

# A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography

Muhammad Umair Ali<sup>a</sup>, Hafiz Farhaj Khan<sup>b</sup>, Manzar Masud<sup>c</sup>, Karam Dad Kallu<sup>d</sup>,  
Amad Zafar<sup>b,\*</sup>

<sup>a</sup> School of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, South Korea

<sup>b</sup> Electrical Engineering Department, Wah Engineering College, University of Wah, Wah Cantt 47040, Pakistan

<sup>c</sup> Mechanical Engineering Department, Wah Engineering College, University of Wah, Wah Cantt 47040, Pakistan

<sup>d</sup> School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST) H-12, Islamabad, Pakistan

## ARTICLE INFO

### Keywords:

Classification  
Hotspot detection  
Infrared thermography (IRT)  
Optimized hybrid features  
Photovoltaic (PV) modules  
Support vector machine

## ABSTRACT

In this paper, a hybrid features based support vector machine (SVM) model is proposed using infrared thermography technique for hotspots detection and classification of photovoltaic (PV) panels. A novel hybrid feature vector consisting of RGB, texture, the histogram of oriented gradient (HOG), and local binary pattern (LBP) as features is formed using a data fusion approach. A machine learning algorithm SVM is employed to classify the obtained thermal images of PV panels into three different classes (i.e., healthy, non-faulty hotspot, and faulty). The comparison of different machine learning algorithms and datasets is also carried out to validate the superiority of the proposed model and hybrid feature dataset. The experimental results reveal that the proposed hybrid features with SVM resulted in 96.8% training accuracy and 92% testing accuracy with lesser computational complexity and storage space than other machine learning algorithms. The proposed approach is easily implementable for efficient monitoring and fault diagnosis of PV panels.

## 1. Introduction

The global demand in electrical energy production and consumption is exponentially increasing due to economic and population growth. According to recent reports (McKinsey, 2019; Ram et al., 2019), the annual global electrical consumption doubles in 2050. Electricity production is currently led by conventional sources of energy, such as fossil fuels and coal, which have an adverse impact on global warming. According to the report published by the international energy agency (IEA, 2019), the annual two-third of the carbon dioxide emission in the atmosphere occurs due to the use of conventional energy sources. On the other hand, traditional energy sources are also limited in nature. Therefore, to cater to the electrical energy requirement for sustainable and eco-friendly energy production, the utilization of renewable and low carbon energy sources can play a vital role (Nengroo et al., 2019; Varun et al., 2009). According to the latest report (Ram et al., 2019), renewable energy sources such as photovoltaic (PV) and wind energy systems are expected to generate 88% of worldwide electricity demand till 2050. In renewable energy sources, PV based energy systems are prominent

owing to many advantages such as noiseless operation, global availability, easy installation, the declining price of PV modules, pollution-free energy generation, and its reliability (Mellit et al., 2018; Nengroo et al., 2018; Niazi et al., 2018a, 2018b). In addition, the PV modules are easily installable at the building rooftop and can work standalone. Therefore, the installation of PV generation plants has been rapidly increasing every year, as shown in Fig. 1 (Kaizuka et al., 2020). In the year 2019, the capacity of global PV energy generation was increased to 117 GW.

The global increase in the installation capacity of PV systems resulted in the increase of PV manufacturers. Therefore, the manufacturer has to guarantee the long-term performance of PV electricity generation system. Although the manufacturer ensures the 80% output power of PV system at the maximum power point (MPP) after 25 years (Jordan et al., 2016), the output performance of the PV system slowly reduces with age. The decrease in the efficiency and performance of the PV systems depends on the degradation as mentioned by the manufacturers and on the outdoor environmental conditions (Kumar and Kumar, 2017). Therefore, the reliability and durability of the PV system become a hot

\* Corresponding author.

E-mail addresses: [umair.ali@sejong.ac.kr](mailto:umair.ali@sejong.ac.kr) (M.U. Ali), [farhaj.khan@wecuw.edu.pk](mailto:farhaj.khan@wecuw.edu.pk) (H.F. Khan), [manzar.masud@wecuw.edu.pk](mailto:manzar.masud@wecuw.edu.pk) (M. Masud), [karamdad.kallu@smme.nust.edu.pk](mailto:karamdad.kallu@smme.nust.edu.pk) (K.D. Kallu), [amad.zafar@wecuw.edu.pk](mailto:amad.zafar@wecuw.edu.pk) (A. Zafar).

<https://doi.org/10.1016/j.solener.2020.08.027>

Received 21 May 2020; Received in revised form 31 July 2020; Accepted 10 August 2020

Available online 19 August 2020

0038-092X/© 2020 International Solar Energy Society. Published by Elsevier Ltd. All rights reserved.

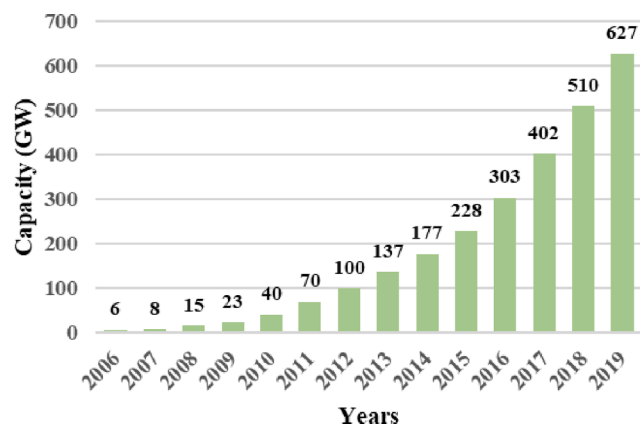


Fig. 1. Globally installed PV capacity (Kaizuka et al., 2020).

research topic in recent years.

The energy losses and output power failures in the PV system can be influenced by different factors such as wear and tear defects (Madeti and Singh, 2017a), maximum power point tracking errors (Hohm and Ropp, 2003), electrical wiring losses (Woyte et al., 2013), manufacturing defects, and overheating (Woyte et al., 2003). These different factors can reduce the output power of the PV system by 50% (Tsanakas et al., 2016). Therefore, the elements adversely affecting the PV output production must be prevented, and the timely determination and classification of faults must be carried out to restrain the power loss. Consequently, it implies that the fault diagnostic tool is needed for the reliability and durability of the PV system. In all issues mentioned above, the overheating of PV modules is one of the critical concerns, which is likely to arise owing to snail tracks (Livera et al., 2019), uneven shading (Garoudja et al., 2017; Niazi et al., 2019c), hotspot (Ahsan et al., 2018), cracks (Wang et al., 2016), aging (Tina et al., 2016), etc. Therefore, effective inspection is necessary to prolong the lifecycle and preemptive maintenance of the PV system (Niazi et al., 2019b). It is impossible to manually inspect the large-scale PV generation plants; therefore, automation is needed to classify the PV condition (defective or healthy). Numerous methodologies have been reported in the literature to address this issue. These methodologies can be divided into two main categories, (i) electrical signal (voltage and current) based (Breitenstein et al., 2009; Madeti and Singh, 2017b), and (ii) image processing-based such as photoluminescence imaging (Trupke et al., 2012), electroluminescence images (Deutsch et al., 2019), fluorescence techniques (Aghaei et al., 2015), and infrared thermography (IRT) (Dunderdale et al., 2019; Niazi et al., 2018a, 2018b). All image processing-based techniques have used machine learning algorithms to differentiate between healthy and faulty PV panels. Among all image processing-based methodologies, IRT is the most acknowledged technique to categorize the PV panels (Berardone et al., 2018; Niazi et al., 2019a). Deutsch et al. (2019) proposed a convolutional neural network to categorize the PV faults using electroluminescence images. Although the image quality of the electroluminescence images is very high, it takes too much time and requires expertise to differentiate between different faults. Jaffery et al. (2017) designed a fuzzy rule-base to classify the thermal images of the PV panel, their proposed technique shows some promising results. Still, the approach may not generalize for every PV panel. Liu et al. (2017) used IRT with an extreme learning machine and moving least square regression method for the detection of fault position. Vergura and Marino (2017) used infrared (IR) images to detect the hotspot in the PV module up to cell level, but they did not classify the PV panel into different classes. Niazi et al. (2019a) addressed the issue of panel classification using the Naive Bayes (NB) technique and classified the PV panel into three different classes. They calculated the histogram of oriented gradient (HOG) and texture features (contrast, energy, correlation, and homogeneity) to train their proposed model.

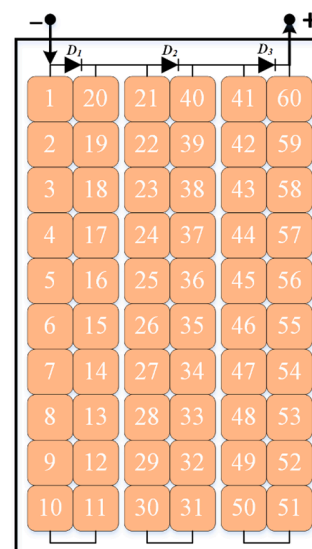


Fig. 2. Internal schematic diagram of the crystalline silicon PV panel.

The cell size of  $4 \times 4$  was used to get the HOG feature for an image. The principal component analysis was utilized to extract the useful features of HOG. The training vector size was 182 features. As the size of the training vector increases, the computational cost and the storage size of the system increases.

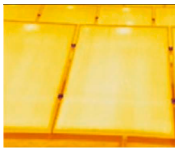
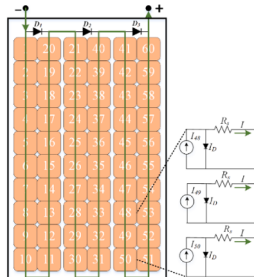
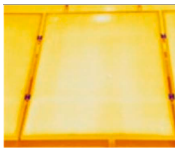
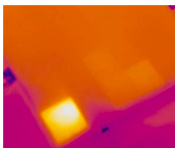
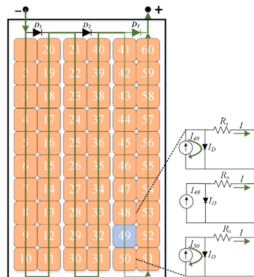

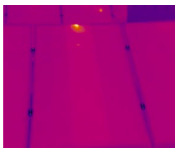
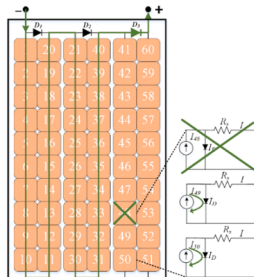
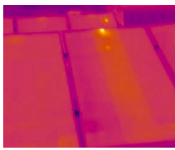
Therefore, in this study, a novel hybrid feature vector consists of color histograms, second-order co-occurrence matrix, HOG, and local binary pattern (LBP) was proposed to train the support vector machine (SVM) model. The proposed methodology classified the thermal images into three different classes (healthy, non-faulty hotspot, and faulty). The different feature vectors and machine learning algorithms were compared to validate the dominance of the proposed training features and the model. Results confirmed that the proposed technique requires less storage size and low computational time with high accuracy. The paper is structured as follows. Section 2 briefly provides the faults description and class labels of solar panels. The proposed hybrid feature extraction and the SVM model is presented in Section 3, whereas experimental results are shown in Section 4. Finally, Section 5 discusses the results, and the conclusion is presented in Section 6.

## 2. Description and classification of PV panel

Nowadays, the monocrystalline silicon (mono-Si) and polycrystalline silicon (poly-Si) PV panels dominate the global PV market with almost 90% share (Lupangu and Bansal, 2017). The variants of thin film PV (TFPV) system such as cadmium telluride (CdTe), copper indium gallium diselenide (CIGS), and amorphous thin-film silicon (a-Si, TF-Si) has the market share of 10% (i.e., reduces from 16% in 2009). Continuous investigation on the high-concentrator PVs (HCPVs) resulted in the development of panels that have almost 43% efficiency, but it still has less than 1% share in the PV market (Khan et al., 2018; Market, 2018). Due to the high share of crystalline silicon PVs in the market, the durability and reliability of the crystalline silicon PV panels need to be investigated. The internal schematic diagram of a single crystalline silicon PV panel is shown in Fig. 2. The crystalline silicon PV panel consists of 60 cells connected in series to boost up the panel voltage, the diode ( $D_1$ ,  $D_2$ , and  $D_3$ ) are used for bypassing of submodule. The bypassing diodes are connected in anti-parallel to the group of 20 crystalline silicon PV cells, as shown in Fig. 2.

In this work, the PV panels categorized as: (i) healthy, (ii) non-faulty hotspot, and (iii) faulty panel. The PV panels with the uniform solar irradiance profile are labeled as the healthy panel. In healthy PV panels, all cells are working properly and contributing in the overall output of

**Table 1**  
Detail of PV panel during different conditions.

Condition and temperature range	Cause	Thermal Image	Temp-erature	Operation schematic diagram of PV panel
Healthy 50–60 °C	No shading and no damage		57 °C	
			52 °C	
Hotspot 65–80 °C	Shaded		72 °C	
	Bird drop		70 °C	
Faulty above 80 °C	Damage		85 °C	
			82 °C	

the PV module, whereas all bypassing diodes remain inactive during the whole process. The working temperature of the healthy PV panel remains within the limit of 50–60 °C. As an example of the healthy PV panels, the operational schematic diagram and corresponding working temperature with actual images is given in Table 1. In the non-faulty hotspot panels, the submodule's overall current reduces due to bird drop or shading of some part of the submodules (cell or multiple cells). The shaded part of the submodules reduces the overall current of the submodules and results in the reduction of the overall efficiency of the whole PV panel. The shaded part of the PV panel becomes brighter due to the high temperature value at that point because all the generated energy starts dissipating at that point. The PV panel with the delamination, soiling, snail tracks, micro-cracks, burnt marks, glass breakage, bypass diode failure, and solder interconnects are grouped as faulty panels. In the case of faulty panels, the overall submodule is bypassed through the diode ( $D_3$ ), as shown in Table 1. In this case, the whole submodule becomes brighter due to reverse voltage. (Kumar and Kumar, 2017) discussed these defects in more detail. Table 1 summarized the causes, working operation, and temperature of PV panels with actual images during all three conditions.

### 3. Methods

#### 3.1. Proposed methodology

In this study, the machine learning framework based on SVM classifier is proposed for accurate classification and monitoring of PV panels. The SVM algorithm's classification and regression attributes have been already used to approximate any multivariate function in several

engineering problems (Ali et al., 2019; Sundermann et al., 2017). The process of the proposed methodology can be broken into five main steps, as shown in Fig. 3. Thermal images of solar panels were acquired using a thermal camera, which were then passed through the pre-processing phase to enhance the image quality and noise removal. After pre-processing, the different image features were calculated to construct the training dataset. The training dataset trained the SVM model using a 5-fold cross-validation methodology (Stone, 1974). This method randomly divides the training feature data into five folds (almost 20% data for each fold). In all five iterations, four-folds (almost 80%) were utilized for training the model, and the remaining one-fold was used for validation of the model to check the classification accuracy. The block diagram of the proposed methodology is shown in Fig. 3.

#### 3.2. PV dataset

The object with temperature more than the absolute zero emits infrared radiations, which have the wavelength in the range of 8 to 12  $\mu\text{m}$  (Minkina and Dudzik, 2009). IRT cameras were used to examine the heat dissipation in many industrial applications. The thermal images of solar PV panels can be acquired using FLIR thermal camera from PV panels. In this study, a previously published dataset was utilized to validate the proposed technique (Niazi et al., 2019a). The installed PV system has a rating of 42.24-kWp and located in Lahore, Pakistan. Eight PV strings with each string consisting of 22 modules (i.e., 240 Wp rating per module) coupled in series were connected to the AC main through inverters. Thermal images of solar panels were acquired using handheld, horizontally aligned thermal camera (FLIR VUE-Pro 640) and were taken from different distances at several angles. The thermal images

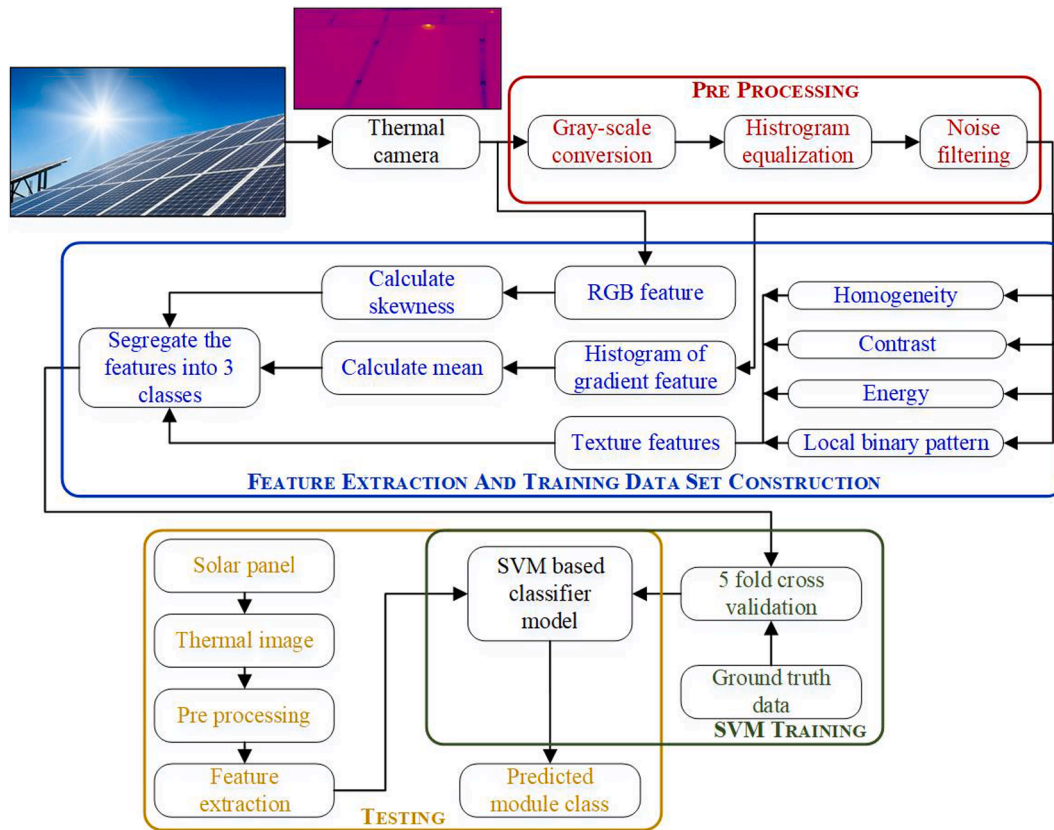


Fig. 3. Working cycle of the proposed methodology.

**Table 2**  
Labeling of acquired images using ground-truth dataset.

PV panel categorization	Number of images
Healthy class	102
Non-faulty hotspot class	99
Faulty class	114

were captured in the environment having a temperature between 32 and 40 °C, average wind speed of around 6.9 m/s, and irradiance level above 700 W/m<sup>2</sup>. The thermal images have a bit depth of 8 bits and a spatial resolution of 640 × 512 per pixel.

The category and number of images for each category is listed in Table 2. More details about experimental setup and image capturing can be found in (Niazi et al., 2019a).

### 3.3. Pre-processing

In the pre-processing phase, the PV panel's thermal RGB image was converted to the gray-scale image to calculate the amount of light at each pixel in a single sample. Additionally, histogram equalization was utilized to improve the contrast of the PV panel's gray-scale thermal image. Furthermore, the data was passed through the Gabor noise filter to remove the noise component to enhance the image's quality to calculate HOG and texture features.

### 3.4. Feature extraction and training dataset construction

In spectrally heterogeneous image processing architect, the appropriate feature extraction is one of the most crucial parts for monitoring and fault diagnosis using machine learning algorithms. In this work, the hybrid feature dataset consists of first-order statistical indicator (color

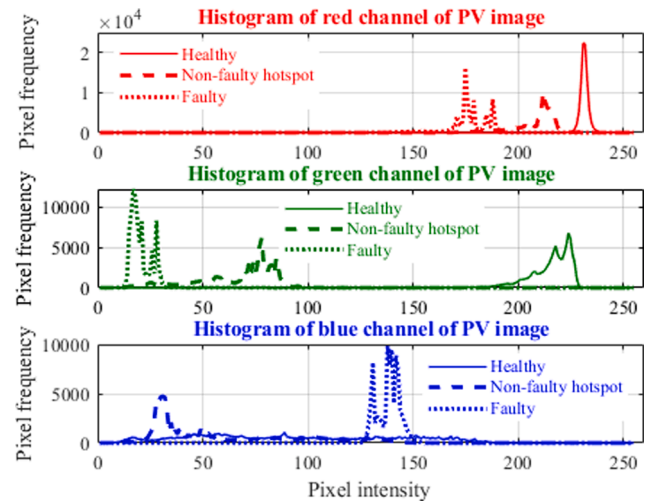


Fig. 4. RGB histograms under different PV module conditions.

histograms), second-order co-occurrence matrix, HOG, and LBP were extracted for PV thermal images. The training, analysis, and validation were carried out using statistic and machine learning toolbox of MATLAB 2020a with the system specification of Intel® Core (TM) i5-3210 M CPU @ 2.50 GHz processor with 10 GB RAM, 250 GB SSD, and a 64 bit Windows 10 education operating system (OS).

#### 3.4.1. Color feature

Color is considered one of the most crucial and distinguishable low-level visual features for classification and segmentation of images in many fields (Allan et al., 2009). Due to simplicity and low complexity,



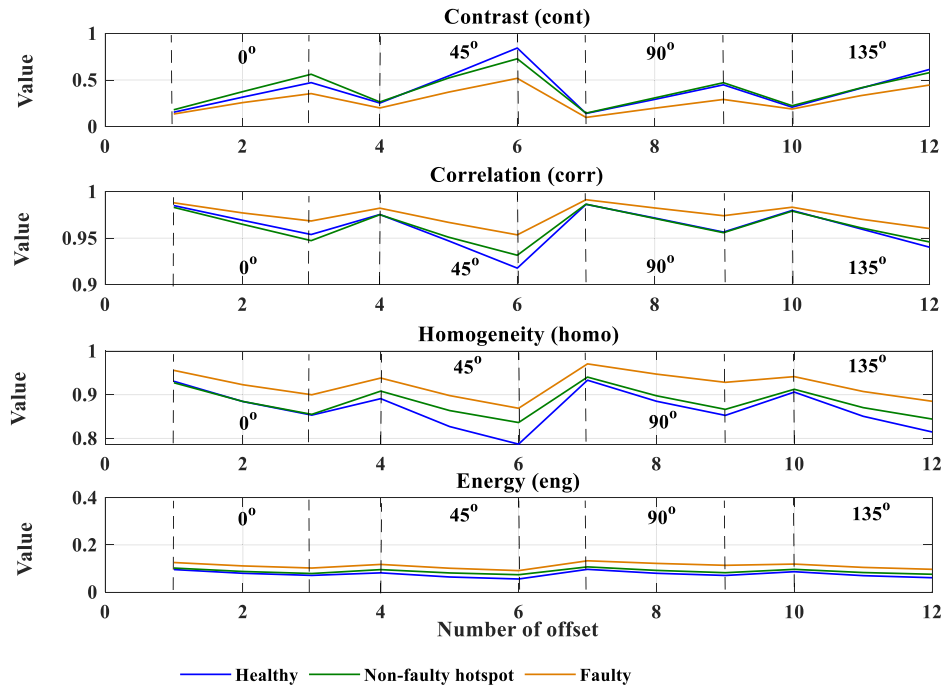


Fig. 5. Behavior of different texture feature different condition of PV panel.

the color histogram techniques have shown some promising results. The three-dimensional (3-D) histograms corresponding to a position in a color space was used to represent a color. The red, green, and blue (RGB) are the most widely used color features found in the literature. The feature of any thermal PV image can easily be represented by using the RGB coefficients. For an 8-bit image, the range of 0 to 255 was used to describe the RGB values, where 0, 127, and 255 means the off (black), half intensity, and full intensity of the colors, respectively. In a thermal PV image, the color of the region changes effectively as the result of an anomaly in any part of the image, which means that the intensity of the RGB changes with the change in the color of the thermal PV image, as shown in Fig. 4.

### 3.4.2. Second-order co-occurrence matrix feature

Haralick et al., (1973) introduced a co-occurrence matrix to extract the texture features, also known as the second-order co-occurrence matrix. Texture features for thermal images of PV modules can be calculated using the gray-level co-occurrence matrix (GLCM). These features mainly depend upon the angle and inter-sample distance between pixels. In this study, four second-order texture features such as contrast (cont), correlation (corr), homogeneity (homo), and energy (eng) were acquired to check and analyze the impact on classification accuracy (Armi and Fekri-Ershad, 2019). 3-pixel sample spacing (1, 2, and 3) and four angles (0°, 45°, 90°, and 135°) were used to compute the texture features. After evaluating the GLCM, the second statistic's textual features can be modeled using the following equations (Armi and Fekri-Ershad, 2019).

$$\text{Contrast (cont)} = \sum_{a,b}^{n-1} |a - b|^2 p_{a,b} \quad (1)$$

$$\text{Correlation (corr)} = \sum_{a,b}^{n-1} \frac{(a - \mu_a)(b - \mu_b)p_{a,b}}{\sigma_a \sigma_b} \quad (2)$$

$$\text{Homogeneity (homo)} = \sum_{a,b}^{n-1} \frac{p_{a,b}}{1 + |a - b|} \quad (3)$$

$$\text{Energy (eng)} = \sum_{a,b}^{n-1} p_{a,b}^2 \quad (4)$$

where  $p_{a,b}$  is the pixel value,  $a$  and  $b$  are the pixel location, and  $n$  is the number of gray level.  $\mu_a \mu_b$  and  $\sigma_a \sigma_b$  are the mean and variance along with rows and column.

### 3.4.3. HOG feature

The HOG descriptor represents the feature by weighing the occurrence of gradient orientation in an image. The HOG calculates the histogram of gradient orientations by dividing the image into small cells. The HOG has been already applied for classification in different fields and has shown promising results (Banerji et al., 2013; Dalal and Triggs, 2005). It has already been proven that the HOG features maintain high selectivity during non-rigid alterations in the body shape (Fleyeh and Roch, 2013). Therefore, the HOG features were also utilized to classify the PV thermal images. The choice of cell size choice is very significant for encoding the shape information. The small cell size has the advantage of encoding more information, but the dimensionality of the feature vector also increases. So, it is a trade-off between the cell size and the dimension of the HOG feature vector.

### 3.4.4. LBP feature

The LBP descriptor is used to define the local contrast and local spatial structure of an image (Armi and Fekri-Ershad, 2019). Resilience to light variations makes it one of the most adopted methods for image classification. In this method, a neighborhood window was selected for each point. Then the intensity of central and neighboring pixels was compared to calculate the LBP matrix. The following equations were used to calculate the LBP matrix of an image.

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} S(m_p - m_c) 2^p \quad (5)$$

$$S(x) = \begin{cases} 1 & m_p - m_c > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where  $m_p$  and  $m_c$  represents the neighbor and central pixel values.  $P$  and

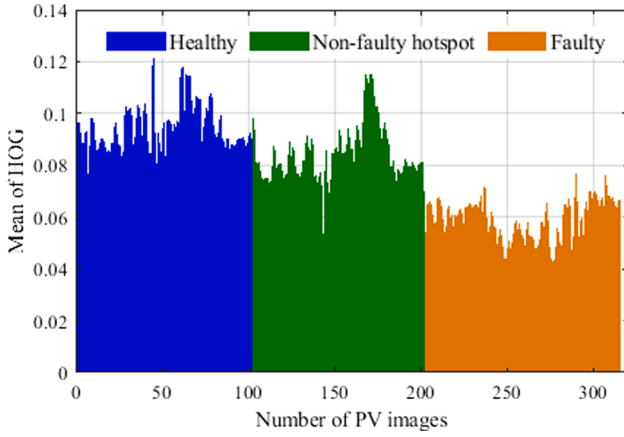


Fig. 6. Plot of HOG mean under different condition of PV panel.

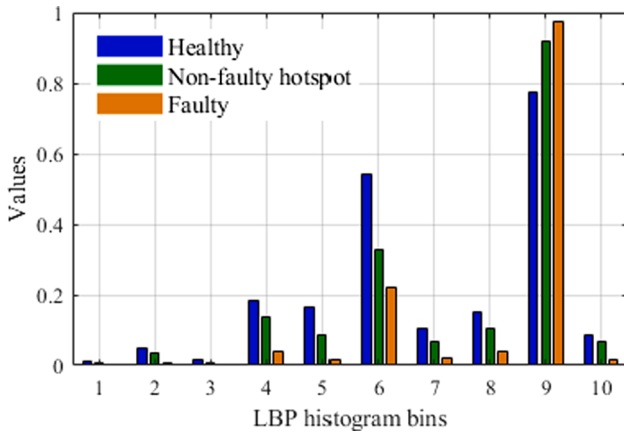


Fig. 7. Bar plots of LBP features.

$R$  represents the number of neighbors and the distance of neighborhood, respectively.  $S(x)$  is the sign function. The MATLAB command of “extractLBPFeatures” was used to extract the LBP features.

#### 3.4.5. Training dataset feature selection

The selection of sensitive critical features is one of the most critical phases in the training of machine learning algorithms. The selected features should show significant and relevant changes with the change in the input. The intensity of the color features (RGB) changes with the change in the condition of PV modules, as shown in Fig. 4. In this study, the skewness index (SI) of RGB feature is calculated to classify the PV images. The SI is defined in the below equation (Naeem et al., 2020).

$$\text{Skewness index (SI)} = \frac{\sum_{\epsilon=0}^{n-1} (y_{\epsilon} - \mu)^3 \Psi(\epsilon)}{\sigma_{\epsilon}^3} \quad (7)$$

where  $y_{\epsilon}$  and  $\Psi(\epsilon)$  are the input intensity range of the image and histogram probability, respectively. The MATLAB commands of “graycomatrix and graycoprops” were used to extract the texture feature with 3-pixel sample spacing (1, 2, and 3) and four uniformly distributed angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ ), as discussed in Section 3.4.2. Fig. 5 shows the values as features of contrast, correlation, homogeneity, and energy for different spacing and distributed angles, respectively.

It is evident from Fig. 5 that the contrast, correlation, and homogeneity have shown reasonable variation in all four angles, whereas energy has almost similar behavior for all four directions, also observed by (Niazi et al., 2019a). It can also be noted that the correlation and homogeneity possess similar behavior under different conditions of PV panels. For HOG feature extraction, (Niazi et al., 2019a) used  $4 \times 4$  cell

size to encode image information and handle dimensionality issues. They used principal component analysis to reduce the dimension of the HOG features. They used 130 HOG features to train their proposed model. The smaller size of the cell encodes more information of the image, as discussed in Section 3.4.3.

In this work, the mean of the HOG features with  $2 \times 2$  cell size is considered to reduce the dimensionality issues without affecting the training accuracy. The behavior of the mean HOG feature under different PV panel conditions is shown in Fig. 6. The 10-image feature of LBP algorithm is shown in Fig. 7.

The final used feature vector space contains 3, 12, 12, 3, 1, and 10 (total 41) features of RGB, contrast, correlation, energy, HOG, and LBP for training, respectively.

#### 3.5. Support vector machine (SVM)

In this section, the SVM classification model’s design is presented in detail (Cristianini and Shawe-Taylor, 2000; Herbrich, 2002). The SVM classifier adjudicates the maximum margin hyperplane depending upon the number of classes. The SVM mapped highly nonlinear input data into a high dimensional feature space, separated using decision (linear/nonlinear) boundaries. SVM strives to enhance the gap between the training data points and hyperplanes, also known as support vectors (SVs). In this work, the binary classification SVM algorithm presented here for simplicity.  $A = (x_a, y_a; a = 1, 2, 3, \dots, n)$ ,  $x_a \in \mathbb{R}^n$ , and  $y_a = \pm 1$  are the data set, input variable and class labels, respectively. The splitting hyperplane for input data  $A$  can be written as following equation.

$$y(x) = \omega^T \cdot \psi(x) + b_1 = 0 \quad \psi(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_h} \quad (8)$$

$\psi(\cdot)$  converts the input data into high dimensional feature space.  $\omega^T$  and  $b_1$  are the weight of  $\psi(x)$  and deviation, respectively. To following cost function should be minimized by increasing the gap between hyperplanes and data points for optimum solution.

$$\min R(\omega, b_1, \xi_n) = 1/2 \omega^T \omega + D \sum_{j=1}^m \xi_n \quad (9)$$

subject to

$$\begin{aligned} y_a [\omega^T \cdot \psi(x_a) + b_1] &\geq 1 - \xi_n \\ \xi_n &\geq 0, \quad a = 1, 2, 3, \dots, n \end{aligned} \quad (10)$$

where  $D$  is the non-negative user defined value ( $D > 0$ ) and  $\xi_n$  is the positive measure of misclassification training error. To solve this optimization problem Lagrange multiplier method can be used.

$$\mathcal{L}(\omega, b_1, \xi, \beta, \vartheta) = R(\omega, b_1, \xi_n) - \sum_{a=1}^n \beta_a (y_a [\omega^T \cdot \psi(x_a) + b_1] - 1 + \xi_a) - \sum_{a=1}^n \vartheta_a \xi_a \quad (11)$$

where  $\beta$  and  $\vartheta$  are the Lagrange multiplier, these values are zero exclusive of small input subset, these small input subset is known as SVs.

The resultant optimized target function can be written in the following form.

$$y(x) = \sum_{a=1}^n \beta_a K(x, x_a) + b_1 \quad (12)$$

where  $K(x, x_a)$  is the kernel function. Different kernel functions can be found in literature such as linear  $K(x, x_a) = (x_a^T x)$ , polynomial  $K(x, x_a) = (x_a^T x + 1)^d$ , radial basis function (RBF)  $K(x, x_a) = \exp(-\|x - x_a\|^2 / 2\gamma^2)$ , and multi-layer perceptron (MLP)  $K(x, x_a) = \tanh(\lambda_{ia} / 2\sigma^2)$  (Cristianini and Shawe-Taylor, 2000; Herbrich, 2002).

To calculate the values of hyperparameters, different methods can be found in the literature (Picard and Cook, 1984; Shawe-Taylor and Cristianini, 2004; Steinwart and Christmann, 2008). Some of them are

**Table 3**

The performance comparison of different classifiers for training dataset I (total 192 features).

Classifiers	Class	TPR	FNR	PPV	FDR	ROC	F-Measures	Sensitivity	Precision	Training time (ms)	Accuracy(%)
<b>QDA</b>	Healthy	80.4	19.6	87.2	12.8	0.95	83.66	80.4	87.2	2900	87.9
	Non-faulty hotspot	86.9	13.1	83.5	16.5	0.94	85.17	86.9	83.5		
	Faulty	95.6	4.4	92.4	7.6	0.99	93.97	95.6	92.4		
<b>n-Bayes</b>	Healthy	91.2	8.8	95.9	4.1	0.99	93.49	91.2	95.9	9400	94.6
	Non-faulty hotspot	93.9	6.1	92.1	7.9	0.98	92.99	93.9	92.1		
	Faulty	98.2	1.8	95.7	4.3	0.99	96.93	98.2	95.7		
<b>KNN</b>	Healthy	82.4	17.6	84.0	16.0	0.95	83.19	82.4	84.0	900	81.3
	Non-faulty hotspot	82.8	17.2	75.2	24.8	0.95	78.82	82.8	75.2		
	Faulty	78.9	21.1	84.9	15.1	0.91	81.79	78.9	84.9		
<b>BE</b>	Healthy	82.8	17.6	90.3	9.7	0.95	86.39	82.8	90.3	3820	88.9
	Non-faulty hotspot	87.9	12.1	79.1	20.9	0.94	83.27	87.9	79.1		
	Faulty	95.6	4.4	97.3	2.7	0.99	96.44	95.6	97.3		
<b>SVM</b>	Healthy	96.1	3.9	93.3	6.7	1.00	94.68	96.1	93.3	1670	95.2
	Non-faulty hotspot	92.9	7.1	94.8	5.2	0.98	93.84	92.9	94.8		
	Faulty	96.5	3.5	97.3	2.7	1.00	96.9	96.5	97.3		

**Table 4**

The performance comparison of different classifiers for training dataset II (total 28 features).

Classifiers	Class	TPR	FNR	PPV	FDR	ROC	F-Measures	Sensitivity	Precision	Training time (ms)	Accuracy(%)
<b>QDA</b>	Healthy	81.4	18.6	87.4	12.8	0.95	84.29	81.4	87.4	1820	88.6
	Non-faulty hotspot	86.9	13.1	81.1	18.9	0.96	83.9	86.9	81.1		
	Faulty	96.5	3.5	96.5	3.5	0.99	96.5	96.5	96.5		
<b>n-Bayes</b>	Healthy	71.6	28.4	70.9	29.1	0.99	71.25	71.6	70.9	6360	69.5
	Non-faulty hotspot	62.6	37.4	59	41	0.98	60.75	62.6	59		
	Faulty	73.7	26.3	78.5	21.5	0.99	76.02	73.7	78.5		
<b>KNN</b>	Healthy	82.4	17.6	84.0	16.0	0.96	83.19	82.4	84.0	310	84.8
	Non-faulty hotspot	80.8	19.2	79.2	20.8	0.93	79.99	80.8	79.2		
	Faulty	90.4	9.6	90.4	9.6	0.98	90.4	90.4	90.4		
<b>BE</b>	Healthy	82.4	17.6	79.2	20.8	0.93	80.77	82.4	79.2	2810	83.8
	Non-faulty hotspot	79.8	20.2	82.3	17.7	0.94	81.03	79.8	82.3		
	Faulty	88.6	11.4	89.4	10.6	0.96	89	88.6	89.4		
<b>SVM</b>	Healthy	92.1	7.9	91.2	8.2	0.94	91.65	92.1	91.2	660	91.7
	Non-faulty hotspot	91.9	8.1	89.2	10.8	0.91	90.53	91.9	89.2		
	Faulty	91.2	8.8	94.5	5.5	0.96	92.82	91.2	94.5		

random search (RS), grid search (GS), Nelder-Mead search (NMS), particle swarm optimization (PSO), pattern search (PS), cross-validation (CV), genetic algorithm (GA), etc. In this work, the GS method was adopted to obtain the hyperparameters of SVM.

#### 4. Results

In this section, the performance comparison of different training features with multiple machine learning algorithms is investigated for the classification of PV thermal images. Three separate experiments were performed to train the models using three different training datasets (dataset I, dataset II, and dataset III). The dataset I consists of 40

texture features (12 contrast, 12 correlation, 12 homogeneity, and 4 energy) and 152 HOG features (for  $2 \times 2$  cell size using principal component analysis) (Niazi et al., 2019a). For dataset I, the optimized values of SVM hyperparameters  $D$  and  $\gamma$  are 1 and 0.004, respectively. The dataset II has 27 texture features (12 contrast, 12 correlation, and 3 energy) and 1 HOG mean feature (for  $2 \times 2$  cell size). The optimized values of SVM hyperparameters  $D$  and  $\gamma$  for dataset II are 0.46 and 0.1, respectively. The proposed hybrid feature dataset (dataset III) contains total 41 features for training, which has 3 RGB, 12 contrast, 12 correlation, 3 energy, 1 HOG, and 10 LBP features in it. The optimized values of 0.2 and 0.46 were obtained for SVM hyperparameter  $D$  and  $\gamma$ , respectively. Different machine learning algorithms such as quadratic

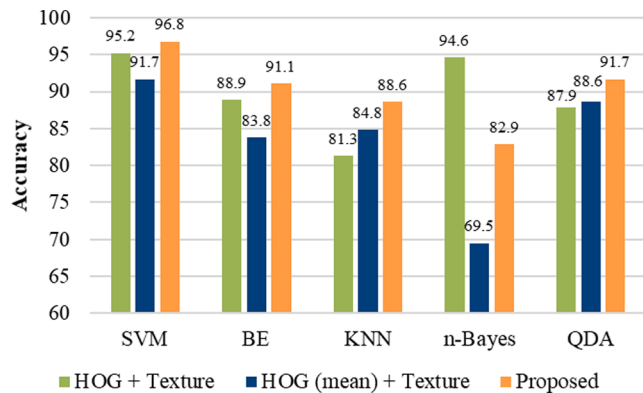
**Table 5**

The performance comparison of different classifiers for training dataset III (total 41 features).

Classifiers	Class	TPR	FNR	PPV	FDR	ROC	F-Measures	Sensitivity	Precision	Training time (ms)	Accuracy(%)
<b>QDA</b>	Healthy	86.3	13.7	89.8	10.2	0.91	88.02	86.3	89.8	480	91.7
	Non-faulty hotspot	88.9	11.1	85.4	14.6	0.93	87.11	88.9	85.4		
	Faulty	99.1	0.9	99.1	0.9	1.00	99.10	99.1	99.1		
<b>n-Bayes</b>	Healthy	77.5	22.5	77.5	29.1	0.90	77.50	77.5	77.5	7790	82.9
	Non-faulty hotspot	73.7	26.3	77.7	22.3	0.86	75.65	73.7	77.7		
	Faulty	95.6	4.4	91.6	8.4	0.97	93.56	95.6	91.6		
<b>KNN</b>	Healthy	85.3	14.7	87.9	12.1	0.96	86.58	85.3	87.9	1100	88.6
	Non-faulty hotspot	82.8	17.2	73.7	16.3	0.91	77.99	82.8	73.7		
	Faulty	96.5	3.5	93.2	6.8	0.99	94.82	96.5	93.2		
<b>BE</b>	Healthy	93.1	6.9	93.1	6.9	0.98	93.10	93.1	93.1	2810	91.1
	Non-faulty hotspot	83.8	16.2	87.4	12.6	0.92	85.56	83.8	87.4		
	Faulty	95.6	4.4	92.4	7.6	0.99	93.97	95.6	92.4		
<b>SVM</b>	Healthy	96.1	3.9	96.1	3.9	1.00	96.1	96.1	96.1	700	96.8
	Non-faulty hotspot	95.9	4.1	96.9	3.1	0.99	96.4	95.9	96.9		
	Faulty	98.2	1.8	97.3	2.7	1.00	97.75	98.2	97.3		

**Table 6**  
Confusion matrix using proposed hybrid features (dataset III).

Class	Classified as		
	Healthy	Non-faulty hotspot	Faulty
Healthy	<b>98</b>	2	2
Non-faulty hotspot	3	<b>95</b>	1
Faulty	1	1	<b>112</b>



**Fig. 8.** Comparative analysis of the different classifiers under different training dataset.

discriminant analysis (QDA), naïve-Bayes (n-Bayes), k nearest neighbor (KNN), bagging ensemble (BE), and SVM were employed to compare the results. The true positive rate (TPR), false-negative rate (FNR), positive predictive value (PPV), false discovery rate (FDR), receiver-operating characteristic (ROC), F-Measures, and training time of all the algorithms with different training dataset are summarized in Tables 3–5.

The confusion matrix of the proposed SVM classifier using training dataset III is presented in Table 6. The bold diagonal elements show the correctly predicted values. The proposed SVM model accurately predicts 305 out of 315 PV thermal images. The performance comparison of all training algorithms for different training datasets is shown in Fig. 8.

To check the testing accuracy of the proposed method, a new dataset of 75 samples (i.e., not included in the training dataset) was formed. The new dataset consisting of 27 healthy, 26 non-faulty hotspot, and 22 faulty were then used to test the proposed SVM model trained on dataset III. Table 7 reports the result of the testing dataset using SVM based model trained on dataset III. It is noted that the proposed method correctly classified 69 out of 75 PV thermal images.

## 5. Discussion

Infrared thermal imaging is the most widely adopted non-destructive methodology to detect the degradation and faults in the PV system. The short circuit and shunted cell/modules generated the heat due to the Joule heating effect. The thermal imaging can easily detect brighter spots due to temperature gradients in PV panels. The early detection of hotspots is essential to ensure the reliability and durability of the PV systems.

In this work, the PV thermal images classification performance of

QDA, n-Bayes, KNN, BE, and SVM algorithms was analyzed using different training datasets. The dataset I have 192 features (152 HOG and 40 texture) for the classifiers (Niazi et al., 2019a). The n-Bayes and SVM classifier shows high classification accuracy with long training time, as shown in Table 3. The KNN method takes less training time but has a high misclassification rate. The mean of the HOG feature combined with 27 texture features (dataset II) was used for classification to reduce the computational complexity and storage size. It results in less training time, but the classification accuracy of all algorithm also reduces except KNN. The SVM shows the best classification rate, with only 26 misclassified samples (see Table 4). Afterward, the LBP and RGB feature separately used for the training of the algorithms. Among all the implemented techniques, SVM showed the best classification accuracy of 87.9% and 89.2% for LBP and RGB features, respectively. To enhance the classification accuracy, the features of LBP and RGB were fused with the dataset II. The fused training dataset (dataset III) has a total of 41 features for the classification of PV images. The results of the proposed training dataset (dataset III) increased the SVM classifier accuracy up to 96.8% (see Table 5). It is noted that only 10 samples were misclassified using the proposed classifier with the hybrid training dataset. The proposed classifier utilized only 0.7 s to train the model. Using hybrid features (dataset III), all other classifiers also show their individual best accuracy except n-Bayes. The accuracy of the n-Bayes classifier reduced due to the mean HOG feature. Tables 3–5 and Fig. 8 clearly showed that the SVM has the best classification under all training datasets. The hybrid dataset has the best accuracy with low computational complexity and required almost five-time lesser storage space. Furthermore, the proposed features with SVM also showed good results for the testing dataset. 69 samples out of 75 (see Table 7) of the testing data were correctly classified in their correct classes, which is good enough and highly acceptable.

## 6. Conclusion

In this work, an IRT based hotspot detection and classification approach for PV modules were proposed. Different training feature vectors (dataset I, dataset II, and dataset III) were used and analyzed to discriminate between healthy, non-faulty hotspot, and faulty PV panels. After extensive training and analysis, the hybrid feature dataset III (3 RGB, 12 contrast, 12 correlation, 3 energy, 1 HOG, and 10 LBP) was constructed using a data fusion approach. The proposed hybrid features (dataset III) correctly classified 305 out of 315 PV thermal samples using the proposed SVM model. The proposed hybrid feature dataset with 96.8% training accuracy required almost five-time lesser storage space than the previously used dataset (dataset I) and has low computational complexity. The proposed trained model also shows high accuracy of 92% when tested with a new dataset. The proposed hybrid dataset SVM model with drone IRT can be utilized to identify the defective PV panel to reduce the plant operator efforts. In future work, this work can be extended to detect and classify the specific PV defect using IRT and advanced machine learning techniques.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Table 7**  
Performance of SVM based model for the testing data.

Class	Classified as			TPR	FNR	PPV	FDR	Accuracy(%)
	Healthy	Non-faulty hotspot	Faulty					
Healthy	<b>26</b>	1	0	96.29	3.71	89.65	10.35	92.00
Non-faulty hotspot	2	<b>23</b>	1	88.46	11.54	92.00	8.00	
Faulty	1	1	<b>20</b>	90.90	9.10	95.23	4.77	



## References

- Aghaei, M., Grimaccia, F., Gonano, C.A., Leva, S., 2015. Innovative automated control system for PV fields inspection and remote control. *IEEE Trans. Ind. Electron.* 62 (11), 7287–7296.
- Ahsan, S., Niazi, K.A.K., Khan, H.A., Yang, Y., 2018. Hotspots and performance evaluation of crystalline-silicon and thin-film photovoltaic modules. *Microelectron. Reliab.* 88–90, 1014–1018.
- Ali, Zafar, Nengroo, Hussain, Park, Kim, 2019. Online remaining useful life prediction for lithium-ion batteries using partial discharge data features. *Energies* 12 (22), 4366.
- Allan, D., Jepson, F., Estrada, J., 2009. Benchmarking image segmentation. *Int. J. Comput. Vis* 85, 167–181.
- Armi, L., Fekri-Ershad, S., 2019. Texture image analysis and texture classification methods-A review. *Arxiv preprint arxiv:1904.06554*.
- Banerji, S., Sinha, A., Liu, C., 2013. Haarhog: Improving the hog descriptor for image classification. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4276–4281.
- Berardone, I., Lopez Garcia, J., Paggi, M., 2018. Analysis of electroluminescence and infrared thermal images of monocrystalline silicon photovoltaic modules after 20 years of outdoor use in a solar vehicle. *Sol. Energy* 173, 478–486.
- Breitenstein, O., Bauer, J., Lotnyk, A., Wagner, J.M., 2009. Defect induced non-ideal dark – characteristics of solar cells. *Superlattices Microstruct.* 45 (4–5), 182–189.
- Cristianini, N., Shawe-Taylor, J., 2000. An introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893.
- Deitsch, S., Christlein, V., Berger, S., Buerhop-Lutz, C., Maier, A., Gallwitz, F., Riess, C., 2019. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Sol. Energy* 185, 455–468.
- Dunderdale, C., Brettigny, W., Clohessy, C., Dyk, E.E., 2019. Photovoltaic defect classification through thermal infrared imaging using a machine learning approach. *Prog. Photovolt.* 28 (3), 177–188.
- Fleisch, H., Roch, J., 2013. Benchmark Evaluation of HOG Descriptors as Features for Classification of Traffic Signs. Working papers in transport, tourism, information technology and microdata analysis. Höögskolan Dalarna, Borlänge, p. 18.
- Garoudja, E., Harrou, F., Sun, Y., Kara, K., Choudier, A., Silvestre, S., 2017. Statistical fault detection in photovoltaic systems. *Sol. Energy* 150, 485–499.
- Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 3 (6), 610–621.
- Herbrich, R., 2002. Learning Kernel Classifiers: Theory and Algorithms. MIT Press, Cambridge, MA, USA.
- Hohm, D.P., Ropp, M.E., 2003. Comparative study of maximum power point tracking algorithms. *Prog. Photovoltaics Res. Appl.* 11 (1), 47–62.
- IEA, 2019. Global Energy & CO2 Status Report 2019. IEA (International Energy Agency): Paris, France.
- Jaffery, Z.A., Dubey, A.K., Irshad, Haque, A., 2017. Scheme for predictive fault diagnosis in photovoltaic modules using thermal imaging. *Infrared Phys. Technol.* 83, 182–187.
- Jordan, D.C., Kurtz, S.R., VanSant, K., Newmiller, J., 2016. Compendium of photovoltaic degradation rates. *Prog. Photovolt.* 24 (7), 978–989.
- Kaizuka, I., Jäger-Waldau, A., Donoso, J., 2020. Snapshot of Global PV Markets – 2020. In: Masson, G. (Ed.). International Energy Agency Photovoltaic Power System Programme.
- Khan, M., Zeb, K., Uddin, W., Sathishkumar, P., Ali, M., Hussain, S., Ishfaq, M., Himanshu, Subramanian, A., Kim, H.-J., 2018. Design of a building-integrated photovoltaic system with a novel bi-reflector PV system (BRPVS) and optimal control mechanism: an experimental study. *Electronics* 7(7), 119.
- Kumar, M., Kumar, A., 2017. Performance assessment and degradation analysis of solar photovoltaic technologies: A review. *Renew. Sustain. Energy Rev.* 78, 554–587.
- Liu, H., Zhang, C., Huang, D., 2017. Extreme learning machine and moving least square regression based solar panel vision inspection. *J. Elect. Comp. Eng.* 2017, 1–10.
- Livera, A., Theristis, M., Makrides, G., Georgiou, G.E., 2019. Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems. *Renew. Energy* 133, 126–143.
- Lupangu, C., Bansal, R.C., 2017. A review of technical issues on the development of solar photovoltaic systems. *Renew. Sustain. Energy Rev.* 73, 950–965.
- Madeti, S.R., Singh, S.N., 2017a. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Sol. Energy* 158, 161–185.
- Madeti, S.R., Singh, S.N., 2017b. Monitoring system for photovoltaic plants: A review. *Renew. Sustain. Energy Rev.* 67, 1180–1207.
- Market, A., 2018. Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2016–2024. URL: <https://www.transparencymarketresearch.com/logistics-market.html> (Last accessed: 16.01. 2018).
- McKinsey, 2019. Global Energy Perspective 2019: Reference Case – McKinsey, 2019.
- Mellit, A., Tina, G.M., Kalogiou, S.A., 2018. Fault detection and diagnosis methods for photovoltaic systems: A review. *Renew. Sustain. Energy Rev.* 91, 1–17.
- Minkina, W., Dudzik, S., 2009. Infrared Thermography: Errors and Uncertainties. John Wiley & Sons.
- Naem, S., Ali, A., Qadri, S., Mashwani, W.K., Tairan, N., Shah, H., Fayaz, M., Jamal, F., Chesneau, C., Anam, S., 2020. Machine-learning based hybrid-feature analysis for liver cancer classification using fused (MR and CT) images. *Appl. Sci.* 10 (9), 3134.
- Nengroo, S., Kamran, M., Ali, M., Kim, D.-H., Kim, M.-S., Hussain, A., Kim, H., 2018. Dual battery storage system: an optimized strategy for the utilization of renewable photovoltaic energy in the United Kingdom. *Electronics* 7 (9), 177.
- Nengroo, S.H., Umair Ali, M., Zafar, A., Hussain, S., Murtaza, T., Junaid Alvi, M., Raghavendra, K.V.G., Jee Kim, H., 2019. An optimized methodology for a hybrid photo-voltaic and energy storage system connected to a low-voltage grid. *Electronics* 8 (2), 176.
- Niazi, K., Akhtar, W., Khan, H.A., Sohaib, S., Nasir, A.K., 2018. Binary classification of defective solar PV modules using thermography. In: IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), pp. 0753–0757.
- Niazi, K., Khan, H.A., Amir, F., 2018b. Hot-spot reduction and shade loss minimization in crystalline-silicon solar panels. *J. Renew. Sustain. Energy* 10 (3), 033506.
- Niazi, K.A.K., Akhtar, W., Khan, H.A., Yang, Y., Athar, S., 2019a. Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier. *Sol. Energy* 190, 34–43.
- Niazi, K.A.K., Yang, Y., Sera, D., 2019b. Review of mismatch mitigation techniques for PV modules. *IET Renew. Power Gener.* 13 (12), 2035–2050.
- Niazi, K.A.K., Yang, Y., Spataru, S.V., Mutarraf, M., Séra, D., 2019c. Experimental benchmarking of partial shading effect on thin-film and crystalline-silicon solar photovoltaic modules. 36th European Photovoltaic Solar Energy Conference and Exhibition.
- Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. *J. AM. Stat. Assoc.* 79 (387), 575–583.
- Ram, M., Bogdanov, D., Aghahosseini, A., Gulagi, A., Oyewo, A., Child, M., Caldera, U., Sadovskaia, K., Farfan, J., Barbosa, L., 2019. Global energy system based on 100% renewable energy-power, heat, transport and desalination sectors. Study by Lappeenranta University of Technology and Energy Watch Group, Lappeenranta, Berlin.
- Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.
- Steinwart, I., Christmann, A., 2008. Support Vector Machines. Springer Science & Business Media.
- Stone, M., 1974. Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Series B Stat. Methodol.* 36 (2), 111–133.
- Sundermann, B., Bode, J., Lueken, U., Westphal, D., Gerlach, A.L., Straube, B., Wittchen, H.U., Strohle, A., Wittmann, A., Konrad, C., Kircher, T., Arolt, V., Pfeleiderer, B., 2017. Support vector machine analysis of functional magnetic resonance imaging of interoception does not reliably predict individual outcomes of cognitive behavioral therapy in panic disorder with agoraphobia. *Front. Psychiatry* 8, 99.
- Tina, G.M., Cosentino, F., Ventura, C., 2016. Monitoring and diagnostics of photovoltaic power plants. In: Sayigh, A. (Ed.) Renewable Energy in the Service of Mankind Vol. II: Selected Topics from the World Renewable Energy Congress WREC 2014. Springer International Publishing, Cham, pp. 505–516.
- Trupke, T., Mitchell, B., Weber, J.W., McMillan, W., Bardos, R.A., Kroeze, R., 2012. Photoluminescence imaging for photovoltaic applications. *Energy Procedia* 15, 135–146.
- Tsanakas, J.A., Ha, L., Buerhop, C., 2016. Faults and infrared thermographic diagnosis in operating c-Si photovoltaic modules: A review of research and future challenges. *Renew. Sustain. Energy Rev.* 62, 695–709.
- Varun, Prakash, R., Bhat, I.K., 2009. Energy, economics and environmental impacts of renewable energy systems. *Renew. Sustain. Energy Rev.* 13(9), 2716–2721.
- Vergura, S., Marino, F., 2017. Quantitative and computer-aided thermography-based diagnostics for PV devices: Part I—framework. *IEEE J. Photovolt.* 7 (3), 822–827.
- Wang, W., Liu, A.C.-F., Chung, H.S.-H., Lau, R.W.-H., Zhang, J., Lo, A.W.-L., 2016. Fault diagnosis of photovoltaic panels using dynamic current-voltage characteristics. *IEEE Trans. Power Electron.* 31 (2), 1588–1599.
- Woyte, A., Nijss, J., Belmans, R., 2003. Partial shadowing of photovoltaic arrays with different system configurations: literature review and field test results. *Sol. Energy* 74 (3), 217–233.
- Woyte, A., Richter, M., Moser, D., Mau, S., Reich, N., Jahn, U., 2013. Monitoring of photovoltaic systems: good practices and systematic analysis. In: Proc. 28th European Photovoltaic Solar Energy Conference, pp. 3686–3694.