



---

RESEARCH ARTICLE

---

ELECTRICAL CHARACTERISTIC CLASSIFICATION OF THE PV'S USING SUPPORT VECTOR MACHINES

Fatih SERTTAŞ<sup>1,\*</sup> , Fatih Onur HOCAOĞLU<sup>1,2</sup> 

<sup>1</sup> Electrical Engineering Department., Engineering Faculty, Afyon Kocatepe University, Afyonkarahisar, Turkey

<sup>2</sup> Solar and Wind Energy Application Center, Afyon Kocatepe University, Afyonkarahisar, Turkey

ABSTRACT

Photovoltaics (PVs) have more attraction day by day due to increasing technologies and decreasing prices. However, in practice, the applicants encounter fundamental problems such as shadow effects and degradation. Due to these problems, the amount of produced annual electricity decreases considerably. Moreover, it is not easy to understand if the company's PV satisfies the requirements or is appropriate for the application. In order to solve such problems, it is instructed to examine the electrical characteristics of the modules. An initial task should be the classification of PV modules according to performance results. Here in this study, it is aimed to classify different PV modules, including different output characteristics. It is aimed to show whether it is possible to classify different PV modules including the same output power or not? To find the answer to this question, a test platform is built. 4 different panels are tested on the platform. While the test, the panels produced by different companies split into two groups, each has the same output power. Under a different insulation condition, the test is performed, and the cells' current-voltage curves are constructed. Different statistics are extracted by using this information. Different variations of these statistics are presented to multi-Support Vector Machine (Multi-SVM). Finally, accurate classification results are obtained.

**Keywords:** Photovoltaic, Current-voltage (IV) curve, Classification, Electrical characteristics, Support vector machines

---

1. INTRODUCTION

Photovoltaics (PVs) differ in their cell type, electrical and thermal characteristics, size, and place of use. Electrical characteristics carry essential information about the performance of the PV [1]. Many studies focus on current-voltage (IV) curves, the PV modules' main electrical characteristic element [2]. Schill et al. (2015) trace the characteristic current-voltage curve of photovoltaic panels every 10 minutes in the Fraunhofer testing facility [3]. An individual electronic load that is switched to maximum power point tracking mode between the I-V measurements exists in the electronic setup. Decreased power output and time efficiency are recorded to investigate the pollution effect. Efficiency falls within five months of 20% of initial values. Velilla et al. (2017) aim to identify the specifications from commercially available PV's [4]. Their study analyzes 1016 modules and six leading solar technologies (m-Si, p-Si, a-Si, CIGS, CIS, and CdTe). Parameters such as power, maximum power point, short-circuit (SC) current, efficiency, open-voltage circuit, and fill factor at standard test conditions are considered. Two-cluster analyses are performed with these variables to find the groups with similarities between modules independently of the technology. Electrical characteristic (I-V curve) is affected by partial shading and damaged by-pass diodes in the junction box. Ko et al. (2017) propose a photovoltaic module containing a diode-cell model to evaluate the current-voltage curve of each module [5]. The problematic by-pass diode is replaced with an ohmic impedance in the software-based simulation. They analyze the electrical characteristics of a shaded PV module theoretically and empirically with a damaged by-pass diode.

Chen et al. (2017) contribute to the literature with the following three aspects: (1) an efficient feature vector of seven dimensions is proposed as the input of the fault diagnosis model by extracting model

---

\*Corresponding Author: [serttasf@gmail.com](mailto:serttasf@gmail.com)

parameters from IV curves; (2) the emerging kernel-based extreme learning machine (KELM) is used to establish the fault diagnosis model by itself. In addition to this, the Nelder-Mead Simplex (NMS) optimization technique is chosen to optimize the KELM parameters that affect the classification performance; (3) a Simulink-based PV modeling approach is proposed for a laboratory PV array to facilitate the fault simulation and data sample acquisition [6].

Dash et al. (2017) have proposed seven climatic zones for India, based on the daytime radiation and environmental temperature precisely for the deployment of a PV system. PV power plants' performance varies with the location's environmental conditions technology of the PV module [7]. They also investigate these parameters' effect on I-V curves (current-voltage curve) of the PV modules.

The current-voltage curve is also used in performance analysis of PV modules. Jones et al. stated in their study that losses affecting PV plant performance can be classified with data-based approaches. It has shown that the classification made is 8% better than only the maximum power point when both the maximum power point and the I-V curve data set are used [8]. Another exemplary study in the literature regarding the use of current-voltage curve in the examination of PV performance was published by Hopwood et al. (2020) [9].

Another important problem that is studied in the literature is fault detection in PV modules. Some classification methods such as fuzzy [10] and hybrid approaches [11] successfully diagnose the failure in the electrical characteristics of PV modules. Dhimish et al. (2017) propose an error detection algorithm based on the analysis of the theoretical curves that describe an existing PV system [12]. They simulate the PV panel parameters by using LabView software. Photovoltaic faults correspond to a short circuit on the photovoltaic panel. They claim that the fault detection algorithm able to detect different faults in photovoltaics. Classification methods are often utilized in the annual energy harvesting calculation studies [13]. Economic viability and supervising are the main issues to do estimating. Zhang et al. (2016) develop a novel hierarchical classification method to facilitate PV profiling.

Furthermore, based on the classification results, a weather-based PV profiling guideline is created. PV cell operating temperature is a significant input parameter in the models for photovoltaic cells' electrical characterization and modules [14]. There are many calculating methods for determining the PV cell temperature. The methods based on direct measurements on the PV panel are more accurate than approaches only based on atmospheric parameters. Fouad et al. (2017) aim to introduce the integration of the diverse factors such as environmental, installation, and cost factors that affect PV modules' performance and system performance [15]. Catelani et al. (2013) propose a Failure Modes, Effects and Criticality Analysis methodology (FMECA) in order to fulfill the requirements of the PV producers or owner [16]. They aim to classify the occurrence, the severity, and the impact of all possible failure mechanisms on the PV module. The approach's main objective is to reduce or eliminate the effect of potential faults before failures occur in the field.

In the light of these literature studies, this paper, it is aimed to show whether it is possible to classify different PV modules including the same output power due to their electrical characteristics since it is difficult to understand if a panel constructed is newly produced or old or if it is appropriate for the application. This is a common problem, especially in underdeveloped countries. In most cases, it is difficult to obtain the specs of the modules. Therefore it is essential to set a scientific way to understand the performance of the module by test. A preliminary study of ongoing research aims to obtain I-V curves of the modulus under different insulation conditions. Therefore a test setup is built at Solar and Wind Research and Application Afyon Kocatepe University, as mentioned in Section 2. In this study, the electrical properties of the panels were obtained by using the data obtained from the established test setup. In the use of panels, not only the output power, but also electrical properties such as current voltage changes should be taken into account. There are many types of panels with the same output power, and higher performance systems can be obtained by selecting the panel suitable

for the intended use. This study was carried out in order to reveal the importance of electrical properties in the classification of panels. The SVM method determined as the classification method is explained in detail in Section 3. In Section 4, the data matrix to be used in classification was created by using statistical properties. Finally, classification results are presented and discussed in Section 5.

## 2. EXPERIMENTAL SETUP

Experiments have been realized using the newly-calibrated PV test device in Afyon Kocatepe University Solar-Wind Energy Implementation Research Department. Four kinds of photovoltaic modules from different labels are studied in the measurements.

### 2.1. Measuring PV Modules

The electrical and thermal values of each module are presented as inputs to the PV test device. According to these values, the PV test device generated standard test condition (STC) parameters compared to the measured values. The PV test device has environmental sensors, reference PV and temperature sensors for the PV module, current clamp, and voltage probes. Reference PV measures the irradiation near the module at the moment of the measurement. Photovoltaics have high-temperature values compared to the outside temperature, which directly affects the module's electrical performance. To account for this, temperature sensors are used and connected to the PV module's backside. PV test device behaves like a variable load during the test and measures the current with the current clamp. Open circuit (OC) voltage and maximum power (MP) voltage are also measured in the measuring period with unique voltage probes.

In experiments, four different PV modules are chosen from four different PV module producers. Their label parameters are given in Table 1. Experimental setup including one of the four tested PV modules, measuring device of the environmental sensors, and PV test device is illustrated in Figure 1.

**Table 1.** Label parameters of the PV modules

	PV 1	PV 2	PV 3	PV 4
<b>Output Power (W)</b>	100	60	60	100
<b>MP Voltage - <math>V_{mpp}</math> (V)</b>	17,7	18,2	17,3	19,8
<b>MP Current - <math>I_{mpp}</math> (A)</b>	5,65	3,3	3,47	5,06
<b>OC Voltage - <math>V_{oc}</math> (V)</b>	22	22,2	21,6	24,35
<b>SC Current - <math>I_{sc}</math> (A)</b>	6,21	3,53	3,73	5,51



**Figure 1.** Utilized technical devices in the experimental study.

### 2.2. Acquisition of Electrical Characteristics Data

PV test device is able to collect approximately 150 points of data for each parameter in one measurement process. These measurement parameters are voltage, current, and power. The data

format and data type are arranged to implement the test. The voltage, current, and power values of the first measurement are presented in Figure 2.

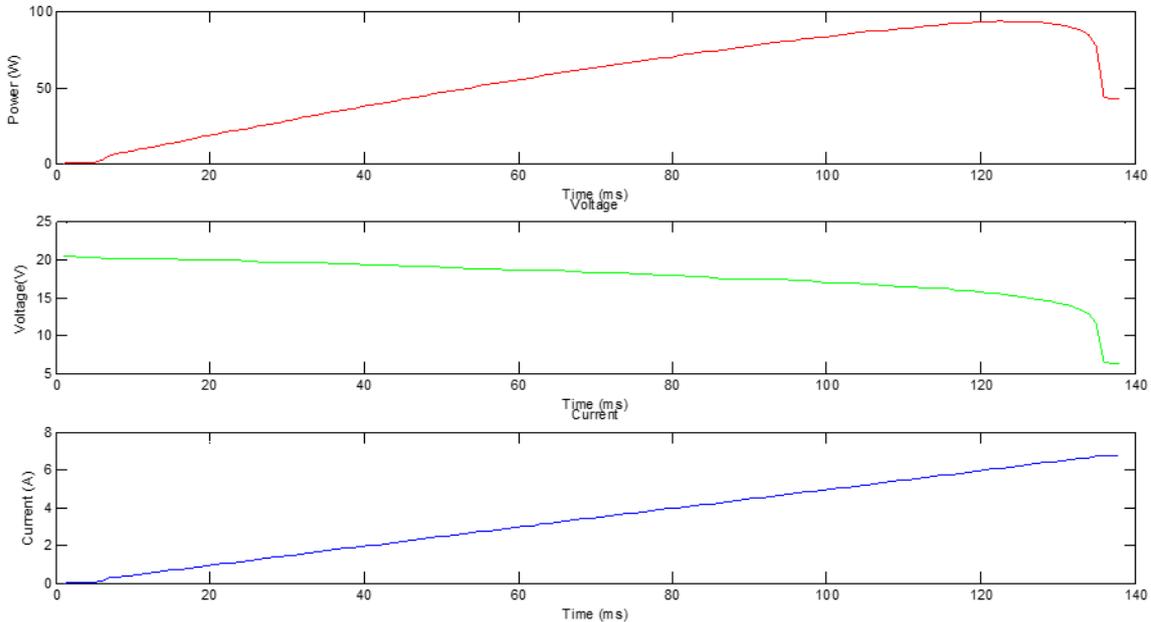


Figure 2. Measured data of one of the test experiments

In Figure 3, I-V curves of the PV modules used in the experiments are presented for comparison. Here, Module 1 and Module 4 are 100 Watts power outputs, whereas; Module 2 and 3 are 60 Watts.

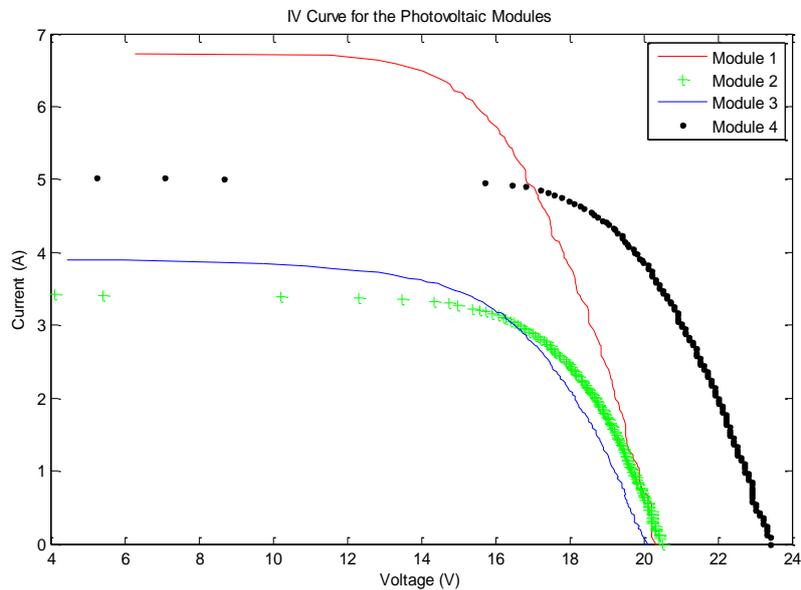


Figure 3. Obtained IV Curves of the PV modules.

As seen from Figure 3, even the modules' output is the same, the IV characteristics are different. The solar panel measuring device allows measurement when there is only solar radiation higher than 1000 W/m<sup>2</sup> in the environment. For this reason, experiments were carried out at time intervals when these factors were equivalent so that the experiments did not vary depending on environmental conditions such as temperature and pressure.

The data set obtained from the experiments is divided into test and training data set. Provided that data belonging to each panel are included in these data sets, the method's applicability has been tried to be revealed by changing the training and test data sets. While classification was made with 100% accuracy in some tests, panels could be classified with 75% accuracy. Since all test combinations are tried, a total of 16 test results are obtained.

### 3. SUPPORT VECTOR MACHINES

Classification is necessary for many electronic and signal processing applications. One of the most popular methodologies is the support vector machine. It was developed as a statistical learning method by VN Vapnik in 1995. The basic principle of the SVM is to form an optimal canonical hyperplane to separate the series of labeled learning (training) data.

w: decision hyperplane normal vector,  
 $x_i$ : data point i,  
 $y_i$ : class of data point i (+1 or -1)  
 $\xi_i$ : slack variables  
 b= bias

In Eq. 1, the hyperplane obtained by training data has to have a maximal M margin to classify the new data with better accuracy.

$$w^T x + b = 0 \tag{1}$$

It should be in a canonic form owing to make it easier to find the critical pattern. Furthermore, calculations will be more straightforward with the canonic form of the hyperplane. Through the training period, the weight norm of canonical hyperplane  $|w|$ , minimizing the Vapnik-Chervenenkis (VC) dimensions by maximizing the margin called Eq. 2. There is a canonical hyperplane because it is numbered Eq. 3 while looking at the state of the two linearly separable classes.

$$M = \frac{2}{|w|} \tag{2}$$

$$\min |w^T x + b| = 1 \quad x_i \in X \tag{3}$$

#### 3.1. Multi-class Support Vector Machines (Multi-SVM)

Support vector machines usually are two-class classifiers. It is not a very suitable approach to solve multi-class problems. A preferable option is provided by the structure of Multi-class Support Vector Machines, which is built with a two-class classifier over a feature vector  $\varphi(x,y)$  reproduced from the pair consisting of the input features and the group of the data [17].

During the test, multi-SVM selects the cluster Eq. 4. The edge while training is the space between these parameters for the appropriate group and the closest other class. Therefore the quadratic program equation requires that Eq. 5. The usual methodology may be extended to give a multi-class equation of several types of linear classifiers. Furthermore it is a basic example of a clustering generalization where classes are not just a set of independent, categorical tags. However, it might be arbitrarily structured objects with defined relationships.

$$y = \operatorname{argmax}_{y_i} w^T \varphi(x, y') \tag{4}$$

$$\forall i \forall y \neq y_i w^T \varphi(x_i y_i) - w^T \varphi(x_i, y) \geq 1 - \xi_i \tag{5}$$

## 4. MATERIAL METHODS

Four different photovoltaic solar modules are studied, and their electrical characteristics under different irradiation conditions are measured and collected. Six different tests for each module are performed during experiments. Some statistical features are extracted using the data measured and collected. After the statistical features are extracted, these features are presented to Multi-SVM to classify the modules. It should be mentioned here that the difference between SVM and Multi-SVM is that: It is possible to produce only two results with SVM, whereas the Multi-SVM has more. These steps are explained in the following subsection.

### 4.1. Statistical Features of the IV and PV Curve Data

Statistical distribution of the leading electrical parameters is required to evaluate mismatch losses using simulation software [18]. Furthermore, it is also essential to verify if a module produced by the manufacturer satisfies the requirements. In order to test the availability, firstly, raw electrical characteristic data is tried to classify using the SVM. However, it is obtained that the results of classifications non-accurate not logical. Therefore, it is concluded that more features needed to be extracted before the application of the classification. It is thought that statistical values can carry information about the modules' characteristics and can be used as the features. Therefore, mean (average) value, maximum value, standard deviation, variance, and median are calculated for each experiment's power and voltage data. Feature equations are given in the (6-9), respectively.

**Feature 1:**

$$\text{Mean Value} = \frac{\sum V_i}{n} \text{ or } \frac{\sum P_i}{n} \quad (6)$$

**Feature 2:**

$$\text{Max of } V_i \text{ or } P_i$$

**Feature 3:**

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (V_i - \bar{V})^2} \quad (7)$$

or

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (P_i - \bar{P})^2}$$

**Feature 4:**

$$\text{Variance} = \frac{1}{n-1} \sum_{i=1}^n (V_i - \bar{V})^2 \quad (8)$$

or

$$\text{Variance} = \frac{1}{n-1} \sum_{i=1}^n (P_i - \bar{P})^2$$

**Feature 5:**

$$\text{Median} = V_i, i = \frac{n+1}{2} \quad (9)$$

or

$$\text{Median} = P_i, i = \frac{n+1}{2}$$

Consequently, these parameters are converted into a matrix. The matrix of an arbitrary experiment is illustrated in Figure 4. Here, the x-axis (right-bottom) shows the number of the parameters. The first

five parameters correspond to the power measurement statistics, whereas; the second five correspond to the voltage statistics.

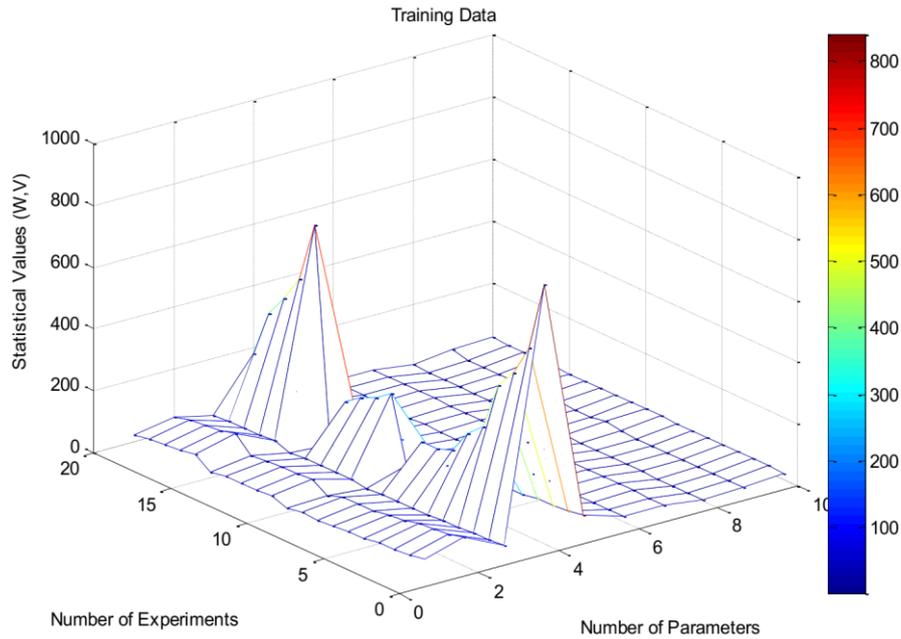


Figure 4. Mesh plot of the training matrix.

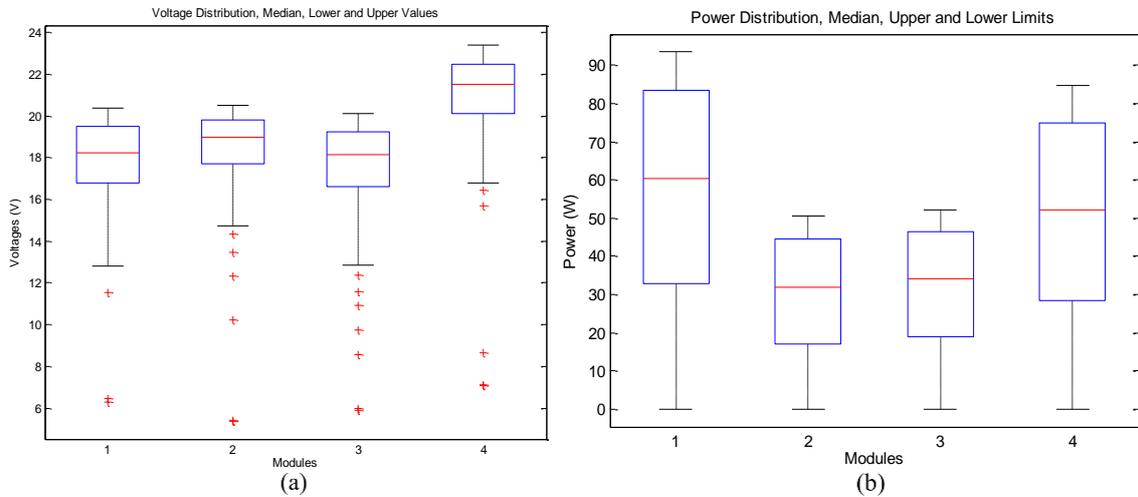
## 5. RESULTS and DISCUSSION

It is obtained that the sensitivity is low, which means that the classification's accuracy and feasibility are high. For this reason, in the study, various test and training sets are formed from the experimental data, and the classification accuracy rate is determined. In order to find the same number of data belonging to each PV module in the training set, four classification panel classification was made in each classification test. Since the upper and lower limits of PV modules are considered, instead of 24 scrolling, 16 scrolling results are listed for 16 different sets and training sets.

On the other hand, voltage and power distributions of the modules are presented in Figure 5. It can be seen from Figure 5 that some data values are the same for different panels. For instance, there are points where the voltage and power values of module 2 and module 3 coincide. Therefore, it is impossible to separate these panels from each other by direct use of these data. For this reason, attributes were derived from the statistical properties of these data, and the parameters for the classification are obtained. These attributes are listed in Table 2, and calculated values of parameters for the first test of the modules are also presented herein.

Table 2. Statistical Parameters used in the classification

Statistical Parameter	Module 1		Module 2		Module 3		Module 4	
	Voltage(V)	Power(W)	Voltage(V)	Power(W)	Voltage(V)	Power(W)	Voltage(V)	Power(W)
Average	17,74	56,17	18,38	29,75	17,41	31,62	20,88	49,84
Maximum	20,40	93,50	20,50	50,40	20,10	52,20	23,40	84,70
Standard Deviation	2,48	29,00	2,37	15,64	2,78	16,07	2,67	26,22
Variance	6,14	841,15	5,60	244,60	7,71	258,08	7,12	687,66
Median	18,25	60,45	19,00	32,00	18,17	34,05	21,50	52,20



**Figure 5.** a) Voltage Distribution of Modules b) Power Distribution of Modules

The classical support vector machine method classifies only two species. However, the multi-SVM method is a developed methodology to classify more species than two. In this study, four different modules are classified. Therefore multi-SVM is implemented to the electrical characteristics data of the modules. The modules' power and voltage statistics data take as test and training parameters for every test by turns. It is evident from Table 3 that, Multi-SVM with mentioned features performs very accurate classification results.

**Table 3.** Classification Loop Results

Test Number	Panel No		Panel No		Panel No		Panel No		Class Accuracy
	Real	C.R*	Real	C.R	Real	C.R	Real	C.R	
1st test	1	1	2	2	3	3	4	4	100%
2nd test	1	1	2	2	3	3	4	4	100%
3rd test	1	1	2	2	3	3	4	4	100%
4th test	1	1	2	2	3	3	4	4	100%
5th test	1	1	2	2	3	3	4	4	100%
6th test	1	1	2	2	3	3	4	4	100%
7th test	1	1	2	2	3	3	4	4	100%
8th test	1	1	2	2	3	2	4	4	75%
9th test	1	1	2	2	3	2	4	4	75%
10th test	1	1	2	2	3	2	4	4	75%
11th test	1	1	2	2	3	3	4	4	100%
12th test	1	1	2	2	3	3	4	4	100%
13th test	1	1	2	2	3	3	4	4	100%
14th test	1	1	2	2	3	3	4	4	100%
15th test	1	1	2	3	3	3	4	4	75%
16th test	1	1	2	2	3	3	4	4	100%
<b>Percentage</b>	<b>100%</b>		<b>93,75%</b>		<b>81,25%</b>		<b>100%</b>		
	<b>Total Percentage</b>								<b>93,75%</b>

\*C.R= Classification Result (Estimated Data)

## **6. CONCLUSION**

This study claims that PV panel electrical characteristics can be classified even the modules have the same power or same voltage value. In addition to classification, there are several significant points extracted, presenting in the following sentences.

- Support vector machine method does not classify by using only power parameters in behalf of similarities between other same powered PV modules. Meanwhile, it is the same for other parameters when they are selected to use alone. Although the current is also tried and used in the data, classification accuracy is reduced. It may be commented as similar current values cause this result. Then current parameters are extracted from the data.
- An accurate classification is helpful for the data collecting and saving process. Through the electrical characteristic monitoring with the PV test device, experiments are plague because of the external conditions, lots of cabling, and measuring devices. This situation causes some errors, such as unnamed data files. Data recording with unnamed files is an important issue. Classification helps to separate and identify the data.
- Choosing different test and training data in the SVM classification method gives more accurate results even if the data number is limited. By turning the data for training and test, measurement data number virtually increases. Without turning data, only if the first classification case was used would produce incorrect results (produce 100% accuracy) shown in Table 3.

According to the study results, it is anticipated that the old (aged) and newly-produced photovoltaic modules will be distinguished in the next studies. A user-interface software will be developed to analyze all the data of the PV measurements directly. Therefore, it will be possible to see older or badly affected PV's in the solar energy sector on the customer side. In addition to this, by measuring photovoltaic panels in a solar power plant and analyzing the data using this SVM-based methodology, it can be found how many PV classes (PV module types) are actually available.

## **CONFLICT OF INTEREST**

The author stated that there are no conflicts of interest regarding the publication of this article.

## **REFERENCES**

- [1] Fares MA, Atik L, Bachir G, Aillerie M. Photovoltaic panels characterization and experimental testing. *Energy Procedia* 2017;119:945–52.
- [2] Abdulrazzaq AK, Bognár G, Plesz B. Accurate method for PV solar cells and modules parameters extraction using I–V curves. *Journal of King Saud University - Engineering Sciences* 2020.
- [3] Schill C, Brachmann S, Koehl M. Impact of soiling on IV-curves and efficiency of PV-modules. *Sol Energy* 2015; 112:259–62.
- [4] Velilla E, Restrepo S, Jaramillo F. Cluster analysis of commercial photovoltaic modules based on the electrical performance at standard test conditions. *Sol Energy* 2017; 144:335–41.
- [5] Ko SW, Ju YC, Hwang HM, So JH, Jung Y-S, Song H-J, et al. Electric and thermal characteristics of photovoltaic modules under partial shading and with a damaged by-pass diode. *Energy* 2017; 128:232–43.

- [6] Chen Z, Wu L, Cheng S, Lin P, Wu Y, Lin W. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics. *Appl Energy* 2017; 204:912–31.
- [7] Dash PK, Gupta NC, Rawat R, Pant PC. A novel climate classification criterion based on the performance of solar photovoltaic technologies. *Sol Energy* 2017;144:392–8.
- [8] Jones CB, Theristis M, Stein JS, Hansen C. Feature Selection of Photovoltaic System Data to Avoid Misclassification of Fault Conditions. 47th IEEE Photovoltaic Specialists Conference (PVSC) 2020;1357-1362.
- [9] Hopwood MW, Gunda T, Seigneur H, Walters J. Neural Network-Based Classification of String-Level IV Curves From Physically-Induced Failures of Photovoltaic Modules. *IEEE Access* 2020; 8:161480-161487.
- [10] Dhimish M, Holmes V, Mehrdadi B, Dales M, Chong B, Zhang L. Seven indicators variations for multiple PV array configurations under partial shading and faulty PV conditions. *Renew Energy* 2017;113:438–60.
- [11] Wu Y, Chen Z, Wu L, Lin P, Cheng S, Lu P. An Intelligent Fault Diagnosis Approach for PV Array Based on SA-RBF Kernel Extreme Learning Machine. *Energy Procedia* 2017;105:1070–6.
- [12] Dhimish M, Holmes V, Mehrdadi B, Dales M, Mather P. Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system. *Energy* 2017;140:276–90.
- [13] Rus-Casas C, Aguilar JD, Rodrigo P, Almonacid F, Pérez-Higueras PJ. Classification of methods for annual energy harvesting calculations of photovoltaic generators. *Energy Convers Manag* 2014;78:527–36.
- [14] Rodrigo P, Fernández EF, Almonacid F, Pérez-Higueras PJ. Review of methods for the calculation of cell temperature in high concentration photovoltaic modules for electrical characterization. *Renew Sustain Energy Rev* 2014;38:478–88. doi:10.1016/j.rser.2014.06.008.
- [15] Fouad MM, Shihata LA, Morgan EI. An integrated review of factors influencing the performance of photovoltaic panels. *Renew Sustain Energy Rev* 2017;80:1499–511.
- [16] Catelani M, Ciani L, Cristaldi L, Faifer M, Lazzaroni M. Electrical performances optimization of Photovoltaic Modules with FMECA approach. *Meas J Int Meas Confed* 2013;46:3898–909.
- [17] Stanford University. Multi-class SVMs. Cambridge Univ Press 2009. <https://nlp.stanford.edu/IR-book/html/htmledition/multiclass-svms-1.html>.
- [18] Gasparin FP, Bühler AJ, Rampinelli GA, Krenzinger A. Statistical analysis of I–V curve parameters from photovoltaic modules. *Sol Energy* 2016;131:30–8.