelligence methods, Random Forests technique, and the numerical methods where are used to obtain numerical solutions of a mathematical problem such as Levenberg–Marquardt method(LM), Newton–Raphson method (NRM),

Chapter One

Introduction

1.1 Background

Climate change is the most common phenomenon in these and next days, and is one of the most controversial environmental phenomena and the most widely circulated around the world because of its negative effects on human and organisms. The most negative effects of it is the increasing proportion of greenhouse gas emissions (GHG) especially carbon gas.

Scientists believe that this increase is the result of the industrial revolution that began in 1888. The concentration of carbon dioxide in the atmosphere has been estimated at 380 parts per million since the beginning of the 21st century. (Karzm.J 2012)

This increase is a serious indication of the constant increase in global temperatures estimated by scientists (1.4 - 5.8) for the period (1990 - 2100), which will lead to melting snow at the poles and rising water level in the oceans and sinking many cities and coastal countries. (Karzm.J, 2012)

Electric power generation is the most productive of greenhouse gases due to the use of fossil fuels (oil, coal, gas) in power plants. Where primary energy consumption grew at a rate of 2.9% last year, almost double its 10-year average of 1.5% per year, and the fastest since 2010.The Carbon emissions grew by 2.0%, the fastest growth in seven years. By fuel, energy consumption growth was driven by natural gas, which contributed more than

40% of the increase. (BP Statistical Review of World Energy2019 | 68th edition).

BP Statistical Review of World Energy Report (2019 | 68th edition) shows that the growth of the renewable power grew by 14.5%, slightly below its historical average, although its increase in energy terms (71 mtoe) was close to the record-breaking increase of 2017. Solar generation grew by 30 mtoe, wind (32 mtoe), and provided more than 40% of renewable growth, Hydroelectric generation increased by an above-average 3.1%, and Nuclear energy rose by 2.4%, its fastest growth since 2010.

- **Solar Energy**

  Solar energy is one of the most important sources rich in the huge amount of energy that is sufficient to meet the world's needs of electric energy and all the human activities that it needs. Where it can be used anywhere by receiving solar radiation. But there are many factors that affect the amount of energy that can Harness it for electrical energy or heating purposes such as geographical location, time of day, and weather conditions. Solar energy can be captured for electricity production using a solar or photovoltaic cell, which is a technology that converts solar radiation into electricity using the photoelectric effect. Typically, photovoltaic cells are located on the roofs of residential, commercial, or industrial buildings to cover the electrical energy needs of these facilities or a portion of them. In addition, many countries built large photovoltaic plants (greater than 100 megawatts) that require large areas depending on the technologies used to be an additional solution to obtain electrical energy. Also, part of these stations was dependent on the technology of concentrating solar energy (CSP), which uses lenses or mirrors to focus sunlight in a narrow beam that heats the liquid, to produces steam then drive the turbine that generates electricity. CSP projects are larger than residential or commercial PV projects and are often owned and operated by electrical facilities. The figures below show the last update of information for global renewable energy consumption, the solar energy consumption by the region, and the installed solar photovoltaic (PV) capacity, respectively.

(https://ourworldindata.org/renewable-energy).

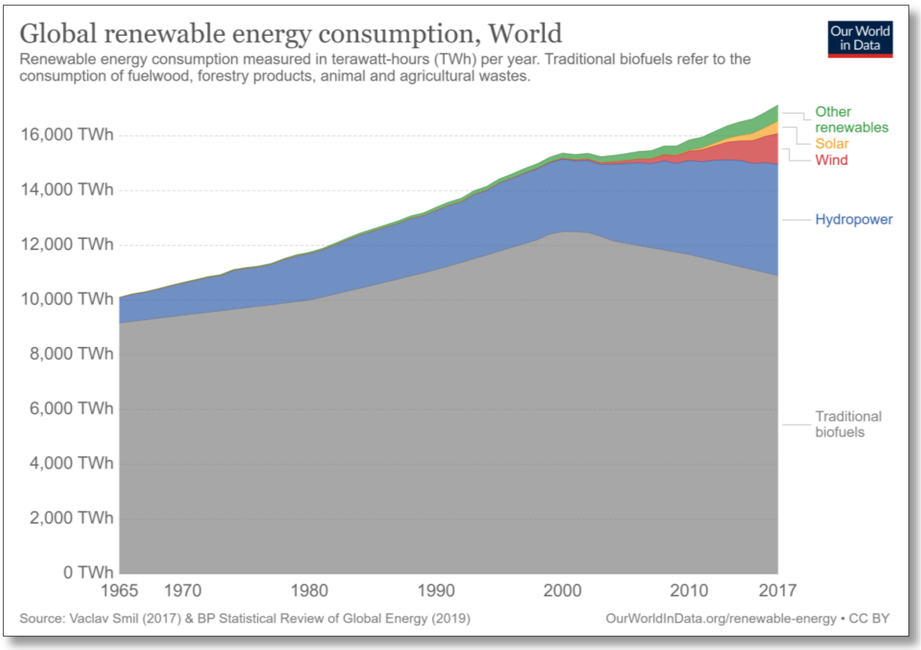


Figure 1.1: The global renewable energy consumption (1965 -2017)

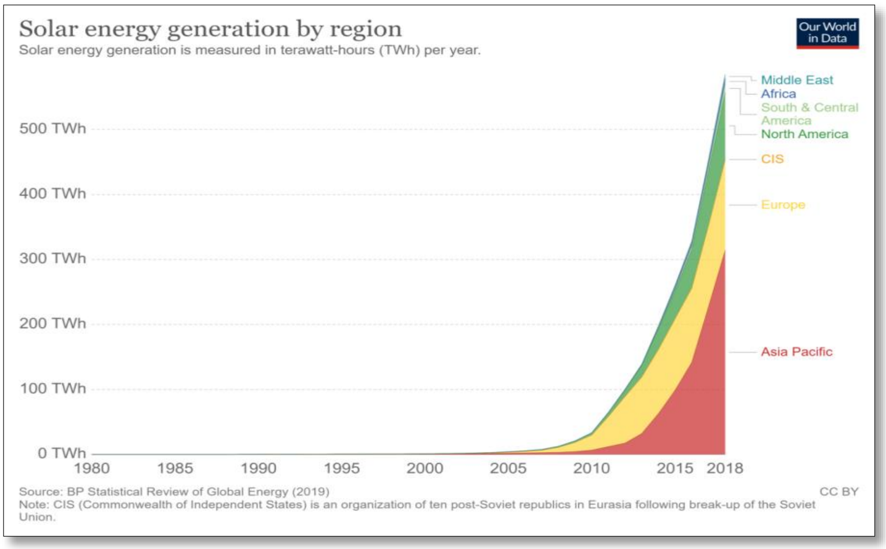


Figure 1.2: The solar energy consumption by the region

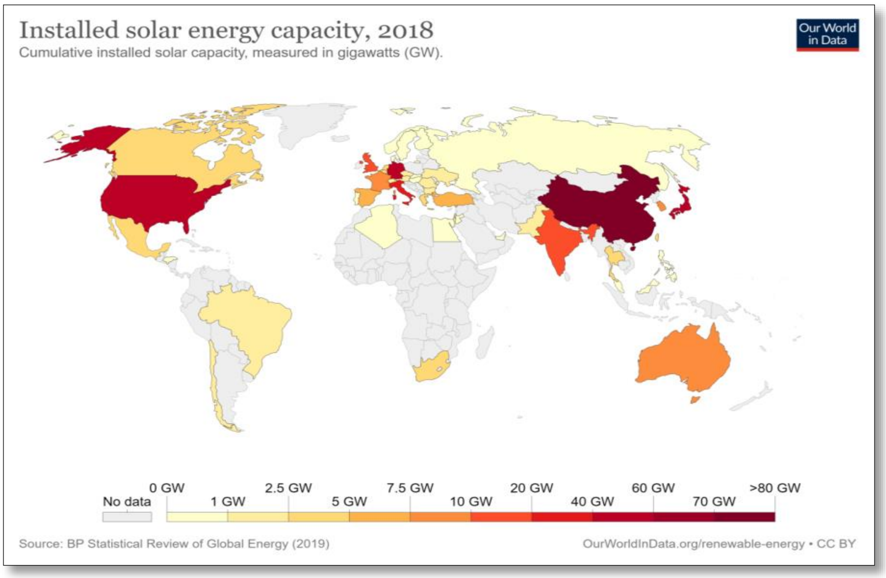


Figure 1.3: The installed solar photovoltaic (PV) capacity

- **Solar photovoltaic system**

  Solar photovoltaic systems are one of the most important sources of renewable energy after increasing interest in renewable energy in general and solar energy in particular, where Solar photovoltaic systems have been able to convert various buildings from energy-consuming buildings to energy-producing buildings due to their ability to generate electrical energy and use it directly through a one stage by converting the DC current from the solar PV modules into AC current through the electrical inverter and use it directly to supply the various buildings with the electrical current needed to cover their needs. Also photovoltaic systems are one of the simplest types of modern technology that do not need many complex stages to obtain the electrical current .and the most important features of solar energy systems: this systems have a long operating life around 30 operating years, low negative impact on the environment, the maintenance of these systems is simple and low cost compared to traditional systems for generating electric energy such as (oil, natural gas,...), and these systems can be used all over the world, whether in the regions with high levels of solar radiation or in the regions with a few sunshine hours.

- **Photovoltaic system component**

  Photovoltaic systems are divided into three main types which are: on grid system, off grid system, and hybrid system. Which the main components for it are:

  - PV modules

  - Inverter

  -Batteries (for off grid, and hybrid systems)

  -Charge controller (for off grid, and hybrid systems)

For all types of photovoltaic systems, the photovoltaic modules are the main component of the greatest importance, being the part responsible for converting the solar radiation into an electric current. And the system depends on the conversion efficiency for PV modules to supply the buildings with electrical current. In terms of the importance of PV modules; a study has started on improving and predicted the I-V curve using many techniques as (online method and off line method). In online methods, the real devices as (resistors, capacitors,) were used to measure the current and the voltage values for PV module or solar cell to extract the I-V curve by construct a specific electric circuit. In offline methods, the artificial intelligence (AI) or empirical mathematical methods were used to extraction of I-V curve depend on historical experimental data.

For this thesis, The Offline method which represent by Random Forest technique was used to predicted the I-V curve .A random forest is one of the method of artificial intelligence to analyze data and extract useful information .This technique relies on decision trees (classification trees and regression trees), which are a non-supervised learning method that aims to create a model for predicting the value of a target variable by learning simple decision rules extracted from the data features.

1.2 problem statement

As a result of the rapid development in the production of solar cells and the urgent need to use them to reduce the use of traditional methods such as (oil, coal ...) in the production of electrical energy, and the ability of PV systems to cover the energy demand for many buildings in various sectors , It is necessary to intensify studies related to the solar cell and its performance, especially the study and analysis of current-voltage curves for the PV modules ; to obtain high-precision curves during practical experiments.

During previous studies and experiments, some of which were mentioned in this work, extracting I-V curves for the PV modules had some determinants such as the inability to obtain the value of the short circuit current and open circuit voltage values when using offline method as using resistors, and the cost of using off-line line methods are high compared to the online methods. And also, some online methods require a large amount of data to analyze and extract I-V curves. Hence the idea of using a new method to extract the current voltage curve of the PV unit, based on the technique of random forest; To increase the accuracy of the results and compare them to the previous methods, Where the random forest technique is characterized by the following:

- Easy of interpretation and the ability to process data that interacts in a non-linear or hierarchical manner.
- It is classified as a modern technology that has gained great importance in various scientific fields, including solar cells, but has been used in a few special studies in extracting solar cell curves.
- The ability to deal with a small number of data and analyze it to get the required results.

1.3 objectives

The main objectives of this thesis are to predict the current voltage curve to predict the performance of the PV module by using random forest technology (RFs) which is one of the proposed new technologies to develop a new model to predict the current output of the PV module to overcome limitations of previous optimization methods. Therefore, the objectives of the research work are described as follows:

- -To develop a new prediction technique for predicting the output current of a PV module using Random Forests (RFs) technique.
- To validate the obtained results from the proposed RFs model by comparing them with the previous research works to show the accuracy of the proposed model.

1.4 Methodology and Scope of work

The methodology of this research is starts by Define the location in Malaysia where the experiments were made, then define the type of PV module was used in the study (STP -120Wp) .after that collected the data for the inputs parameter, where it represented in (ambient temperature, solar radiation, PV DC voltage, short circuit current, and open circuit voltage) by using I-V400 - SOLAR I-V analyzer , in different solar radiation and cell temperatures, to developed a new RFs model by training and testing the data . Finally, a validation for the obtained results by comparing them with other previous research works to show the accuracy of the proposed models. The outline of research methodology represents in figure 1.4.

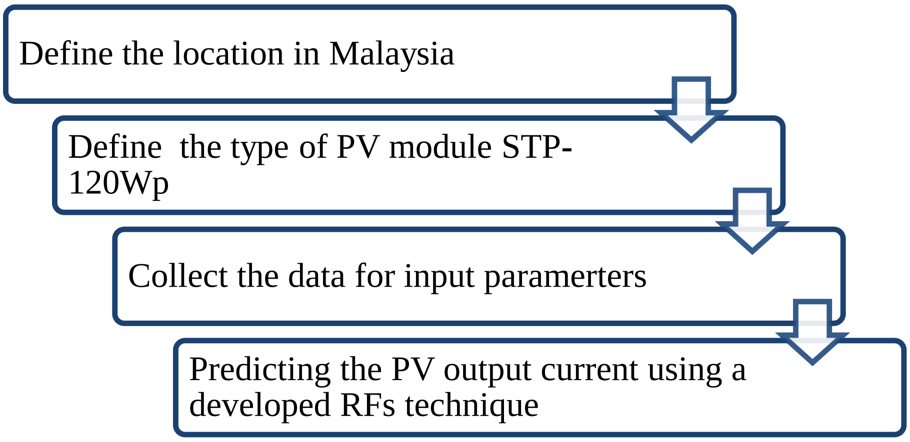


Figure 1.4: The outline of research methodology

Chapter II

Literature review

2.1 Introduction

solar energy (called photons) is converted into electrical energy in a process called the photovoltaic effect, effective devices are used and the main component of this process are solar cells that have semiconductor properties encapsulated inside a material to protect them from the environment. It requires the development of many semiconductor processing technologies to manufacture them at low cost and high efficiency. Three types of PV systems are used: Grid-connected PV system, stand-alone PV system, hybrid PV system. Grid- connected PV system: produced electrical power by converted the direct solar radiation and injected it to grid without storage. Stand-alone PV system: A system that is designed to work independently of the grid and uses a battery to store excess energy and use it when needed.

This chapter will emphasize the characteristics of PV module in order to pave the way for predicting I-V curve in the following chapters.

2.2 Solar Cells

The solar photovoltaic system is the energy generator whose production depends on the solar cells as main component of it, these cells connect with each other in series and in parallel combination, to form the required power for a PV module, and then these PV modules are connected as well in series to form the PV array then multi PV arrays connected in parallel to achieve the required capacity of the PV system.

The Solar cells can be classified into three generations; the first generation of solar cells is characterized by being a high-quality single junction device because it contains a large area. This generation includes high energy and labor inputs which prevent any significant progress in reducing production costs. The second-generation cells are a mixture of silicon with other materials that increase the production efficiency for it, to be able to achieve the required energy. Much types of materials have made great progress in the efficiency of the cells as calcium telluride (CdTe) and silicon copper indium gallium and amorphous silicon and formed silicon. Where these materials were used in thin film substrate such as glass or ceramic, which reduced the mass of materials and thus reduce production costs. The third-generation cells have been developed to be more productive than previous generation cells, by using some modern technologies such as multi-junction photovoltaic cell, modifying incident spectrum (concentration), and Use of excess thermal generation to enhance voltages or carrier collection. (Tiwari, M.N, and Dubey, S., 2010)

2.3 Photovoltaic Module

Photovoltaic modules are one of the latest technologies used to generate electricity in a clean, quiet and reliable way for a long time. Photovoltaic system consists of a light-cell, and devices convert radiation directly into electricity, where the sun is usually that energy source. PV cell made of at least two layers of semiconductor material, one of them has a negative charge, and the other positive. When the radiation reaches the surface of the cell, the semiconductor atoms absorbed some photons to release the electrons and then passes through the external circuit and reach the positive layer to produce the electrical current (Tiwari, M.N, and Dubey, S.,2010)

2.3.1Types of photovoltaic module:

1- **Mono crystalline**: This type of solar cell made by cells cut from a single cylindrical crystal of silicon. The efficiency of it is the highest compare with other types (approximately 18-19.8% conversion of incident sun light). The manufacturing process of this cell is relatively complex which makes them slightly more expensive than others in manufacturing. Mono crystalline silicon is used in the manufacturing of high-performance solar cells. (Tiwari, M.N, and Dubey, S.,2010).

2- **Poly-crystalline silicon**: This type of solar cell made by cutting micro-fine wafers from ingots of molten and re-crystallized silicon. The efficiency of this type is less than Mono crystalline and higher than Thin film (approximately 14-16% conversion of incident sunlight). , so polycrystalline cells are cheaper in produce (Tiwari, M.N, and Dubey, S., 2010).

3- **Thin film**: These are made by depositing an ultrathin layer of photovoltaic material onto a substrate. The most common type of thin-film PV is made from the material a-Si (amorphous silicon), but numerous other materials such as CIGS (copper indium/gallium selenide) CIS (copper indium selenide), CdTe (Cadmium Teluride). The efficiency of thin film solar cell is lower than Mono-crystalline Poly-crystalline approximately (2%- 14%).(Tiwari, M.N ,and Dubey, S.,2010).

2.4 I-V Characteristic Curve

2.4.1 Solar cell modeling

The PV cells depend on the photovoltaic effect process to generate the electric current from the sun radiation. The PV cells consist of two different types of thin layer semiconductors materials (p-type and n-type) to create the P-N junction. When solar cells are exposed to solar radiation, an electric field is generated, the electrons transfer to the positive field (P-type) and the holes transfer to the negative field (N-type), and when the solar radiation photons have a suitable wavelength for these cells, the energy inherent in the solar radiation photon is transferred to an electron for the semiconductor material, which causes it to transfer to a higher energy state is known as the (conduction band). The electrons in the conduction band are free to move through matter and as a result, the electric current is generated. Figure (2.1) shows the equivalent circuit of the solar cell.

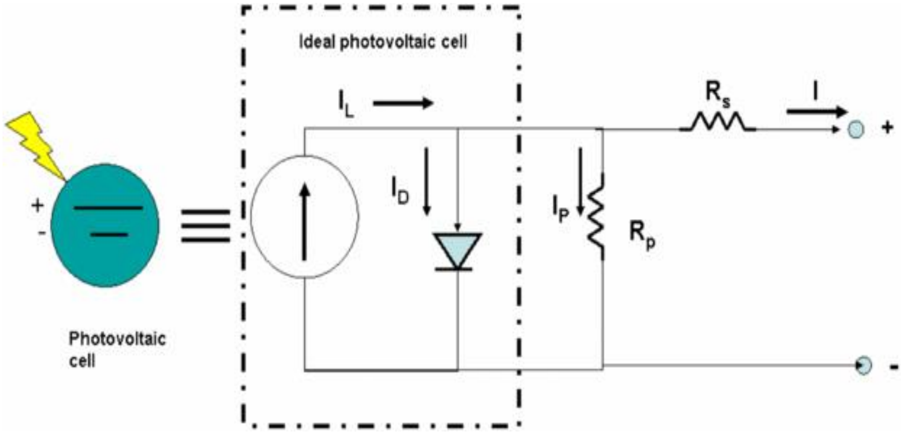


Figure 2.1: Equivalent model of a photovoltaic cell

The PV cell output current depends linearly on the light intensity, the equations (2.1), (2.2) represent the General state for PV cell output current:

$$I = I_L - I_D - I_P \tag{2.1}$$

$$I = I_L - I_o \left[\exp\left(\frac{q(V+R_s I)}{(nkT)}\right) - 1\right] - \left(\frac{V+R_s I}{R_P}\right) \tag{2.2}$$

Where I_L is the photocurrent, I_D is the Shockley diode equation, I_o is the reverse saturation current of the diode (A), q is the electron charge ($1.6 * 10^{-19}$C), k is the Boltzmann constant ($1.38 * 10^{-23}$J/K), T is the temperature of the p-n junction (K), n is ideality factor of the diode, R_P is the shunt resistor in(Ω),R_s is the series resistor in (Ω).(Tiwari, M.N, and Dubey, S.,2010)
The PV cell output current affected by the variation of solar radiation, cell temperature, series resistor, Shunt resistor, and reverse saturation current of the diode. The equations in following points represent the effected of solar PV cell model. These equations are quoted from (Salmi, T. et al, 2012).

- **Effects of Solar Radiation Variation**

The photocurrent affected by the variation of solar radiation value. Equation (2.3) represent that the value of the current increases When the value of solar radiation increases.

$$I = [\ I_{sc} + K_i\ (T- 298)\ ]\frac{\beta}{1000}$$

(2.3)

Where K_i is the cell's short circuit current temperature coefficient (0.0017 A/C°) and β is the solar radiation (W/m^2).

The variation of the solar radiation effects on the output result of the solar PV cell, when the value of the solar radiation increases, the output current of the PV cell increases and the open circuit voltage increases logarithmically, while the short-circuit current increases linearly with solar radiation variation, as well as its value is affected by the area of the PV cell. And this increasing of solar radiation increases the heat of the PV cell.

- **Effect of Varying Cell Temperature**

Solar cells are affected by variation of ambient temperature and cell temperature which known as nominal operating cell temperature (NOCT). This affected represented as decreased the value of open circuit voltage and increase the value of short circuit current when the ambient temperature increases. The Nominal Operating Cell Temperature (NOCT) is defined as the temperature reached by open circuited cells in a module under the conditions as solar irradiance on cell surface 800 W/m^2, Air Temperature 20°C, and Wind Velocity = 1 m/s. The equation (3.4) represent this effect:

$$I_o(T) = I_o \left(\frac{T}{T_{nom}}\right)^3 \exp\left[\left(\frac{T}{T_{nom}}-1\right)\frac{Eg}{N.Vt}\right] \quad (2.4)$$

Where I_o is the diode reverse saturation current, T_{nom} is the nominal temperature, E_g is the band gap energy of the semiconductor and V_t is the thermal voltage.

- **Effect of Varying R_s & R_{sh}**

  The series resistance for solar PV cell is low, and it can be neglected in some calculations, the slope angle of the I-V curves affected by changing the value of this resistance so the maximum power point decreased when the resistance value increases. While the shunt resistance for PV cell must be large enough to obtain higher output power and fill factor.

2.4.2 Electrical Characteristics of Solar Cells

i. Open-circuit:

The open-circuit voltage is a maximum voltage available from a solar cell, and this occurs at zero current. And the equation (2.5) IS represented it. (Jain, F., 2016)

$$V_{oc} = \frac{kT}{q}\ln\left[\frac{I_L}{Io}+1\right] \quad (2.5)$$

ii. Short-circuit:

The short circuit current occurs on a point of the curve where the voltage is zero. At this point, the power output of the solar cell is zero. This case of circuits can be achieved by connecting the positive and negative terminals by copper wire.

iii. Maximum power point

Maximum power point (MPPT) is the operating point in which the power dissipated in the resistive load at maximum value. At this point, the maximum values of current and voltage is obtained. MPPT affected by the variation of the series and shunt resistances. The equations (2.6.1, 2.6.2, and 2.6.3) represent the relationship of Maximum power point (MPPT):(Jain, F., 2016)

$$Vm = Voc - \frac{kT}{q} \ln \left[1 + \frac{qVm}{kT}\right] \quad (2.6.1)$$

$$I_m = I_o \, [e^{\frac{qVm}{kT}} - 1] - I_L \quad (2.6.2)$$

$$P_{max} = V_m \times I_m \quad (2.6.3)$$

2.4.3 The efficiency of solar cell:

The efficiency of a solar PV cell is the ratio between the maximum power and the input power from incident light. Solar PV cells are relatively low in efficiency comparing with other methods of generating electrical energy, these cells reached a maximum during 2019 to 19%, and this value has increased at a good rate during the past years through improving various production factors that affect the high efficiency of solar PV cells such as reflection efficiency, Thermodynamics efficiency, charge separation efficiency, and conductivity values. The equations (2.7) represent the efficiency of a solar PV cell.

$$\eta = \frac{P_{max}}{P_{in}} = \frac{I_{max} \times V_{max}}{A \times G_s} \qquad (2.7)$$

Where A is the area of the solar PV cell and G_s is incidence solar radiation. (Jain, F., 2016)

2.4.5 The fill factor:

The fill factor is a measure of the quality of the solar cells in terms of the extent of the junction of PV cells and the reach of the decrease in the series resistance. The fill factor of a PV cell can be measured as a ratio of the maximum power of a solar PV cell to a multiple of (V and Isc). For ideal PV cell the value of FF is unity and for efficient solar cells the value of FF should be more than (0.7).(Jain, F., 2016)

$$FF = \frac{P_{max}}{I_{sc} \times V_{oc}} = \frac{I_m \times V_m}{I_{sc} \times V_{oc}} \qquad (2.8)$$

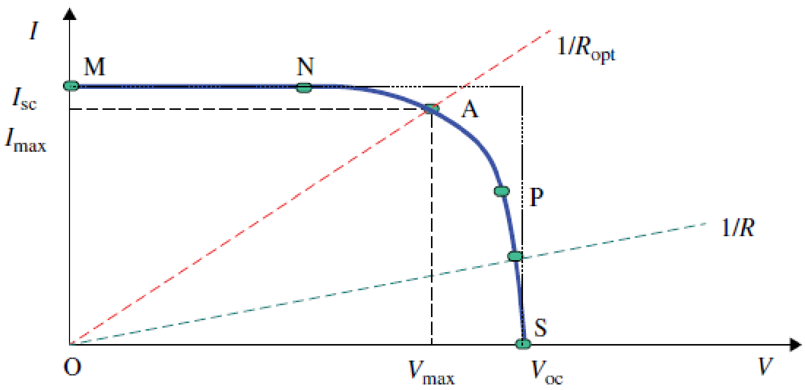


Figure 2.2: The I-V curve Characteristic of a photovoltaic cell

2.5 PV Module and Array Modeling

In recent years, the pace of development in the manufacture of solar PV module has been high, moving from 36 cells consisting of one PV module to 60, and a greater development reaching 72 cells per module, This affected on the capacity of the PV module (voltage and current) as the solar PV module in 2019 reached 72 cells, means that the current of the module increase to 10 amperes approximately, and this value is high comparing with previous version of PV module. However, one PV module alone is not sufficient to feed a solar energy system or production plant.

The PV modules are connected in series or in parallel to obtain the required power for system, provided that the voltage for one string (more one module connected in series) does not exceed more than 1000 volts

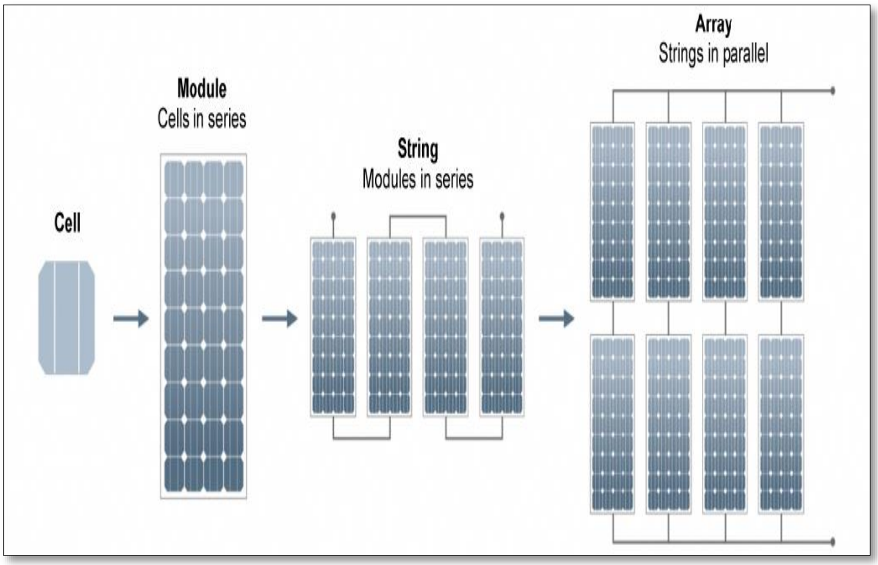


Figure.2.3: Schematic diagram of a PV module

The equations (2.9, 2.10, and 2.11) represent the relations between the cell's voltage (V_C) and current (I_C) and the module's voltage (V_M) and current (I_M):

$$I_M = N_P\ I_C \tag{2.9}$$

$$V_M = N_S\ V_C \tag{2.10}$$

$$R_{SM} = \frac{Ns}{NP}\ Rsc \tag{2.11}$$

Where N_S is the number of PV cells connected in series, N_P is the number of PV cells connected in parallel, and R_{SM} is the equivalent series resistance of the PV module. (Khatib, T. and Elmenreich, W., 2016)

2.6 Methods for I-V Curve Extraction

In photovoltaic technology there are many methods to extract the curve, these methods can be classified into online and offline used to make measurement or extraction I-V curve, which the main principle methods.

2.6.1 Online Methods:

There are many methods used to measure the current-voltage curve of the solar PV cell, Which the controlling of PV module current from zero current (V_{oc}) to the short circuit current (I_{sc}) a major factor in the extract of I-V curve .this methods use the real devices to perform the I-V curve extraction task as(capacitor, resistor, inductor, and switches),Where the basic rule for it is to set a variables ranging from a very large value to a very small value to measure current and voltage values, also short-circuit current and open-circuit voltage.

- **Variable Resistor**

  Using variable resistors is one of the simplest methods to measure the current-voltage curve of PV module. The variable resistance is connected to the PV module then value of this resistance is changed from zero to infinity and the current and voltage are measured in each change step. However, there are some limitations for this method where it can be used just for small PV module capacity, because it is difficult to provide variable resistors for high capacity. By used variable resistors the short-circuit current and open circuit voltage cannot possible to determine, Also, the manual change of resistance is slow, thus the solar radiation and the thermal effect are variable during measurement. (E.E. Van dykn et al.) are one of researchers used this method to monitoring the I-V curve characteristic for seven PV Modules under realistic outdoor conditions for one year, Where the main objective of this study was to demonstrate the value of the low-cost I–V sequencer. This research successfully emphasizes the importance of outdoor performance monitoring as there was a significant difference between indoor and outdoor measurements. (Van Dk, E.E et al., 2005)

- **Capacitive Load**

  During this method a large capacitor is used to measure the I-V curve of the PV cell. The start point of measurement is short -circuit form for the capacitor, and when the switch between the capacitor and PV module closed the loading starts. Through the loading the charge of capacitor increases so the voltage increases and the current decreases. At the end of charge, the open circuit phase is achieved that is mean the current of module will be zero. By this method a reliable I-V curve can be obtained through high quality capacitors with few losses. Many researchers used this method to extract the I-V curve as (Muñoz,J. ,and Lorenzo,E., 2006) the authors in this research used the capacitor load for PV array from approximately 80 A as short circuit current and 800V as open circuit voltage ,based on insulated gate bipolar transistors (IGBTs) , (Mahmoud,M.M., 2006) used the PV generator by a capacitor and to charge it fully from short circuit to open circuit, and to record the respective voltage and current by X–Y recorder, where these method can be used for PV generators of peak power up to 10kW.or a computerized data acquisition system (CDAS).this method can be used for PV generators of higher power. The I–V curve obtained by this method is much more accurate and uniform since it is measured in a very short time.

- **Electronic load**

The electronic load method, used to achieve the I-V curve for PV module. Usually this method used (MOSFETs) as load .for the purpose of extract IV curve, three modes of operation must be operate (cut-off, active and ohmic region), where the current supplied from PV module is flow, When the resistance between (source and drain) is set through (the gate-source. Voltage). In a (An electronic load for testing photovoltaic panels) research, the authors used several MOSFETs for testing PV module to extracted the I-V curve. Where a linear MOSFET serves as an electronically controlled load that moves the operating point of the PV panel over the entire I–V characteristic. In addition to the (I-V) and (P-V) characteristics, the circuit provides the values of the open circuit voltage, short-circuit current, peak power, and the corresponding voltage and current. The advantages of this method are the high testing speed the ability of measuring high currents (kuai, Y., and Yuvarajan, S., 2006).

- **Bipolar Power Amplifier**

  By this method, the Bipolar Power Amplifier device is used and a BJTs transistors are used as load forward and reverse current respectively, to achieved the I-V curve for PV module, where bipolar transistors must operate in three modes of operation (cut-off, active and saturation region).In a research of (Guvench,M.G. et al., 2004) the authors measured the I-V curve characteristics of large area solar cells (up to 8 in diameter) operated under simulated solar irradiation by automated way ,through using standard bench top GPIB instruments interfaced to a PC and by using the function generator as a stepped voltage source. And the high value of test current needed to this solar cell is obtained from a unity gain DC power amplifier driven by the function generator.

- **Four-quadrant power supply**

  By this method, a four-quadrant power supply is a laboratory power supply produced the positive and negative voltage, which is used to extract the I-V curve for PV module in the first quadrant, while the points in second and fourth quadrant are important to detect the mismatched in

PV module operation states Such as total or partial shading of the PV cells in one or more PV module. The I-V curve is detected by connected this device to the PV cell output, where it is forcing the PV cell to the fixed voltage and in the same time the current into this device is measured .the start point represent the short-circuit current where the voltage value is 0v, then the voltage is increasing in steps with constant value and the value of current is measured in each step until the current is equal to zero, where this point represent the open circuit state.(Duran,E. et al.,2008) **.**In (de Blas, M.A. et al., 2001) the authors are extracted the experimental I-V curves for solar module made up of 36 mono-crystalline silicon cells connected in series under different irradiance and temperature conditions with natural sunlight on clear days by a bipolar power supply, where at each point the value of current and voltage of the solar cell is measured. The temperature and irradiance of the module are recorded at the beginning and end of each experiment, and the different magnitude curves are rejected.

- **DC-DC Converter**

  A DC-to-DC converter is an electronic circuit or electromechanical device used to obtain the I-V curve of PV module by applied the property of it of acting as resistor emulators. This converter has three basic configurations (Buck converter, Boost converter and Buck-Boost converter) and similar to a DC transformer in Continuous Conduction Mode and in Discontinuous Conduction Mode. In (Khatib, T., et al.,2017) the I-V curve is extracted using the DC-DC boost converter by adjusting the operating cycle of the control signal boost converter’s switch. The advantage of this method is that they are implemented without need external devices. And in the (Duran, E., et al,2012) the I-V curve of PV model was extracted through a prototype system that is controlled using a microcontroller, where the solutions of electrical circuit depend on DC -DC converters. The results of this experiment are characterized by simple structure, scalability, fast response, and low cost compared to previous methods. However, the researchers found that this method has disadvantages as the current ripple by the inductor due to the switching technique that does not exist in the other methods. Table 2.1 represents a comparison between the characteristics of the previous methods of extraction the I-V curve.

Table 2.1: The comparison between the characteristics of the different methods of extraction the I-V curve

	Flexibility	Modularity	Fidelity	Fast Response	Direct Display	Cost
Variable Resistor	Medium	Medium	Medium	Low	No	Low
Capacitive Load	Low	Low	Medium	Low	No	High
Electronic Load	High	High	Medium	Medium	Yes	High
Bipolar Power Amplifier	High	High	High	Medium	Yes	High
4-Quadrant Power Supply	Low	Low	High	High	No	High
DC-DC Converter	High	High	High	High	Yes	Low

2.6.2 Offline Methods:

This method aims to extract the I-V curve of the solar PV cell based on historical experimental information, where it is based on empirical mathematical methods or artificial intelligence methods, the I-V curve is constructed by a model of five parameters or more of the PV cell, the PV cell output current is defined as a function of this parameters and it is necessary to improve the value of these parameters based on historical empirical information. However, for this method some disadvantages such as the ability to identify any error or abnormal condition in the PV system. The major offline methods are represented in Artificial Intelligence methods, Random Forests technique, and the numerical methods where are used to obtain numerical solutions of a mathematical problem such as Levenberg–Marquardt method (LM), Newton–Raphson method (NRM),

- **Levenberg–Marquardt Method (LM)**

  The Levenberg-Marquardt algorithm (LM) is an iterative method that determines the minimum of a multivariable function to solve the non-linear least squares problems in mathematics, and computer programs as MATLAB program. LM algorithm is a combination of steepest-descent and the Gauss-Newton method. (D.W. Marquardt, 1963).

  This technique has been used by many researchers in researches related to extraction of the I-V curve for solar cells such as (Dkhichi,F. et al., 2014) in this research the authors used the LM method to obtain the I-V curve of 57 mm diameter (R.T.C France)solar cell single diode model through the real values of five parameters, this method achieved more accurate I-V curve, lower values of power error, and a high harmonization between the experimental and calculated curves of I-V and P-V.(Ma.T et al., 2014 ;Tossa,A.K. et al., 2014) used this method to estimate the I-V curve.

- **Newton–Raphson method (NRM)**

  This method is one of the iteration techniques, used to obtain the root of the equation for a nonlinear least-squares optimization algorithm, this algorithm is modified with Levenberg parameter. NRM technique is used to obtain the I-V curve for solar PV cell by five parameters of PV cell from the experimental data. In (Easwarakhanthan et al., 1986) the authors are designed and simulated a nonlinear electrical model for (57 mm diameter (R.T.C France)) solar cell by optimization algorithm, designed for micro-computers, which uses Newton's method to determine the five parameters of the single diode model (V_{oc},I_{sc},I_m,V_m and the slopes at $I = 0$ and $V = 0$) to obtain the minimum value of errors.

- **Artificial Intelligence**

  The emergence of several disadvantages in the previous methods led to think about the development of other methods based on artificial intelligence (AI). This Algorithms were proposed to generate the I-V curve for the solar PV cell, this method is characterized by high efficiency and accuracy, in addition to its ability to estimate the standards of solar PV cell models. The method aims to improve the value of each parameter in the five-parameter model based on empirical information. There are many algorithms used to extract the current-voltage curve of the solar cell, as:

Genetic algorithm (GA) are used to solve constrained and unconstrained optimization problems. These algorithms have been used in many different engineering and practical sectors, where they have been used by many researchers as (Dizqah et al., 2014; Ismail et al., 2013; Jervase et al., 2001), to extract the current-voltage curve of the solar cell, where these algorithms are global methods for the purpose of improvement.

Particle swarm optimization (PSO) algorithm Particle swarm optimization is a computer algorithm. The idea of the algorithm depends on a number of elements spread in a limited search area and randomly moving to search optimization in this area. Whereas, if the number of swarm elements increases and the research area becomes smaller, finding the optimum solution becomes faster. But if the number of elements is smaller and the area of research is large, the less chances of finding the best solution. This algorithm can be used to estimate the parameters of PV cell and extracted the I-V curve for it. Many researchers were used PSO algorithm for PV module as (Khanna et al., 2015), in this research, PSO algorithm has been applied to estimate the solar cell parameters of the two-diode model and the proposed three-diode model from the illuminated I-V characteristics measured using a solar simulator. Parameters were estimated using an iterative PSO approach from the I-V characteristics. Although PSO is a random approach, but in this work, the authors have shown that the multiple PSO runs gave consistent results. PSO was, therefore, found to be a good method for accurate estimation of solar cell parameters, with MAPE not exceeding 0.18% of Isc, for any sample.

Flower pollination algorithm FPA is a nature-inspired algorithm that simulates the pollination behavior of cultivated plants, and it is one of the methods for improving the swarm, as many researchers have used this algorithm to improve results in various fields such as energy, electrical system, signal and image processing, wireless sensor networking, clustering and classification, global function optimization ,computer gaming ,structural and mechanical engineering optimization , This this algorithm is characterized as flexible, adaptable, scalable, and simple optimization method(Alyasseri et al.,2018).In (Alam et al., 2015) The author proposed the FPA algorithm as newly developed optimization technique to extract the optimal parameters of a single diode and a double diode models various types of PV modules. . The proposed extraction technique is tested using three different sources of data, represented by previous literature, the measured data in the laboratory, and the data of the manufacturer’s data sheets. the proposed FPA model achieves the least (RMSE) between the estimated and experimental data and the highest speed of conversion to the optimal solutions with the shortest convergence time.

Shuffled frog leaping algorithm is a meta-heuristic optimization technique. The concept of the SFLA is based on observing, imitating, and modeling the social behavior of a group of frogs when they search for the location of a rich source of food. Several engineering optimization problems have been solved by the SFLA. The SFLA has been successfully applied to solve many power system optimization problems such as transient stability improvement of a grid-connected wind farm, unit commitment problem, harmonic distortion minimization in inverter systems, power system damping, optimal switch placement in a distribution system, and optimal reactive power dispatch. The main advantage of the SFLA is its high-speed convergence where it combines the merits of both GA-based technique and the social behavior of the PSO approach. (Hasanien, 2015) In this research, the SFLA technology used to determining the unknown parameters of the single diode PV model. The SFLA is used to identify the unknown parameters of the PV model such that the maximum power of the model is equal to the maximum experimental power extracted from the datasheet of the PV manufacturer. The validity of the proposed PV model is verified by the simulation results which are performed under different temperature and irradiation conditions. The simulation results are compared with the experimental results of different PV modules such as Kyocera KC200GT and Solarex MSX- 60. The effectiveness of the proposed PV model was evaluated by comparing absolute error of the model with that of other PV models.

In (Siddiqui,M.U, and Abido,M.,2013) the authors in this research explain the performance of different evolutionary algorithms that used to generate the I-V curve of PV cell .This models using the estimated parameters were then used to predict the electrical performance of six PV modules. Artificial neural networks (ANN) are one of (AI) method that used to extract the I-V curve of PV cell by predict the value of model parameters based on historical experimental datasets. These methods are more accurate, where the errors for predicted I-V curve are less than other methods. For more information for predicted errors result regarding these methods, (khatib,T. et al.,2018) research discussed it. However, this method has some challenges as a result of the complex data training process.

Artificial Neural Networks (ANN)

are classified as a kind of mathematical algorithms to solve many problems. These networks are simulating the same principle of functioning for the neural networks in the human brain. This similarity in work among them comes from the fact that the human brain contains parallel connections that achieve a huge amount of tasks in a very short time compared with a computer, just as the human brain contains a very large number of highly complex nonlinear computational elements called (neurons) as well as a huge number of internal connections. The human neurons are distinguished by the way they process data in parallel, which gains them the super speed and this what scientists have tried to apply to artificial neural networks in the computer. Artificial neural networks are used for extracting the I-V curve of solar cells and PV modules by using historical experimental data that depend on five or seven parameters (Karatepe et al., 2006; Celik, 2011; Bonanno et al., 2012, Tamer et al., 2018). These methods are quite accurate. but this method have some limitation ,where that these methods are only able to measure I-V curves for the solar cell parameters which ANN have been trained based on, whereas, the ability of measuring general I-V curve for different solar cells is very limited.

Random Forests Technique (RFs)

The previous methods have some limitation in extraction I-V curve for PV cell such as low accurate result for using online methods (Evolutionary algorithms), Difficulty in finding the model parameters for empirical model and Complexity training process for ANN methods. As a result of these constraints, random forest technique is one of a new method of extracting the current-voltage curve of the solar PV module with greater accuracy. Random forest is an ensemble machine learning method. That the decision trees in these forests were learning to predict the output variable (Goal) instituted on the input variables. Where decision trees models divided to: regression trees models and classification trees models. (Breiman, 2001). In the research of (Ibrahim. I, 2016) a standalone PV system was studied, where the random forest technique was used to developed a new model for predicted a current -voltage curve for PV module. This model was depended on four inputs; solar energy, ambient temperature, day number and daily hours. The metric errors value was adopted to prove the accuracy of this model, where the values of root mean square error, mean absolute percentage error and mean bias error 2.748%, 8.715%, and -2.577%, respectively.

In this work the random forest technique was used to developed a new proposed model based on other inputs; solar radiation, cell temperature, and PV experimental voltage. The mean absolute percentage error value was adopted to verify the accuracy of this model and compare it with the previously mentioned models.

Chapter III

I-V Characteristic Curve Extraction Using Random Forests Technique

3.1 Introduction

In this chapter, the current - voltage curve of a solar PV module will be predicted by performing a set of practical experiments and programming these experiments using the MATLAB.2016 program, where a new code has been developed using random forest technique for regression to predict the I-V curve by growing trees depending on a random vector, to obtained numerical output values. Then the results of this model are compared with other models that use different methods to predict the PV module output current, and verify the results from in terms of accuracy and the values of Mean bias error (MBE), Root mean square error (RMSE), and Mean absolute error (MAPE).

Seven experiments were done on the solar PV module to measure the current and the voltages by using the (I-V CURVE TRACER DEVICE), in each experiment the solar radiation and the temperature changed. The table (3.1) showed Specifications for PV module used in these experiments.

Table 3.1: The Specifications for PV module

	PV Module (STF-120P6)
Rated Maximum Power (P_{max})	120
Open Circuit voltage at STC (V_{oc})	21.5
Short Circuit Current at STC (I_{sc})	7.63
Conversion Efficiency	14%
Maximum Voltage (V_{mp})	17.4
Maximum Current (I_{mp})	6.89

3.2 I-V Characteristic Curve Extraction Using Random Forests Technique

3.2.1 Random Forest Technique

Random Forests are a group of decision trees (classification or regression trees). Each tree in this forest depends on the values of a random vectors, Where the sampled from all trees take independently and in the same distributing. (Breiman,L., 2001).

- **Decision trees**

  Decision trees (classification trees and regression trees) are machine-learning methods (supervised learning model) for constructing prediction models from known responses data. (Shobha, G, and Rangaswamy, S.). In predicting the response, the decision follows each tree from the root (beginning) node down to a leaf node. The leaf node contains the response (MATLAB, 2016).

  Classification tree is a predictive model collected of a weighted combination of varied classification models. Where the performance of the predicted model increasing, when multiple classification models combining with each other. The classification group collected a group of trained weak learner models and data on which these learners were trained, it can predict a response for new data by aggregating predictions, and it can be stored this data and it can be can compute re-replacement predictions. (MATLAB, 2016).

The output response for classification ensembles give as nominal form ('true' or 'false'). (BREIMAN, 2001). Where these outputs are formed in the form of separate responses. (Shobha,G. ,and Rangaswamy,S.). Classification models can be applied in many applications as spam filters, advertisement recommendation systems, and image and speech recognition. as example of classification problem is Predicting whether a patient will have a heart attack within a year, and the possible classes are true and false. Classification algorithms usually apply to nominal response values. However, some algorithms can accommodate ordinal classes. (Matlab, 2016)

Regression tree is a predictive model collected of a weighted combination of varied regression trees, where the performance of the predicted model increasing, when multiple regression tree combining with each other.

Regression models represent the relationship between one dependent output (response variable), where it gives as numeric form, and one or more independent input variables (predictor variables). In regression trees models, the goal is to predict a continuous measurement for an observation. That is, the responses variables are real numbers. (Shobha,G.

, and Rangaswamy,S.) This model can be used in many applications as forecasting stock prices, energy consumption, or disease incidence. The Regression models can be classified into:

– Fit linear regression models, where it represents the linear relation between the output (response) and the inputs (predictors) .and it can be used for predict responses or simulated it, assess this model using hypothesis tests, or use plots to visualize diagnostics, residuals, and interaction effects.
– Generalized linear regression models are a special case of nonlinear models but it using linear methods.
– -Nonlinear regression models, where it describes the relation between a continuous response variable and one or more continuous predictor variables
– -Nonparametric regression models, where it can be used for more complex regression curves without determining the relation between the output (response) and the inputs (predictors) with a predetermined regression function. The output from this model is often described as diagram.
- Gaussian process regression models, where it can be used to compute prediction intervals. (MATLAB, 2016).

- **Random Forest Algorithm**

Random forest is an ensemble machine learning method. That the decision trees in these forests were learning to predict the output variable (Goal) instituted on the input variables. Where decision trees models divided to: regression trees models and classification trees models. Regression trees models predict the target variables based on continuous set of value, while classification trees models take the separate set of value to predict the target variables (Shobha,G. ,and Rangaswamy,S.).And these forests incorporate Bagging; Bagging is a technique to aggregates the fitted values in various ways, which used to reduce the variance of prediction values (Beriman, L., 1996).

Many types of algorithm can be used to develop prediction models such as QUEST, CRUISE, CART, RPART and GUIDE, C4.5. **CART** is a classification algorithm that depends on the Gini index (a generalization of the binomial variance) as node impurity criterion. This algorithm chooses the binary splits that decrease the impurity maximum. A large tree growing instead of (employing stopping rule) to produce a sequence of sub trees, and pruning it until the root node. **PART** is recursive partitioning and regression trees. Based on the target data that the regression trees or classification trees want; the continuous variables or categorical can be used in this type of algorithms. **C4.5** is one of the classification algorithms that use Entropy for impurity function to minimize an estimate of the misclassification error. This algorithm divided the node into two splits of usual form. Dealing with categorical variables to splits is easy in this algorithm. In **GUIDE, CRUISE and QUEST** the variable selection is unbiased. **GUIDE** algorithm is quantile, Poisson and proportional hazards regression method. In this algorithm a simple polynomial tree model is built for least squares. The high accuracy of CRUISE and QUEST can be obtained if these algorithms are using the linear combination splits. (S.Shamy & Dr.J Dheeha).

- **The Random Forests Structure**

The random forest algorithm in the general structural framework for either classification or regression trees is a combination of training and testing stages for the data that is collected. The algorithm begins with entering data into the training phase, where the algorithm draws M_{tree} bootstrap samples from the original training data, and then creates a number of unpruned classification or regression trees for each bootstrap sample, where the best split is chosen from the random sample of the predictors at each node of classification or regression trees. The new data are predicted by aggregating the predictions of the M_{trees} trees. (Liaw & Wiener 2002).

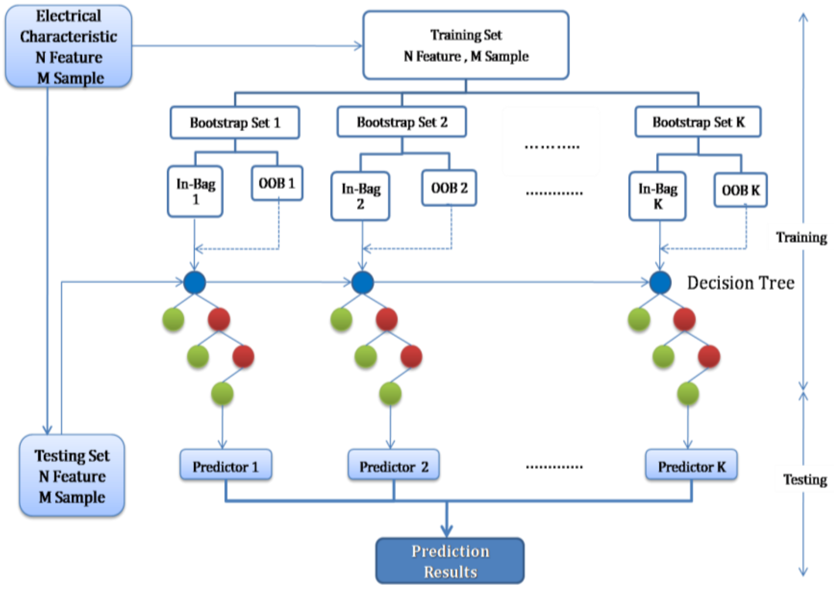


Figure 3.1: The Random Forests Structure

The random forest algorithm involves in two main stages, the Figure 3.1 represent the Random Forests Structure. The RFs methodology for regression can be described as follows:

- **Training stage:**

  firstly; the bootstrap samples (M) created by the original data obtained by the experiments in the same size randomly .Secondly; about 30% from the data were left out of bootstrap sample where it called (out -of bag data OOB), while the remaining data sets were trained in bootstrap samples where it called (In-Bag data) and used it to develop the RFs model.(Beriman,L., 2001).the classification or regression trees grow from each bootstrap samples, at each node from (n) tree, some variables were selected randomly and just this variables are searched through for the best split.

This process is done through Bootstrapping technique (Re-sampling technique) .and it can be used to obtain a best idea of the representation of class labels existing in the dataset. (Shobha,G. ,and Rangaswamy,S.)

In the training stage the bagging is important to use for two reasons; the first one is to improve accuracy when random features are used. The second is to give ongoing estimates of the generalization error of the combined set of trees, estimates the strength and correlation, by aggregate OOB prediction data at each bootstrap iteration (Beriman,L., 2001).In addition to that OOB information can be used to estimate the variable importance measures, Intrinsic proximities between cases, Scaling coordinates based on the proximities, and to detect the outliers .(Beriman,L., 2002).

Based on OOB samples, the variable importance can be obtained randomly for each tree. this measure depends on the permutation importance measure (Breiman,2001). the variable importance measure is calculated as aggregation of the difference between prediction accuracy before and after permuting variable N, averaged over all the trees. The following equation describes the variable importance VI for N variables (Guo,L. et al. 2011):

$$\mathrm{VI}^{(t)}(\mathrm{N}) = \frac{\sum_{xi \in \beta^{c(t)}} \mathrm{I}\left(\mathrm{Lj} = \mathrm{C_i}^{(t)}\right)}{\left|\beta^{c(t)}\right|} - \frac{\sum_{xi \in \beta^{c(t)}} \mathrm{I}\left(\mathrm{Lj} = \mathrm{C}_{i,\pi n^{(t)}}{}^{(t)}\right)}{\left|\beta^{c(t)}\right|} \qquad (3.1)$$

For overall tree the important result can be calculated by:

$$\mathrm{VI}^{(t)}(\mathrm{N}) = \frac{\sum_{T} \left(\mathrm{VI}^{(t)}(\mathrm{N}) \right)}{\mathrm{T}} \qquad (3.2)$$

Where $\beta^{c(t)}$ corresponds to OOB samples for a tree, T is the number of tree (1, 2, 3…. T). $Ci^{(t)}$, And $Ci,\pi n^{(t)}$ are the predicted classes for sample before and after permuting the variable n respectively, xi is the sample value, L_j is the true label; both are in the training stage, *I* represents the importance function that got based on the values of L_j, i is the number of samples per leave in tree and *j* is the number of samples per tree in the forest.

In Final step in training stage the outliers in regression models can be detected by cluster analysis, these analyses depend on the model density. The Outliers in the response variable represent model failure and called observations, and with respect to the predictors are called leverage points. Outliers play important role in regression. Removing these values from data sets will increase the results accuracy.

– **Testing Stage:**

A new data predicted by aggregating the predictions of the (n) trees. Where this predicted data represents as: majority votes for classification, average for regression. (Liaw,A.,and Wiener,M.)

- **The Evaluation for Random Forests Technique**

  In the ensemble machine learning (regression models or classification models), the output (response) get from the input variables (predictors) and to estimate these results; three types of error metrics can used as (MBS, RMSE, MAPE).

  MBS is defined as mean bias error, where it can used to indicate the average deviation for prediction in the model between the forecast value (predictors) and observation value (target value).RMSE is defined as Root Mean Square Error this type of error can used to measure the standard deviation of the predictions, by measuring the difference between the forecast and observation values. MAPE is defined as Mean absolute percentage error, used to indicate the predictions accuracy as percentage.

The following equations represent these errors:

$$MBE = \frac{I}{n}\sum_{i=1}^{n}(Ypi - Yi) \tag{3.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Ypi - Yi)^2 \ldots} \tag{3.4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Yi - Ypi)}{Yi}\right| \tag{3.5}$$

Where Y_i is an observed value (target value), Y_{pi} is a forecast value (predicted value), and n is the number of observations.

3.2.2 Proposed Model for I-V Curve Extraction Using Random Forest Technique

In this study, the output current for PV module predicted by random forests algorithm which is called (Bagger algorithm) in (MATLAB 2016 program). Tree Bagger algorithm is a training algorithm for random forests applies the general technique of bootstrap aggregating, or bagging, Bagging stands for bootstrap aggregation. Every tree in the ensemble is grown on an independently drawn bootstrap replica of input data. Observations not included in this replica are "out of bag" for this tree. the prediction of an ensemble of trees for unseen data can be compute, Tree Bagger takes an average of predictions from individual trees. To estimate the prediction error of the bagged ensemble, the predictions for each tree on its out-of-bag observations can be compute, average these predictions over the entire ensemble for each observation and then compare the predicted out-of-bag response with the true value at this observation.

The RFs proposed model start by setting the input variables which is represented by: ambient temperature, solar radiation, PV DV voltage, short circuit current, and open circuit voltage. For this study the number of trees setting as 500 trees, and the number of leaves is five for regression as set by default by the algorithm (MATLAB, 2016).

The number of leaves and trees in prediction process effect on the accuracy, variance, and the rate of training and error. where if that less than the optimum numbers; this leads to an increase in training, error rate, and variance.

Accordingly, the value of metric errors as root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean bias error (MBE) take to estimate the performance of the developed RFs model. The minimum value for these errors means high accuracy for the developed RFs model, less error rate, and less variance. Figure (3.2) shows the flowchart for predicting PV module output current.

The proposed RFs model represent in three stage as follows:

Stage I:

In this stage, the input data for experiments (ambient temperature, solar radiation, PV DV voltage, short circuit current, and open circuit voltage) prepared in data file to use in training and testing sets. The following steps describe the procedure of the proposed model:

i. The data spilt in predictor array (X) and in response array (Y). The Tree Bagger algorithm uses for training by set the number of trees 500, and the number of leaves 5 per tree as default (MATLAB, 2016).
ii. In this step, the variable importance is estimated by permuting the values of this feature across every observation in the data set for each tree, then measure mean squared error that becomes after the permutation. The variable importance measure is used to improve prediction ability for (RFs algorithm). The MATLAB line code number (41-62) in Appendix C describe the variable importance measuring for proposed RFs model.
iii. The outliers are observations whose value is far from the value of the rest of the data in a same data set, and these values are considered to have a negative impact on estimation and the accuracy of prediction process, so after estimating the important variables, the outlier values are detected and removed from the data set. The MATLAB line code numbers (64-91) in Appendix C describe the outliers detected process for proposed RFs model by using cluster analysis and applied it by computing the proximity matrix (fillProximities) code.

Stage II:

i. After training stage which the important variables were estimated, and the outliers detected and removed, the random forest algorithm train and test using the values of experiment number four, five and seven to predicted the output current for PV module.

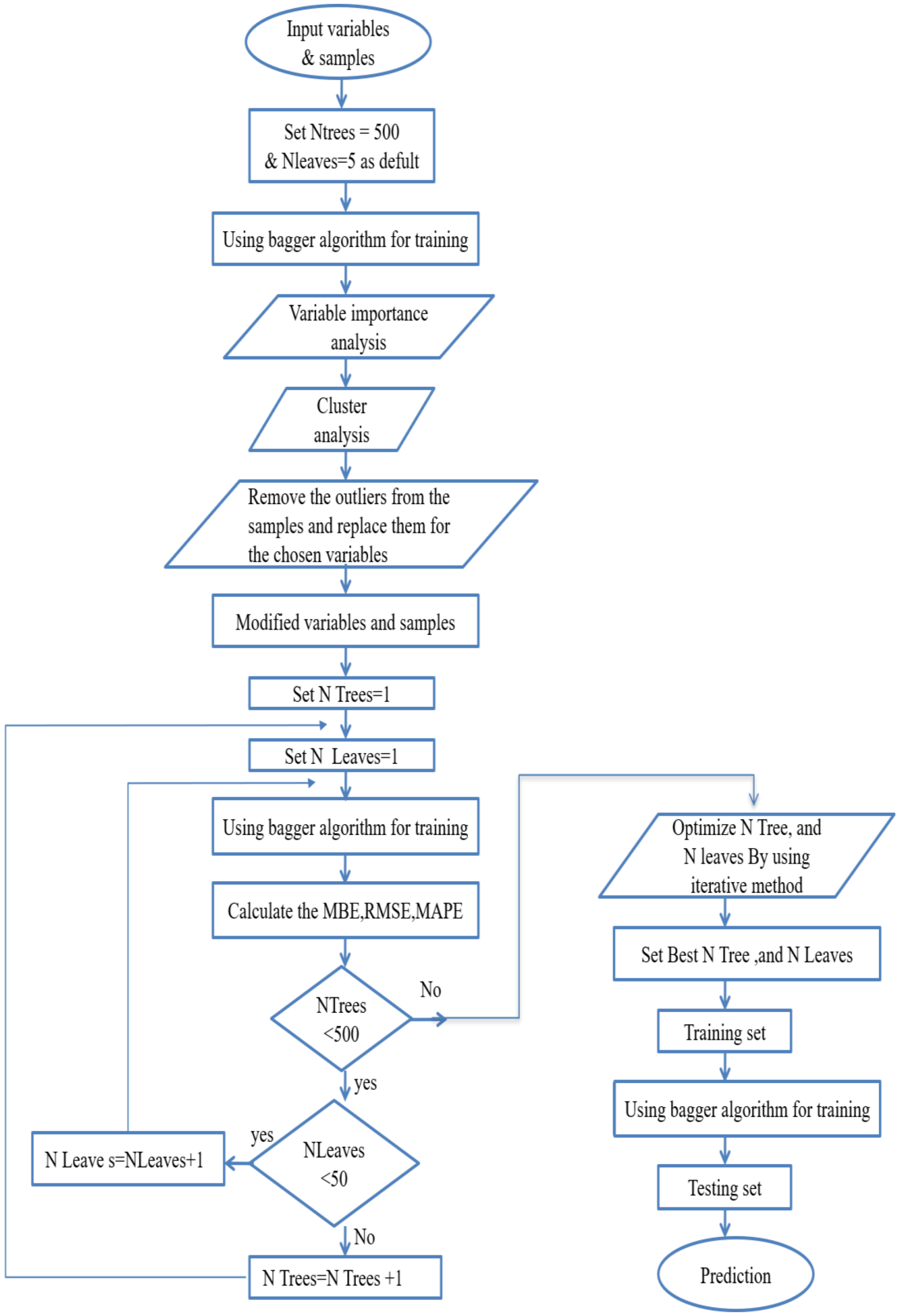
ii. The values of RMSE, MBS, and MAPE are compute for all data, and then the minimum values of metric errors use to evaluate the predicted result for PV module output current. The MATLAB line code numbers (92-255) in Appendix C describe the predicting output current for PV module using developed RFs model.

Stage III:

i. In this stage the optimum number for tree number and minimum leave number was found through the for-loop process. The numbers of trees and leaves are set 500 and 50 respectively. Through this loop the value of metric errors (MSE, MBE, RMSE, and MAPE) was found and reliance on the value of the RMSE in finding the optimum number of leaves. The MATLAB line code number (257-306) in Appendix C describe the process of finding optimal number for Tree number and minimum leave number.

ii. The optimum I-V curve of experiment seven is estimated through the for loop and comparing it with pervious result obtained from stage two. The MATLAB line code numbers (308-337) in Appendix C describe the testing result for this stage.

Finally, The I-V Curve prediction result of developed RFs model at 318.32(K) Temperature value was compared and plot with different solar irradiation levels. The MATLAB line code number (338-351) in Appendix C describe the I-V Curve Prediction Result of RF at T=318.32(K) and different solar irradiation levels. And The I-V Curve prediction result of developed RFs model at 978 (W/m^2) solar radiations was compared and plot with different temperatures levels. The MATLAB line code numbers (353-362) in Appendix C describe the I-V Curve Prediction Result of RF SI=978(W/m^2) and different temperatures levels.



Figure

3.2: The Random Forests Flowchart For proposed model

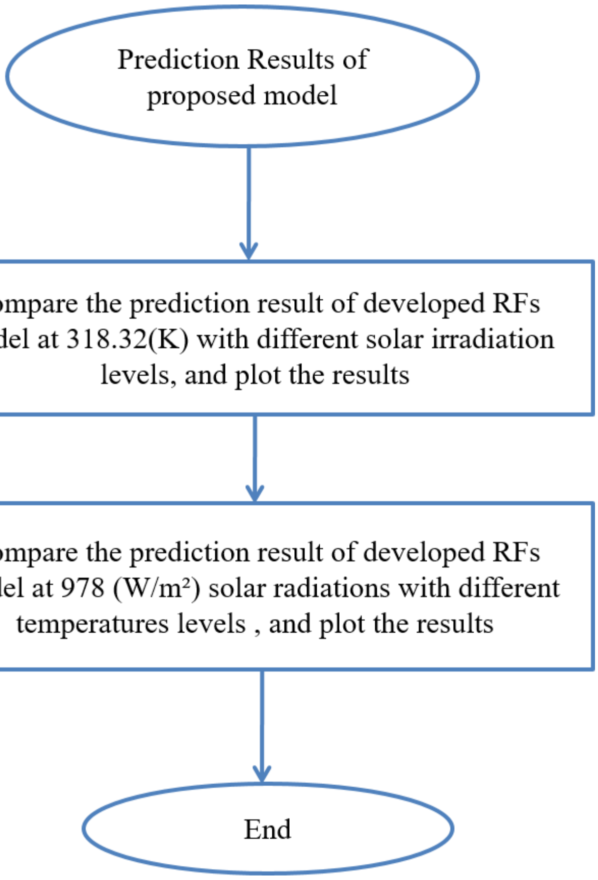


Figure 3.3: The final step in stage three for RFs proposed model

Chapter IV

Results and Discussion

4.1 Introduction

In this research, the database collected for 120Wp polycrystalline PV module by performing seven experiments under different conditions of cell temperatures and solar radiations to measure the current values and voltages values of the solar PV cell, where the number of readings of the seven experiments reached approximately 490 readings recorded by (I-V400-solar PV analyzer)), and approximately 5,000 readings were theoretically calculated.; to extract the I-V curve through a new model based on random forest technique. Table 4.1 represent the most important point for these experiments, where the Short circuit current (Isc) cannot be obtain at V equal zero, so the values in this table represent the maximum values current at minimum values of voltage.

Table 4.1: the most important point of the experiments was obtained by (I-V400-solar PV analyzer)

	Temperature (k)	Solar Radiation (W/m^2)	Open circuit voltage (Voc)	Max. circuit current (Im)
Exp. 1	318.32	118.28	17.5	0.84 A at 0.38 V
Exp.2	321.25	148	17	0.88 A at 9.6699 V
Exp.3	327.7	306	17.6	1.98 A at 0.5874 V
Exp.4	324.21	711	18.3	4.48 A at 0.675 V
Exp.5	329.1	780	18.5	5.2 A at 0.6 V
Exp.6	331.42	840	18.3	5.45 A at 0.4 V
Exp.7	328.56	978	18.5	6.48 A at 0.48 V

4.2 Results of proposed model of I-V curve extraction

In the developed RFs model, the data divided into training phase and testing phase to check the accuracy of the proposed model. Based on the information in chapter 3, the numbers of trees take 500 and the number of leaves take 5 as default value to estimate the variable importance and discover the outliers in the datasets. Variable importance VI represents the statistical significance of each variable in the data with respect to its effect on the resulting model. VI is the classification of each predictor based on the contribution of predictors to the model. This technique helps to dispose some of the predictors who are not contributing anything and that instead add time to processing. Figure 4.1 represent the variable importance for five parameters; ambient temperature, solar radiation, PV DC voltage, short circuit current, and open circuit voltage. The most variable importance is solar radiation with a rate of 1 from 1.5, for ambient temperature 0.95 from 1.5, and 1.45 from

1.5 for PV DV voltage. As this result the ambient temperature, solar radiation, PV DV voltage parameters used as input data for the proposed RFs model.

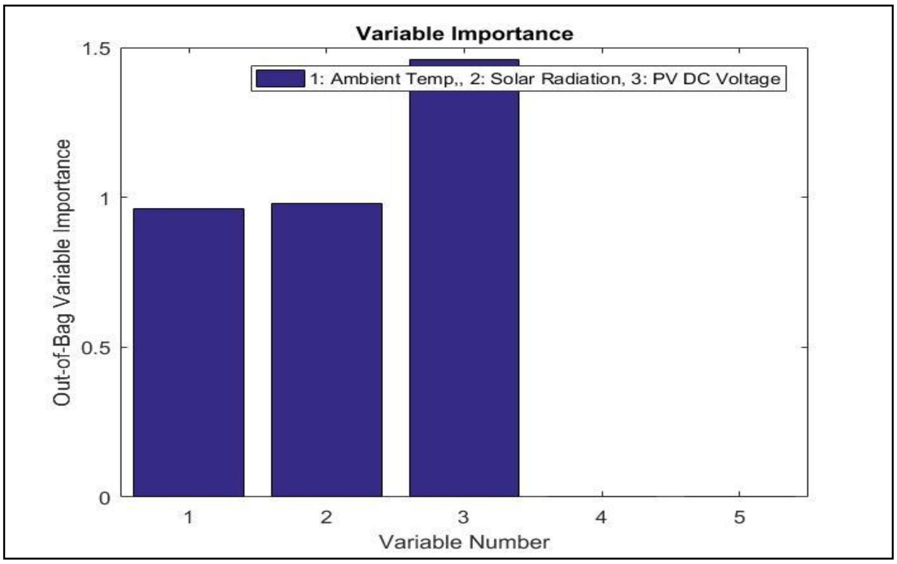


Figure 4.1: The variable importance

In next step, the outliers detected by cluster analysis. Cluster analysis is a type of statistical method that can be applied to data. Clustering analyzes is the task of grouping a set of data in such a way that data in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). It is used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning. Many typical clustering models are used, namely, subspace model, connectivity model, centroid model, density model, group model, distribution model, and graph-based model. A normal distribution model is used to analyze the data set. Figure (4.2) shows the distribution of training data by cluster analysis, Figure 4.3 represent the percentage of outliers for training data group detected by cluster analysis.

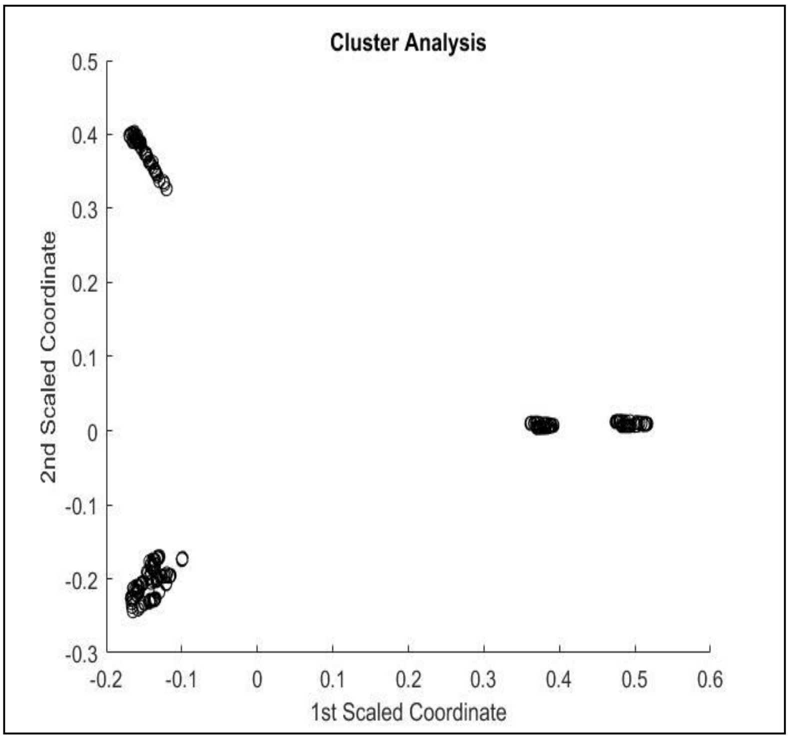


Figure 4.2: The cluster analysis for training data

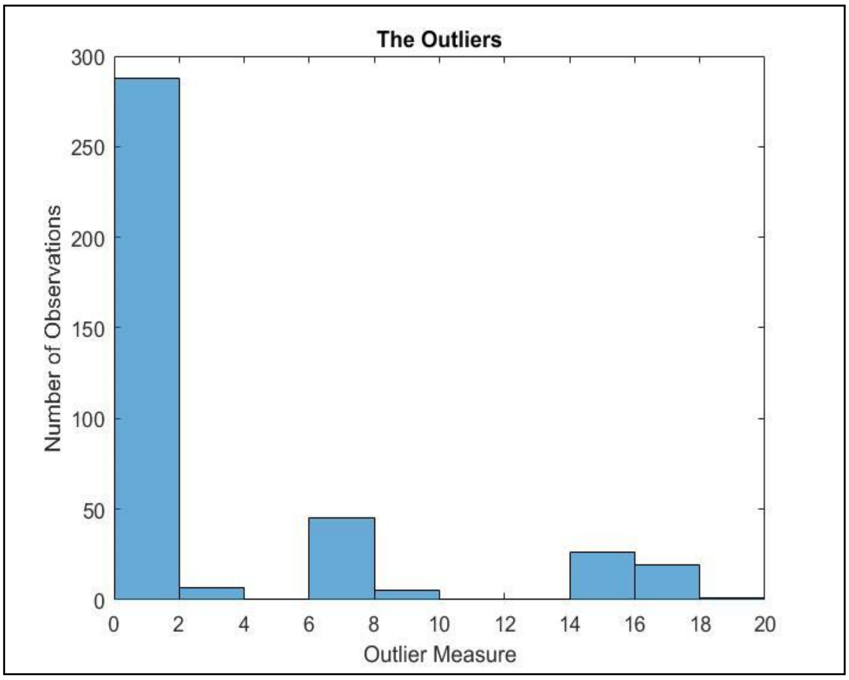


Figure 4.3: The outliers are detected in the training data

Through the RFs algorithm of this proposed model, the number of trees and leaves were improved through an iterative method for a number of trials up to 25,000, through the designation of 500 trees and 50 leaves, where the best number of trees and leaves were searched to obtain the optimized parameters of proposed model based on the result of RMSE, where the best value for the tree and the leaf is equal 3, 6 respectively. .At these values Metric error ((Mean Bias Error (MBE), Root Mean Square Error (RMSE),and Mean Absolute Percentage Error (MAPE)) were calculated which have the best error value and which give an indication of the accuracy of this proposed model for extraction the I-V curve of the solar PV module. Figure 4.4 show the trees grown in the forest. Figure4.4 shows how the trees are built in the forest by using Tree Bagger algorithm for regression, which grows the decision trees in the ensemble using bootstrap samples of the data. this algorithm selects a random subset of predictors to use at each decision split to obtain the PV module

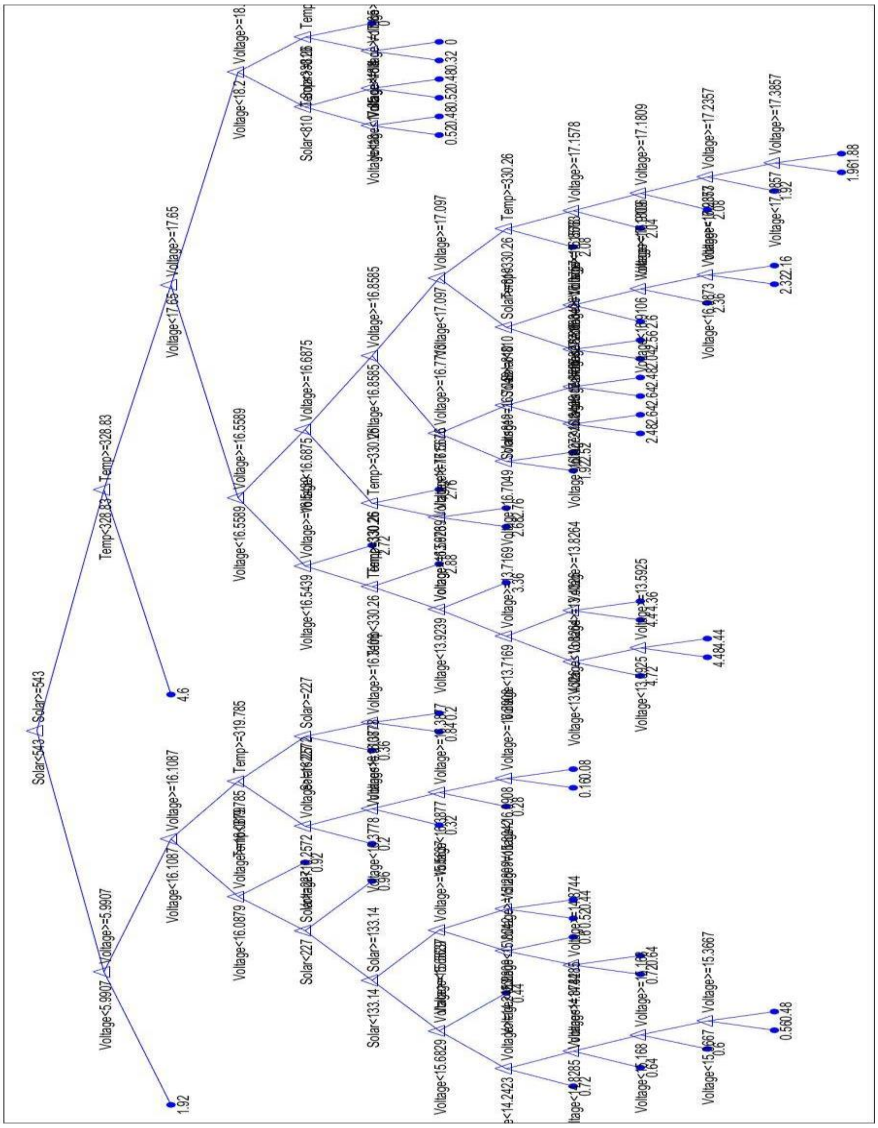


Figure 4.4: The trees grown in the forest

- **Testing Results**

In the proposed model, the datasets of experiment number five, and seven was tested which is considered part of the training data, and the dataset of experiment number four were used for the testing stage only. The selected of the experiments are randomly from all experiment to shows the performance of the MATLAB code. Figure 4.5 represents the I-V curve of the PV module was obtain from RFs proposed model and the actual I-

V curve by the theoretical data for experiment number four. Where this experiment was performed at solar radiation 711 w/m2, and 324

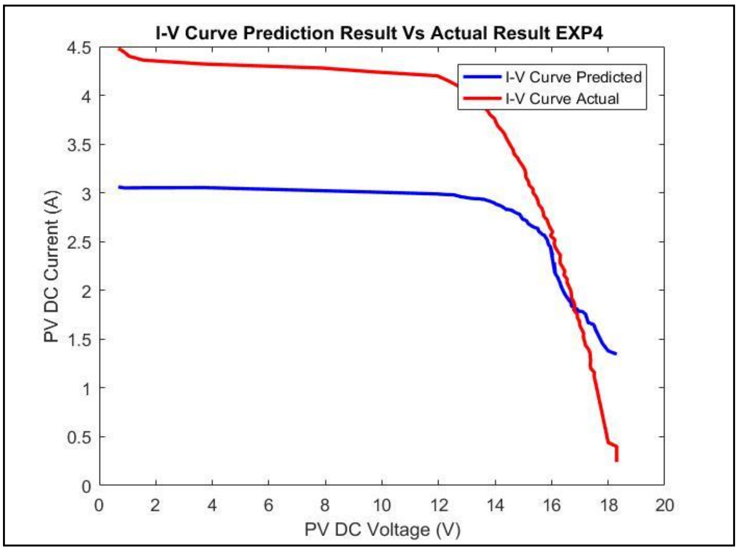


Figure 4.5: The I-V curves of the PV module at experiment four

Figure 4.6 represents the predicted I-V curve and actual I-V curve for the PV module at experiment number five, where this experiment was performed at solar radiation 780 w/m^2, and 329.1 k as cell temperature, and the actual I-V curve by the theoretical data.

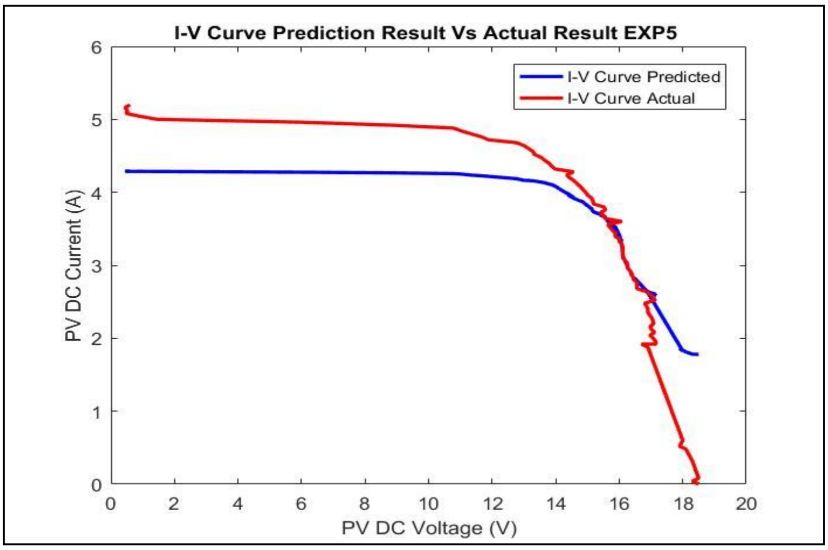


Figure 4.6: The I-V curves of the PV module at experiment five

The last test was done on experiment number seven. Figure 4.7 represents the predicted I-V curve and actual I-V curve for the PV module at experiment number seven, where this experiment was performed at solar radiation 978 w/m^2, and 328.56 k as cell temperature, and the actual I-V curve by the theoretical data.

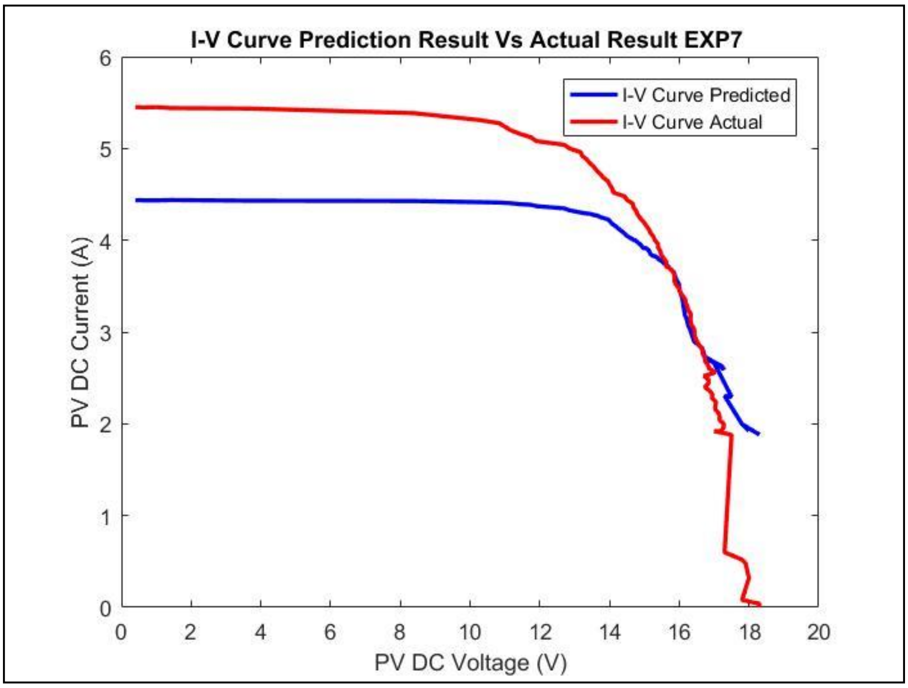


Figure 4.7: The I-V curves of the PV module at experiment Seven

In the last step of the testing stage as mentioned in Chapter 3, the data of experiment number seven was entered into the for loop to extract the optimum current-voltage curve at the optimum number of leaves. Figure 4.7 represents the optimum I-V curve of the PV module at experiment seven and comparing it with actual curve and the curve was extracted in the testing phase.

Whereas, if the number of readings for experiments increases, there is a greater convergence between the actual curve and the predicted curve of the proposed model.

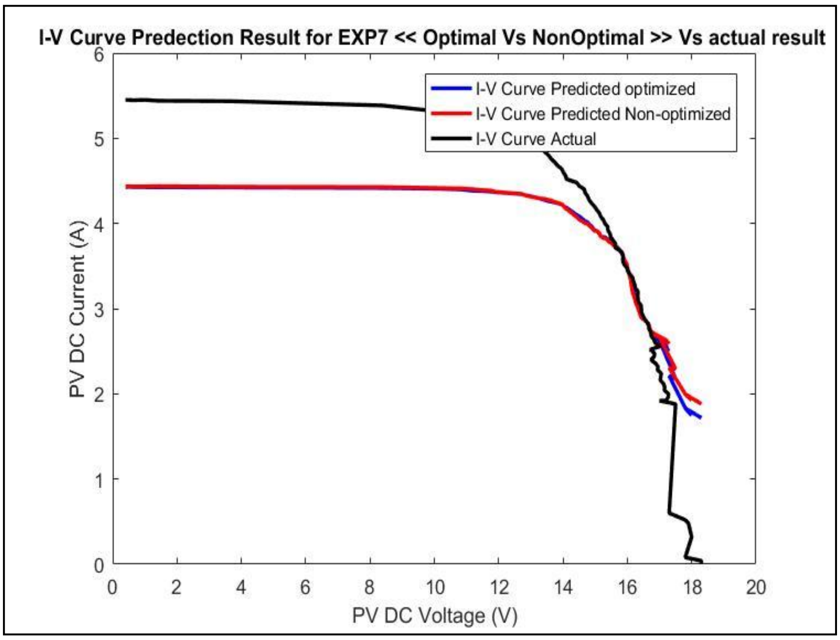


Figure 4.8: The optimum I-V curve of the PV module at experiment seven compared with actual curve and the curve was extracted in the testing phase.

Through the previous results, we note that the experiments no. (5,7) on which the training and testing processes were performed on it ; the I-V curves of the PV Module were near, and the difference between them were less than the results of Experiment No. 4 ,where it underwent the test process only. One of the characteristics of the random forest algorithm is that the differences in the results become very small When the number of training times for the data set reached a certain limit, this appeared when the number of training times was increased for experiment # 7.

In the final step in the developed RFs model, I-V curve was predicted at cell temperature equal to 318.32 k with different values of solar radiation, figure 4.9 represented it .And at solar radiation equal to 978 w/m^2 with different temperatures levels, the I-V curve of PV module was predicted as shown in figure 4.10.

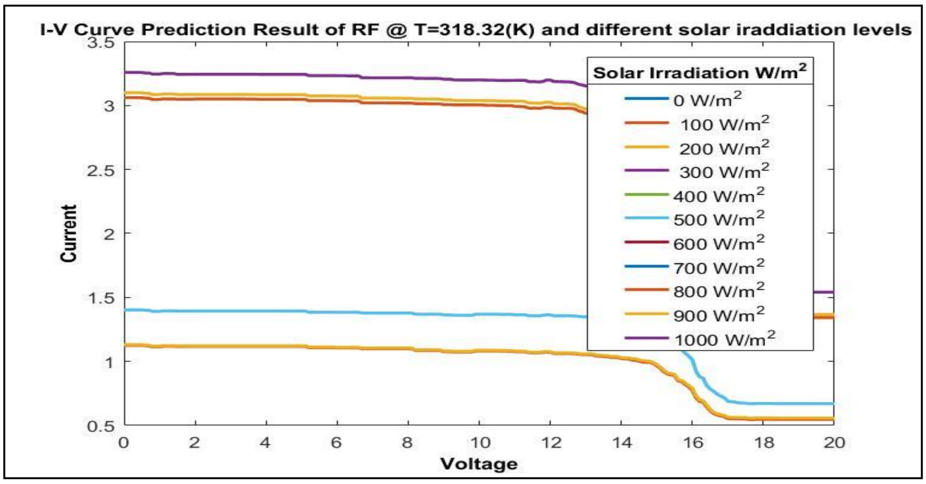


Figure 4.9: The predicted I-V curve at T=318.32 k with different solar radiation

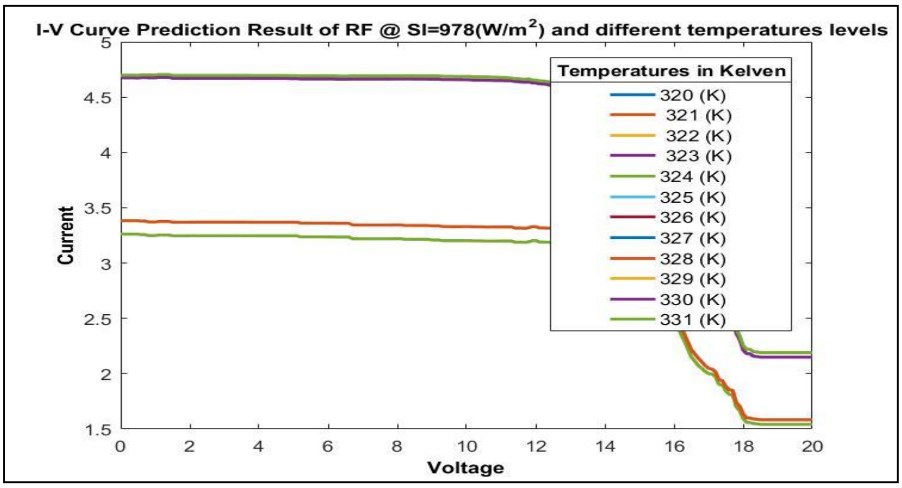


Figure 4.10: The predicted I-V curve at RS=978 W/m^2 with different cell temperature

- **Testing Result with Correction factor**

For each trial, we have some error between the results of predicted I-V curve and actual I-V curve for the PV module. In this thesis we have decrease the error by use correction factor techniques. A correction factor defined as is any mathematical modification made to a calculation to account for deviations in either the sample or the method of measurement. A correction factor was used as a factor multiplied with the result of an equation to correct for a known amount of systemic error. The MATLAB line code in Appendix D describe the RFs proposed model with correction factors are used to minimize the error between the predicted I-V curve and actual I-V curve for PV module. The figures (4.11, 4.12, 4.13, and 4.14) represent the results of RFs proposed model with correction factor.

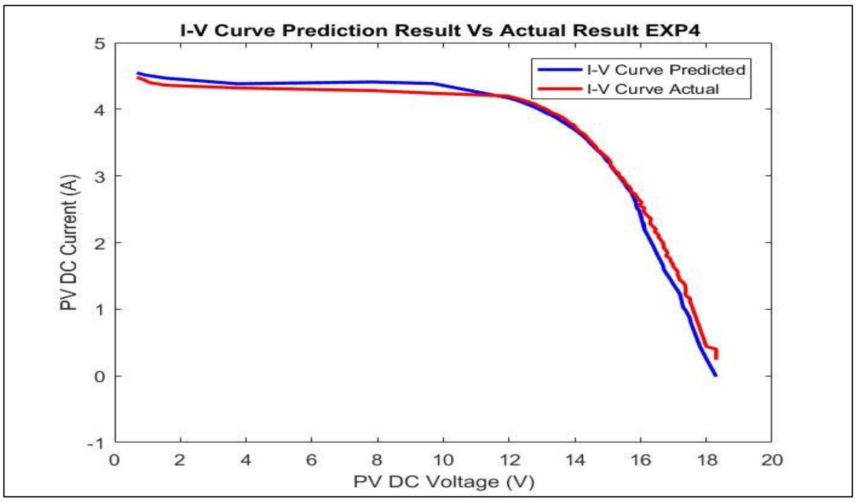


Figure 4.11: The I-V curve of the PV module at experiment four with correction factor

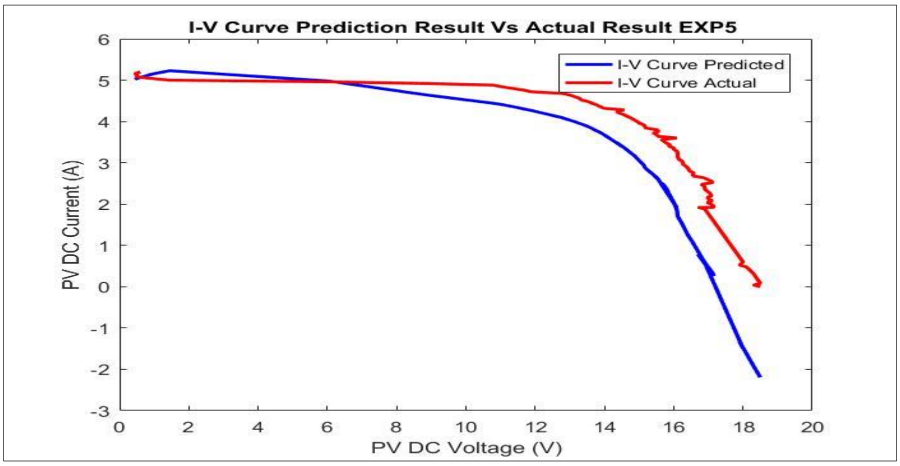


Figure 4.12: The I-V curve of the PV module at experiment five with correction factor

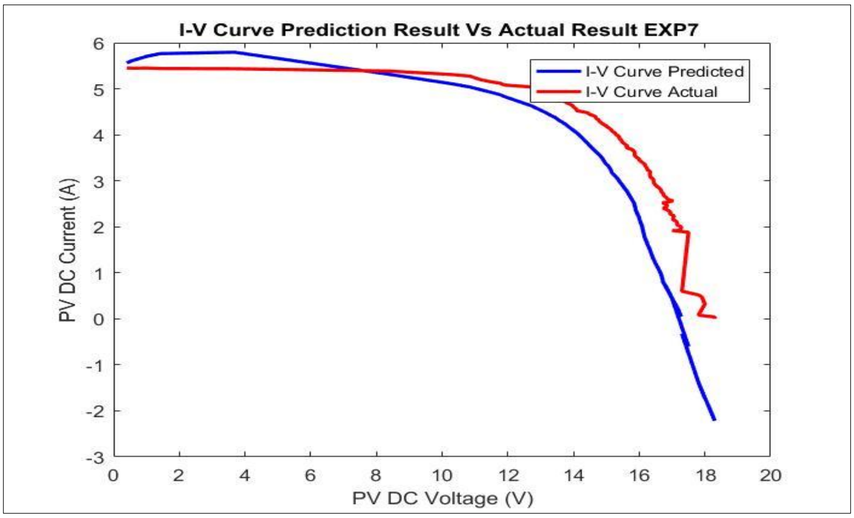


Figure 4.13: The I-V curve of the PV module at experiment seven with correction factor

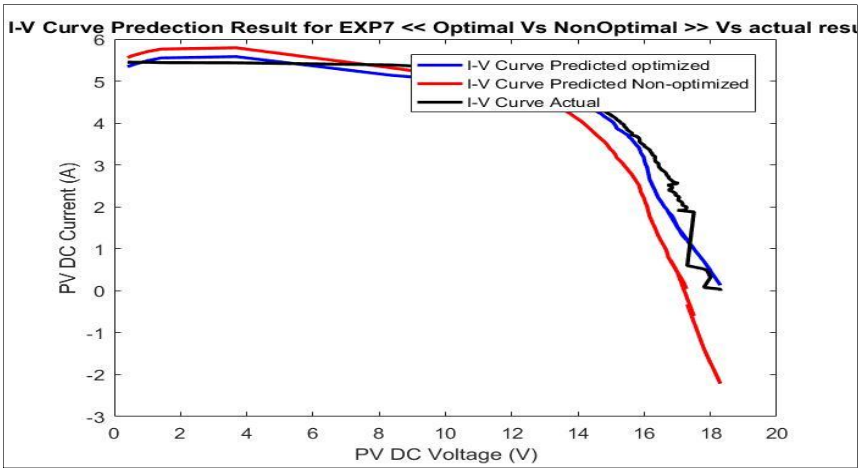


Figure 4.14: the optimum i-v curve of the pv module at experiment seven compared with actual curve and the curve was extracted in the testing phase with correction factor

4.3 Evaluation of the proposed model

The proposed model for extracting the current-voltage curve of the solar PV module was evaluated based on the values of metric errors (MBE, RMSE, and MAPE) resulting after the training and testing stages of the datasets. Table 4.2 represent the best results of metric errors for the proposed model, where these values accrued in best number of trees a best number of leaves.

Table 4.2: the result of metric errors for the proposed model.

	MBE (%)	RMSE (%)	MAPE (%)
RFs proposed model	-0.3959	0.04251	4.31509

Table 4.3, table 4.4, and table 4.5 show the result of metric errors (MBE, RMSE, and MAPE, respectively to 25,000 trials for 500 trees with 50 leaves in each tree.

Table 4.3: the result of Mean Bias Error (MBE) for RFs proposed model

Leaf / Tree	1	2 ….	5	**6** ….	10 ….	20 ….	50
1	0.18896	-0.08076	-0.17902	0.48033	-0.05771	-0.10043	0.05663
3	0.214677	0.308013	0.29900	-0.3959	0.04980	-0.08539	-0.09927
⋮							
10	-0.07209	-0.05179	0.00771	0.04949	0.06128	0.13115	0.15310
⋮							
100	0.03941	0.10096	0.02513	0.07523	0.05163	-0.00876	0.01481
⋮							
300	0.04997	0.02055	0.01544	0.03386	0.03934	0.01328	0.01351
⋮							
500	0.02929	0.04009	0.03944	0.04106	0.04679	0.03048	0.01845

Table 4.4: the result of Root Mean Square Error (RMSE) for RFs proposed model

Leaf / Tree	1	2 ….	5	**6** ….	10 ….	20….	50
1	1.88368	1.85450	0.81762	2.07870	1.85131	0.50525	0.61937
⋮							
3	1.89406	0.992791	0.598618	0.04251	1.014668	0.285986	0.473203
⋮							
10	0.51075	0.530405	0.251615	0.511941	0.534778	0.560651	0.848940
⋮							
100	0.36303	0.42492	0.28712	0.40440	0.36265	0.358138	0.582668
⋮							
300	0.33245	0.28839	0.32164	0.41448	0.41718	0.414510	0.616914
⋮							
500	0.320275	0.318481	0.371127	0.320936	0.333158	0.409864	0.592151

Table 4.5: the result of Mean Absolute Percentage Error (MAPE) for RFs proposed model

Leaf / Tree	1	2 ….	5	**6** ….	10 ….	20 ….	50
1	Inf	31.9164	16.6976	34.2800	31.9230	13.6579	17.8608
3	Inf	24.0772	17.2095	4.31509	21.9908	10.0069	14.2238
⋮							
10	Inf	15.8529	10.8672	15.2129	15.1176	16.0436	19.9223
⋮							
100	Inf	14.3310	10.7645	13.2961	12.1858	11.5250	14.6171
⋮							
300	Inf	10.9092	11.3401	13.2350	13.3008	12.6556	15.0241
⋮							
500	Inf	11.4029	12.6062	11.5126	11.7666	12.7097	14.7072

After obtaining the results of the proposed model, these results of the model were compared with some previous models to extract the I-V curve of the solar PV module using different methods. The MAPE value was used as a measure of the accuracy of these models. Table 4.6 shows the results of this comparison. This table take from (Khatib, T et al,2018)

Table 4.6: The Comparison between Different Methods of Extracting the I-V curve of the Solar PV Cell

I-V curve extraction Method	Accuracy	Type
variable resistance (Van Dyk et al., 2005)	78%	Online
capacitive load (Mahmoud, 2006; Muñoz and Lorenzo, 2006)	80%	Online
MOSFETs (Kuai and Yuvarajan, 2006)	91.6%	Online
Boost converter (Khatib et al., 2017)	(61-67)%	Online
Artificial neural network (Karatepe et al., 2006; Celik, 2011; Bonanno et al., 2012;khatib,T. et al.,2018)	(85.3-99.5)%	Offline
Numerical methods (Bai et al., 2014; Ma et al., 2014; Tossa et al., 2014; Easwarakhanthan et al., 1986; Navabi et al., 2015; Hejri et al., 2014;Villalva et al., 2009;Dkhichi,F. et al., 2014)	(90.5-99)%	Offline
Evolutionary algorithms (Sharma et al., 2012; Dizqah et al., 2014; Moldovan et al., 2009; Ismail et al., 2013; Zagrouba et al., 2010; AppelbaumandPeled, 2014; Jervase et al., 2001; Khanna et al., 2015; Ye et al., 2009; Jing Jun and Kay-Soon, 2012; Alam et al., 2015; Hasanien, 2015;Ishaque and Salam, 2011; Ishaque et al., 2011; Ishaque et al., 2012; Jiang et al., 2013; Gong and Cai, 2013; Siddiqui and Abido, 2013; Muhsenet al., 2016; Muhsen et al., 2015)	(78-98.6)%	Offline
Random Forest Technique (Ibrahim,I.A et al.,2018)	91.28%	Offline
Proposed Model	95.68%	Offline

4.4 Chapter Summary

In this chapter, the results of the proposed model for predicting the current-voltage curve for PV module were clarified. This Results showed the low values for errors. The RMSE, MAPE, and MBE values are equal to (0.04251, 4.315097, -0.3959), respectively of proposed RFs model. The value of MAPE was adopted to compare the proposed RFs model with other models. As these results clarified the Relatively high accuracy of the proposed model compared to previous models from online method and good accuracy compared with other offline method. And because of lack of data, correction factor was used to reduce the difference between the actual I-V curve and predicted I-V curve of proposed model, to obtain better result for the RFs proposed model.

Chapter V

Conclusion and future work

5.1 Conclusion

In this thesis, the aim was to predict the current-voltage curve of the photovoltaic module and to obtain the optimum state through random forests technique as new technique, then compare this model with other previous prediction models and verify the accuracy of this proposed model. MATLAB, 2016 program was used to develop this model and to train and test the data sets.

The proposed model was presented for prediction through a set of experiments on a solar PV module (STF - 120P6), and use the parameters (ambient temperature, solar radiation, PV DV voltage, short circuit current, and open circuit voltage) as inputs for this model.

The accuracy and success of the proposed model were verified by comparison with the actual model resulting from the theoretical calculations and then finding the values of metric errors (RMSE, MAPE, and MBE), where these values to (0.04251%, 4.315097%, - 0.3959%), respectively. A value of (MAPE) was relied on to verify the accuracy of the proposed model which is equal to (4.315097%) By comparison with previous results, it is clear that the proposed model is more accurate than others RFs model, But there are still other, more accurate methods of extracting the I-V curves of a solar PV module, such as artificial neural network techniques .. After obtaining the previous results, the correction factor was used to reduce the difference between the actual I-V curve and the predicted I-V curve of the proposed model.

5.2 Suggestions for Future Work

This research work included presenting a new method for predicting the PV module output current. the proposed model was achieved high accuracy compared to the results of other models. However, for further improvement and development of this research, the following suggestions are presented:

- Developing the proposed model by make a greater number of experiments at different temperatures and solar radiation to increase the number of data in the training phase and use other factors such as the inputs of meteorological variables.
- Apply the proposed model on more than one type of new solar PV module with different rated power.
- Development of the proposed model and use it in simulating solar energy systems such as On-grid system and Off-grid system.

References

- Karzam,J.(2012) ***Climate Change in the Arab World: Mechanisms of Defense and Confrontation.*** The Palestinian Case.*Palestine: Development work center*.
- BP p.l.c.(2019) ***BP Statistical Review of World Energy.68th***.
- https://ourworldindata.org/renewable-energy/(Accessed: 19 Nov 2019)
- Tiwari, G. N.,Dubey,S.(2010) ***Fundamentals of Photovoltaic Modules and Their Applications.****UK: The Royal Society of Chemistry*
- Salmi,T , Bouzguenda,M, Gastli,A, Masmoudi,A .(2012***) 'MATLAB/ Simulink Based Modelling of Solar Photovoltaic Cell'*** , *International Journal Of Renewable Energy Research*,2(2), p213-218.
- Jain, F.(2016) '**Chapter 6 :Solar Cells,Department of Electrical and Computer Engineering'**,University of Connecticut,29/11/2016
- Khatib,T. , Elmenreich,W. , (2016) **Modeling Of Photovoltaic Systems Using Matlab**.Canada: Wiley
- vanDyk, E.E. , Gxasheka,A.R.,Meyer,E.L., (2005), ***'Monitoring current–voltage characteristics and energy output of silicon photovoltaic modules',*** *Renewable Energy* ,30,p399-411.
- Muñoz,J. ,Lorenzo,E.(2006***) 'Capacitive load based on IGBTs for on-site characterization of PV arrays'***, *Solar Energy*, 80,p 1489-1497.
- Mahmoud,M.M , (2006) ***'Transient analysis of a PV power generator charginga capacitor for measurement of the I–Vcharacteristics'***,*Renewable Energy* ,31 ,p 2198–2206.

- kuai,Y. , Yuvarajan,S. , (2006) ***'An electronic load for testing photovoltaic panels'***, *Journal of Power Sources*, 154 ,p 308–313.
- Guvench, M.G., Gurcan ,C., Durgin,K. ,MacDonald, D. ,(2004), ***'Solar Simulator and I-V Measurement SystemFor Large Area Solar Cell Testing'***,*Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*.
- Durán ,E., Piliougine,M. , Sidrach-de-Cardona,M., Galan,J. ,Andújar,J. M., (2008) ***'Different methods to obtain the I– M.V curve of PV modules: A review'*** *,33rd IEEE Photovoltaic Specialists Conference*, San Diego, CA, USA, 2008.
- Durán ,E., Andújar,J. M., Enrique, J.M., Pērez-Oria, J.M.,(2012), ***'Determination of PV Generator I-V/P-V Characteristic CurvesUsing a DC-DC Converter Controlled by a Virtual Instrument'*** *,International Journal of Photoenergy*,p1-13.
- De Blas, M.A. , Torres,J.L. , Prieto,E. ,andGarcı´a, A., (2002) , ***'Selecting a suitable model for characterizingphotovoltaic devices',*** *Renewable Energy,* 25, p 371–380.
- Khatib,T. , Elmenreich,W. ,Mohamed,A. , (2017)'***Simplified I-V Characteristic Tester for PhotovoltaicModules Using a DC-DC Boost Converter',*** *Sustainability*, 9(4).
- Durán, E. ,Andújar, J.M., Enrique, J.M. , Pérez-Oria,J.M. ,(2012) ***'Determination of PV Generator I-V/P-V Characteristic CurvesUsing a DC-DC Converter Controlled by a Virtual*** Instrument',*International Journal of Photoenergy*.

- Marquardt, D.W., (1963) ***'An Algorithm for Least-Squares Estimation of Nonlinear Parameters',*** *Journal of the Society for Industrial and Applied Mathematics*, 11(2), p431-441.

- Dkhichi,F.,Oukarfi,.B , (2016) ***'Levenberg-Marquardt algorithm for parameter identification of solar cell model'*** , *XXIth Annual Seminar on Automation, Industrial Electronics and Instrumentation (SAAEI'14)*, Tangier,June 2014.

- Ma,T., Yang,H.,Lu.L,(2014)'***Development of a model to simulate the performance characteristics of crystalline silicon photovoltaic modules/strings/arrays',*** *Solar Energy* ,100,p 31–41.

- Tossa,A.K., Soro Y.M. , Azoumah, Y., Yamegueu,D.,(2014) ***'A new approach to estimate the performance and energy productivity of photovoltaic modules in real operating conditions',*** *Solar Energy,* 110 ,p543–560.

- Easwarakhanthan, T., Bottin, J., Bouhouch, I. and Boutrit, C. ,(1986) ***'Nonlinear Minimization Algorithm for Determining the Solar Cell Parameters with Microcomputers'***,*International Journal of Solar Energy*, 4(1),p1-12.

- Dizqah, AM, Maheri, A & Busawon, K 2014, ***'An accurate method for the PV Model identification based on a genetic algorithm and the interior-point method'***, *Renewable Energy*, vol. 72, p212-222.

- Ismail, M.S., Moghavvemi, M., Mahlia, T.M.I., 2013. ***Characterization of PV panel and global optimization of its model parameters using genetic algorithm***. *Energy Convers Manage*. 73, p10–25.

- Jervase, J.A. & Bourdoucen, H. & Al-lawati, Ali. (2001). ***Solar cell parameter extraction using genetic algorithms.*** *Measurement Science and Technology*. 12. 1922-1925.
- Siddiqui,M.U, and Abido,M. ,(2013) ***'Parameter estimation for five- and seven-parameter photovoltaicelectrical models using evolutionary algorithms'*** *,Applied Soft Computing*,13 (12),p 4608–4621.
- Khanna, V., Das, B.K., Bisht, D., Vandana, Singh, P.K., 2015. ***A three diode model for industrial solar cells and estimation of solar cell parameters using PSO algorithm.*** *, Renewable Energy* 78, 105–113.
- Alyasseri, Zaid and Khader, Ahamad Tajudin and Al-Betar, Mohammed and Awadallah, Mohammed and Yang, Xin-She,2018***, 'Variants of the Flower Pollination Algorithm: A Review'***, *Studies in Computational Intelligence,* P 91-118.
- Alam, D.F., Yousri, D.A., Eteiba, M.B., 2015. ***Flower pollination algorithm based solar PV parameter estimation.*** *Energy Conversion Manage*. 101, P 410–422.
- khatib,T., Ghareeb,A., Tamimi,M. , Jaber,M., Jaradat,S. ,(2018) ***'new offline method for extracting I-V characteristic curve for photovoltaic modules using artificial neura***l networks',*Solar Energy*, 173 , p 462–469.
- Karatepe, E., Boztepe, M., Colak, M., 2006. ***Neural network based solar cell model.*** *Energy Convers. Manage*. 47, P1159–1178.
- Bonanno, F., Capizzi, G., Napoli, C., Graditi, G., Tina, G., 2012. ***A radial basis function neural network approach for the electrical***

characteristics estimation of a photovoltaic module. *Applied Energy* 97, P 956–961.

- Celik, A.N., 2011. ***Artificial neural network modelling and experimental verification of the operating current of mono-crystalline photovoltaic modules.*** *Solor Energy* 85,P 2507–2517.
- Breiman, L. (2001) ***'Random Forests'***, *Machine Learning*, 45, p5–32.
- Shobha,G. ,and Rangaswamy,S. ,(2018) ***'Chapter 8 - Machine Learning'***, *Handbook of Statistics*,38,p197-228.
- MATLAB,(2016) , ***Random Forests mathwork*** , (11 Sep 2019).
- Breiman, L. (1996) ***'Bagging predictors***'*Machine Learning*, 24,p123–140.
- Breiman,L. (2002) ***Manual on setting up, using, and understanding random forests*** v3. 1. Statistics Department University of California Berkeley, CA, USA.
- Guoa,L. , Chehata ,N., Mallet,C. , Boukir,S. (2011) ***'Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests'***,*ISPRS Journal of Photogrammetry and Remote Sensing*,66 ,p56–66.
- Liaw, A., and Wiener, M (2002***)'Classification and Regression by RandomForest',*** *R News*, 2(3), p18-22.
- Ibrahim,I.A,Khatib,T.,Mohamed,A., and Elmenreich,W.,(2018), ***'Modeling of the output current of a photovoltaic grid-connected system using random forests technique'***, *Energy Exploration & Exploitation* ,36(1) , P132-148.

- Shamy ,S., Dheeba,J. , (2016) ***, 'Review of QUEST, GUIDE, CRUISE, C4.5 and RPART Classification Algorithms'*** , *International Journal of Advanced Technology in Engineering and Science* , 4 (6) , P116-123 .

Appendix A

Samples of experimental data used in the developed RFs model for predicting I-V curve for photovoltaic modules

EXP#	TEMPERATURE	SOLAR RADIATION	VOLTAGE	CURRENT
1	318.32	118.28	17.5	0
	318.32	118.28	17.202	0.04
	318.32	118.28	16.917949	0.08
	318.32	118.28	16.6	0.12
	318.32	118.28	16.863636	0.16
	318.32	118.28	16.145455	0.2
	318.32	118.28	16.505263	0.24
	318.32	118.28	16.386667	0.28
	318.32	118.28	16.369014	0.32
	318.32	118.28	16.388679	0.36
	318.32	118.28	16.053968	0.4
	318.32	118.28	15.816667	0.44
	318.32	118.28	15.549091	0.48
	318.32	118.28	15.336	0.52
	318.32	118.28	15.184285	0.56
	318.32	118.28	15.151665	0.6
	318.32	118.28	14.505357	0.64
	318.32	118.28	14.041026	0.68
	318.32	118.28	13.979268	0.72
	318.32	118.28	11.913636	0.81
	318.32	118.28	12.033333	0.8
	318.32	118.28	0.38	0.84
2	321.25	148	17	0
	321.25	148	17.0688	0.04
	321.25	148	16.9625	0.08
	321.25	148	16.818	0.12
	321.25	148	16.6939	0.16
	321.25	148	16.5526	0.2
	321.25	148	16.5079	0.24
	321.25	148	16.4356	0.28
	321.25	148	16.3479	0.32

	321.25	148	16.2391	0.36
	321.25	148	16.1125	0.4
	321.25	148	16.0339	0.44
	321.25	148	15.8708	0.48
	321.25	148	15.7375	0.52
	321.25	148	15.5468	0.56
	321.25	148	15.3899	0.6
	321.25	148	15.1718	0.64
	321.25	148	14.9697	0.68
	321.25	148	14.5769	0.72
	321.25	148	14.3078	0.76
	321.25	148	13.3289	0.8
	321.25	148	12.878	0.84
	321.25	148	9.6699	0.88
	321.25	148	11.347	0.92
⋮	⋮	⋮	⋮	⋮
7	328.56	978	18.5	0
	328.56	978	18.6	0.04
	328.56	978	18.3	0.32
	328.56	978	18.4	0.4
	328.56	978	18.1	0.56
	328.56	978	18.2	0.64
	328.56	978	18.3	0.88
	328.56	978	17.9	2.04
	328.56	978	17.7	2.12
	328.56	978	17.8	2.16
	328.56	978	17.7	2.24
	328.56	978	17.3	2.88
	328.56	978	17.2571	2.92
	328.56	978	17.2125	2.96
	328.56	978	17.2429	3
	328.56	978	17.2269	3.04
	328.56	978	17.1438	3.08
	328.56	978	17.108	3.12
	328.56	978	17.0239	3.16
	328.56	978	16.9958	3.2
	328.56	978	17.0286	3.24

	328.56	978	16.9169	3.28
	328.56	978	16.8667	3.32
	328.56	978	16.8617	3.36
	328.56	978	16.8586	3.4
	328.56	978	16.7228	3.44
	328.56	978	16.6462	3.48
	328.56	978	16.5982	3.52
	328.56	978	16.6909	3.56
	328.56	978	16.5854	3.6
	⋮			
	328.56	978	14.944	5.04
	328.56	978	14.8707	5.08
	328.56	978	14.7861	5.12
	328.56	978	14.7451	5.16
	328.56	978	14.704	5.2
	328.56	978	12.8598	5.88
	328.56	978	12.7618	5.92
	328.56	978	12.5309	5.96
	328.56	978	12.34	6
	328.56	978	12.0302	6.04
	328.56	978	11.8319	6.08
	328.56	978	11.3095	6.12
	328.56	978	10.8729	6.16
	328.56	978	10.0281	6.2
	328.56	978	6.7812	6.24
	328.56	978	2.3873	6.28
	328.56	978	1.2065	6.32
	328.56	978	0.6901	6.36
	328.56	978	0.5299	6.4
	328.56	978	0.475	6.44
	328.56	978	0.48	6.48

Appendix B

Samples of theoretical data used in the developed RFs model for predicting I-V curve for photovoltaic modules

MATH #	TEMPERATURE	SOLAR RADIATION	VOLTAGE	CURRENT
M1	318	118.28	17.4102	0
	318	118.28	17.4082	0.001
	318	118.28	17.4062	0.002
	318	118.28	17.4042	0.003
	318	118.28	17.4021	0.004
	318	118.28	17.4001	0.005
	318	118.28	17.3981	0.006
	318	118.28	17.3961	0.007
	318	118.28	17.3941	0.008
	318	118.28	17.3921	0.009
	318	118.28	17.39	0.01
	318	118.28	17.388	0.011
	318	118.28	17.386	0.012
	318	118.28	17.3839	0.013
	318	118.28	17.3819	0.014
	318	118.28	17.3799	0.015
	318	118.28	17.3778	0.016
	318	118.28	17.3758	0.017
	318	118.28	17.3737	0.018
	318	118.28	17.3717	0.019
	318	118.28	17.3696	0.02
	⋮	⋮	⋮	⋮
	318	118.28	7.6226	0.83
	318	118.28	7.2522	0.831
	318	118.28	6.8006	0.832
	318	118.28	6.2218	0.833
	318	118.28	5.4149	0.834
	318	118.28	4.0733	0.835
	318	118.28	-0.9754	0.836
M2	321.25	148	18.3005	0
	321.25	148	18.2985	0.005

	321.25	148	18.2966	0.01
	321.25	148	18.2946	0.015
	321.25	148	18.2926	0.02
	321.25	148	18.2906	0.025
	321.25	148	18.2886	0.03
	321.25	148	18.2866	0.035
	321.25	148	18.2846	0.04
	321.25	148	18.2826	0.045
	321.25	148	18.2807	0.05
	321.25	148	18.2787	0.055
	321.25	148	18.2767	0.06
	321.25	148	18.2747	0.065
	321.25	148	18.2727	0.07
	321.25	148	18.2707	0.075
	321.25	148	18.2687	0.08
	321.25	148	18.2666	0.085
	321.25	148	18.2646	0.09
	321.25	148	18.2626	0.095
	321.25	148	18.2606	0.1
	⋮	⋮	⋮	⋮
	321.25	148	8.5672	5.385
	321.25	148	8.3775	5.39
	321.25	148	8.1694	5.395
	321.25	148	7.9391	5.4
	321.25	148	7.6812	5.405
	321.25	148	7.3882	5.41
	321.25	148	7.0489	5.415
	321.25	148	6.646	5.42
	321.25	148	6.1501	5.425
	321.25	148	5.5047	5.43
	321.25	148	4.5789	5.435
	321.25	148	2.9206	5.44
	321.25	148	-1.0368	5.445
	⋮	⋮	⋮	⋮
M7	328.56	978	18.4009	0
	328.56	978	18.3979	0.01
	328.56	978	18.395	0.02
	328.56	978	18.3921	0.03

	328.56	978	18.3892	0.04
	328.56	978	18.3862	0.05
	328.56	978	18.3833	0.06
	328.56	978	18.3804	0.07
	328.56	978	18.3774	0.08
	328.56	978	18.3745	0.09
	328.56	978	18.3715	0.1
	328.56	978	18.3685	0.11
	328.56	978	18.3656	0.12
	328.56	978	18.3626	0.13
	328.56	978	18.3596	0.14
	328.56	978	18.3566	0.15
	328.56	978	18.3536	0.16
	328.56	978	18.3506	0.17
	328.56	978	18.3476	0.18
	328.56	978	18.3446	0.19
	328.56	978	18.3416	0.2
	⋮	⋮	⋮	⋮
	328.56	978	10.5631	6.17
	328.56	978	10.4054	6.18
	328.56	978	10.2351	6.19
	328.56	978	10.0498	6.2
	328.56	978	9.8469	6.21
	328.56	978	9.6225	6.22
	328.56	978	9.3716	6.23
	328.56	978	9.0872	6.24
	328.56	978	8.7587	6.25
	328.56	978	8.3703	6.26
	328.56	978	7.8949	6.27
	328.56	978	7.2821	6.28
	328.56	978	6.4189	6.29
	328.56	978	4.9458	6.3
	328.56	978	-5.8886	6.31

Appendix C

MATLAB code used to develop RFs model for predicting I-V curve for photovoltaic modules

```
%%Random Forests-Regression --- Prediction - First Stage%
clear; clc;
 load('DATA.mat');load('NEW_DATA.mat');load('Input_EXP.mat');load('TEST1.mat');
%((TRAINING STAGE))%
CURRENT_EXP1_7 = [Output_EXP1, Output_EXP2, Output_EXP3,Output_EXP4, Output_EXP5
Output_EXP6,Output_EXP7];
INPUT_EXP1_7 = [Input_EXP1, Input_EXP2, Input_EXP3,Input_EXP4, Input_EXP5
Input_EXP6,Input_EXP7];
S=INPUT_EXP1_7(2,:); % Solar Radiation
T=INPUT_EXP1_7(1,:); % Ambient Temp (°C)
V=INPUT_EXP1_7(3,:); % experimantal PV DC Voltage (V)
I=CURRENT_EXP1_7;    % experimantal PV DC Current (A)
%Delete EXP4 From training stage
CURRENT_EXP1_7(:,[97:187])=[];
INPUT_EXP1_7(:,[97:187])=[];
DATA_EXP([97:187],:)=[];
Out_EXP1_7([97:187],:)=[];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%RF_Training Code%
Y=table2array(Out_EXP1_7);               % Split data into response array
ticID=tic;
X=[DATA_EXP(:,1:3),DATA_EXP(:,5:6)];     % Split data into predictor array
t=500;                                   % Trees Number
%B=TreeBagger(t,X,Y,'method','regression','oobpred','on');
B=TreeBagger(t,X,Y,'method','classification','oobpred','on');
view(B.Trees{t},'Mode','Graph');
%Estimating Variable Importance
B=TreeBagger(t,X,Y,'method','regression','oobvarimp','on');
figure('name','Estimating Variable Importance','NumberTitle','off');
plot(oobError(B),'LineWidth',2)
xlabel('Number of Grown Trees');
ylabel('Out-of-Bag Mean Squared Error');
%%Most Important Variables
figure('name','Most Important Variables','NumberTitle','off');
bar(B.OOBPermutedVarDeltaError)
title('Variable Importance');
xlabel('Variable Number');
ylabel('Out-of-Bag Variable Importance');
legend({'1: Ambient Temp,, 2: Solar Radiation, 3: PV DC Voltage'});
nidx = find(B.OOBPermutedVarDeltaError<0.65);%Imposing an arbitrary cutoff at
0.65 - Not Important Variables
%%Fraction of in-Bag Observation "Which observations are out of bag for whic
```

```
trees"
finbag = zeros(1,B.NTrees);

for t=1:B.NTrees
finbag(t)=sum(all(~B.OOBIndices(:,1:t),2));
end

finbag = finbag/size(X,1);
figure('name','Which observations are out of bag for which trees','NumberTitle','off')
plot(finbag,'LineWidth',2)
xlabel('Number of Grown Trees');
ylabel('Fraction of in-Bag Observations');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Finding The Outliers

BI=fillProximities(B); %Proximity Matrix that used

figure('name','The Outliers','NumberTitle','off')
%hist(BI.OutlierMeasure)
histogram(BI.OutlierMeasure)
title('The Outliers');
xlabel('Outlier Measure');
ylabel('Number of Observations');

%%Discovering Clusters in the data

figure('name','Discovering Clusters in the data','NumberTitle','off')
[~,e] = mdsProx(BI,'colors','K');
title('Cluster Analysis');
xlabel('1st Scaled Coordinate');
ylabel('2nd Scaled Coordinate')

%%Assess the Relative Importance of the scaled axes by plotting the first 20 eigenvalues
figure('name','Assess the Relative Importance of the scaled axes by plotting the first 20 eigenvalues','NumberTitle','off')
bar(e(1:20));
xlabel('Scaled Coordinate Index');
ylabel('Eigen Value');

%Saving The compact version of the Ensemble
compact(B);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% ((TESTING STAGE))%%%

%Testing data - For EXP4

Vtest4 = Input_EXP4(3,:)';
Itest4 = Output_EXP4';
%RF_Testing Code
Xtest4=Input_EXP4';
```

```
[Yfit4,~]= predict(B,Xtest4);
figure('name','EXP4_IV_Curves','NumberTitle','off')
plot (Vtest4,Yfit4,'blue','LineWidth',2)
hold on
plot (Vtest4,Itest4, 'red','LineWidth',2)
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
title ('I-V Curve Prediction Result Vs Actual Result EXP4 ','LineWidth',14,'FontWeight','bold','Color','k')
hold off

figure('name','EXP4-ERROR','NumberTitle','off')
E = abs(Itest4-Yfit4);
plot(E,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Average Deviation Indicator
MBE=(sum(Itest4(:)-Yfit4(:)))./numel(Itest4);
if (MBE<0)
F='Over forecasted';
elseif (MBE>0)
F='Under Forecasted';
elseif (MBE==o)
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest4(:)-Yfit4(:))./Itest4(:)))./numel(Itest4))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest4(:)-Yfit4(:)).^2)/numel(Itest4);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency Indicator}'];
disp(n1)
disp(n2)
disp(n3)
disp(n4)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Testing data - For EXP5

Vtest5 = Input_EXP5(3,:)';
```

```
Itest5 = Output_EXP5';
%load('Data_EXP1_6.mat');
%RF_Testing Code
Xtest5=Input_EXP5';
[Yfit5,~]= predict(B,Xtest5);
figure('name','EXP5_IV_Curves','NumberTitle','off')
plot (Vtest5,Yfit5,'blue','LineWidth',2)
hold on
plot (Vtest5,Itest5, 'red','LineWidth',2)
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
title ('I-V Curve Prediction Result Vs Actual Result EXP5 ','LineWidth',14,'FontWeight','bold','Color','k')
hold off

figure('name','EXP5-ERROR','NumberTitle','off')
E = abs(Itest5-Yfit5);
plot(E,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Average Deviation Indicator
MBE=(sum(Itest5(:)-Yfit5(:)))./numel(Itest5);
if MBE<0
F='Over forecasted';
elseif MBE>0
F='Under Forecasted';
elseif MBE==o
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest5(:)-Yfit5(:))./Itest5(:)))./numel(Itest5))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest5(:)-Yfit5(:)).^2)/numel(Itest5);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency Indicator}'];
disp(n1)
disp(n2)
disp(n3)
disp(n4)
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% ((TESTING STAGE))%%
%Testing data - For EXP7
Vtest7 = Input_EXP7(3,:)';
Itest7 = Output_EXP7';

%RF_Testing Code
Xtest7=Input_EXP7';
[Yfit7,node]= predict(B,Xtest7);
figure('name','EXP7_IV_Curves','NumberTitle','off')
plot (Vtest7,Yfit7,'blue','LineWidth',2)
hold on
plot (Vtest7,Itest7, 'red','LineWidth',2)
title ('I-V Curve Prediction Result Vs Actual Result EXP7 ','LineWidth',14,'FontWeight','bold','Color','k');
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
hold off
figure('name','EXP7-ERROR','NumberTitle','off')
E = abs(Itest7-Yfit7);
plot(E,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Averag Deviation Indicator
MBE=(sum(Itest7(:)-Yfit7(:)))./numel(Itest7);
if MBE<0
F='Over forecasted';
elseif MBE>0
F='Under Forecasted';
elseif MBE==o
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest7(:)-Yfit7(:))./Itest7(:)))./numel(Itest7))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest7(:)-Yfit7(:)).^2)/numel(Itest7);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency Indicator}'];
disp(n1)
```

```
disp(n2)
disp(n3)
disp(n4)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

MSE2=[];
MBE2=[];
MAPE2=[];
RMSE1=[];
RMSE2=[];
v=[];

for t=1:1:500
    for l=1:1:50
tic;
G=TreeBagger(t,X,Y,'method','regression','oobpred','on','oobvarimp','on','minleaf',l);
G.NumTrees;
%%Saving The compact version of the Ensemble
compact(G);

%((TESTING STAGE))%%
%RF_Testing Code
[YfitO,node]= predict(G,Xtest7);
%v(t,l)=toc;
v(t,l)=toc;
Eo = abs(Itest7-YfitO);
%RF-Performance
%Mean B ias Error (MBE) or Mean Forecasting Error (MFE) in Amp. // Average Deviation Indicator
MBE2(t,l)=(sum(Itest7(:)-YfitO(:)))./numel(Itest7);
if MBE2<0
F='Over forecasted';
elseif MBE2>0
F='Under Forecasted';
elseif MBE2==0
F='Ideal Forecasted';
end

% Mean Square Error
MSE1(t,l) = mse(Eo);

% Root Mean Square Error
RMSE1(t,l) = sqrt(MSE1(t,l));

%RMSE2(t,l)=sum((Itest7(:)-YfitO(:)).^2)/numel(Itest7);

% Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE(t,l)= mean((abs(Eo(2:end)./Itest7(2:end))));
%MAPE2(t,1) = (abs((sum((Itest7(:)-YfitO(:))./Itest7(:)))./numel(Itest7))).*100;
%MAPE2(t,l)=(sum(abs(Eo(:))./(sum(Itest7(:))))).*100;

```

```

end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% finding optimal number for Tree number and minleave number
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

mi=min(RMSE1(t,1));
[A,Z] = find(RMSE1(t,1)==mi)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Testing Result
%%%%%%%%%%%%%%%%%%%%5%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Go=TreeBagger(t,X,Y,'method','regression','oobpred','on','oobvarimp','on','minleaf',1);
[Yfito7,node]= predict(Go,Xtest7);
%view(Go.Trees{500},'Mode','Graph');
figure('name','Optimizing Result IV_Curve','NumberTitle','off')
plot (Vtest7,Yfito7,'blue','LineWidth',2)
hold on
plot (Vtest7,Yfit7,'red','LineWidth',2)
plot (Vtest7,Itest7,'k','LineWidth',2)
title ('I-V Curve Predection Result for EXP7 << Optimal Vs NonOptimal >> Vs actual result ','LineWidth',14,'FontWeight','bold','Color','k');
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted optimized ','I-V Curve Predicted Non-optimized','I-V Curve Actual'});
hold off


figure('name','EXP7 Optimize Vs Non-Optimze ERROR','NumberTitle','off')
E =abs(Itest7-Yfit7);
plot(E,'r','LineWidth',2)
hold on
Eoo = abs(Itest7 - Yfito7);
plot(Eoo,'k','LineWidth',2)
xlabel('pv dc cureent');
ylabel('Magnitude (A)');
title('Error');
legend({'Exp7 Error ','Exp7 Error Optimize'});
hold off

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%TEST DATA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Yt = predict(B,Test1');
figure('name','Test1','NumberTitle','off')
Yt1=Yt';
plot(Test1(3,1:201),Yt1(1:201),Test1(3,202:402),Yt1(202:402),Test1(3,403:603),Yt1(403:603),Test1(3,604:804),Yt1(604:804),Test1(3,805:1005),Yt1(805:1005),Test1(3,1006
```

```
1206),Yt1(1006:1206),Test1(3,1207:1407),Yt1(1207:1407),Test1(3,1408:1608),Yt1(1408
1608),Test1(3,1609:1809),Yt1(1609:1809),Test1(3,1810:2010),Yt1(1810:2010),Test1
(3,2011:2211),Yt1(2011:2211),'LineWidth',2)
title ('I-V Curve Prediction Result of RF @ T=318.32(K) and different sola
iraddiation levels ','LineWidth',14,'FontWeight','bold','Color','k')
xlabel('Voltage','LineWidth',14,'FontWeight','bold','Color','k')
ylabel('Current','LineWidth',14,'FontWeight','bold','Color','k')
lgd=legend({'0 W/m^2',' 100 W/m^2',' 200 W/m^2', ' 300 W/m^2', '400 W/m^2', '500
W/m^2', '600 W/m^2','700 W/m^2','800 W/m^2','900 W/m^2','1000 W/m^2'},'FontSize',
10,'TextColor','black');
title(lgd,'Solar Irradiation W/m^2')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Ytt=predict(B,Test3');
figure('name','Test3','NumberTitle','off')
Yt3=Ytt';
plot(Test3(3,1:201),Yt3(1:201),Test3(3,202:402),Yt3(202:402),Test3(3,403:603),Yt3
(403:603),Test3(3,604:804),Yt3(604:804),Test3(3,805:1005),Yt3(805:1005),Test3(3,1006
1206),Yt3(1006:1206),Test3(3,1207:1407),Yt3(1207:1407),Test3(3,1408:1608),Yt3(1408
1608),Test3(3,1609:1809),Yt3(1609:1809),Test3(3,1810:2010),Yt3(1810:2010),Test3
(3,2011:2211),Yt3(2011:2211),Test3(3,2212:2412),Yt3(2212:2412),'LineWidth',2)
title ('I-V Curve Prediction Result of RF @ SI=978(W/m^2) and differen
temperatures levels ','LineWidth',14,'FontWeight','bold','Color','k')
xlabel('Voltage','LineWidth',14,'FontWeight','bold','Color','k')
ylabel('Current','LineWidth',14,'FontWeight','bold','Color','k')
lgd=legend({'320 (K) ',' 321 (K) ',' 322 (K) ', ' 323 (K) ', '324 (K) ', '325
(K)', '326 (K) ','327 (K) ','328 (K) ','329 (K) ','330 (K) ','331 (K) '},'FontSize',
10,'TextColor','black');
title(lgd,'Temperatures in Kelven')
```

Appendix D

MATLAB code used to develop RFs model for predicting I-V curve for photovoltaic modules with correction factor

```
%%Random Forests-Regression --- Prediction - First Stage%

clear; clc;
load('DATA.mat'); load('NEW_DATA.mat'); load('Input_EXP.mat'); load('TEST1.mat');

%((TRAINING STAGE))%

CURRENT_EXP1_7 = [Output_EXP1, Output_EXP2, Output_EXP3,Output_EXP4, Output_EXP5, Output_EXP6,Output_EXP7];
INPUT_EXP1_7 = [Input_EXP1, Input_EXP2, Input_EXP3,Input_EXP4, Input_EXP5, Input_EXP6,Input_EXP7];

S=INPUT_EXP1_7(2,:); % Solar Radiation
T=INPUT_EXP1_7(1,:); % Ambient Temp (°C)
V=INPUT_EXP1_7(3,:); % experimantal PV DC Voltage (V)
I=CURRENT_EXP1_7;    % experimantal PV DC Current (A)

%Delete EXP4 From training stage
CURRENT_EXP1_7(:,[97:187])=[];
INPUT_EXP1_7(:,[97:187])=[];
DATA_EXP([97:187],:)=[];
Out_EXP1_7([97:187],:)=[];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%RF_Training Code%

Y=table2array(Out_EXP1_7);               % Split data into response array
ticID=tic;
X=[DATA_EXP(:,1:3),DATA_EXP(:,5:6)];     % Split data into predictor array
t=500;                                   % Trees Number


%B=TreeBagger(t,X,Y,'method','regression','oobpred','on');
B=TreeBagger(t,X,Y,'method','classification','oobpred','on');
view(B.Trees{t},'Mode','Graph');

%Estimating Variable Importance
B=TreeBagger(t,X,Y,'method','regression','oobvarimp','on');
figure('name','Estimating Variable Importance','NumberTitle','off');
plot(oobError(B),'LineWidth',2)
xlabel('Number of Grown Trees');
ylabel('Out-of-Bag Mean Squared Error');

%%Most Important Variables
figure('name','Most Important Variables','NumberTitle','off');
bar(B.OOBPermutedVarDeltaError)
title('Variable Importance');
xlabel('Variable Number');
ylabel('Out-of-Bag Variable Importance');
legend({'1: Ambient Temp,, 2: Solar Radiation, 3: PV DC Voltage'});
nidx = find(B.OOBPermutedVarDeltaError<0.65);%Imposing an arbitrary cutoff at 0.65 - Not Important Variables

%%Fraction of in-Bag Observation "Which observations are out of bag for which trees"
```

```
finbag = zeros(1,B.NTrees);

for t=1:B.NTrees
finbag(t)=sum(all(~B.OOBIndices(:,1:t),2));
end

finbag = finbag/size(X,1);
figure('name','Which observations are out of bag for which trees','NumberTitle','off')
plot(finbag,'LineWidth',2)
xlabel('Number of Grown Trees');
ylabel('Fraction of in-Bag Observations');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%Finding The Outliers

BI=fillProximities(B); %Proximity Matrix that used

figure('name','The Outliers','NumberTitle','off')
%hist(BI.OutlierMeasure)
histogram(BI.OutlierMeasure)
title('The Outliers');
xlabel('Outlier Measure');
ylabel('Number of Observations');

%%Discovering Clusters in the data

figure('name','Discovering Clusters in the data','NumberTitle','off')
[~,e] = mdsProx(BI,'colors','K');
title('Cluster Analysis');
xlabel('1st Scaled Coordinate');
ylabel('2nd Scaled Coordinate')

%%Assess the Relative Importance of the scaled axes by plotting the first 20 eigenvalues
figure('name','Assess the Relative Importance of the scaled axes by plotting the first 20 eigenvalues','NumberTitle','off')
bar(e(1:20));
xlabel('Scaled Coordinate Index');
ylabel('Eigen Value');

%Saving The compact version of the Ensemble
compact(B);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% ((TESTING STAGE))%%%

%Testing data - For EXP4

Vtest4 = Input_EXP4(3,:)';
Itest4 = Output_EXP4';
%RF_Testing Code
Xtest4=Input_EXP4';
[Yfit4,~]= predict(B,Xtest4);
```

```
YC4 = -4e-05*(Vtest4.^4) - 0.0003*(Vtest4.^3) + 0.0162*(Vtest4.^2) - 0.1142*
(Vtest4) + 1.633;
Yfit4 =Yfit4+YC4;
figure('name','EXP4_IV_Curves','NumberTitle','off')
plot (Vtest4,(Yfit4),'blue','LineWidth',2)
hold on
plot (Vtest4,Itest4, 'red','LineWidth',2)
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
title ('I-V Curve Prediction Result Vs Actual Result EXP4 ','LineWidth',
14,'FontWeight','bold','Color','k')
hold off

figure('name','EXP4-ERROR','NumberTitle','off')
E = abs(Itest4-Yfit4);
plot(E,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Averag
Deviation Indicator
MBE=(sum(Itest4(:)-Yfit4(:)))./numel(Itest4);
if (MBE<0)
F='Over forecasted';
elseif (MBE>0)
F='Under Forecasted';
elseif (MBE==o)
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest4(:)-Yfit4(:))./Itest4(:))).numel(Itest4))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest4(:)-Yfit4(:)).^2)/numel(Itest4);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy
Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency
Indicator}'];
disp(n1)
disp(n2)
disp(n3)
disp(n4)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Testing data - For EXP5
```

```

Vtest5 = Input_EXP5(3,:)';
Itest5 = Output_EXP5';
%load('Data_EXP1_6.mat');
%RF_Testing Code
Xtest5=Input_EXP5';
[Yfit5,~]= predict(B,Xtest5);

YC5 = -0.0003*(Vtest5.^4) + 0.0095*(Vtest5.^3) - 0.1066*(Vtest5.^2) + 0.3832*(Vtest5) + 0.5514;
Yfit5 = Yfit5 + YC5
figure('name','EXP5_IV_Curves','NumberTitle','off')
plot (Vtest5,(Yfit5),'blue','LineWidth',2)
hold on
plot (Vtest5,Itest5, 'red','LineWidth',2)
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
title ('I-V Curve Prediction Result Vs Actual Result EXP5 ','LineWidth',14,'FontWeight','bold','Color','k')
hold off

figure('name','EXP5-ERROR','NumberTitle','off')
E5 = abs(Itest5-Yfit5);
plot(E5,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Average Deviation Indicator
MBE=(sum(Itest5(:)-Yfit5(:)))./numel(Itest5);
if MBE<0
F='Over forecasted';
elseif MBE>0
F='Under Forecasted';
elseif MBE==o
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest5(:)-Yfit5(:))./Itest5(:)))./numel(Itest5))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest5(:)-Yfit5(:)).^2)/numel(Itest5);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency
```

```
Indicator}'];
disp(n1)
disp(n2)
disp(n3)
disp(n4)


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%

% ((TESTING STAGE))%%
%Testing data - For EXP7
Vtest7 = Input_EXP7(3,:)';
Itest7 = Output_EXP7';

%RF_Testing Code
Xtest7=Input_EXP7';
[Yfit7,node]= predict(B,Xtest7);
figure('name','EXP7_IV_Curves','NumberTitle','off')

Yc = -0.0003*(Vtest7.^4) + 0.009*(Vtest7.^3) - 0.0979*(Vtest7.^2) + 0.3513*Vtest
+ 0.9562 %Excel equation
Yfit7 = Yfit7 + Yc;
plot (Vtest7,(Yfit7),'blue','LineWidth',2)
hold on
plot (Vtest7,Itest7, 'red','LineWidth',2)
title ('I-V Curve Prediction Result Vs Actual Result EXP7 ','LineWidth',
14,'FontWeight','bold','Color','k');
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted','I-V Curve Actual'});
hold off
figure('name','EXP7-ERROR','NumberTitle','off')
E = abs(Itest7-Yfit7);
plot(E,'LineWidth',2)
xlabel('pv dc current');
ylabel('Magnitude (A)');
title('Error');
toc(ticID);
%RF-Performance
%Mean Bias Error (MBE) or Mean Forecasting Error (MFE) in Amp.// Averag
Deviation Indicator
MBE=(sum(Itest7(:)-Yfit7(:)))./numel(Itest7);
if MBE<0
F='Over forecasted';
elseif MBE>0
F='Under Forecasted';
elseif MBE==o
F='Ideal Forecasted';
end
%Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE =(abs((sum((Itest7(:)-Yfit7(:))./Itest7(:)))./numel(Itest7))*100);
%Root Mean Square Error (RMSE) in Amp. // Efficiency Indicator
RMSE=sum((Itest7(:)-Yfit7(:)).^2)/numel(Itest7);
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Outputs
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n1=['Mean Bias Error(MBE): ',num2str(MBE),'(A)','{Average Deviation Indicator}'];
n2='Forecasting Status:F';
n3=['Mean Absolute Percentage Error (MAPE):',num2str(MAPE),'%','{Accuracy Indicator}'];
n4=['Root Mean Square Error (RMSE): ',num2str(RMSE),'(A)','{Efficiency Indicator}'];
disp(n1)
disp(n2)
disp(n3)
disp(n4)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Mean Bias Error
MBE_Values=[];
% Mean Square Error
MSE_Values=[];

% Root Mean Square Error
RMSE_Values=[];
% Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE_Values=[];

MSE2=[];
MBE2=[];
MAPE2=[];
RMSE1=[];
RMSE2=[];
v=[];

for t=1:1:500
    for l=1:1:50
tic;
G=TreeBagger(t,X,Y,'method','regression','oobpred','on','oobvarimp','on','minleaf',l);
G.NumTrees;
%%Saving The compact version of the Ensemble
compact(G);

%((TESTING STAGE))%%
%RF_Testing Code
```

```
[YfitO,node]= predict(G,Xtest7);
%v(t,l)=toc;
v(t,l)=toc;
Eo = Itest7-YfitO;
%RF-Performance
%Mean B ias Error (MBE) or Mean Forecasting Error (MFE) in Amp. // Average Deviation Indicator
MBE2(t,l)=(sum(Itest7(:)-YfitO(:)))./numel(Itest7);
if MBE2<0
F='Over forecasted';
elseif MBE2>0
F='Under Forecasted';
elseif MBE2==0
F='Ideal Forecasted';
end


% Mean Bias Error
MBE_Values(t,l) = mean(Eo);
% Mean Square Error
MSE_Values(t,l) = mse(Eo);
% Root Mean Square Error
RMSE_Values(t,l) = sqrt(MSE_Values(t,l));
% Mean Absolute Percentage Error (MAPE) // Accuracy Indicator
MAPE_Values(t,l)= mean((abs(Eo(2:end)./Itest7(2:end))));
%MAPE2(t,l) = (abs((sum((Itest7(:)-YfitO(:))./Itest7(:)))./numel(Itest7))).*100;
%MAPE2(t,l)=(sum(abs(Eo(:))./(sum(Itest7(:))))).*100;
%RMSE2(t,l)=sum((Itest7(:)-YfitO(:)).^2)/numel(Itest7);
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% finding optimal number for Tree number and minleave number
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

mi=min(RMSE_Values(t,1));
[A,Z] = find(RMSE_Values(t,1)==mi)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Testing Result
%%%%%%%%%%%%%%%%%%%5%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Go=TreeBagger(t,X,Y,'method','regression','oobpred','on','oobvarimp','on','minleaf',l);
[Yfito7,node]= predict(Go,Xtest7);
Y7O = -0.0003*(Vtest7.^4) + 0.0095*(Vtest7.^3) - 0.1036*(Vtest7.^2) + 0.3686(Vtest7) + 0.9635;

Yfito7 = Yfito7 +Y7O
%view(Go.Trees{500},'Mode','Graph');
figure('name','Optimizing Result IV_Curve','NumberTitle','off')
plot (Vtest7,(Yfito7),'blue','LineWidth',2)
hold on

```

```
plot (Vtest7,(Yfit7),'red','LineWidth',2)
plot (Vtest7,Itest7,'k','LineWidth',2)
title ('I-V Curve Predection Result for EXP7 << Optimal Vs NonOptimal >> V
actual result ','LineWidth',14,'FontWeight','bold','Color','k');
xlabel('PV DC Voltage (V)');
ylabel('PV DC Current (A)');
legend({'I-V Curve Predicted optimized ','I-V Curve Predicted Non-optimized','I-V
Curve Actual'});
hold off


figure('name','EXP7 Optimize Vs Non-Optimze ERROR','NumberTitle','off')
E =abs(Itest7-Yfit7);
plot(E,'r','LineWidth',2)
hold on
Eoo = abs(Itest7 - Yfito7);
plot(Eoo,'k','LineWidth',2)
xlabel('pv dc cureent');
ylabel('Magnitude (A)');
title('Error');
legend({'Exp7 Error ','Exp7 Error Optimize'});
hold off

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%TEST DATA
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Yt = predict(B,Test1');
figure('name','Test1','NumberTitle','off')
Yt1=Yt';
plot(Test1(3,1:201),Yt1(1:201),Test1(3,202:402),Yt1(202:402),Test1(3,403:603),Yt
(403:603),Test1(3,604:804),Yt1(604:804),Test1(3,805:1005),Yt1(805:1005),Test1(3,1006
1206),Yt1(1006:1206),Test1(3,1207:1407),Yt1(1207:1407),Test1(3,1408:1608),Yt1(1408
1608),Test1(3,1609:1809),Yt1(1609:1809),Test1(3,1810:2010),Yt1(1810:2010),Test
(3,2011:2211),Yt1(2011:2211),'LineWidth',2)
title ('I-V Curve Prediction Result of RF @ T=318.32(K) and different sola
iraddiation levels ','LineWidth',14,'FontWeight','bold','Color','k')
xlabel('Voltage','LineWidth',14,'FontWeight','bold','Color','k')
ylabel('Current','LineWidth',14,'FontWeight','bold','Color','k')
lgd=legend({'0 W/m^2',' 100 W/m^2',' 200 W/m^2', ' 300 W/m^2', '400 W/m^2', '500
W/m^2', '600 W/m^2','700 W/m^2','800 W/m^2','900 W/m^2','1000 W/m^2'},'FontSize',
10,'TextColor','black');
title(lgd,'Solar Irradiation W/m^2')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Ytt=predict(B,Test3');
figure('name','Test3','NumberTitle','off')
Yt3=Ytt';
plot(Test3(3,1:201),Yt3(1:201),Test3(3,202:402),Yt3(202:402),Test3(3,403:603),Yt
(403:603),Test3(3,604:804),Yt3(604:804),Test3(3,805:1005),Yt3(805:1005),Test3(3,1006
1206),Yt3(1006:1206),Test3(3,1207:1407),Yt3(1207:1407),Test3(3,1408:1608),Yt3(1408
1608),Test3(3,1609:1809),Yt3(1609:1809),Test3(3,1810:2010),Yt3(1810:2010),Test
(3,2011:2211),Yt3(2011:2211),Test3(3,2212:2412),Yt3(2212:2412),'LineWidth',2)
title ('I-V Curve Prediction Result of RF @ SI=978(W/m^2) and differen
```

```
temperatures levels ','LineWidth',14,'FontWeight','bold','Color','k')
xlabel('Voltage','LineWidth',14,'FontWeight','bold','Color','k')
ylabel('Current','LineWidth',14,'FontWeight','bold','Color','k')
lgd=legend({'320 (K) ',' 321 (K) ',' 322 (K) ', ' 323 (K) ', '324 (K) ', '325 (K)', '326 (K) ','327 (K) ','328 (K) ','329 (K) ','330 (K) ','331 (K) '},'FontSize',10,'TextColor','black');
title(lgd,'Temperatures in Kelven')
```

جامعة النجاح الوطنية
كلية الدراسات العليا

توقع منحى التيار – الفولتية للخلايا الشمسية باستخدام تقنية الغابات العشوائية

إعداد
أريج أحمد حسن عليا

إشراف
د. تامر الخطيب

قدمت هذه الأطروحة استكمالا لمتطلبات الحصول على درجة الماجستير في هندسة الطاقة النظيفة وترشيد الاستهلاك بكلية الدراسات العليا في جامعة النجاح الوطنية، نابلس – فلسطين

2020

توقع منحى التيار – الفولتية للخلايا الشمسية باستخدام تقنية الغابات العشوائية

إعداد

أريج أحمد حسن عليا

إشراف

د.تامر الخطيب

الملخص

تعد دراسة المنحنيات الخاصة للخلايا الشمسية ذات أهمية كبيرة في تطوير الخلايا و زيادة قدرتها، من هنا جاءت فكرة هذه الرسالة التي كان الهدف منها التنبؤ بمنحى التيار- الفولتية للخلية الشمسية من خلال تطوير نموذج جديد يعتمد على تقنية الغابات العشوائية في تدريب واختبار البيانات باستخدام برنامج الماتلاب .تعد تقنية الغابة العشوائية هي طريقة للتعلم الآلي، حيث تعتمد هذه التقنية على أشجار القرار (أشجار التصنيف، أشجار الانحدار). تم الاعتماد على تقنية الغابات العشوائية باستخدام أشجار الانحدار في النموذج المقترح الجديد للتنبؤ بمتغير الإخراج (التيار الناتج للوحدة الكهروضوئية)، اعتمادًا على مجموعة من المدخلات ممثلة بخمس معلمات (درجة الحرارة ، الإشعاع الشمسي، جهد الخلية الشمسية PV DV Voltage ، تيار الدائرة القصيرة، جهد الدائرة المفتوحة). تم الحصول على مجموعات البيانات عن طريق إجراء العديد من التجارب على الوحدة الكهروضوئية متعددة البلورية (- STF 120P6) ذات كفاءة تقدر بـ 14.0٪. تم إجراء سبع تجارب على الوحدة الكهروضوئية (- STF 120P6) عند قيم مختلفة لدرجة حرارة الخلية والاشعاع الشمسي لقياس التيار والفولتية باستخدام جهاز (I-V CURVE TRACER DEVICE).

من خلال تدريب واختبار هذه البيانات ، تم الحصول على نتائج عالية الدقة للنموذج المقترح ، حيث تم الحصول على قيم أخطاء القياس (RMSE، MAPE ، و MBE) ، والتي تساوي (0.04251٪ ، 4.315097٪ ، -3959.0٪) ، على التوالي. تم استخدام قيمة (MAPE) لتقييم هذا النموذج ومقارنته مع النماذج السابقة التي اعتمدت على أساليب مختلفة للتنبؤ بمنحنى I-V للوحدة الشمسية الكهروضوئية. يمكن تصنيف هذه الطرق إلى طرق (offline and online). تعتمد الطرق (online)على الأجهزة الحقيقية لاستخراج منحنى IV مثل (المكثف ، والمقاوم ، والمحث ، والمفاتيح) ، أما طرق (offline) فإنها تتمثل بطرق الذكاء الاصطناعي ، وتقنية الغابات العشوائية ، والطرق العددية حيث يتم استخدامها للحصول على حلول رقمية لمشكلة رياضية مثل طريقة Levenberg – Marquardt (LM) ، طريقة Newton – Raphson (NRM) ،