Order number:

TUNISIAN REPUBLIC
****
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
****
UNIVERSITY OF KAIROUAN
****
HIGHER INSTITUTE OF APPLIED SCIENCE AND TECHNOLOGY IN KASSERINE

MEMORY

Presented for graduation from:

Research Master

Specialty:

Nanomaterial and Embedded Electronics

Option:

Embedded electronics

By:

MEHER BOULAABI

Subject:

Dashboard for Systems Fault Prediction and Classification

Presented to the jury on 07/03/2024, consisting of:

Dr.	**Saidi BOUZIDI (Associate Professor at ISSAT-KASSERINE)**	**President**
Dr.	**Wajdi SAADAOUI (Assistant Professor at ISSAT-GAFSA)**	**Reporter**
Dr.	**Mansour HAJJI (Assistant Professor at ISSAT-Kasserine)**	**Supervisor**
Dr.	**Vicenç PUIG (Full Professor at UPC–Barcelona)**	**Co-Supervisor**
Dr.	**Majdi MANSOURI (Associate Research Scientist at Texas A&M University at TAMU-QATAR)**	**Co-Supervisor**

DEDICATIONS

To my dear parents : Noureddine BOULAABI and Anes NASRI

For all your sacrifices, unwavering love, boundless tenderness, steadfast support, and heartfelt prayers that have guided me through my educational journey..

To my dear brothers : Mourad BOULAABI and Anis BOULAABI

For your constant encouragement and unwavering moral support, helping me overcome challenges and reach for the stars.

To my dear friends : Houssem MNASSRI , Ghada RHIMI and Hasna MASSAOUDI

For your enduring friendship, support, and motivation that have been my pillars of strength, inspiring me to push my boundaries and achieve my goals.

To all my family,

For standing by my side, offering endless support, and believing in my dreams throughout my university career. Your collective presence has been my greatest source of strength and inspiration.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank **Dr. Sayed MISSAOUI** (Associate Professor at ISSAT-KASSERINE), serving as the President of the project, and **Dr. Wajdi SAADAOUI** (Assistant Professor at ISSAT-GAFSA), acting as the reporter of the project, for their invaluable contributions and support throughout this endeavor.

I am also indebted to **Dr. Vicenç PUIG** (Full Professor at UPC–Barcelona), my co-supervisor based in Barcelona, for his valuable input and collaboration throughout the research process. His international perspective and technical guidance enriched the scope of my work, and I am grateful for the opportunity to benefit from his mentorship.

Special thanks are extended to **Dr. Majdi MANSOURI** (Associate Research Scientist at Texas A&M University at TAMU-QATAR), for his unwavering support during the stage in Spain. His practical advice, moral encouragement, and assistance during challenging times were invaluable. His mentorship played a crucial role in shaping my understanding and approach to the project.

I would like to express my sincere gratitude to my primary supervisor, **Dr. Mansour HAJJI** (Assistant Professor at ISSAT-Kasserine), whose guidance, support, and expertise played a pivotal role in the successful completion of this report. His dedication to fostering a conducive research environment and his insightful feedback significantly contributed to the quality of this work.

I would like to express my gratitude to the **University of Kairouan** for providing me with the opportunity to undertake this internship. Additionally, I extend my heartfelt thanks to the members there for their assistance during the preparation of the documents.

Your support has been invaluable throughout my academic journey, and I am deeply thankful for the enriching experiences and lessons learned.

ABSTRACT

This study investigates grid-connected photovoltaic systems and fault diagnosis using the Random Forest Classifier. The project initially classified signals in the dataset to detect faults, demonstrating the classifier's effectiveness. To enhance efficiency, a novel approach focused solely on the transitory regime for each fault class, resulting in a remarkable computation time of ***0.008 seconds*** and ***100% accuracy***. This targeted analysis improves real-time fault detection, emphasizing the importance of optimized tools like the Random Forest Classifier. The findings contribute to advancing fault diagnosis techniques, enhancing the safety and efficiency of the photovoltaic industry by preventing potential damage and interruptions.

Keywords: Photovoltaic systems, Fault diagnosis, Random Forest Classifier, Transitory regime, GCPV

RÉSUMÉ

Cette étude explore les systèmes photovoltaïques connectés au réseau et le diagnostic de panne en utilisant le classificateur Random Forest. Le projet a initialement classifié les signaux dans l'ensemble de données pour détecter les défauts. Une approche novatrice s'est concentrée uniquement sur le régime transitoire de chaque classe de défaut, aboutissant à un temps de calcul remarquable de ***0,008 secondes*** et ***une précision de 100%***. Cette analyse ciblée améliore la détection en temps réel des défauts. La résultat contribuent à faire progresser les techniques de diagnostic de panne en améliorant la sécurité et l'efficacité de l'industrie photovoltaïque en évitant les dommages potentiels et les interruptions.

Mots clés :Systèmes photovoltaïques, Diagnostic de panne, Classificateur Random Forest, Régime transitoire, GCPV

CONTENTS

Dedications i

Acknowledgments ii

Abstract iii

List of Tables viii

List of Figures x

List of Abbreviations xi

General Introduction 1

1 APPLICATION FOR GRID-CONNECTED PHOTOVOLTAIC 3

1.1 INTRODUCTION 3

1.2 SYSTEM CONFIGURATION 3

1.3 GCPV SYSTEM COMMON FAULTS 5

1.4 FAULT DIAGNOSIS TECHNIQUES ADOPTED IN LITERATURE 6

1.5 FAULT DIAGNOSIS TECHNIQUES USED FOR GCPV SYSTEM 6

1.5.1 MODEL-BASED FAULT DIAGNOSIS APPROACHES 7

1.5.2 DATA-DRIVEN FAULT DIAGNOSIS APPROACHES 11

1.6 SYSTEM DESCRIPTION 13

1.7 PHASOR MEASUREMENT UNIT (PMU) 14

1.8 INPUT DATA DESCRIPTION 14

1.9 HISTORTY-BASED FAULT DIAGNOSIS 16

1.9.1 THE DIFFERENT STEPS OF DIAGNOSIS 17

1.10 TRANSITORY REGIME 17

1.11 CONCLUSION 19

2 PYTHON FOR AI & RANDOM FOREST CLASSIFICATION **20**

2.1 INTRODUCTION 20

2.2 INTRODUCTION TO PYTHON 21

2.2.1 PYTHON'S POPULARITY IN THE AI COMMUNITY 21

2.3 LIBRARIES USED IN THE PROJECT 22

2.3.1 OVERVIEW OF RELEVANT PYTHON LIBRARIES FOR AI 22

2.3.2 EXPLANATION AND FUNCTIONALITY OF EACH LIBRARY IN OUR APPLICATION 23

2.4 DATA PREPARATION 25

2.4.1 DESCRIPTION OF THE DATASET USED IN THE PROJECT 25

2.4.2 PREPROCESSING STEPS FOR DATA CLEANING AND FORMATTING . 25

2.4.3 BENEFITS OF USING RANDOM FOREST WITHOUT FEATURE EXTRACTION TECHNIQUES 25

2.5 RANDOM FOREST ALGORITHM 27

2.5.1 ENSEMBLE LEARNING TECHNIQUE 27

2.5.2 RANDOM FOREST ALGORITHM 28

2.5.3 RANDOM FOREST AS A CLASSIFICATION ALGORITHM 29

2.5.4 DIFFERENCE BETWEEN DECISION TREE AND RANDOM FOREST . 29

2.5.5 EXPLANATION OF THE RANDOM FOREST TECHNIQUE AND PROCESS 31

2.5.6 IMPORTANT HYPERPARAMETERS 32

2.6 MODEL INTERPRETATION AND FEATURE IMPORTANCE 33

2.6.1 EXPLAINING THE IMPORTANCE OF FEATURE IMPORTANCE IN RANDOM FOREST 33

2.6.2 TECHNIQUES FOR INTERPRETING RANDOM FOREST MODEL RESULTS 33

2.6.3 ANALYSIS OF FEATURE IMPORTANCE IN THE SIGNAL CLASSIFICATION TASK 34

2.6.4 ADVANTAGES AND DISADVANTAGES OF THE RANDOM FOREST . . 35

2.7 CONCLUSION 36

3 EVALUATING TWO RANDOM FOREST SCENARIOS **37**

3.1 INTRODUCTION 37

3.2 SCENARIO 1: COMPREHENSIVE FAULT DETECTION WITH LARGE DATASETS . . . 38
3.2.1 PREPROCESSING THE DATA . . . 38
3.2.2 MODEL TRAINING AND EVALUATION . . . 40
3.2.3 RESULTS AND DISCUSSION . . . 42
3.3 SCENARIO 2: FOCUS ON TRANSITORY REGIME AND COMPUTATIONAL EFFICIENCY . . . 44
3.3.1 RESULTS AND DISCUSSION . . . 47
3.4 DISCUSSION OF THE DIFFERENCES METRICS . . . 53
3.5 CONCLUSION . . . 55

4 BUILDING THE dashboard 57
4.1 INTRODUCTION . . . 57
4.2 Dashboard designs . . . 58
4.3 DESCRIPTION OF PAGES . . . 58
4.3.1 HOME PAGE . . . 59
4.3.2 TEST ONE SIGNAL . . . 59
4.3.3 TEST MORE SIGNAL . . . 61
4.3.4 COMPUTATION TIME . . . 64
4.4 CREATING THE DEVELOPMENT ENVIRONMENT . . . 64
4.4.1 SETTING UP WITH ANACONDA . . . 64
4.4.2 SETTING UP WITH CMDER . . . 65
4.4.3 ATOM AS AN IDE . . . 65
4.5 LOADING THE SAVED MODEL . . . 66
4.5.1 OVERVIEW OF THE MODEL SAVED . . . 66
4.5.2 ENSURING COMPATIBILITY AND DEPENDENCIES WITH THE SAVED MODEL . . . 67
4.5.3 A COMPREHENSIVE STEP-BY-STEP INSTRUCTIONAL REPORT . . . 67
4.6 CONCLUSION . . . 77

Conclusion and perspectives 78

Bibliography 83

LIST OF TABLES

1.1 The labeling and description of the measured and monitored system data variables . . 15
1.2 Description and characteristic of the different labeled faults injected 16

2.1 Python in the AI community 21
2.2 Comparison of Decision Trees and Random Forest 30
2.3 Random Forest Parameters 32

3.1 Distribution of Instances Across Classes 39
3.2 Properties and characteristics of the dataset. 40
3.3 Statistics for each numerical column. 40
3.4 Performance Metrics of the RF model in Scenario 01 42
3.5 Classification Report 43
3.6 Data Reduction 47
3.7 Balanced Data 47
3.8 Classification Report of scenario 02 51
3.9 Differences metrics in our application 54

4.1 Features of Atom Code Editor 66

LIST OF FIGURES

1.1 Synoptic of the GCPV system under study 4
1.2 Grid-Connect Solar Power System 4
1.3 Different fault types for GCPV system Diagram 5
1.4 Classification of Fault Diagnosis Methods 8
1.5 General Block diagram circuit of the Grid-Connected Photovoltaic system 13
1.6 Block diagram circuit of the PMU data generator implemented inside the DSpace 1104. 15
1.7 The different steps of a diagnosis system 18

2.1 Preprocessing steps 26
2.2 Bagging vs Boosting 28
2.3 Random Forest overview 28
2.4 Decision tree vs Random Forest 30
2.5 Random Forest process steps 32
2.6 Feature Importance of RF 34

3.1 Confusion Matrix 42
3.2 Extraction of the transitory regime of X1. 45
3.3 Extraction of the transitory regime of X2. 46
3.4 Extraction of the transitory regime of X3. 46
3.5 First decision tree from the forest 49
3.6 Confusion Matrix 50
3.7 Check the Outliers 52

4.1 Home Page 59
4.2 Upload Button 60
4.3 Table Visualization 60
4.4 Majority Class Prediction 60
4.5 Specific Part Prediction 61

4.6 Adjusting Random Factor 61
4.7 Visualization and Plotting of one variable 62
4.8 Visualization and Plotting of multiple variables 62
4.9 Multiple File Uploads 63
4.10 Summary Table: Class Results and Computation Time for Each Signal 63
4.11 Filter by Class Name 64
4.12 Home Page 68
4.13 Test one signal page 69
4.14 Read the data of the new signal 70
4.15 Final Predicted Class 71
4.16 Probability of each class 72
4.17 Plot one single variable 73
4.18 Plot multiple variable 74
4.19 Adjust random factor 75
4.20 The class of each signal 76

LIST OF ABBREVIATIONS

FD	**F**ault **D**etection
FDD	**F**ault **D**etection and **D**iagnosis
FC	**F**ault **C**lassification
ML	**M**achine **L**earning
DL	**D**eep **L**earning
PV	**P**hoto**V**oltaic
GCPV	**G**rid- **C**onnected **PV**
AI	**A**rtificial **I**intelligence
CNN	**C**onvolutional **N**eural **N**etwork
AC	**A**lternating **C**urrent
DC	**D**irect **C**urrent
RF	**R**andom **F**orest
SGD	**S**igned **D**irected **G**raph
ODEs	**O**rdinary **D**ifferential **E**quations
MPPT	**M**aximum **P**ower **P**oint **T**racking
SVPWM	**S**pace **V**ector **P**ulse **W**idth **M**odulation
LSTM	**L**ong- **S**hort **T**erm **M**emory
PMU	**P**hasor **M**easurement **U**nit