

**ENHANCED IMAGE CLASSIFICATION
FOR DEFECT DETECTION ON SOLAR
PHOTOVOLTAIC MODULES**

NINUK WILIANI

ASIA e UNIVERSITY

2023

ENHANCED IMAGE CLASSIFICATION FOR DEFECT DETECTION ON
SOLAR DETECTION ON SOLAR PHOTOVOLTAIC MODULES

NINUK WILIANI

A Thesis Submitted to Asia e University in
Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

October 2023

ABSTRACT

Solar photovoltaic modules are a technology that utilizes solar energy. Solar photovoltaic modules have many advantages, such as clean electric energy without pollution, very simple to channelling energy, and the most important is that it does not produce greenhouse gas emissions and can be built in remote areas because it doesn't require energy transmission. In actuality, solar photovoltaic module systems are minimal maintenance and do not require any moving parts, but they still have more chances to get various defects by the environment or human beings. Once PV modules are electrically linked, the performance of the entire system might be impacted by any problem between them. Error-prone areas may be difficult to locate or recognise in a big solar photovoltaic module. A solar photovoltaic modules system can hide it until the whole system collapse or breakdown. On the surface of the photovoltaic modules, solar cell defects are identified based on cell shapes and textures. However, high similarity of characteristics among the shapes and textures has been a major challenge in defect classification process. The objective of this research was to develop and analyse feature extraction used for classification techniques for defect detection of solar photovoltaic modules surfaces. Methodologically, the entire study used a quantitative experiment technique. This research uses the Gaussian Naïve Bayes Algorithm using a ratio of training data and testing data of 70:30 resulting in an accuracy value of 46%. The second algorithm uses K-Nearest Neighbour using a ratio of training data and testing data of 95:05 resulting in an accuracy value of 62%. Both methods combine Statistical Feature Extraction and GLCM. Statistical tools provide quantitative information about the intensity distribution of pixels in an image, capturing important statistical properties such as mean, standard deviation, skewness and kurtosis. GLCM, on the other hand, analyses the spatial relationship between pixel pairs and extracts texture features such as contrast, correlation, energy and homogeneity. The accuracy value shows that the KNN algorithm is better when compared to the Naïve Bayes algorithm. Using the same data, these results are compared again using Convolutional Neural Network. The architecture used uses Le Net which is then modified into 3 2D layers and 1 Maxpooling screen. Experiments also compare the size of the image as input, using relu activation and adam optimization. The experiment results in the highest accuracy value at a ratio of 70:30 for training data and test data, which is 68%.

Keywords: Enhancing, Naïve Bayes, K-Nearest neighbour, Convolutional Neural Network, feature extraction

APPROVAL

This is to certify that this thesis conforms to acceptable of scholarly presentation and is fully adequate, in quality and scope, for the fulfilment of the requirements for the degree of Doctor of Philosophy

The student has been supervised by: **Professor Dr. Titik Khawa Abdul Rahman**

The thesis has been examined and endorsed by:

Associate Professor Dr. Shahril Irwan Sulaiman

UiTM

Examiner 1

Professor Dr. Azman Yassin

UUM

Examiner 2

This thesis was submitted to Asia e University and is accepted as fulfilment of the requirements for the degree of Doctor of Philosophy.



Professor Dr Siow Heng Loke

Asia e University

Chairman, Examination Committee

(4 October 2023)

DECLARATION

I hereby declare that the thesis submitted in fulfilment of the PhD degree is my own work and that all contributions from any other persons or sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university. In making this declaration, I understand and acknowledge any breaches in this declaration constitute academic misconduct, which may result in my expulsion from the program and/or exclusion from the award of the degree.

Name: Ninuk Wiliani

A handwritten signature in black ink, appearing to be 'Ninuk Wiliani', written in a cursive style.

Signature of Candidate:

Date: 4 October 2023

Copyright by Asia e University

ACKNOWLEDGEMENTS

I would Thank to God for his mercy to this day in launching all the affairs in completing my education this year. My deepest gratitude to Dr Titik Khawa Abdul Rahman for her great patience and guidance in this research. Without your help and guidance, I would not have been able to complete this research properly. I am very grateful for the time you took to guide me.

I would also like to thank Associate Professor Dr Suzaimah Ramli for her willingness to guide and continue to encourage me until this research was completed. Your support and encouragement provided additional motivation that means a lot to me. I would also like to thank my parents and those who have always provided the greatest support for me, my beloved children's, Rana Kamilah Arif, Abiyyu Muhammad Arif, and Adzraa Luthfiyah Arif. Their presence has given me infinite strength and inspiration on this journey.

Thank you also to several other people who have provided support, although I cannot mention them one by one. Every contribution and assistance provided is greatly appreciated and has had a positive impact on this research, also like to thank my institution for supporting my educational activities.

Finally, thank you to all those who have been involved and contributed to this research. All the help and support provided has played an important role in the success of my research, and may all the kindness that has been given to me also be returned to all of you.

TABLE OF CONTENTS

ABSTRACT	ii
DECLARATION	iii
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xv
LIST OF ABBREVIATION	xx
CHAPTER 1 INTRODUCTION	1
1.0 Background of the Study	2
1.1 Problem Statement	4
1.2 Objectives	7
1.3 Research Question	8
1.4 Justification and Significance of the Study	10
1.5 Theoretical Contribution	10
1.6 Practical Contribution	11
1.7 Contribution to Methodology	12
1.8 Scope of the Study	13
1.9 Chapter Summary	14
CHAPTER 2 REVIEW OF LITERATURE	16
2.0 Introduction	16
2.1 Solar Photovoltaic Modules	16
2.1.1 System of Solar Photovoltaic Modules	17
2.1.2 Ingredients of Solar Photovoltaic Modules	18
2.1.3 Types of Defect Solar Photovoltaic Modules	19
2.2 Overlapping Fields with Image Processing	24
2.3 Digital Image Processing	24
2.4 Computer Vision	27
2.5 Image Pre-processing	28
2.5.1 RGB to Grayscale	28
2.5.2 Histogram	29
2.6 Feature Extraction	32
2.6.1 Feature Extraction by Color	34
2.6.2 Feature Extraction by Geometry	38
2.6.3 Feature Extraction Based on Statistical	41
2.6.4 Feature Extraction Based on Texture	51
2.6.5 Feature Extraction Based on Shape	56
2.7 Image Classification	59
2.7.1 Naïve Bayes Classifier	60
2.7.2 K-Nearest Neighbour Algorithm	61
2.7.3 Convolutional Neural Network	63
2.8 Google Colab	67
2.9 Confusion Matrix	68
2.10 Area Under the Curve (AUC)	69
2.11 Mean Absolute Percentage Error (MAPE)	71

2.12	Mean Average Precision (mAP)	72
2.13	Previous Research of Defect on Solar Photovoltaic Modules	73
2.14	Chapter Summary	105
CHAPTER 3 METHODOLOGY		106
3.0	Introduction	106
3.1	Research Design	107
3.2	Dataset of Solar Photovoltaic Modules	108
3.3	Instrumentation	109
3.4	Development of Feature Extraction Technique for Solar Photovoltaic Modules Image	110
3.4.1	Image to Grayscale	111
3.4.2	Histogram	112
3.4.3	Feature Extraction Process	113
3.4.4	Statistical	115
3.4.5	Gray Level Co-Occurrence Matrix (GLCM)	117
3.5	Analysis of Feature Extracted According to Non Defect, Crack, Scratch and Spot	118
3.6	Development of Classification Defect Solar Photovoltaic Modules Image	120
3.6.1	Classification with Naïve Bayes Algorithm	122
3.6.2	Classification with K-Nearest Neighbour Algorithm	124
3.6.3	Classification with Convolutional Neural Network (CNN)	126
3.7	Chapter Summary	127
CHAPTER 4 RESULTS AND DISCUSSION		129
4.0	Introduction	129
4.1	Results for Feature Extraction Technique for Solar Photovoltaic Modules	130
4.1.1	RGB to Grayscale	130
4.1.2	Histogram of Solar Photovoltaic Modules Image	131
4.1.3	Feature Extraction by Statistical Texture Analysis	135
4.1.4	Feature Extraction with GLCM	137
4.2	Analysis of Features Characteristics from Solar Photovoltaic Modules Images for Determining the Type of Defects	141
4.2.1	Feature Extraction by Statistical Method	141
4.2.2	Feature Extraction by GLCM Method	171
4.3	Result of Classification Technique	206
4.3.1	Integration of Statistical Feature Extraction and GLCM Features.	207
4.3.2	Classification Technique by Naïve Bayes Algorithm	208
4.3.3	Classification Technique by K-Nearest Neighbour Algorithm (KNN)	224
4.3.4	Classification Technique by Convolution Neural Network (CNN)	238
4.4	Comparison for Classification	256
4.5	Chapter Summary	257

CHAPTER 5	CONCLUSION	106
5.0	Conclusion	106
5.1	Recommendation for Feature Research	110
REFERENCES		265

LIST OF TABLES

Table		Page
1.1	Mapping for research objective	9
2.1	Various color based feature extraction	35
2.2	Various texture based feature extraction	51
2.3	The summary of on previous research of defect detection in solar photovoltaic modules	75
3.1	Image of data	108
4.1	RGB to grayscale	131
4.2	Real image and the respective grayscale and histogram	133
4.3	Feature extraction by statistical of grayscale image	135
4.4	Feature extraction using statistical	136
4.5	Matrix of image	137
4.6	Four principal degree is used	138
4.7	Co-occurrence matrix	138
4.8	Co-occurrence into normalize matrix	139
4.9	Feature extraction with GLCM	139
4.10	Feature extraction using GLCM	140
4.11	Statistical data for non defect images	142
4.12	Maximum and minimum values of statistical data for non-defect image by statistical	146
4.13	Statistical data for crack defect images	147
4.14	Maximum and minimum values of statistical data for crack image by statistical	153
4.15	Statistical data for scratch defect image	154

4.16	Maximum and minimum values of statistical data for scratch image by statistical	158
4.17	Statistical data for spot defect image	159
4.18	Maximum and minimum values of statistical data for Spot image by Statistical	163
4.19	Combination maximum and minimum feature extraction by statistical	164
4.20	Performance comparison of various type with statistical method	170
4.21	Value of GLCM for non-defect solar photovoltaic modules image	172
4.22	Maximum and minimum values non-defect	177
4.23	Crack using GLCM	179
4.24	Crack defect of maximum values	185
4.25	Scratch using GLCM	186
4.26	Scratch defect maximum and minimum values	192
4.27	Table of spot using GLCM	193
4.28	Spot of defect maximum and minimum values	199
4.29	Maximum and minimum value of characteristics features of solar photovoltaic modules calculated by GLCM	200
4.30	Performance comparison of various type with GLCM Method	205
4.31	Combination for two feature extraction defect with statistic and GLCM	207
4.32	Result of classification statistical feature extraction using Bernaulli Naive Bayes	209
4.33	Result of classification with statistical feature extraction using Gaussian Naive Bayes	210

4.34	Result of classification with GLCM feature extraction with Bernaulli Naive Bayes	210
4.35	Result of classification with GLCM with Gaussian Naive Bayes	211
4.36	Result of classification with GLCM feature extraction using Multinomial Naive Bayes	212
4.37	Result of classification with statistical and GLCM using Binomial	212
4.38	Result of classification with statistical and GLCM with Gaussian	213
4.39	Actual versus detected type of defect by Naïve Bayes algorithm	214
4.40	Confusion matrix of the chosen models	215
4.41	Calculation of Recall	217
4.42	Measurement of specificity	218
4.43	Interpretation of typical MAPE values	220
4.44	Actual versus Detected	221
4.45	Result of MAPE calculation	222
4.46	Precision of testing models	222
4.47	Result of mAP calculation	224
4.48	Value of accuracy for $K = 1$	228
4.49	Actual versus prediction type of defect	229
4.50	Value of evaluation with KNN Models	230
4.51	Measurement of Recall with KNN algorithm	231
4.52	Measurement of specificity with KNN Algorithm	233
4.53	Actual versus detected	234
4.54	Result of MAPE calculation	235
4.55	Precision of testing models from KKN algorithm	236
4.56	Result of mAP calculation	237

4.57	LeNet-5 architecture	238
4.58	Result of classification with Le-Net architecture – Experiment A	240
4.59	Result of classification with modification layer – Experiment B	240
4.60	Result of classification with activation and optimization - Experiment C -	241
4.61	Result of classification with average pooling - Experiment D	243
4.62	Result of classification with ratio data set - Experiment E	244
4.63	Actual versus prediction type of defect of CNN method	246
4.64	Result of confusion matrix of CNN models	247
4.65	Calculation of Recall in CNN	249
4.66	Measurement of specificity of KNN algorithm	250
4.67	Actual versus prediction	252
4.68	Result of MAPE calculation	253
4.69	Precision of testing models	254
4.70	Calculation	256
4.71	Comparison for classification	256

LIST OF FIGURES

Figure		Page
1.1	Fire hazard in 383 kW PV arrays in Bakersfield, California in 2009	3
2.1	Image of crack defect	20
2.2	Defect glass (crack) makes solar photovoltaic modules more prone to future weather damage	20
2.3	Figure of scratch defect	21
2.4	Snail trails can be signs of microcracks in the underlying solar photovoltaic modules and external particles inside the solar module	22
2.5	Figure of spot defect	23
2.6	Dust tends to build up at of the module frame after rain	23
2.7	Block of image processing	24
2.8	Computer vision pipeline	28
2.9	Global histogram	29
2.10	Dark image and histogram	30
2.11	Bright image and histogram	30
2.12	Low contrast image and histogram	31
2.13	High contrast image and histogram	31
2.14	Feature extraction process	33
2.15	Feature extraction method	34
2.16	Comparison of dark vs bright image (a) dark image (b) dark image histogram (c) bright image (d) bright image histogram	35
2.17	Grey level histogram that can be partitioned by (a) single threshold and (b) multiple thresholds	36
2.18	CNN architecture	65

3.1	Research design	107
3.2	Example datasets	109
3.3	Hardware overview	110
3.4	Flowchart of feature extraction	111
3.5	Flowchart of image to grayscale	111
3.6	Process of histogram	112
3.7	Feature extraction process	114
3.8	Feature extraction by statistical	116
3.9	Gray level co-occurrence matrix	117
3.10	Flowchart of analyse of feature extraction data	119
3.11	Diagram of proposal feature extraction combination	121
3.12	Feature extraction combination with classification technique	122
3.13	Flowchart of classify with Naive Bayes	123
3.14	Flowchart of solar photovoltaic modules defect detection using K-NN	125
3.15	Flowchart of Convolutional Neural Network	127
4.1	Non-defect (a - d), crack defect (e – h), spot defect (i – l), and scratch image (m – p)	129
4.2	(a) Convert from image to (b) histogram of color image	132
4.3	Histogram for grayscale images	133
4.4	Histogram of mean values for non defect images	142
4.5	Histogram of variance value for non-defect images	143
4.6	Histogram of standard deviation for non defect images	144
4.7	Histogram of skewness for non-defect images	144
4.8	Histogram of Kurtosis for non-defect images	145
4.9	Histogram of entropy for non-defect images	146

4.10	Histogram of mean for crack defect images	148
4.11	Histogram of variance for crack defect images	148
4.12	Histogram of standard deviation for crack defect images	149
4.13	Histogram of skewness for crack defect images	150
4.14	Histogram of kurtosis for crack defect images	151
4.15	Histogram of entropy for crack defect images	152
4.16	Histogram of mean for scratch defect image	154
4.17	Histogram of variance for scratch defect images	155
4.18	Histogram of standard deviation for scratch defect images	156
4.19	Histogram of skewness for scratch defect images	156
4.20	Histogram of kurtosis for scratch defect images	157
4.21	Histogram of entropy of scratch defect images	158
4.22	Histogram of spot for defect images	159
4.23	Histogram of variance for spot defect image	160
4.24	Histogram of standard deviation for spot defect images	161
4.25	Histogram of skewness for spot defect images	161
4.26	Histogram of kurtosis for spot defect images	162
4.27	Entropy histogram of spot defect images	163
4.28	Diagram chart of maximum and minimum of means	165
4.29	Diagram chart of max and min of variance	166
4.30	Diagram chart of maximum and minimum of standard deviation	167
4.31	Diagram chart of maximum and minimum of skewness	167
4.32	Diagram chart of maximum and minimum of kurtosis	168
4.33	Diagram chart of maximum and minimum of entropy	169
4.34	Graph showing feature extraction with statistical vs defect of type	170

4.35	Chart diagram with energy 0, 45, 90 and 135 degree	173
4.36	Chart diagram with homogeneity 0, 45, 90 and 135 degree of non defect images	174
4.37	Chart diagram with entropy 0, 45, 90 and 135 degree	175
4.38	Chart diagram with contrast 0, 45, 90 and 135 degree of non defect images	176
4.39	Chart diagram with energy 0, 45, 90 and 135 degree of crack images	180
4.40	Chart diagram with homogeneity 0, 45, 90 and 135 degree of crack images	181
4.41	Chart diagram with entropy 0, 45, 90 and 135 degree of crack images	182
4.42	Chart diagram with contrast 0, 45, 90 and degree of crack images	184
4.43	Chart diagram with energy 0, 45, 90 and 135 degree of scratch images	187
4.44	Chart diagram with homogeneity 0, 45, 90 and 135 degree of scratch images	188
4.45	Chart diagram with entropy 0, 45, 90 and 135 degree of scratch images	189
4.46	Chart diagram with contrast 0, 45, 90 and 135 degree of scratch images	190
4.47	Chart diagram with energy 0, 45, 90 and 135 degree of spot images	194
4.48	Chart diagram with homogeneity 0, 45, 90 and 135 degree of spot images	195
4.49	Chart diagram with entropy 0, 45, 90 and 135 degree of spot images	196
4.50	Chart diagram with contrast 0, 45, 90 and 135 degree of spot images	197

4.51	Chart of maximum and minimum energy values of defect by GLCM	201
4.52	Chart of maximum and minimum homogeneity values	202
4.53	Chart of maximum and minimum entropy values of defect by GLCM	203
4.54	Chart of contrast of defect	204
4.55	Graph showing feature extraction with GLCM vs defect of type	206
4.56	Visualises dataset using K Nearest Neighbour (KNN)	225
4.57	Visualises dataset using K-Nearest Neighbour (KNN)	226
4.58	Visualises dataset using K-Nearest Neighbour (KNN)	226
4.59	Visualises dataset using K-Nearest Neighbour (KNN)	226
4.60	Value of K with range 1 - 500	227
4.61	Value of K with range 1 - 100	227
4.62	mAP based on K-Value	228

LIST OF ABBREVIATION

PV	Photo Voltaic
KNN	K- Nearest Neighbour
CNN	Convolutional Neural Network
GLCM	Grey Level Cooccurrence Matrix
DC	Direct Current
AC	Alternating Current
GA	Genetic Algorithms
DIP	Digital Image Processing
DE	Differential Evolution
PSO	Particle Swarm Optimization
HSA	Harmony Search Algorithms
ANN	Artificial Neural Networks
BA	Bat Algorithms
FA	Firefly Algorithm
ACO	Ant Colony Optimization
NSO	Negative Selection Algorithm
CSO	Cuckoo search algorithm
BDE	Beta Differential Evolution
SVM	Support Vector Machine
ROC	Receiving Operating Characteristics
PCA	Principal Component Analysis
ICA	Independent Component Analysis
AI	Artificial Intelligent
RGB	Red Green Blue

STD	Standard Deviation (STD)
ASM	Angular Second Moment
KNN	K-Nearest Neighbour
MAPE	Mean Absolute Percentage Error
mAP	Mean Average Precision
AP	Average Precision

CHAPTER 1

INTRODUCTION

Presenting the study findings on the design and development of solar photovoltaic module fault detection is the primary goal of this thesis. Defect detection systems for solar PV modules have been widely applied to identify installation errors (Naveen & Sugumaran, 2021). Some of these studies convey that solar photovoltaic modules defects depend on many parameters consisting of installation methods such as voltage, current, power and series resistance; in addition, there are many environmental factors such as temperature and environmental conditions and maintenance methods.

Defects on the surface of solar photovoltaic modules have been well studied and classified according to their origin, nature and effect on solar photovoltaic modules performance. The most common types are scratches (Chen et al., 2020), edge-dipped panels leaving small invisible holes (Buerhop et al., 2017), micro-cracks that are not immediately visible (Mahdavi-pour, 2017), chemical residues left behind during the maintenance process to clean the panel surface (Li et al., 2019).

The main focus of this research is to investigate the use of intelligent image processing techniques for defect classification in solar photovoltaic modules. To achieve this goal, several different methods were tested and compared to determine the most effective approach. In this chapter, the results of these experiments are presented and discussed in detail.

Overall, the results of this study demonstrate the feasibility of using feature extraction for defect classification in solar photovoltaic modules. Choosing the correct method and carefully optimising various parameters makes it possible to achieve a high level of accuracy in classifying different defect types. These results have important implications for developing algorithms used and data implemented for more

efficient and effective maintenance procedures for solar photovoltaic modules and demonstrate that image feature extraction techniques have a potentially important role in this field.

1.0 Background of the Study

The world still has to rely on petroleum as the primary buffer for energy needs. Meanwhile, this energy source is becoming increasingly scarce and expensive. The need for electrical energy continues to increase beyond expectations. This is due to the increase in all life activities that use electricity. Therefore, there is a need for energy derived from nature to be an alternative energy known as renewable energy that is environmentally friendly and contributes to addressing global warming and reducing carbon dioxide emissions.

The sun is the primary energy source that emits the most significant energy to the earth. Under normal conditions, the earth's surface can receive solar energy of around 1000 watt/m^2 (Mokhtari, & Kimour, 2019). The solar energy that can reach the earth's surface is about 207.898 MW (4,80 kWh/m²/day). Less than 30% of this energy is reflected in the atmosphere (Purwoto et al., 2018). Utilising energy from the sun would be very beneficial (Siregar & Wardana, 2017).

Solar photovoltaic modules *are* a technology that utilizes solar energy. Solar photovoltaic modules have many advantages, such as clean electric energy without pollution, being accessible to move, being very simple to channelling energy, and, most importantly, not producing greenhouse gas emissions. It can be built in remote areas because it does not require energy transmission.

The fact is that solar photovoltaic modules systems have no moving parts and generally require low maintenance. However, they still have more chances to experience various failures in the solar photovoltaic modules array, battery cables, and

power conditioning unit. It is difficult to completely shut down solar photovoltaic modules in the event of a malfunction because the modules are energised by sunlight during the day. Solar photovoltaic modules are modular technology power plants that can be made by connecting many solar photovoltaic modules in series and parallel. Once the solar photovoltaic modules are electrically connected, any fault between them can affect the entire performance of the system. It can be complicated to correctly identify or detect faults in a large solar photovoltaic modules array. A faulty solar photovoltaic modules system can go unidentified because it is hidden until the entire system collapses or breaks down (Khan et al., 2017).

Figure 1.1: Fire hazard in 383 kW PV arrays in Bakersfield, California in 2009



Source: (Khan et al., 2017)

Several hazards have been reported in Sudanese solar photovoltaic modules stacking due to undetected faults. Figure 1.1 shows a fire hazard on a 383 kW PV module in Bakersfield, California, 2009. Another fire hazard occurred at a 1 MW PV power plant in Mount Holly, California, in 2011, shown in Figure 1.2. In both cases, unknown faults in the system caused the fire and shut down the system completely. This fire hazard points out the weaknesses in conventional fault detection and protection schemes in solar photovoltaic modules arrays and reveals the urgent need for better ways to prevent such problems. Failure-free, excellent efficiency, fast,