Photovoltaic Reliability Literature Review and Solar Image Classification Analysis

Qing Guo, Yueyao Wang

1 Introduction

The solar energy generation is becoming more important worldwide. The research about photovoltaic (PV) reliability also receives increasing attentions. For customers and investors, minimizing the cost while improving the durability of PV system is always attractive. This is the main driver behind the growing interest in the accurate calculation and prediction of degradation of PV under real operating conditions. The photovoltaics technology converts of light into electricity using semiconducting materials. A photovoltaic system usually contains solar panels to convert sunlight into electricity, a solar inverter to convert the output from direct to alternating current, as well as other electrical accessories. Each of these component may introduce system faults. In this report, we did a literature review regarding the photovoltaics (PV) reliability study. We covered the following areas: the PV degradation mechanism, potential data type and source, and techniques to analysing PV related data. Besides we also present a data analysis that uses machine learning models to analysis solar image data.

The overview of the rest of the report is as follow: Section 2 introduced common failure modes that leads to the system failure. Section 3 presents different data types that can be used to analysis the PV reliability and the performance metrics of PV degradation. Section 4 shows some other statistical models that are commonly used to analysis the PV reliability data. Section 5 presents a machine learning model to do image classification using Electroluminescence Image Data.

2 Photovoltaic Common Failure Modes

In a PV system, the solar panel and inverter are main component that receives attention when study its reliability. Photovoltaic cells or solar cells are the smallest units that are connected electrically in series and/or parallel circuits to produce higher voltages, currents and power levels. Photovoltaic modules consist of PV cell circuits sealed in an environmentally protective laminate, and are the fundamental building blocks of PV systems. And a photovoltaic array is the complete power-generating unit, consisting of any number of PV modules and panels.

Performance degradation can happen at all levels, (i.e, cell, module, array and system) with different factors and degradation mechanisms apparent at each level. There are both intrinsic and extrinsic factor that leads to the performance degradation. The main extrinsic factors related to the degradation includes temperature, humidity, precipitation, dust, snow and solar irradiation. At the array level, shading and module mismatches are additional factors that contribute to degradation.

For the intrinsic factors, at the PV cell level, possible failures may caused by corrosion, light induced degradation and cracked cells. At the module level, glass breakage, busbar failure and broken interconnects can contribute to the degradation. At the system level, potential induced degradation [2] is the most recent reported reason that caused by the high voltage stress.

3 PV Reliability Data

3.1 Data Types

Due to relative new of PV technology, the failure time data in time is rare. Degradation data and image data are two common data types that can be used to perform the reliability analysis.

3.1.1 Degradation Data

In PV reliability study, the extent by which that different degradation mechanism affect the PV system is of interest. Degradation is usually a time series data. To evaluate the degradation performance of a PV system, the following measurements are often collected as the metrics: (1) Direct current and voltage at the maximum power point (MPP) of the module or array: I_{MPP} and V_{MPP} and power P_{MPP}; (2) Alternating current power of PV system P_{AC}; (3) Meteorological measurements such as the temperatures, wind speed, relative humidity. The metrics can be eletrical parameters from IV curves; regression models such as the Photovoltaics for Utility Scale Applications (PVUSA) and normalized or scaled ratings, such as P_{MPP}/P_{MAX}.

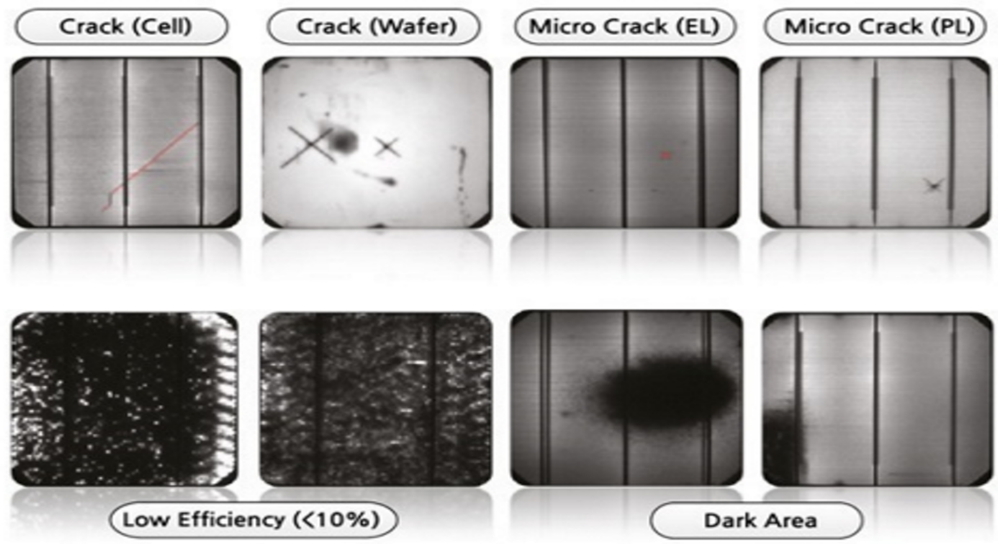

Figure 1: An example of EL image data

3.1.2 Electroluminescence Image Data

Visual insepct of solar panels is another important source to gather failure information about PV system. Some failure modes can not easily detect by human visual inspection, thus, thermographic and luminescence based imaging are used to detect those faults. Electroluminescence (EL) imaging is a non-destructive technology for failure analysis of PV modules with high resolution. To obtain an EL image, current is applied to a PV module, which induces EL emission which can be imaged by a silicon Charge-coupled Device (CCD) sensor. In EL images, defective cells appear darker, because disconnected parts do not irradiate. Figure 1 provide an example of the EL image, where we can tell different failure type by the dark area in the photo.

3.2 Potential Data Sources

There are two potential PV reliability in literature. The first is the github repository that provide EL image data [3]. This repo (`https://github.com/zae-bayern/elpv-dataset`) provides a dataset of solar cell images extracted from high-resolution electroluminescence images of photovoltaic modules. And another data source is from the Regional Test Center. The Regional Test Centers are a group for testing photovoltaic systems and components related to photovoltaic systems. It collects and provides PV quality data system [1].

4 Models

The measurements of degradation rates are important in assessing the effective lifetime of a PV module. Some manufactures start to include the maximum degradation rate in warrant. Therefore, the annual degradation rate, in units of %/year is usually of interest in literature [7]. Since the PV performance degradation data is often in the format of time-series, calculating the trend of the performance time series becomes the goal of the statistical analysis. In literature, methods used to calculate the degradation rate includes linear regression, classical seasonal decomposition and auto regressive integrated moving average (ARIMA) model.

4.1 Statistical Models

The most commonly used methods is lienar regression. In linear regression model, the performance metrics is assumed to linearly related to the time t, which is

$$y_i = \alpha t_i + \beta + \epsilon_i.$$

Where y_i represents the ith time slot measured performance metric, $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, for $i = 1, \ldots, n$. And α is the slope of the trend, which is the desired degradation rate. However, the estimation obtained by ordinary least squares is sensitive to outliers and may have large uncertainty. Some other time series models are also used to overcome the limitations.

The classical seasonal decomposition (CSD) model assumes the performance metrics can be decomposed into seasonal trend and overall trend. So the model is expressed as

$$y_i = T_i + S_i + e_i,$$

where T_i is the overall performance degradation trend, and S_i accounts for the seasonal trend, and e_i is the residual component. The seasonal component is assumed to be stable over the years.

ARIMA model is also used in the literature [5]. Degradation rates, i.e. the change of power with time, is proportional to the power and a random error term:

$$\frac{dP}{dt} = c + P + \epsilon.$$

However, the power change is observed in very large time intervals, so the above differential can be written as

$$P_t = \Phi P_t + \delta + \epsilon_t.$$

An ARIMA model predicts the current value in a time series as a linear combination of past values (the autoregressive part) and past random fluctuations (the moving average part).

An example of ARIMA with seasonality period of 12 months can be expressed as

$$P_t + P_{t-12} - \Phi P_{t-1} + \Phi P_{t-13} = \delta + \epsilon_t + \theta \epsilon_{t-12},$$

where Φ and θ are autoregressive and moving average coefficient.

4.2 Machine Learning Models

When dealing with the image data, machine learning models [4] [6], can be used to perform feature extraction and classification. The commonly used model in literature includes support vector machine (SVM), Random Forest (RF) and convolutional neural networks (CNN).

Most recent work about EL image classification including Sun et al. [8], Mantel et al. [6] and Deitsch et al. [4]. Mingjian Sun et al. present an automatic fault detection using deep learning CNN on EL images. Due to the dataset size limitation, data augmentation is applied. In Deitsch et al. work, both CNN and SVM are stuided.

Mantel et al. work uses SVM and Random Forest to identify PV defects or faults category. In this work, the EL images are automatically split into cells unlike the above two studies. To perform cell extraction, all cell borders are assumed to be parallel to the row and column plane of the input image. Then the region with potential fault are extracted using Hough transformation.

Bus bars and cell boundaries are removed before the image feature extraction. Then 25 features including the geometry of the regions and the statistical characteristics of the pixel values are extracted from the processed image. The SVM and RF models are used to classify images into finger failures and three types of cracks.

5 Real Data Analysis

5.1 Data Introduction

A public dataset (`https://github.com/zae-bayern/elpv-dataset`) of solar cells extracted from high resolution EI images of monocrystalline and polycrystalline PV modules.

The dataset contains 2,624 samples of 300x300 pixels 8-bit grayscale images of functional and defective solar cells with varying degree of degradation extracted from 44 different solar moduleswhere 18 modules are of monocrystalline type, and 26 are of polycrystalline type. The defects in the annotated images are either of intrinsic or extrinsic type and are known to reduce the power efficiency of solar modules.

All images are normalized with respect to size and perspective. Additionally, any distortion induced by the camera lens used to capture the EL images was eliminated prior to solar cell extraction.

Then we will talk about how to process these data. They hired some evaluators and ask them two questions. The first is how do they think the function of the solar cells, functional or defective; the second is how they are confident about their assessments. Then we can get four different situations as Table 1

According to the probability how likely the product is defective, we assign four labels to sollar cells, 0%, 33.33%, 66.66%, 100%. CNN model is a very popular machine learning method to deal with image data. We split our data into training data and validation data to avoid overfiting problem. Table 2 presents one split situation of our data. We can find there are much more data labeled 0% or 100% and only a small amount of data are assigned to 33.33% and 66.66% in both training and testing data.

Table 1: Partitioning of solar cells into functional and defective, with an additional self-assessment on the rater's confidence after visual inspection. Non-confident decisions obtain a weight lower than 100% for the evaluation of the classifier performance

Condition	Confident?	Label p	Weight w
functional	✓	functional	100%
	✗	defective	33%
defective	✓	defective	100%
	✗	defective	67%

Table 2: The distribution of the total number of solar cell images in the dataset depending on sample label p and the PV module type from which the solar cells were originally extracted. The numbers of solar cell images are given for one training /test split.

Solar water	Train				Test				Total
	0%	33%	67%	100%	0%	33%	67%	100%	
Monocrystalline	438	87	41	249	150	30	15	64	1074
Polycrystatline	683	132	37	301	237	46	13	101	1550
Total	1221	219	78	550	387	76	28	165	2624

5.2 Results

We split our data to training data (80%) and test data (20%), then fitting a CNN model. In CNN model, we use Relu activation function and set the kernel size 3×3 and stride 1×1. We also allow our model to pad. After ten epochs, the model converges. We use the fitted model to calculate confusion matrix in both training and testing data. Figure 3 and Figure 4 shows the result. It displays that model have a good prediction in label 0% and 100% in both training and testing data. The reason is that these two labels have more data which can offer much more information, then the prediction will be high.

There are some metrics to assess the effect of binary model, such as precision, recall, F1, and so on. Here we have four labels in our model, thus choosing accuracy and Cohen's Kappa. Next, we change the proportion of our training set and want to explore how the proportion have an influence on model model prediction. We select the proportion from 0.05 to 0.9 and the step size is 0.05. After fitting models, we calculate the accuracy and kappa

Table 3: Confusion Matrix on Train Set with 80% Training Samples

Prediction \ Truth	0%	33.33%	66.66%	100%
0%	1069	150	43	88
33.33%	4	35	0	1
66.66%	0	0	14	0
100%	133	51	28	483
Total	1206	236	85	572

Table 4: Confusion Matrix on Test Set with 80% Training Samples

Prediction \ Truth	0%	33.33%	66.66%	100%
0%	249	40	11	39
33.33%	0	2	2	0
66.66%	0	0	14	0
100%	53	17	8	104
Total	302	59	21	143

on testing data. We are able to get 18 results in total. The accuracy of CNN under different proportion of training set is shown in Figure 2. Here we can find after proportion bigger than 0.15, the proportion does not have much effect on model accuracy. However, in Figure 3, it shows that the proportion of training set indeed have effects on Kappa, because it is harder to converge. Also, Kappa cannot be improved over 0.5, but accuracy can reach out 0.7.

5.3 Conclusion

We tried a well-known model CNN to analysis image data. In the model assessments, confusion matrix, accuracy, and Cohen's kappa are all considered. We find that the accuracy in the experiment is not sensitive to the proportion of training set, while Cohen's Kappa is significantly influenced by the proportion.

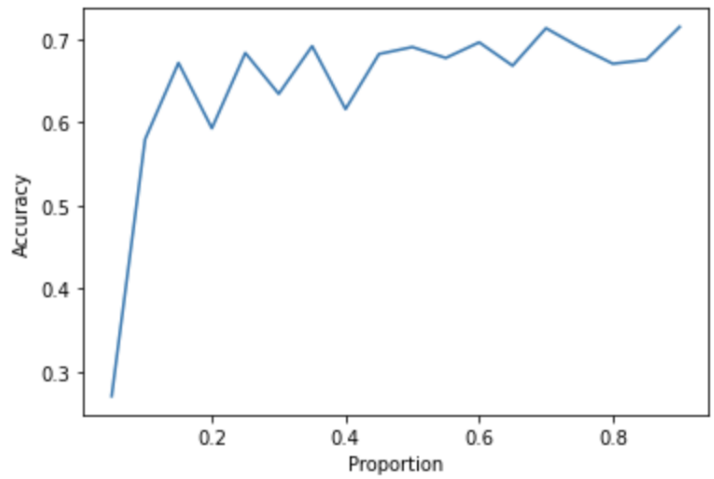

Figure 2: Accuracy of CNN model under different proportion training data

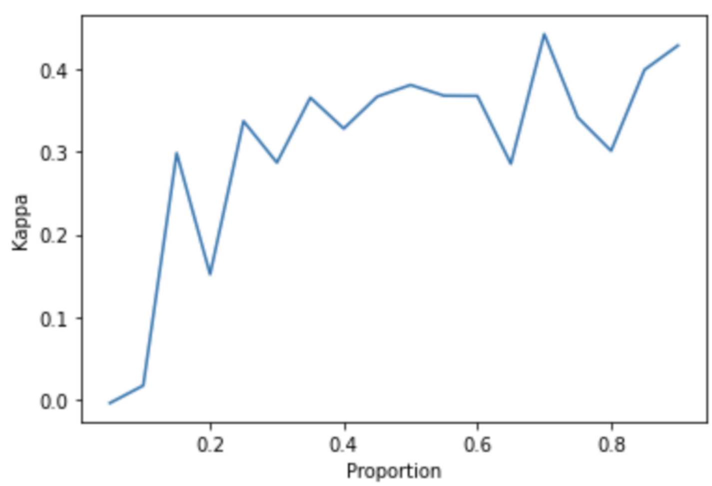

Figure 3: Kappa of CNN model under different proportion training data

There are some strategies we can explore to improve the model performance. Firstly, adjusting model parameters can probably be very useful. Secondly, add more hidden layers to model. Also, we can try other activation function, such as tanh, softplus, logistic and so on.

References

[1] Reginal Test Centers monitoring and measurements. `https://rtc.sandia.gov/about-pvrtc/design-of-the-regional-test-centers/`.

[2] J Berghold, O Frank, H Hoehne, S Pingel, B Richardson, and M Winkler. Potential induced degradation of solar cells and panels. *25th EUPVSEC*, pages 3753–3759, 2010.

[3] Claudia Buerhop-Lutz, Sergiu Deitsch, Andreas Maier, Florian Gallwitz, Stephan Berger, Bernd Doll, Jens Hauch, Christian Camus, and Christoph J. Brabec. A benchmark for visual identification of defective solar cells in electroluminescence imagery. In *European PV Solar Energy Conference and Exhibition (EU PVSEC)*, 2018.

[4] Sergiu Deitsch, Vincent Christlein, Stephan Berger, Claudia Buerhop-Lutz, Andreas Maier, Florian Gallwitz, and Christian Riess. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*, 185:455–468, June 2019.

[5] Dirk C Jordan and Sarah R Kurtz. Analytical improvements in pv degradation rate determination. In *2010 35th IEEE Photovoltaic Specialists Conference*, pages 002688–002693. IEEE, 2010.

[6] Claire Mantel, Frederik Villebro, Gisele Alves dos Reis Benatto, Harsh Rajesh Parikh, Stefan Wendlandt, Kabir Hossain, Peter Poulsen, Sergiu Spataru, Dezso Sera, and Søren Forchhammer. Machine learning prediction of defect types for electroluminescence images of photovoltaic panels. In *Applications of Machine Learning*, volume 11139, page 1113904. International Society for Optics and Photonics, 2019.

[7] Alexander Phinikarides, Nitsa Kindyni, George Makrides, and George E Georghiou. Review of photovoltaic degradation rate methodologies. *Renewable and Sustainable Energy Reviews*, 40:143–152, 2014.

[8] Mingjian Sun, Shengmiao Lv, Xue Zhao, Ruya Li, Wenhan Zhang, and Xiao Zhang. Defect detection of photovoltaic modules based on convolutional neural network. In *International Conference on Machine Learning and Intelligent Communications*, pages 122–132. Springer, 2017.