

**ADVANCED DEEP LEARNING FRAMEWORK
FOR PREDICTIVE MAINTENANCE OF RENEWABLE
ENERGY PLANTS**

BY

ABDULBASIT HASSAN

A Thesis Presented to the
**COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

SYSTEMS AND CONTROL ENGINEERING

DECEMBER 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **ABDULBASIT HASSAN** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN SYSTEMS AND CONTROL ENGINEERING**.

26/12/24

Dr. Mujahed Mohammad Al Dhaifallah
Department Chairman

Prof. Suliman Saleh Al-Homidan
Dean of Graduate Studies

31/12/2024

Date



Shafiullah

Dr. Md Shafiullah
(Advisor)

Dr. Asif Islam
(Co-Advisor)

Dr. Ali Nasir
(Member)

Dr. Muhammad Majid Gulzar
(Member)

Dr. Syed Muhammad Amrr
(Member)

© Abdulbasit Hassan

2024

This work is dedicated to Almighty Allah.

ACKNOWLEDGEMENTS

In the name of Allah, the Beneficent, the Most Merciful. This work would not have been possible without Almighty Allah's boundless help and guidance (SWT). Alhamdulillah! I want to express my heartfelt gratitude to my family for their support and encouragement throughout my academic journey. My sincere appreciation also extends to the management of King Fahd University of Petroleum & Minerals for providing the funding and fostering an environment conducive to educational excellence.

I am deeply indebted to my thesis advisor, Dr. Md Shafiullah, for his invaluable mentorship and continuous support, as well as to my thesis committee members, Dr. Asif Islam, Dr. Ali Nasir, Dr. Muhammad Majid Gulzar, and Dr. Syed Muhammad Amrr, whose constructive feedback and expertise significantly contributed to the success of this research. My gratitude further extends to the Chairman and esteemed faculty members of the Department of Control and Instrumentation Engineering, whose knowledge and dedication have profoundly shaped the direction of this work.

I sincerely acknowledge the Renewable Energy Technical Incubator (RETI), KFUPM, for their generous sponsorship and invaluable support of this thesis.

I greatly appreciate my fellow graduate students and colleagues, Mr. Ghali Ahmad, Mr. Ibrahim Kabir, Mr. Abdullahi Jibril, Mr. Ridwanllahi Isa, and others for your support. Your shared experiences have enriched my academic journey and personal growth.

Abdulbasit Hassan

December 2024

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	X
LIST OF FIGURES	XI
LIST OF ABBREVIATIONS	XIV
ABSTRACT.....	XVI
المخلص الرسالة	XVIII
CHAPTER 1 INTRODUCTION.....	1
1.1 Background of Wind Turbine Energy System (WTES).....	1
1.2 Background on PV Solar Farm	4
1.3 Motivation.....	6
1.3.1 Wind Farm	6
1.3.2 Solar Farm.....	6
1.4 Objectives.....	7
1.5 Scope of the study.....	8
1.6 Thesis Organization	8
CHAPTER 2 LITERATURE REVIEW	10
2.1 Introduction.....	10

2.2 Gearbox Temperature Prediction-NBM.....	11
2.3 Ensemble Gearbox Temperature Prediction NBM	12
2.4 Ensemble NBM for PM with Corresponding Statistical Process Control.....	14
2.5 Summary of Related Work on Wind Turbine Fault Prediction.....	17
2.6 Solar Farm Overview	19
2.7 Summary.....	23
CHAPTER 3 FAULT PREDICTION FRAMEWORK	24
3.1 Introduction.....	24
3.2 Data Acquisition and Pre-processing.....	25
3.2.1 Data Cleaning.....	25
3.2.2 The interquartile Range (IQR) method	26
3.2.3 Z-score.....	27
3.2.4 Normalization.....	27
3.2.5 Wind Power Curve (WPC) Filtering.....	27
3.2.6 Feature Extraction/Selection.....	28
3.3 Ensemble NBM Processing	31
3.3.1 Multiple Linear Regression.....	34
3.3.2 Random Forest	34
3.3.3 Artificial Neural Network	35
3.4 Model Performance Metrics	36

3.5 Fault Prediction/ Alarm Generation System Approach.....	37
3.6 Solar Farm Defect Prediction Framework	38
3.6.1 EL Image Acquisition and Preprocessing	39
3.6.2 Deep Learning Model Construction.....	40
3.6.3 Deep Learning Model Evaluation.....	41
3.7 Summary.....	42
CHAPTER 4 MODEL CONSTRUCTION AND EVALUATION	44
4.1 Introduction.....	44
4.2 NBM Data Preparation	44
4.2.1 Data Acquisition.....	44
4.2.2 Pre-processing	45
4.2.3 Feature Extraction/Selection.....	47
4.3 Ensemble Model Construction.....	49
4.4 Solar Defect Prediction Model	59
4.4.1 EL Images Preparation	60
4.4.2 CNN-Based Defect Classification Using EL Images	61
4.5 Solar Defect Classification Result Discussion	71
4.6 Summary.....	73
CHAPTER 5 NBM FAULT PREDICTION/ALARM GENERATION SYSTEM ..	74
5.1 Introduction.....	74

5.2 Fault Alarm System	74
5.3 Experimental Setup	75
5.4 Online Condition Monitoring Simulation Components	81
5.4.1 Data Acquisition System	82
5.4.2 Buzzer - PKM22EPP – 40	82
5.4.3 Raspberry Pi Pico	82
5.5 Result Discussion.....	83
5.6 Summary.....	85
CHAPTER 6 CONCLUSION AND RECOMMENDATION.....	86
6.1 Summary of the Work	86
6.2 Limitations and Recommendations for Future Work.....	90
6.2.1 Wind Turbine Gearbox Fault Prediction	90
6.2.2 Solar Farm Photovoltaic Modules Defect Prediction	91
REFERENCES.....	93
VITAE.....	115

LIST OF TABLES

Table 2. 1: Summary of Data-driven NBM for bearing fault prediction.	17
Table 2. 2 Recent studies on solar farm PV defect classification.....	21
Table 4. 1 EDP T06 Wind turbine data description	46
Table 4. 2 Most relevant features description	48
Table 4. 3 Optimal hyperparameters for RF and DNN architecture.....	51
Table 4. 4 NBM performance metric comparison	52
Table 4. 5 CNN-based model architecture for defect classification.	64
Table 4. 6 Training parameters of the defect detection CNN model.....	64
Table 5. 1 Standalone and ensemble SCC threshold and fault signal alarm.....	75

LIST OF FIGURES

Figure 1. 1: Global renewable energy investment and capacity for wind and solar energy [6].	2
Figure 1. 2: A typical Wind Turbine Drivetrain [11]	3
Figure 1. 3: Downtime caused by turbine subsystems [17].	4
Figure 2. 1: Maintenance strategies [37], [38]	11
Figure 2. 2: A 60-cell PV module with a crack defect [25].	21
Figure 3. 1: Wind turbine fault prediction framework.	25
Figure 3. 2: Illustrates the methodology approach.	33
Figure 3. 3: Structure of the parallel ensemble framework.	33
Figure 3. 4: Process of developing solar PV defect prediction framework.	39
Figure 4. 1: Time series variation of features in T06 - wind turbine.	46
Figure 4. 2: (a) Original wind power curve (b) Result of the first preprocessing step of the wind power curve.	47
Figure 4. 3: Filtered wind power curve.	48
Figure 4.4: Feature ranking of the three FSA.	48
Figure 4. 5: (a) Shallow network configuration, (b) Two hidden layers DNN, and (c) Deep neural network configuration (multiple hidden layers with multiple neurons in each layer).	50
Figure 4. 6: Training error and validation loss curve for (a) Shallow neural network, (b) 2 - hidden layer network, (c) 5 - hidden layer network, (d) 10 - hidden layer network.	52

Figure 4. 7: DNN evaluation result for MAE, WIA, R^2, and MSE	53
Figure 4. 8: Standalone and ensemble NBM evaluation result.....	53
Figure 4. 9: Prediction of Gearbox bearing actual and predicted temperature for (a) MLR first 100th data point, (b) MLR first 1000th data point, (c) MLR Test set.....	55
Figure 4. 10: Prediction of Gearbox bearing actual and predicted temperature for (a) RF first 100th data point, (b) RF first 1000th data point, (c) RF Test set..	56
Figure 4. 11: Prediction of Gearbox bearing actual and predicted temperature for (a) DNN first 100th data point, (b) DNN first 1000th data point, (c) DNN Test set.....	57
Figure 4. 12: Ensemble NBM prediction of Gearbox bearing actual and predicted temperature for (a) first 100th data point, (b) first 1000th data point, and (c) Ensemble NBM Test set.....	59
Figure 4. 13: Samples of the EL cell image classes [1].....	60
Figure 4. 14: (a) Original, (b) 180 degrees flip, (c) Horizontal flip, and (d) Vertical flip.....	61
Figure 4. 15: Samples of EL cell images based on resolution according to their pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200, and (2) 250 x 250.....	63
Figure 4. 16: Presents the training and validation loss based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (2) 250 x 250.....	67
Figure 4. 17: Presents the training and validation accuracy based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (2) 250 x 250.	69

Figure 4. 18: The CNN model classification accuracy based on the pixel size.	70
Figure 4. 19: Presents the confusion matrix based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (2) 250 x 250.	71
Figure 4. 20: Samples of misclassified classes.	72
Figure 5. 1: NBM prediction with the measured gearbox bearing temperature from 13th to 18th October and the corresponding prediction error for (a) MLR and (b) RF	76
Figure 5. 2: (a) Gearbox bearing temperature prediction versus the measured SCADA data with SCC fault alarm generation for (a) Shallow network, (b) 2 hidden layer network, (c) 5 hidden layer network, (d) 10 hidden layer network and (e) ensemble model.	81
Figure 5. 3: Experimental setup framework	81
Figure 5. 4: Piezoelectric buzzer [131]	82
Figure 5. 5: Raspberry Pi Pico ports [130]	83

LIST OF ABBREVIATIONS

O&M	:	Operation and Maintenance
WT	:	Wind Turbine
NBM	:	Normal Behaviour Model
MLR	:	Multiple Linear Regression
RF	:	Random Forest
SCADA	:	Supervisory Control and Data Acquisition
PM	:	Predictive Maintenance
ANN	:	Artificial Neural Network
ML	:	Machine Learning
MI	:	Mutual Information
PCA	:	Principal Component Analysis
LASSO	:	Least Absolute Shrinkage and Selection Operator
DNN	:	Deep Neural Networks
EWMA	:	Exponential Weighted Moving Average
SCC	:	Shewhart Control Chart
WPC	:	Wind Power Curve
UCL	:	Upper Control Limit
LCL	:	Lower Control Limit
CM	:	Condition Monitoring
WIA	:	Willmott's index of agreement
SPC	:	Statistical Process Control
LSTM	:	Long Short-Term Memory

IEC	:	International Electrotechnical Commission
PV	:	Photovoltaic
CNN	:	Convolutional Neural Network
EL	:	Electroluminescence

ABSTRACT

Full Name : Abdulbasit Hassan

Thesis Title : Advanced Deep Learning Framework for Predictive Maintenance of
Renewable Plants

Major Field : Systems and Control Engineering

Date of Degree : December 2024

The transition from traditional fossil fuel-based resources to renewable and clean energy requires significant investments and addressing the challenges of ensuring reliability and stability in complex electricity grids. Over the past decade, renewable energy resources (RER) have gained substantial popularity due to their potential to enhance energy sustainability and reduce energy generation costs. Solar and wind are the two most dominant RERs; wind turbines (WT) have a greater advantage over other renewable energy sources in producing energy around the clock. A fundamental problem these renewable sources face is the energy cost, thus challenging their efficient operation and maintenance (O&M). Among the prominent maintenance strategies, Corrective maintenance (failure-based) for wind turbines (WTs) poses significant challenges due to limited accessibility and adverse weather conditions, which prolong outage durations and increase costs. Additionally, preventive maintenance (time-based) can often be unnecessary, as maintenance activities are scheduled before they are required. Predictive maintenance (PM) has been developed to be an economical solution for actively overcoming problems, such as wind turbine gearbox bearing failures that seriously affect the performance and stability of renewable energy systems (RES). This study proposes an intelligent PM

approach using three standalone normal behaviour models and their Multi-model ensemble technique for gearbox fault prediction. It uses the data acquired through the WT supervisory control and data acquisition (SCADA) system. The efficiency of the developed model was evaluated using statistical indices (SI), with satisfactory SI values demonstrating the effectiveness and robustness of the proposed models. A post-processing analysis is also carried out using the Shewhart control chart (SCC) to determine the deviation of predictions from the actual SCADA data measurement to identify the PM schedule. The developed models accurately predict the failure event at least 72 hours before it occurs, as reported in the wind farm maintenance logbook. Similarly, a solar photovoltaic (PV) defect classification model was developed using the convolutional neural network (CNN) and electroluminescence (EL) cell images under step-wise damp heat exposures from 500h to 3000h. The model's accuracy was evaluated based on EL cell image pixel resolution, ranging from lower to higher resolutions. The model classification results indicated that a medium pixel size of 100×100 and a high pixel size of 250×250 achieved the highest classification accuracy, with no misclassified EL cell images. Therefore, a medium pixel size of 100×100 is more suitable due to its lower pixel values, which has a lower risk of overfitting and low computational requirements for training and inference. This predictive maintenance approach has the potential to generate substantial cost savings by enhancing the economic viability of wind turbine systems and solar PV operations. Early fault detection gives wind and solar farm operators adequate time to implement corrective actions, allowing for timely repair or replacement of faulty gearboxes. This proactive maintenance strategy has the potential to minimize unnecessary costs and reduce downtime effectively.

الملخص الرسالة

الاسم الكامل: عبد الباسط حسن

عنوان الرسالة: إطار متقدم للتعلم العميق للصيانة التنبؤية لمحطات الطاقة المتجددة

مجال التخصص: هندسة الأنظمة والتحكم

تاريخ الحصول على الدرجة: نوفمبر 2024

يتطلب الانتقال من الموارد التقليدية المعتمدة على الوقود الأحفوري إلى الطاقة المتجددة والنظيفة استثمارات كبيرة، بالإضافة إلى مواجهة التحديات المتعلقة بضمان موثوقية واستقرار الشبكات الكهربائية المعقدة. وخلال العقد الماضي، شعبية كبيرة نظراً لإمكاناتها في تعزيز استدامة الطاقة وتقليل تكاليف توليدها. (RER) اكتسبت مصادر الطاقة المتجددة مميزة (WT) تُعتبر الطاقة الشمسية وطاقة الرياح من أكثر مصادر الطاقة المتجددة هيمنة، حيث تتميز توربينات الرياح كبيرة مقارنة بمصادر الطاقة المتجددة الأخرى بقدرتها على إنتاج الطاقة على مدار الساعة. ومع ذلك، تواجه هذه من بين استراتيجيات. المصادر المتجددة تحديات كبيرة تتعلق بتكاليف الطاقة، مما يُصعب تشغيلها وصيانتها بكفاءة الصيانة البارزة، تشكل الصيانة التصحيحية (المعتمدة على الفشل) لتوربينات الرياح تحديات كبيرة بسبب صعوبة الوصول إليها والظروف الجوية السيئة، مما يؤدي إلى إطالة مدة التوقف وزيادة التكاليف. بالإضافة إلى ذلك، غالباً ما تكون الصيانة الوقائية (المجدولة زمنياً) غير ضرورية، حيث يتم تنفيذ أنشطة الصيانة قبل الحاجة الفعلية لها. لذلك، تم كحل اقتصادي للتعامل مع المشكلات الناشئة، مثل أعطال محامل صندوق التروس (PM) تطوير الصيانة التنبؤية تقترح هذه الدراسة نهجاً. (RES) لتوربينات الرياح، التي تؤثر بشكل كبير على أداء واستقرار أنظمة الطاقة المتجددة ذكياً للصيانة التنبؤية باستخدام ثلاثة نماذج سلوك طبيعي مستقلة وتقنية تجميع متعددة النماذج للتنبؤ بأعطال صندوق التروس. ويعتمد النموذج على البيانات التي يتم جمعها من خلال نظام الإشراف والتحكم في توربينات الرياح ، حيث أظهرت القيم المرضية (SI) تم تقييم كفاءة النموذج المطور باستخدام المؤشرات الإحصائية. (SCADA) Shewhart للمؤشرات فعالية وقوة النماذج المقترحة. كما تم إجراء تحليل ما بعد المعالجة باستخدام مخطط التحكم أظهرت. لتحديد جدول الصيانة التنبؤية SCADA لتحديد الانحرافات بين التوقعات والقياسات الفعلية لبيانات (SCC) النماذج المطورة بدقة عالية في التنبؤ بحدوث الأعطال قبل 72 ساعة على الأقل من وقوعها، كما هو موضح في سجل

صيانة مزرعة الرياح. وبالمثل، تم تطوير نموذج تصنيف عيوب الألواح الشمسية الكهروضوئية باستخدام الشبكة تحت تعريض متدرج (EL) وصور الخلايا الكهروضوئية المأخوذة بتقنية التألق الكهربائي (CNN) العصبية الالتفافية للرطوبة والحرارة لفترات زمنية تتراوح بين 500 ساعة و3000 ساعة. تم تقييم دقة النموذج بناءً على دقة وضوح من الدقة المنخفضة إلى العالية. وأظهرت نتائج التصنيف أن حجم البكسل (EL) بكسل صور الخلايا الكهروضوئية EL المتوسط (100×100) وحجم البكسل العالي (250×250) حققا أعلى دقة تصنيف، مع عدم وجود صور خلايا مصنفة بشكل خاطئ. وبالتالي، يُعتبر حجم البكسل المتوسط (100×100) أكثر ملاءمة نظرًا لقيمته الأقل، مما يقلل يمتلك هذا النهج في الصيانة. من خطر الإفراط في التكيف مع النموذج ويقلل من متطلبات الحوسبة للتدريب والاستنتاج التنبؤية القدرة على تحقيق وفورات كبيرة في التكاليف من خلال تعزيز الجدوى الاقتصادية لتشغيل أنظمة توربينات الرياح والألواح الشمسية. كما أن الكشف المبكر عن الأعطال يوفر لمشغلي مزارع الرياح والطاقة الشمسية وقتًا كافيًا لاتخاذ الإجراءات التصحيحية اللازمة، مما يتيح إصلاح أو استبدال المعدات المعيبة في الوقت المناسب. ويمتلك هذا النهج الاستباقي للصيانة القدرة على تقليل التكاليف غير الضرورية وتقليل فترات التوقف بشكل فعال.

CHAPTER 1

INTRODUCTION

1.1 Background of Wind Turbine Energy System (WTES)

The availability of sustainable energy for quality of life and economic development is among the most formidable issues confronting nations in our century [2], [3], [4]. A faster transition to renewable and clean energy from traditional fossil fuel-based resources requires tremendous investments and overcoming complex electricity grids' reliability and stability issues. Over the last couple of decades, renewable energy resources have gained popularity for their overall reduction of energy generation costs. The solar photovoltaic and wind energy are the predominant modern renewable energy resources [5], [6]. In 2023, the global installation capacity for wind energy systems surpassed 1000 GW, which was only around 250 GW in 2012, Figure 1.1 illustrates the global renewable investment and capacity for wind and solar energy [7]. Achieving low operation and maintenance (O&M) costs in the wind energy system is seen as one of the significant operational challenges. The wind farm O&M costs approximately account for one-third of the total energy generation costs due to their remote location, accessibility to the generation station, transportation of the components, and downtime after being subject to any failure [8], [9]. The share could account for up to 15% of the newly installed systems, which can rise to 35% by the end of the plant lifespan [10].

Figure 1.2 illustrates the skeletal view of the wind turbine drivetrain consisting of the rotor hub, main bearing, gearbox, and generator. The gearbox operates in harsh environmental conditions. Their most important task is to amplify the speed delivered by the lower-speed shaft (LSS) so that the velocity demand of the High-speed shaft (HSS) driving the generator is met. Consequently, they experience all vibrations originating from the turbine-side components, wind, and other oscillations imposed by the load via the generator. The gearbox of the wind turbine is an essential element susceptible to deterioration. Gearbox malfunctions may manifest in several components, which include ring gears, tooth fractures, housings, internal shafts, high-speed and low-speed shaft bearings, helical gears, and inadequate lubrication [11].

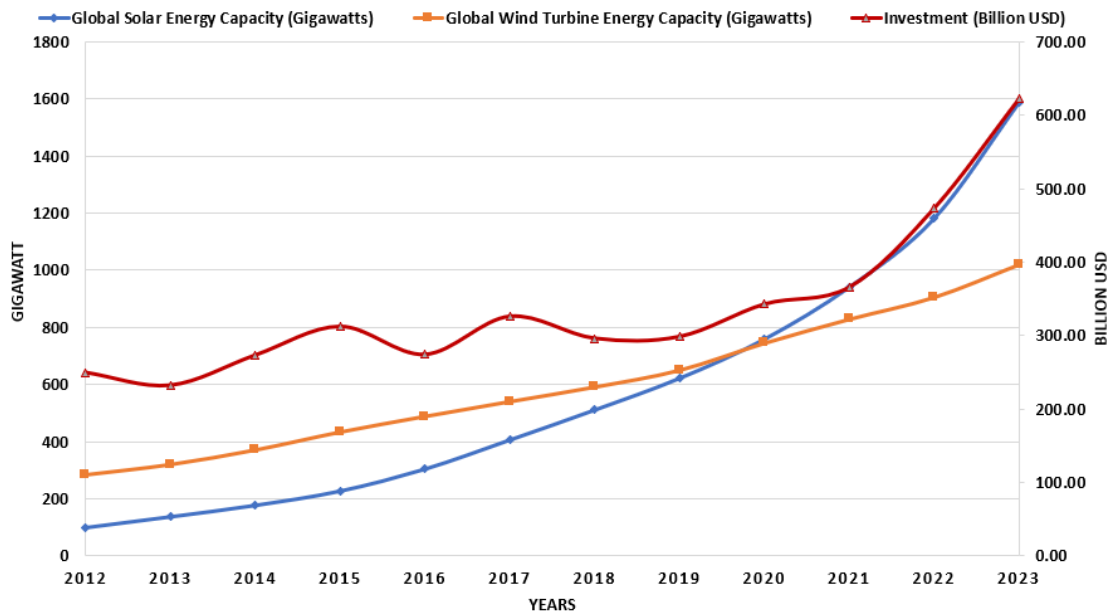


Figure 1. 1: Global renewable energy investment and capacity for wind and solar energy [6].

It is worth noting that the industry classified wind turbine (WT) gearbox and main bearings failure as the utmost operational reliability challenge due to the complex repair procedures, higher replacement costs, and longer downtime duration [12], [13]. Recent data has shown

that the failure rate of main bearings (over a 20-year lifetime) can reach as much as 30% [14]. The annual downtime of the wind turbine gearbox may reach a maximum of 60 days, which affects the overall system efficiency [15]. According to [16] maintenance of a gearbox is a laborious process, requiring an average time frame of 256 hours, with 59% of failures of the system ascribed to the malfunctions of the gearbox. Thus, the repair time required is considerably long [17]. Figure 1.3 describes a survey of data from 27,000 wind turbines, with capacities ranging from 500 kW to 5 MW and covering both onshore and offshore applications, revealing that gearbox failures accounted for the highest WT downtime hours [18]. To ensure the Wind energy system (WES) remains competitive, the cost of renewable power generation in 2023 report that the Overall Cost of Energy (OCOE) for WES has decreased significantly between 2010 and 2023, with a reduction of 70% for onshore wind turbines (from USD 0.111/ kilowatt-hour (kWh) to USD 0.033/kWh) and 63% for offshore wind turbines (from USD 0.203/kWh to USD 0.075/kWh) [19]. Additionally, O&M costs account for 26% of the OCOE in onshore wind projects and 26.2% in offshore wind projects [20].

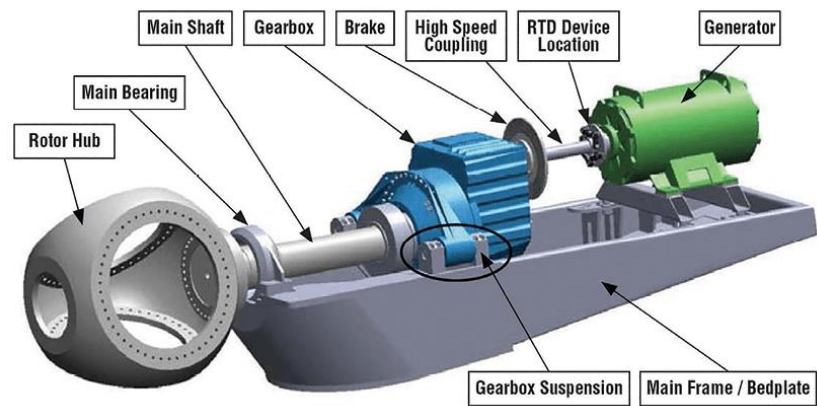


Figure 1. 2: A typical Wind Turbine Drivetrain [11]

1.2 Background on PV Solar Farm

Solar energy is available in vast amounts. It is one of the most potent forms of energy and is also vastly used by many [21]. Figure 1.1 shows the capacity of solar energy worldwide, and it proves that it has been practiced and proven effective in many places [22]. Such systems are widely used today and recognized as an essential alternative for solar energy exploitation [23] due to their wide availability, low cost, and share in the market [24]. The high growth in solar PV is particularly driven by the lower price of generated electricity from newly established solar farms in contrast to newly constructed power plants utilizing alternative technologies [25].

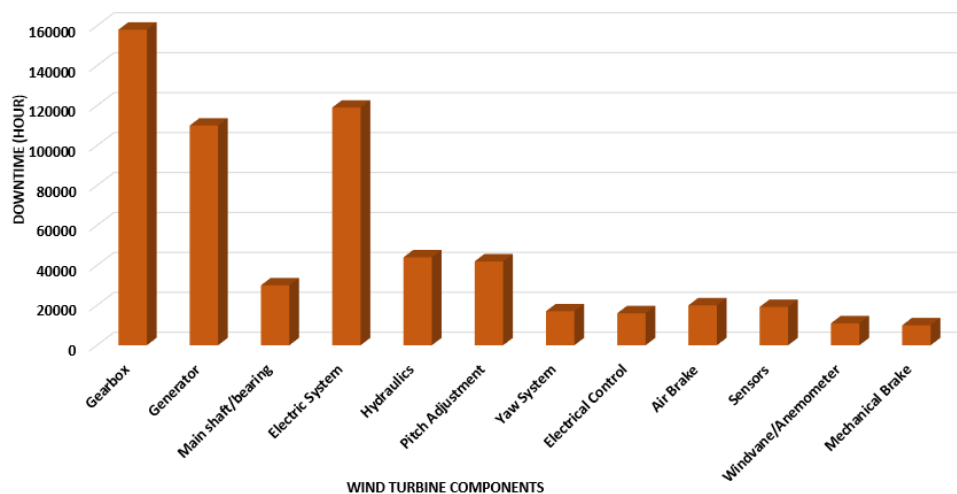


Figure 1. 3: Downtime caused by turbine subsystems [17].

In 2023, renewable energy capacity experienced the largest ever increase, with 473 gigawatts (GW) added; this accounted for 86% of global power additions, mainly due to massive solar and wind power growth. Nearly three-fourths of renewable energy additions were solar power, with 346GW, and 116GW from wind [26].

The maintenance schedule carried out in solar farms includes corrective, preventive, and predictive is found to be economical (Renewable Energy Laboratory et al., 2018) (Wang et al., 2023). Predictive maintenance (PM) models are an unfailing, reliable substitute for on-site maintenance methods employed in solar farms. They offer tools for independently diagnosing and preventing sudden failures in components or systems. These models rely on gathering and scrutinizing data to contrast the projected performance with the actual system energy output, thereby determining its operational condition [29]. PM models are designed to schedule cost-effective maintenance before components become inoperable [30].

Electroluminescence (EL) images EL can be used to estimate PV outpower [31]. Furthermore, it can subsequently be used to analyze patterns of solar cell defects and finally decide on the maintenance schedule. The primary function of the PV system is to generate electricity from solar energy. The array of solar cells is the fundamental part of the photovoltaic cells that help generate electricity. The solar cell's crystalline structure is fragile, which sometimes causes microcracking during the manufacturing or mounting process [32]. The cracks will degrade solar cell performance by disrupting the electricity supply in the interconnected solar cells. PV modules constitute the most significant expensive for PV systems of any size, accounting for approximately 25-35% of the total expenditure for utility-scale solar projects within the 5-100 MW range [33].

The energy generated depends upon several factors, including the panel state and the prevailing climate conditions, which consist of the meteorological data. Solar modules are also susceptible to factors that can diminish or impact performance. These challenges

necessitate implementing a system to monitor and optimize power generation because manual labour alone is inadequate as the number of modules grows.

1.3 Motivation

The reliability of renewable energy plants is essential in power plants, as unexpected failures can lead to power interruptions due to downtime, thereby substantially impacting the economy.

1.3.1 Wind Farm

Wind turbines are characterized by their rapid growth in size, market share, and technological advancements. Unlike other power generation plants, operational and maintenance costs significantly influence their efficiency in energy production. The gearbox accounts for the highest percentage of total downtime among all components, directly influencing operational and maintenance costs. Inspired by the above statement, this thesis investigates the development of three standalone normal behaviour models (NBMs) and their Multi-model ensemble data-driven ensemble machine-learning model for wind turbine gearbox bearings predictive maintenance.

1.3.2 Solar Farm

Solar power plant installations, called solar farms, consist of many PV modules made of PV cells. These modules generate electricity by converting photons into electrons within the module. PV systems comprise thousands of PV modules, each comprising several PV

cells. These numbers depend on the configuration of the module. Due to the increased installation, solar farm maintenance has become a complex task [34].

A variety of maintenance solutions can be used to prevent sudden failure in systems. The PV module failure rate is reported to be between 0.025%/year and 0.1%/year [35]. In contrast, the degradation rate is 0.5% annually, exhibiting greater consistency across various types and manufacturers [36]. Consequently, it has become imperative to reduce energy generation costs by employing O&M strategies incorporating state-of-the-art deep learning technology.

1.4 Objectives

This thesis aims to create an efficacious data-driven ensemble machine-learning model for wind turbine gearbox bearings predictive maintenance of wind farms and a defect prediction deep learning model for solar farms using EL images. To achieve this, the following specific objectives are established:

- To develop a data-driven normal behaviour model for wind turbine gearbox bearing temperature prediction employing various advanced machine learning models.
- To develop a post-processing strategy using statistical process control (SPC) techniques.
- To build a fault prediction alarm prototype system using an embedded system.
- To develop an intelligent framework for estimating panel state using electroluminescence images.

- To establish possible collaboration with industrial counterparts related to the developed frameworks.

1.5 Scope of the study

The scope of this thesis is to develop three standalone NBMs and their Multi-model ensemble model utilizing machine learning techniques to predict gearbox faults in wind turbines. Structured data from the SCADA system of the T06 wind turbine will be used for model development. Additionally, a Raspberry Pi microcontroller will be employed for alarm generation implementation. Similarly, a deep learning framework for the solar farm will be proposed for the automated classification and prediction of defects using electroluminescence images.

1.6 Thesis Organization

The first chapter of this thesis introduces the significance of wind turbine and solar PV energy systems as the two most dominant resources and the challenges they pose in today's renewable energy plants.

The second chapter presents a comprehensive background and a review of existing studies on wind turbine faults and solar PV defect prediction. This chapter is segmented into three sections: gearbox temperature prediction utilizing NBM, Ensemble NBM for temperature prediction, and ensemble fault prediction model with corresponding statistical process control.

The third chapter comprehensively discusses the prediction framework and the proposed methodology's theoretical background for bearing faults and solar PV defects. In contrast,

Chapter 4 discusses the development and performance evaluation of the wind turbine machine learning-based NBM and the solar PV deep learning model.

The fifth chapter focused on developing the alarm generation system, detailing and discussing the experimental results from the wind turbine fault prediction. Similarly, it examines the overall findings of the proposed solar photovoltaic defect prediction framework. Finally, the sixth chapter concludes the thesis by outlining the limitations and future recommendations for the potential commercialization of the developed data-driven ensemble fault prediction model for wind turbines and the solar PV defect prediction framework.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The wind industry is very concerned about reliable monitoring for detecting, diagnosing, and predicting such failures. Three basic maintenance approaches are corrective, preventive, and predictive, as defined in Figure 2.1 [37], [38]. The corrective maintenance strategy acts after the failure occurs, whereas the preventive one acts as scheduled periodic inspections and replacements. A corrective strategy reduces servicing costs a bit, but it comes at the expense of huge revenue loss due to more extended downtime and increased repair costs due to equipment damage. The preventive one reduces the risks of extended downtime by ensuring good equipment conditions. However, it does not protect the system operation from unexpected failures. In addition, the overall costs increase due to unnecessary maintenance schedules. In contrast, predictive maintenance (PM) is a more effective preventive and corrective strategy as it is data-driven and reduces downtime and maintenance costs by predicting equipment failure. Similarly, PM can reduce the overall downtime for wind energy systems by providing diagnostics for the key components, including the bearings (main and gearbox), yaw, pitch, and generator bearings [39]. Also, PM ensures that maintenance actions are only taken when they are needed, thus optimizing resources and extending equipment lifetime [40], [41], [42].

The increasing prominence of intelligent production and maintenance has emerged as a key focal point within Industry 4.0, marking a significant transition from the information age

to the era of advanced smart automation [28]. In response, various data-driven predictive maintenance approaches for the WT systems were reported in the literature, including support vector machines [43], linear regression [44], decision trees [45], random forests (RF) [46], and neural networks [47]. Existing works based on the fault Prediction framework will be examined in this section.

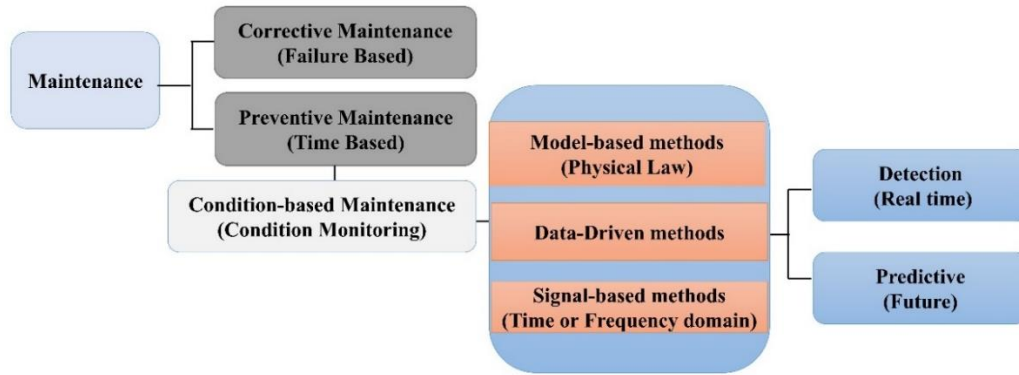


Figure 2. 1: Maintenance strategies [37], [38]

2.2 Gearbox Temperature Prediction-NBM

The CM approach in this article aims to predict the gearbox-bearing temperature of the wind turbine by employing the SCADA system's operational data. Recent studies have shown that it is a proper way to monitor operating status by temperature. The accuracy in temperature prediction can enhance the WT system's reliability. For the previous reasons, several data-driven techniques have been implemented to predict the gearbox temperature, such as artificial neural networks (ANN), grey models, fuzzy systems, and auto-regression models.

A few years ago, the most popular models used for time series prediction fell into two categories: artificial intelligence-based models and statistical predictors. Artificial

intelligence models have gained popularity among researchers due to their superior predictive capabilities [48], [49], [50]. Paul A. Adedeji *et al.* [51] developed a WT gearbox oil temperature prediction model using an adaptive neuro-fuzzy inference system (ANFIS) optimized genetic algorithm hybrid model and ANFIS optimized with PSO hybrid model. The former outperformed the latter with higher accuracy and lower error metrics. Similarly, [52] developed four distinct regression models to represent the typical behaviour of gearbox oil temperature: long short-term memory (LSTM), temporal convolutional network (TCN), XGBoost, and gated recurrent unit (GRU).

Ayush Verma *et al.* (2022) [53] demonstrated the usage of high-frequency SCADA data, instead of the commonly used 10-minute averaged data, by developing a feed-forward ANN-based NBM for predicting gearbox oil temperature. The proposed approach enhances detection capabilities and improves understanding of wind turbine conditions. However, using high-frequency data increases signal noise, which can be mitigated through the averaging process. This smoothing process, while reducing noise, has the potential to cause the loss of critical information and result in false or missed alarms. The temperature trends of the WT gearbox were observed with an NBM based on low-resolution SCADA vibration data and Support Vector Regressor, and the model indicates a minimal overheating trend for the defective wind turbine [54].

In the research by Fotso *et al.* [55], the operating parameters of the gearbox were taken as input variables for the backpropagation neural network, which is NBM, while bearing power loss was the output variable for bearing power loss modeling and prediction.

2.3 Ensemble Gearbox Temperature Prediction NBM

To further improve the bearing temperature prediction, an ensemble NBM is utilized, The Ref. Yan *et al.* [56] proposed model achieved an excellent prediction of gearbox-bearing temperature by integrating the optimization algorithm with an ensemble predictor. Firstly, the data-driven approach utilized the variational mode decomposition (VMD) technique in decomposing the original bearing temperature data into many sub-series, each characterized by a distinct frequency. Following this step, an unsupervised ML using stack autoencoder, derived from the group methodology of data management, acts as the predictor in the subseries. Finally, the weights are optimized for subseries by the competitive algorithm of imperialism, and the results are integrated.

A hybrid model for temperature prediction in the WT gearbox component using LSTM with an adaptive error correction algorithm was developed in [57]. A new signal decomposition technique, the VDM, was applied due to the error series' non-stationary, high frequency, and disordered characteristics for adaptive error correction. The suggested model may be utilized online to alerts about early high-temperature wind turbine gearbox components. This methodology applies only to single-step prediction, while in the real world, there are more requirements for multi-step prediction to provide more extended maintenance time. An ensemble deep-learning regression model was proposed for gearbox temperature prediction using RF to collect features and LSTM for sequential prediction [58]. A hybrid formation for the WT gearbox temperature prediction model was developed by combining a bidirectional LSTM architecture and a one-dimensional CNN. The proposed model combines the strengths of these two models and provides far better results than the other five comparative models [59].

2.4 Ensemble NBM for Predictive Maintenance with Corresponding Statistical Process Control

This part will discuss the existing literature that defines the fault symptoms of the WT gearbox before it occurs while utilizing ensemble NBM. Broadly, wind turbine fault prediction and detection using data mining algorithms can be categorized into two primary approaches: normal behaviour modeling using regression method and classification-based methods [46]. Bearings are a crucial component of the rotation mechanism in wind turbines, including transmission bearings (main and gearbox bearings), as well as yaw, pitch, and generator bearings [39]. The health conditions of the gearbox bearing and generator winding were monitored utilizing the neural network-based models in Refs. [60], [61]. Ref. Sun *et al.* [62] predicted the failure of the main bearing almost seven months ahead, employing the principal component analysis (PCA). However, the field of data-driven intelligent approaches for WT systems condition monitoring is relatively new, and it requires extensive exploration to optimize the generation of clean and sustainable energy.

In a comparative analysis of six data-mining algorithms: k-nearest neighbours (kNN), least absolute shrinkage and selection operator (LASSO), support vector machines (SVM), ridge regression (Ridge), shallow neural networks, and deep neural networks (DNN) for fault prediction in gearbox systems based on lubricant pressure, the proposed methodology utilized an exponential weighted moving average (EWMA) control chart to detect impending failures 2 to 3 days in advance across six commercial wind farms in China. The study demonstrated that monitoring lubricant pressure effectively reduced false alarms compared to monitoring gearbox oil temperature [63].

A fault prediction model was developed utilizing the extreme gradient boosting (XGBoost) algorithm and tested on two wind turbines from separate wind farms. The model successfully identified impending faults 1 and 3 days in advance. The threshold for fault detection was established by integrating multivariate quality control charts (MQCC) with the absolute prediction error (APE) of healthy operational data [64]. Similarly, [65] developed a data-driven predictive model for the WT gearbox and generator using Xgboost and LSTM. The model was tested on two distinct wind farms. The anomaly behaviour was evaluated using the shewhart control chart (SCC). A machine learning-based predictive model was employed to predict gear whine noise, demonstrating superior accuracy compared to the traditional ordinary least squares method [66].

Ensemble-based techniques have been applied in the literature for wind power output forecast [67], and fault diagnosis [68]. In a Classification problem, the actual monitoring data from two megawatts (MW) WT was implemented to predict the WT gearbox operational state using a combined technique based on LSTM, fuzzy synthesis (FS), and transfer learning [69].

The Ref. de Lima Munguba et al. [70] utilized procedures for nineteen ensembles, which further the model combinations to predict wind turbine drivetrain components anomaly using an unsupervised ML approach on vibration data. The fault detection positive and negative cases are evaluated using the receiver operating characteristic curve (AUC) and composite consistency (CC) score. The proposed pipeline attained over 89% AUC for the gearbox high-speed shaft bearing with nearest neighbour stacking. Ref. [45] applied classification and regression approaches for predictive fault analysis on five major wind turbine components, including the gearbox. The model was developed using a decision tree

(DT) within a data-centric methodology and compared with a purely model-centric approach. However, the data-centric approach is noted for its extended data preparation steps, which are considered a drawback within the machine learning pipeline.

Due to persistent false alarms, a health rate utilizing the sliding window approach was implemented to strengthen the condition-monitoring procedure of the gearbox in determining the health status of the gearbox in WT, a bagging ensemble technique, and K-nearest neighbours algorithm enhance the operational efficiency by about 30% through SPC threshold setting up [71]. The effectiveness of the ensemble deep regression framework was observed in monitoring the tooth profile deviation over a single model in a real-time gear machining quality estimation using a multi-layer fusion residual network (MFResNet) and an AlexNet, respectively [72]. In early fault detection, [73] develops a model called SLFormer to solve the prediction problem, which Integrates LSTM into the transformer encoder. The NBM outperformed six other prominent models in accuracy and stability by successfully detecting gearbox anomalies twenty-one days before they occurred using the EWMA control chart and effectively suppressing false fault alarms. A stacking model developed by Yuan et al. [46] utilized XGBoost, gradient boosting decision trees (GBDT), and RF as base models, with XGBoost as the meta-model for NBM in gearbox fault prediction. This model demonstrated the ability to forecast faults 13 to 22 days in advance. Fault symptoms were indicated through the output of a change-point detection algorithm based on the Cumulative Sum (CUSUM) control chart.

2.5 Summary of Related Work on Wind Turbine Fault Prediction

The section presents an overview of relevant existing studies, defining the NBM approach, alarm threshold, and limitations corresponding to each methodology embraced, as tabulated in Table 1.

Table 2. 1: Summary of Data-driven NBM for bearing fault prediction.

Ref.	NBM Method	Alarm Threshold Technique	Limitation	Remark
[74]	PCA (Not Ensemble)	EWMA	Prediction model deployment capability is not guaranteed	Predict bearing failure 7 months beforehand
[9]	LSTM (Not Ensemble)	fault indicator (FI) and Simple moving average (SMA)	The Prediction model is only suitable for small datasets.	Predict bearing failure 4 months before component failure.
[75]	SVM, ANN-MLP, DT, Gaussian Process (GP), and Mixture Discriminant Analysis.	Binary classification	Prediction error among the ML algorithms applied.	Vibration signals
[76]	ANN, SVM, and a logistic regression.	Binary classification	SCADA data frequency	Predict failure up to one month before it happens.
[53]	ANN (Not Ensemble)	one-class SVM classifier.	The prediction model cannot tell the root cause of the gearbox fault	12 weeks before the failure (High-frequency data)
[71]	Bagging Ensemble Strategy and KNN Algorithm	SPC Pearson correlation coefficients	The large training set has slow estimation accuracy	After approximately 2,500 points
[72]	multi-layer fusion residual network, a deep regressor (MFResNet), and an AlexNet	tooth profile deviation $f f \alpha$ (right)	Tested well on a Small dataset	Vibration Signal

	(shallower regressor)			
[70]	Ensemble pipeline of 19 models	Binary classification	Extensive tuning through hyper-parametrization and the combination is required to select an ensemble model.	Fault prediction on Drive train components using vibration data.
[63]	Compares six data-mining algorithms (Not Ensemble).	EWMA control chart	Computational cost	Identify gearbox failure between two and three days ahead. Lubricant pressure prediction (No false alarm), also compared with gearbox oil temperature (found less effective)
[73]	Integrating LSTM into the Transformer encoder – (SLFormer model)	EWMA	Model deployment capability Computational efficiency	Predict gearbox anomalies 21 days in advance
[64]	Xgboost (Not Ensemble)	Multivariate Quality Control Charts (MQCC) + Absolute prediction error	The cyclical and seasonal nature of wind power generation can affect the CM model.	Identified impending faults 1 and 3 days in advance
[46]	Model stacking using RF, GBDT, Xgboost	CUSUM control chart	False alarm.	Predict faults 13 to 22 days in advance.
[45]	DT (Not Ensemble)	Classification and regression task	Prolonged data preparation steps and possible overfitting on the ML pipeline.	Fault prediction on five components.

2.6 Solar Farm Overview

Research work on electroluminescence (EL) images EL image defect predictions of detection cover image pre-processing, object detection, feature extraction, and defect classification. Now, PV modules are laminated against ambient conditions by masking the solar cell surface [25]. Instead of using complete images at the module level, defect detection, and classification pipelines employ EL images focused explicitly on individual cells. This method involves the cropping of module-level EL images into component cells [77], [78], [79], [80]. The images at the cell level offer a more extensive training dataset, given that typical module images contain either 60 or 72 cells per module [25], [79] Smaller images demand fewer computational resources for processing [25]. Additionally, pre-processing techniques such as image scaling [81], image augmentation [77], [82] were implemented.

Figure 2.2 presents a PV module EL image of 60 cells arranged in 6 x 10 grid mono-crystalline silicon wafers. Critical cracks, which would not be visible during routine visual inspection of the module, can be seen on the left side of the EL image [25]. Machine learning has advanced in speed and precision. Computers may now do activities that would often need significant human effort. Previous studies demonstrate that computer vision and machine learning have significant promise for analyzing EL images.

Ref. [25] introduced a pixel-level classification to quantify the solar PV module defect using u-net architecture. Also, Ref. [83] proposed that image reconstruction performs better than feature extraction in solar PV defect detection. In comparing three statistical learning

algorithms RF, SVM, and ANN for Cell degradation type classification into good, crack, and corroded, SVM has the highest model prediction accuracy [1]. Using deep learning techniques by [77], two degradation techniques (cracked and Corroded) with RF, SVM, and CNN, CNN outperforms RF and SVM in supervised PV cell classification. Similarly, Ref. [84] developed an open-source pipeline that sequentially performs preprocessing, classification/detection, and post-processing. The defect identification was developed and compared between RF, ResNet models, and YOLO, and the optimal model was found to be ResNet18 and YOLO. ResNet18 and YOLO models confused images of cracks with images of intact cells. The cracks are tiny, and more training examples are needed to distinguish them from intact cells. In Ref. [32] a model was developed and tested on the standard EL image dataset using the deep Siamese convolutional neural network for defect classification. The simulation results indicate that the suggested model attains superior classification accuracy. Also, [85] proposed a robust model with a classification and segmentation network capable of learning from a small set of labelled and annotated samples. The detection process first performs a real-time micro-crack anomaly discrimination task for the EL image classification. An automatic defect classification with synthetic EL images using CNN and ML models was proposed in [82]. Traditional operations and maintenance of solar PV systems provide significant issues for plant operators, particularly those overseeing LSS, [86] proposed an advanced machine-learning technique for fault detection and classification of a Large-Scale Photovoltaic (LSSPV) plant. Table 2.2 presents the recent studies with each contribution application, result, and drawback.

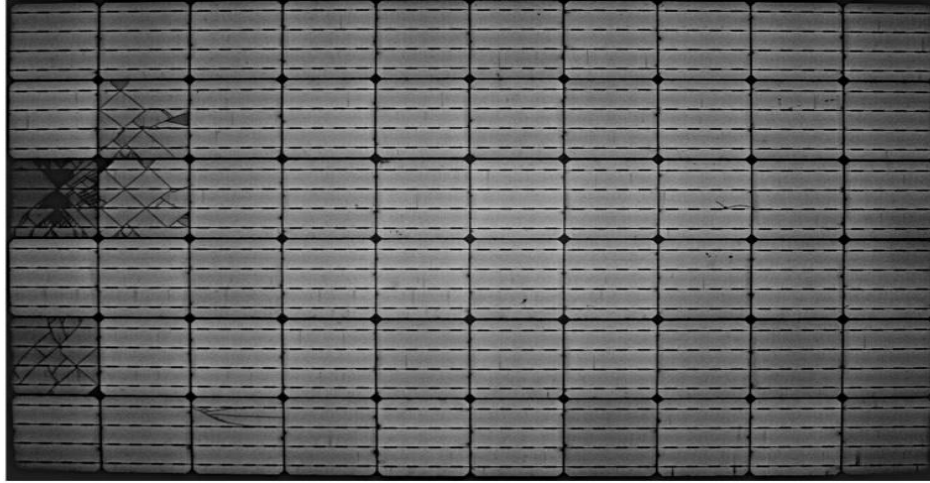


Figure 2. 2: A 60-cell PV module with a crack defect [25].

Table 2. 2 Recent studies on solar farm PV defect classification.

Ref.	Application	Contribution	Result	Drawback
[25]	Defect quantification and detection.	Introduced a pixel-level classification.	Developed a semantic segmentation model using u-net architecture.	The model performs weakly on smaller extrinsic defects like cracks and gridlines.
[83]	Defect Detection in Solar Modules	Employed Independent Component Analysis (ICA) on EL Images.	The approach for image reconstruction performs better than the feature extraction.	Difficulty in classifying the type of defect.
[1]	Busbar cell Degradation Type Classification using EL images (good, cracked, and heavily corroded)	Cell images were augmented, and three statistical learning algorithms, RF, SVM, and artificial neural network (ANN), were employed.	SVM has the highest model prediction accuracy.	Fewer train data affects the classification accuracies.
[77]	Cell degradation Type Classification	Employed three statistical	CNN outperforms RF and SVM in	Data imbalance affects the

	using EL images (cracked and corroded).	learning algorithms: RF, SVM, and CNN.	supervised PV cell classification.	learning algorithms.
[84]	Automated defect detection and classification on EL images	Defects are identified using RF, ResNet models, and YOLO).	The developed model analyzes EL images and detects defects in approximately 5 seconds per PV module.	ResNet18 and YOLO models confused images of cracks with images of intact cells due to data imbalance.
[32]	EL image is preprocessed to predict four types of defects	Deep Siamese CNN is used for the defect classification.	The model achieves better classification accuracy.	Model training, testing, and validation require more time to operate.
[85]	Detection of micro-crack anomalies in photovoltaic module cells	Classification and segmentation network learning from a few labeled and annotated samples.	The segmentation network incorporates an attention mechanism to address the imbalance in distribution between anomaly and background pixels.	Low detection speed.
[82]	Automatic defect classification with synthetic EL images.	CNN was used for deep feature extraction and existing machine learning (ML) models.	Accurate ML classification was achieved using transfer learning as a benchmark for comparison.	The generative adversarial network generates high-resolution EL images that require more computation time.
[86]	Large-scale (LS) PV plant anomaly detection.	The model performed clustering and anomaly detection with K-means and LSTM, respectively,	The model prediction outperformed ANN due to lower relative error.	Possibility of prediction inaccuracies arising from changes made during classification.

		to predict the power of string modules.		
[31]	Optimization of PV cell output power estimation using EL images and I-V curves.	Employed ML methods from Traditional to Deep Learning methods.	Solved the issue of unbalanced data using Synthetic Images using GAN.	Limited amount of data with different defects for improved model performance and generalization.

2.7 Summary

This chapter summarizes relevant studies on the wind turbine (WT) fault prediction framework and the solar photovoltaic defect prediction framework. A summary of recent and relevant literature is presented in tabular form, highlighting the limitations of each methodology, along with their respective applications, results, and contributions.

CHAPTER 3

FAULT PREDICTION FRAMEWORK

3.1 Introduction

This chapter discussed the fault prediction framework comprising the ensemble data-driven NBM and the fault-prediction alarm generation system. The objective is to align the wind turbine's standard operational behaviour with the ensemble model to derive the residuals, representing the discrepancy between the actual measured values from the SCADA system and the output values of the ensemble model. Finally, statistical process control based on the Shewhart control chart is employed for the gearbox fault prediction. Figure 3.1 presents a schematic diagram describing the proposed fault prediction framework.

Similarly, a comprehensive description of the solar photovoltaic module defect prediction framework will be presented, as illustrated in Figure 3.4. The framework outlines the procedure for predicting the health status of PV cells using a deep learning classification model, enabling the identification of cell conditions such as being good, cracked, or corroded.

As shown in Figure 3.1, this section will define the architectural procedure for the ensemble NBM development. The required steps involve data acquisition and preparation, feature selection, NBM construction and evaluation, and finally, the ensemble model technique development.

3.2 Data Acquisition and Pre-processing

Energias De Portugal (EDP) is a global leader in the energy sector that provided the dataset utilized for this study. This dataset is an annual SCADA measurement that offers historical wind turbine monitoring data from January to December 2017 [87], [88].

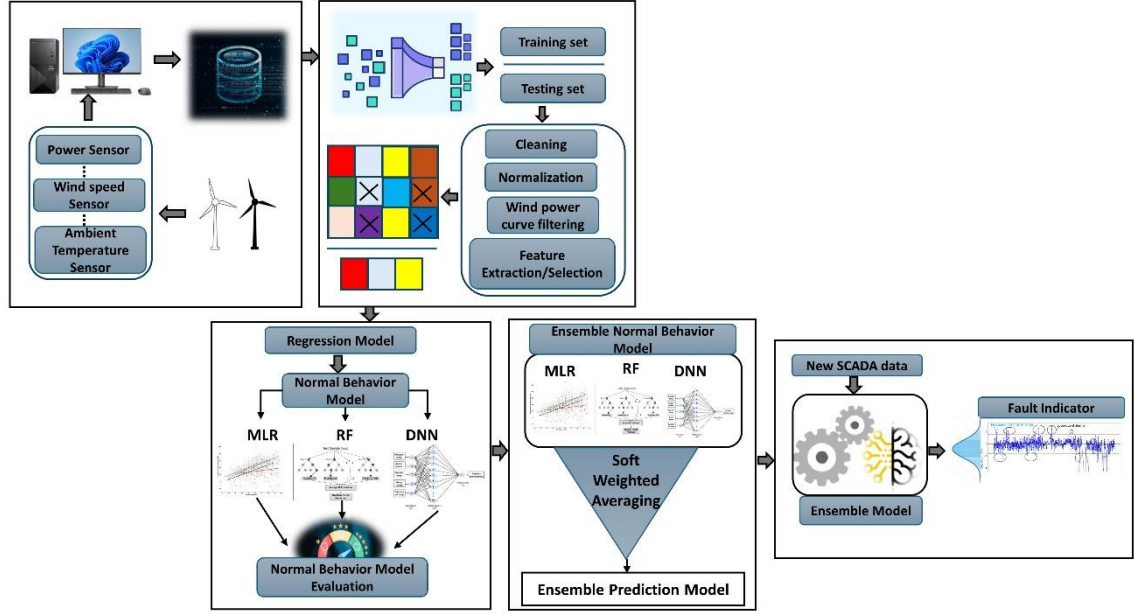


Figure 3. 1: Wind turbine fault prediction framework.

In the pre-processing technique, all outliers were removed to filter out noise within the dataset, after which the SCADA data was utilized to develop the normal behaviour model [89]. Firstly, data pre-processing is conducted to handle outliers, missing data, and different magnitudes of sensors.

3.2.1 Data Cleaning

To ensure that the ensemble NBM accurately depicts the state of the system using the feed data, the quality of the data must be checked. Any outlier data points must be eliminated

to prevent the model from misinterpreting system performance. It is necessary to perform data cleaning procedures to create a model that describes the WT in its healthy state. Consequently, for the data cleaning, the following removal criteria were applied to the identified relevant data samples for the ensemble model development:

- Where sample data has missing input or output (Nan)
- When active power is zero or negative while the wind speed is higher than the cut-in speed
- Outlier removal using the quartile method [46], [90]
- International Electrotechnical Commission (IEC 61400-12-1 2005) Regulation on Wind Turbine Power Curve characteristics [91]
- Data normalization

3.2.2 The interquartile Range (IQR) method

This is a systematic way to detect outliers in a dataset. It computes the first quartile Q_1 and third quartile Q_3 as the values at the 25th and 75th percentiles of the data, respectively, as described in equation (3.1).

$$IQR = Q_3 - Q_1 \quad (3.1)$$

Lower and upper bounds are set as thresholds to detect outliers. The normal data bounds are defined as follows:

$$[K_l, K_u] = [Q_1 - 1.5IQR, Q_3 + 1.5IQR] \quad (3.2)$$

Where K_l and K_u are lower and upper bounds beyond which data points are considered outliers.

3.2.3 Z-score

This computes the mean $X_{(mean)}$ and standard deviation $X_{(std)}$ of the sample data. The formula in (3.2) illustrates how the chosen input data is transformed into a z-score [92]

$$Z = \frac{X_i - X_{(mean)}}{X_{(std)}} \quad (3.2)$$

3.2.4 Normalization

The research utilized the StandardScaler function from the Python scikit-learn library to normalize the input variables. The input variables are distributed across several dimensions and ranges; therefore, transforming them into a standard range is required. The function effectively scales the variables to the range [0,1]. Transforming the values of the input variables into a standard range is required because they are distributed across several dimensions and ranges.

3.2.5 Wind Power Curve (WPC) Filtering

The WT power curve describes the power produced and the wind speed relationship. Therefore, it is the most common tool for performance analysis and anomaly identification (Sun et al., 2016). The power curve predicts how the wind turbine should behave. Therefore, data that describe a healthy operation condition are necessary for the power curve to follow its shape. The objective is to fit a power curve to the data that represents the turbine's normal operational behaviour, enabling the identification and elimination of any data that indicates abnormal performance or underperformance using the IEC standard [91].

To further filter the WPC, Ref. [93] segment the data into intervals corresponding to variations in turbine behaviour. After binning the samples, the quantiles of the data inside each bin are calculated, and outliers are identified by removing data points that deviate significantly, as determined through corresponding boxplots. The flagging criterion for outliers is based on the standard deviation or scalar divergence from the bin center mean. A scalar metric employing a threshold value of 25% from the mean whisker length of the bin center was used to identify outliers.

3.2.6 Feature Extraction/Selection

A wide range of features can form real-world data, many of which may not be informative for the process they aim to represent. Such non-informative features within any dataset increase computation time, adding to overfitting in the resultant models. Therefore, properly selecting the best subset from a dataset can significantly enhance machine learning models' performances. One commonly adopted technique for feature selection is manual selection based on domain knowledge. However, this could lead to losing powerful predictors due to ignorance about their importance [45].

Ref. [52] extracts physical features using energy loss based on the gearbox thermal balance and wind energy conversion techniques. Ref. [94] integrates PCA with radial basis function neural network (RBFNN) to predict output power in early warning. Pearson correlation coefficients were calculated to select the key features of the gearbox-bearing temperature [71]. Another Ref. [70] examines the predictive power of different input features through permutation feature importance (PFI) analysis. An ensemble VIP was developed using the RF Variable Important Plots (VIP) algorithm to provide credit scores on the essential variables for the WT gearbox diagnostics [95]. Three feature selection algorithms (FSA)

were compared, namely the RF, filter, and wrapper models, to obtain the most relevant gearbox oil temperature features [46].

A feature ranking was introduced to select relevant features to strengthen the gearbox fault diagnosis predictive model [96]. These correlated features increase computational time and can lead to overfitting [45]. To mitigate these potential consequences, this study focuses on developing an ensemble NBM for the WT gearbox, with gearbox-bearing temperature selected as the dependent variable. Three feature selection algorithms, including Principal component analysis (PCA) [94] Mutual information (MI) [45], and Lasso regression (LR) [66], were utilized to categorize the most relevant input features for the gearbox-bearing temperature ensemble model.

3.2.6.1 Principal Component Analysis

Principal component analysis is a statistical method that breaks down a data matrix into vectors termed principal components. It is commonly employed for dimensionality reduction, facilitating exploratory data analysis. Principal components serve various purposes, such as compressing information to minimize storage requirements, reducing high-dimensional data to two or three dimensions for visualization and exploration, and obfuscating datasets to enable data sharing without revealing sensitive information or the underlying nature of the data [60]. PCA is a feature selection algorithm with a dimensionality reduction method that converts features into novel collections of orthogonal components. In PCA, features are reduced by identifying components that capture the maximum variance [94].

3.2.6.2 Mutual Information

This method assesses the mutual dependency between two variables and serves as a filtering model in feature selection [97]. The mutual information selector quantifies the reliance between each feature and the target variable, therefore encompassing both linear and non-linear interactions [96]. In contrast to the high-correlation filter, mutual information assesses the reliance between two characteristics. In contrast to the high-correlation filter, mutual information considers linear and non-linear connections. The mutual information for each feature is calculated by rating characteristics pertinent to the goal variable. Subsequently, the most informative features are selected to optimize the target variable prediction [45]. The computational time required for all strategies is quite minimal, except for mutual information [98].

3.2.6.3 Lasso Regressor

Lasso regression is an effective method for feature selection, particularly in high-dimensional datasets, as it identifies the most relevant features by applying a penalty to the absolute values of the regression coefficients. This technique is extensively utilized across various domains, such as spectroscopic applications [99], to improve model accuracy and interpretability. Lasso regressor performs feature selection by penalizing the absolute value of the regression coefficients, effectively driving some of them to zero and thus excluding certain features [66]. Ref. [100] presents a group lasso regularized multiple kernel learning framework, which enforces sparsity at the group or feature level for effective feature extraction. This approach enables the selection of various kernels for each feature type, thereby accommodating the complexity of heterogeneous data while emphasizing relevant features through sparsity across different feature groups.

3.3 Ensemble NBM Processing

Ensemble prediction models (EPM) are a powerful approach in machine learning that combines multiple models to improve predictive performance and robustness. Three widely used ensemble learning methods, bagging, boosting, and stacking, can be employed to enhance the machine learning process [101]. Bagging, also known as bootstrapping and aggregating, is an ensemble learning technique. In the bootstrapping phase, ensemble models are trained on bootstrap replicates of the training dataset. During aggregation, the final prediction is determined through majority voting based on the predictions made by the individual models. Boosting is an ensemble learning technique that involves sequentially adaptively training multiple weak learners. Similar to bagging, boosting can be applied to regression and classification problems. Stacking is an ensemble learning technique that combines information from multiple predictive models to create a new model, known as a meta-model.

While many research studies have been bothered by the local optimal phenomenon, the ensemble learning idea among machine learning methods is to integrate multiple sub-models with comprehensive evaluation, considering every sub-model influence. It has the advantages of avoiding overfitting and getting rid of the local optimal solution, strong expression ability, etc., and can achieve higher accuracy [102]. Ensemble NBM is applied across various fields, including automobile price prediction [103], disease diagnosis [104], [105], environmental weather conditions, flood forecasting [106], handling uncertainty in data [107], and rainfall prediction [108].

In the field of wind energy systems, research on wind power prediction addresses the problem by proposing a multi-network deep ensemble model, utilizing an improved manta ray foraging optimization algorithm alongside PCA for feature selection [109]. Similarly, Ref. [110], [111] works on a similar application, using multi-scale features [112].

Inspired by the above, we proposed a model based on the soft weighted average strategy by integrating three regressors for temperature prediction to leverage the strengths of multiple models. Based on the existing works, this strategy has been widely applied in software defect prediction [113], vehicle detection [114], and prediction of heart disease [115]. Using a softmax function, Soft weighted averaging was applied to the weights for weight normalization before applying the weighted sum of the predictions from the three-base normal behaviour models. This approach ensures that the weights are scaled between 0 and 1, with their sum equal to unity.

Therefore, this article proposes an average weighted strategy for developing a normal behaviour model for the wind turbine's gearbox-bearing temperature. Figure 3.1 illustrates the methodology approach.

This study employs a parallel ensemble technique in which base learners are generated simultaneously without data dependency [101]. In this approach, the errors made by one model are independent of those in other models, allowing the ensemble model to average out these errors effectively [116]. This technique is applied to Multiple linear regression (MLR) [117], RF [58], and DNN [63]. The approach involves training these base models in parallel to generate multiple predictions, which are subsequently combined through a

voting process known as weighted averaging. Details of this process are provided in Figure 3.3.

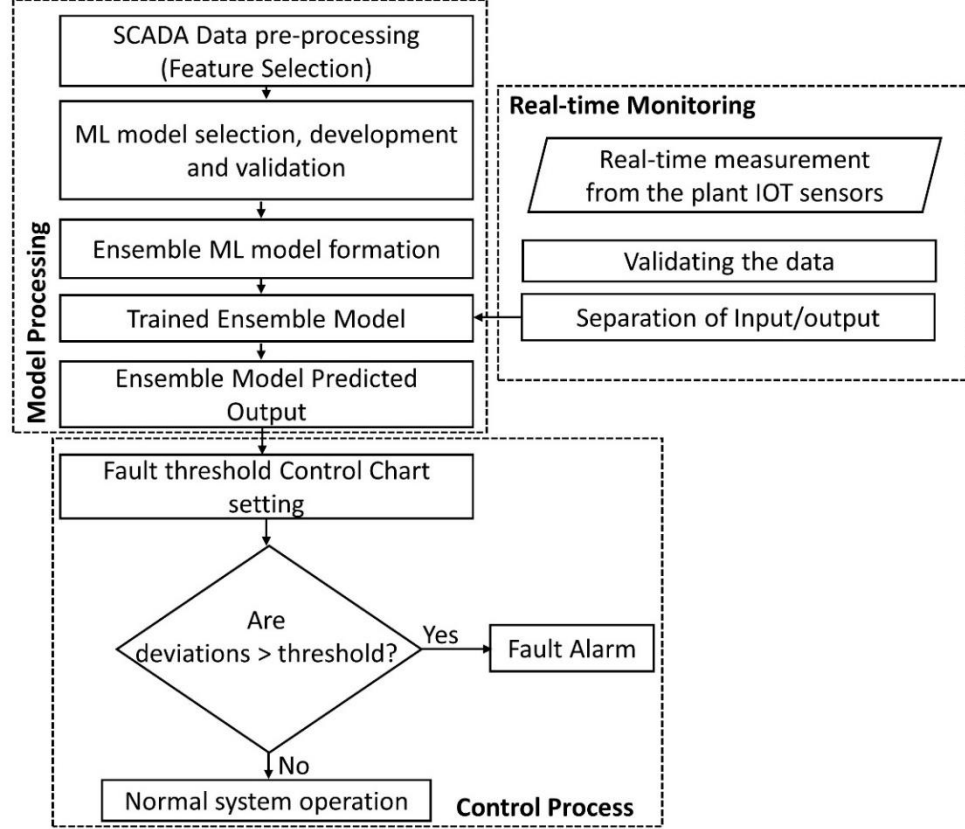


Figure 3. 2: Illustrates the methodology approach.

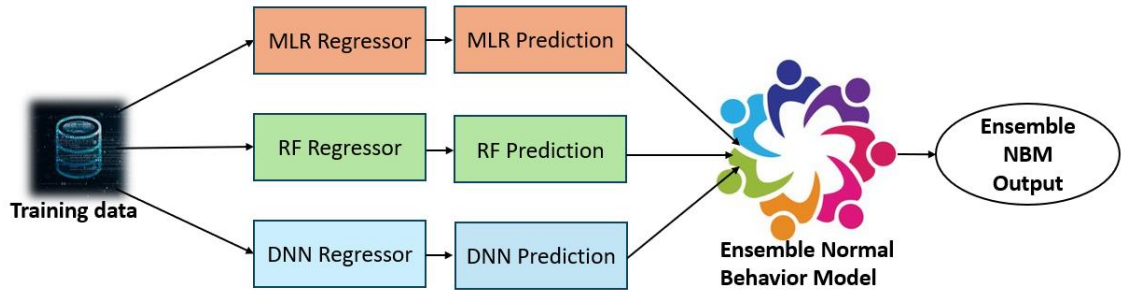


Figure 3. 3: Structure of the parallel ensemble framework.

The fault prediction model of the gearbox temperature in equation (3.3) is constructed with six feature parameters, the active power (A_p), gearbox oil temperature (G_0), rotor speed (R_s), ambient temperature (A_T), ambient wind speed (A_W), and nacelle temperature (N_T).

$$\hat{P} = \hat{f}(A_p, G_o, R_s, A_T, A_w, N_T) \quad (3.3)$$

Where $\hat{P}$ represent the predicted gearbox temperature and $\hat{f}(\cdot)$ is the proposed ensemble model. The ensemble model will employ three data-driven methods, the MLR, RF, and DNN, to predict the gearbox-bearing temperature based on the training dataset.

3.3.1 Multiple Linear Regression

The MLR method determines the relationship between several independent variables, input features, and the dependent variable (output parameter). In this study, the efficiency parameter was attempted to be calculated using the six parameters that were gathered and sent into MLR. The Reference [92] expressed the MLR mathematical relation (13). Where α_0 presents the fitted regression intercept, ε is a fixed integer, X_k' denotes the set of independent variables with k as the index corresponding to the dependent variable Y_i observed in multiple linear regression and α' s are the regression coefficients. The processed T06 WT dataset was inputted into the Scikit-learn linear regression algorithm in Python.

$$Y_i = \alpha_0 + \alpha_1 X_1 + \cdots \alpha_k X_k + \varepsilon, \quad i = 1, 2, \cdots k \quad (3.4)$$

3.3.2 Random Forest

Random forests are composed of many randomly constructed decision trees, each casting a vote for a specific class, such as Class 1 (Yes) or Class 2 (No). Each tree is constructed using a bootstrap sample of the dataset [95]. RF is a supervised ML model that employs an ensemble of decision trees generated from bootstrap samples. It has shown promise as a meta-model due to its ability to capture complex predictive patterns [118]. The model

utilizes the concept of bagging, whereby multiple decision trees are trained on random subsamples of the training dataset to improve prediction accuracy [119]. RF can be viewed as a function that combines weak and strong learners to produce a more robust predictor. The final model output aggregates the predictions from all individual decision trees to generate a more accurate and reliable result [120]. Each tree in the RF is constructed using a subset of the training data, and the optimal split for each node is selected from a randomly chosen subset of candidate variables. This approach helps mitigate overfitting and enhances the model's generalization performance [121].

3.3.3 Artificial Neural Network

An artificial neural network (ANN) is characterized as an intricate system composed of interconnected fundamental computing units (formal neurons) [122]. The neurons are arranged in clusters or layers and can be interconnected in various configurations. The neuronal connection topology delineates the network architecture associated with the job it must do. This job is often defined by a collection of input values and their corresponding output values that the network must learn to provide accurate responses for additional unknown inputs. The learning process is the evaluation of network parameters that fulfill a performance requirement by decreasing the discrepancy between the network output and the actual output [123]. Learning is conducted according to an algorithm specific to the network architecture.

3.3.3.1 Deep Neural Networks

DNN is a type of ANN with multiple layers between input and output layers, which allows it to model more complex and non-linear relationships. DNNs have proven to be effective

modeling tools, particularly in capturing non-linear relationships among parameters in complex engineering and energy systems [53]. They are complex structures based on biological phenomena consisting of layers and neurons; these neurons transform data as it passes through the network to enable the model to learn hierarchical feature representation from raw data. The input features are initially passed to the first layer, the input layer. After processing the input and extracting relevant information, the data are transmitted to one or more hidden layers, each consisting of one or multiple neurons. The final layer, or output layer, generates the model's predictions. Non-linear activation functions are applied to process data received from preceding layers before passing it to subsequent layers [124]. During training, a procedure adjusts the weights and biases in the hidden and output layers to achieve the desired response by minimizing a predetermined loss or error function [125]. Learning rates are part of the DNN architecture, which can be dynamically adjusted during training using learning rate scheduling or adaptive learning rate (Adam).

3.4 Model Performance Metrics

The statistical indices in equations (3.5 – 3.9) include the mean squared error (MSE), mean absolute error (MAE), coefficient of determination (R^2), Root mean squared error (RMSE), and Willmott's index of agreement (WIA), are used to evaluate the prediction performance of individual base models and assign corresponding weights for accurate ensemble model. The metrics are specified in terms of m , the number of data instances in the test set, $\bar{y}_i$ is the average on the measured value, $\hat{y}_i$ is the predicted value, and y_i is the measured value.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (3.5)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \quad (3.6)$$

$$R^2 = \frac{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad (3.7)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (3.8)$$

$$WIA = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (|\hat{y}_i - \bar{y}_i| + |y_i - \bar{y}_i|)^2} \quad (3.9)$$

3.5 Fault Prediction/ Alarm Generation System Approach

Examining the deviations observed from the ensemble's normal behaviour after the model development and performance evaluation is essential. Statistical processing control (SPC) has been identified as a powerful tool for achieving process stability and reduction of variability for improved capability [90]. Therefore, with these notes, the SPC is used to identify anomalies in wind turbines by evaluating the deviation between SCADA-measured values and the ensemble machine learning model predicted values. A shewhart control chart was utilized in this research to identify the deviations. It is derived from a normal distribution, plotting deviations between two control limits to identify the abnormal behaviour. The control limits include the upper control limit (UCL) and the lower control limit (LCL). The alerts and action limits are integrated to form the Upper Control Limit (UCL) and Lower Control Limit (LCL), delineating the chart's sensitivity. [123]. The UCL and LCL are calculated as the multiple of the standard deviation of the distribution (σ) and

the moving range (MR) to determine anomalies [126]. The process described is explained in the following equations:

$$\Delta = y_i - \hat{y}_i \quad (3.10)$$

$$MR = |\Delta_i - \Delta_{i-1}| \quad (3.11)$$

$$\sigma = \frac{\overline{MR}}{1.128} \quad (3.12)$$

$$UCL = +3\sigma \quad (3.13)$$

$$LCL = -3\sigma \quad (3.14)$$

When the system data under analysis exhibits healthy behaviour, aligning with the model's predictions, deviations on the Shewhart control chart will follow a normal distribution with a mean of zero. Conversely, non-conforming data will indicate randomness within the control chart, with points exceeding the control limits and demonstrating a shift in the mean, thereby signalling abnormal behaviour. Upon validating the model by confirming that the detected abnormalities correspond to actual faults, it can be deployed for real-time fault detection, enabling predictive maintenance.

3.6 Solar Farm Defect Prediction Framework

As illustrated in Figure 3.4, this section defines the architectural procedure for solar photovoltaic defect prediction utilizing a convolutional neural network. The required steps include electroluminescence image acquisition and preprocessing, followed by constructing a deep learning model, which primarily involves EL image feature extraction and multi-class classification.

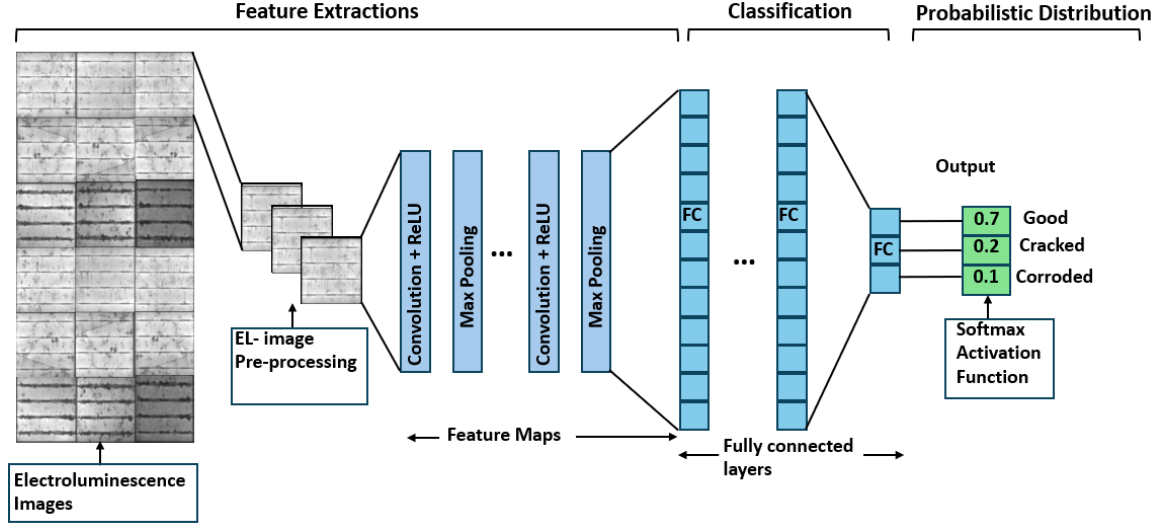


Figure 3. 4: Process of developing solar PV defect prediction framework

3.6.1 EL Image Acquisition and Preprocessing

The EL images are 8-bit grayscale images acquired from Case Western Reserve University [77], this data contains the training and testing set of the image dataset of the PV module under step-wise damp heat exposures from 500h to 3000h. The pixel size is 250 x 250. In the preprocessing step, image normalization Ref. [81] image augmentation [77] and EL image resizing was implemented [1].

3.6.1.1 EL Image Normalization

This preprocessing step scales pixel values to the range [0, 1] by standardizing the pixel intensity values of images. This helps improve the training performance of the deep learning model by ensuring that the data is consistent and scaled appropriately [81]. Equation 3.15 presents the typical expression for image normalization, where I_{pixel} represents the original pixel intensity of the image, and $I_{normalized}$ denotes the normalized pixel intensity.

$$I_{normalized} = \frac{I_{pixel}}{255} \quad 3.15$$

3.6.1.2 EL Image Augmentation

The quantity of images available for deep learning model development is critical. Image augmentation is commonly employed to artificially expand the dataset by applying various transformations to the original images to address limitations in dataset size. This technique is widely utilized to enhance the generalization capability of machine learning models, particularly in scenarios where the dataset is small or imbalanced; corrosion and cracking only become noticeable after 2,500 hours of damp heat exposure, which limits the portion of the dataset that represents these features [77]. Geometry and color transformation are among the key augmentation strategies.

3.6.1.3 EL Image Scaling

Image resizing is scaling an image to a specific width and height while maintaining its quality and aspect ratio. It is a crucial preprocessing step in image analysis tasks. The cell images of 250x250 pixels will be resized to lessen the computation time [25], [77], and also to observe the accuracy of solar PV defect classification.

3.6.2 Deep Learning Model Construction

A CNN is a customized deep neural network that processes structured grid data, including photographs. It is extensively utilized in computer vision applications such as picture

categorization, object identification, and segmentation. [127]. CNN has the potential to extract features in EL images for efficient defect prediction, as proposed in [32], [82].

The application of CNN for defect prediction in solar photovoltaic systems has received considerable attention due to its effectiveness in analyzing complex image data. Numerous studies have highlighted the efficacy of modified CNN architectures in improving defect detection accuracy, a critical factor in ensuring the performance and longevity of solar panels. Ref. [128] overcome the difficulty of EL images in solar fault detection and classification by employing a deep CNN, specifically Xception. The proposed model achieved 94.382% accuracy, 92% precision, and 95.12% recall. Similarly, using an improved VGG-16 model, the defect detection accuracy of solar PV modules increased from 97.65% to 99.13%, compared to the performance of the classical VGG-16 model.

Inspired by the above, we proposed a deep learning model based on the convolutional neural network to extract complex features and classify defects within the electroluminescence solar PV cell images. This multi-classification deep learning model will classify the EL cell images into three main classes, good, cracked, and corroded, based on the model prediction according to the softmax activation function probability distribution.

3.6.3 Deep Learning Model Evaluation

The classification accuracy of the predictive model will be evaluated on the test set using the developed CNN model. Two commonly utilized machine learning metrics will be employed: precision, which measures the model's ability to classify electroluminescence image pixels correctly, and recall, which assesses the model's capability to identify defects

within EL images accurately [1], [32], [84]. The accuracy, precision, and recall equations in a multi-class classification problem are detailed in equation (3.16 – 3.18). Here, C represents the total number of classes, TP_i stands for the correctly predicted samples for class i , FP_i represents the samples incorrectly predicted as class i , and FN_i refers to the actual samples of class i that were misclassified as another class. Similarly, a confusion matrix will be presented to illustrate the misclassified electroluminescence cell images and to provide a clear understanding of the differences between the actual classifications and the model's predicted classifications of the EL cell images.

$$Accuracy = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C (TP_i + FP_i + FN_i)} \quad (3.16)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (3.17)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (3.18)$$

3.7 Summary

The fault prediction framework consists of model development, training, and online monitoring. The model development and training process handles the data pre-processing and feature selection based on the operational data obtained from the SCADA system for the ensemble model development. The real-time SCADA data is passed through the ensemble NBM after the pre-processing and feature selection steps during the online monitoring process. The larger deviations between the predicted value of the ensemble model and the actual values recorded by the SCADA system have high potential fault symptoms.

The solar farm defect prediction framework employs a convolutional neural network for identifying photovoltaic module defects through data acquisition, preprocessing, model development, and evaluation. Electroluminescence images from Case Western Reserve University capture defect evolution under damp heat exposure. Preprocessing involves normalization, augmentation, and scaling to standardize and expand the dataset for effective model training. Evaluation metrics include accuracy, precision, and recall, with a confusion matrix highlighting classification performance. This robust defect prediction approach leverages state-of-the-art CNN architectures to improve fault detection accuracy and support solar panel maintenance and longevity.

CHAPTER 4

MODEL CONSTRUCTION AND EVALUATION

4.1 Introduction

This chapter implements the step-by-step research methodology outlined in Chapter 3, encompassing the construction of the individual NBMs and the Ensemble NBM for wind turbine fault prediction. Similarly, it details the development of the solar photovoltaic cell defect prediction model.

4.2 NBM Data Preparation

This section involves all data preparation steps. Such as SCADA data acquisition and preprocessing analysis.

4.2.1 Data Acquisition

This thesis utilized open-source SCADA data from wind turbine operations managed by EDP in the Gulf of Guinea, West Africa [65]. The data was recorded at 10-minute intervals from January 1, 2017, to December 31, 2017, and includes 79 features of the four wind turbines available (T01, T06, T07, and T011). According to the maintenance log, this study focused on turbine T06, which experienced a failure event in the gearbox bearing. Figure 4.1 shows the time series variation of T06 - wind turbine features.

4.2.2 Pre-processing

Basic statistical measures, including minimum, maximum, average, and standard deviation, were computed for key features of the wind turbine systems according to the SCADA data. However, for simplicity, this study utilized only the average values of the recorded parameters, as they capture most of the relevant information. By selecting the average values of key features, the feature set was reduced to 21. Data points lacking information due to sensor malfunctions or communication errors were removed, and data points significantly deviating from the wind turbine power curves were also filtered out, as these anomalies may have resulted from operational constraints or sensor malfunctions. The SCADA data for T06 WT initially contains 52,346 data points, and 37,335 are valid samples obtained after pre-processing.

Noting that the wind power curve (WPC) characterizes the operational status of a wind turbine, this study focused on establishing a representative trajectory for the WPC.

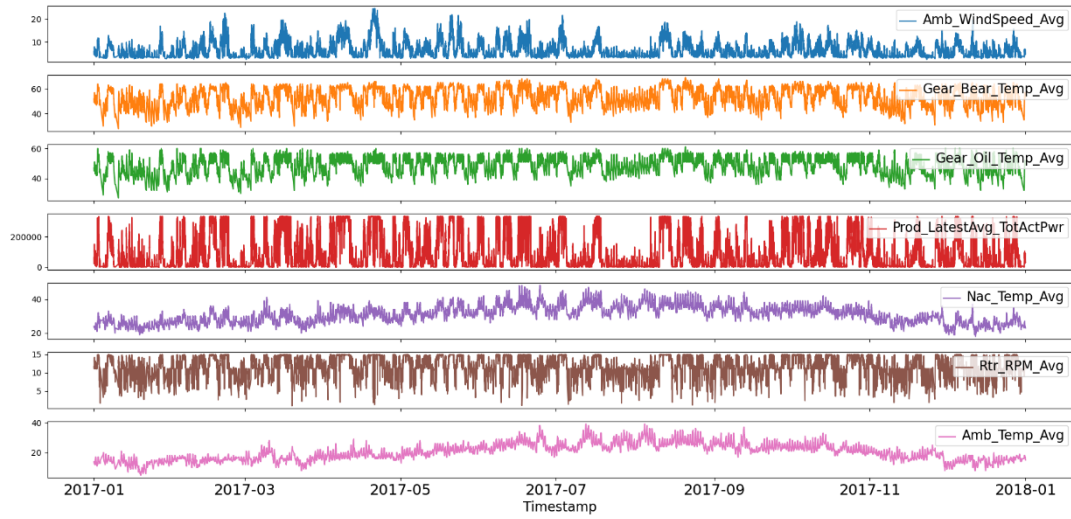


Figure 4. 1: Time series variation of features in T06 - wind turbine.

Table 4. 1 EDP T06 Wind turbine data description

Features	Unit	Mean	Standard Deviation	Minimum	Max
Gearbox oil Temperature	°C	50.6	4.6	27	61
Ambient Temperature	°C	20.6	5.5	5	39
Gearbox Bearing Temperature	°C	56.6	5.8	28	69
Active Power	W	133382.8	116487.2	1	334113
Ambient Wind Speed	m/s	7.5	3.4	2.4	24.5
Nacelle Temperature	°C	31.7	5.1	18	48
Rotor Speed	rpm	12.7	1.94	1	14.9

Figure 9 presents the results obtained before and after the initial pre-processing step of the WPC. In the subsequent pre-processing step, a wind power curve filtering technique was applied to further refine the WPC, with the resulting output displayed in Figure 4.3.

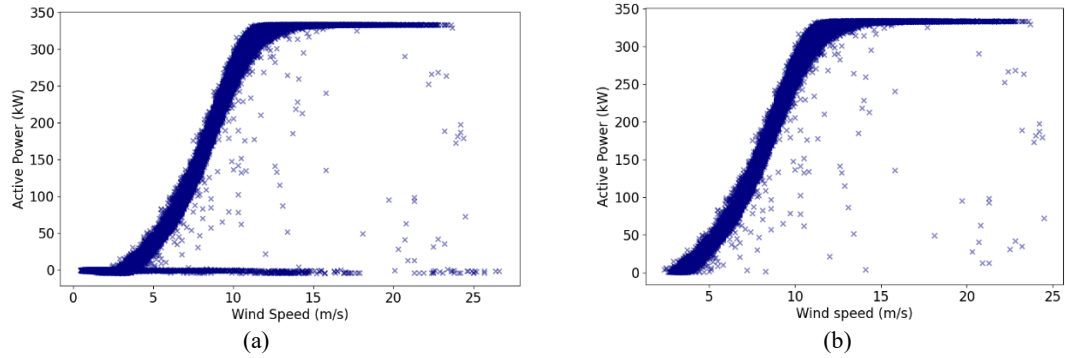


Figure 4. 2: (a) Original wind power curve (b) Result of the first preprocessing step of the wind power curve.

4.2.3 Feature Extraction/Selection

After removing abnormal turbine data, the proposed ensemble technique applies three feature selection algorithms: mutual information selector, principal component analysis, and LASSO Regression. These algorithms were employed to hierarchically identify the most relevant features for model development to predict gearbox-bearing temperature. Each feature selection algorithm has specific tuning parameters. For example, the mutual information selector method ranks features based on their MI scores, with the primary tuning parameter being the number of top-ranked features selected. Principal Component Analysis depends on the number of components chosen to capture maximum variance. At the same time, the LASSO regression method includes a regularization parameter that adjusts the strength of penalization on the absolute values of the regression coefficients. Given the inherent subjectivity, these parameters are often selected intuitively. Consequently, the configuration yielding the best performance was chosen for the proposed ensemble technique.

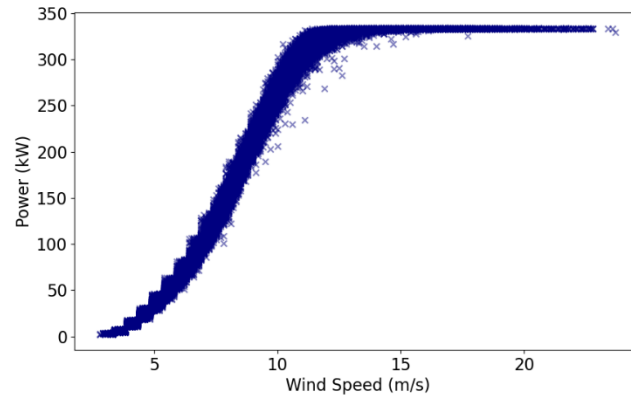


Figure 4.3: Filtered wind power curve.

The initial feature selection method employed was manual, selecting only those features that represented the average values of the variables. Figure 4.4 shows the results from three feature selection algorithms. Six features were ultimately chosen as input variables for predicting gearbox-bearing temperature, as outlined in Table 3.

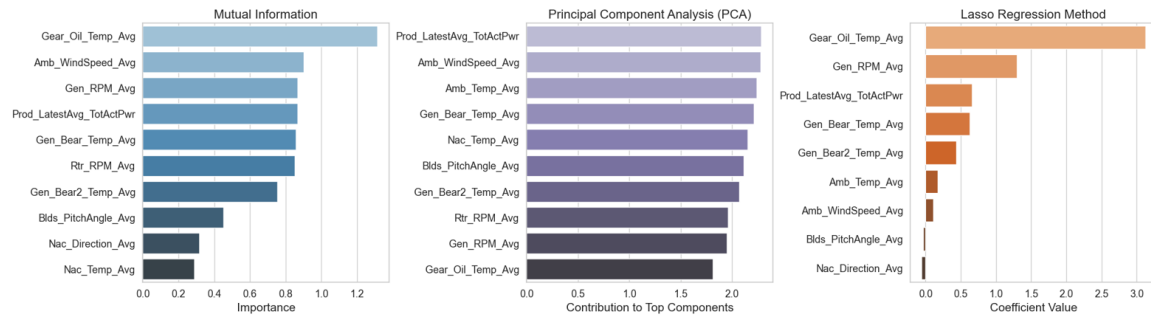


Figure 4.4: Feature ranking of the three FSA.

Table 4. 2 Most relevant features description

Features	Description	Range
Gear_Oil_Temp_Avg	Gearbox Oil Temperature	[27 - 61]
Prod_LatestAvg_TotActPwr	Total active power	[1 – 334113]
Rtr_RPM_Avg	Rotor speed	[1 – 14.9]
Amb_Temp_Avg	Ambient Temperature	[5 – 39]
Amb_WindSpeed_Avg	Ambient Wind Speed	[2.4 – 24.5]
Nac_Temp_Avg	Nacelle Temperature	[18 – 48]

4.3 Ensemble Model Construction

This section presents the development of Individual and ensemble NBM techniques utilizing a voting approach in which the forecasts of distinct anomaly detection algorithms are amalgamated by weighted voting [70]. The Weighted Average ensemble approach allocates weights to distinct models according to performance, utilizing the softmax function. [102].

Three machine learning models, namely multiple linear regression, random forest, and deep neural network, were developed and compared with the ensemble technique based on their performance metrics, as detailed in Chapter 3. In this thesis, the feed-forward neural network models with single to multiple hidden neurons will be explored to map the input-output relationship and determine the health condition of the gearbox bearing. The neural network architecture of the shallow neural network and multiple hidden layers are presented in Figure 4.5.

During model training, 80% (29,868) of the dataset was allocated for training, with the remaining 20% (7,467) used for validation. The training and validation loss with the corresponding training error is illustrated in Figure 4.6. The 5-hidden layer structure was selected as the architecture of the DNN employed for the NBM due to its lowest training error and validation loss.

Results indicate that RF and DNN outperformed the standalone MLR model. In contrast, the ensemble model achieved the best performance. The ensemble assigned the highest weight to the better-performing model, resulting in the lowest error metrics and the highest

WIA and R^2 values. Table 4.3 describes the optimal hyperparameters for RF and DNN architecture, and finally, Table 4.4 presents the model's performance.

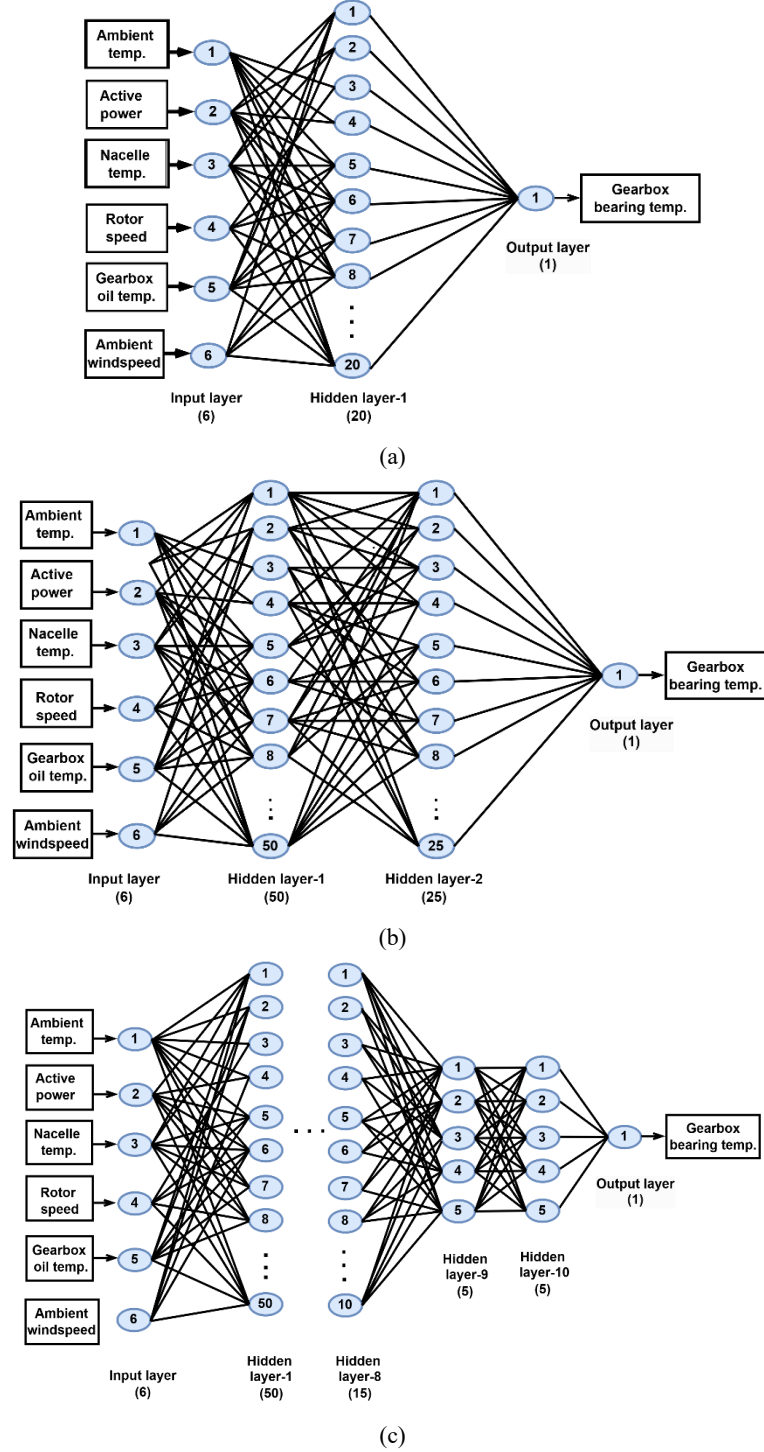
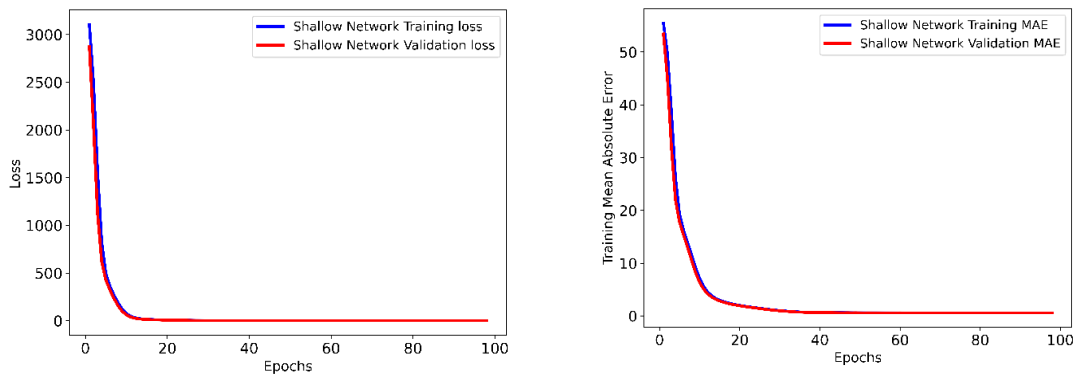


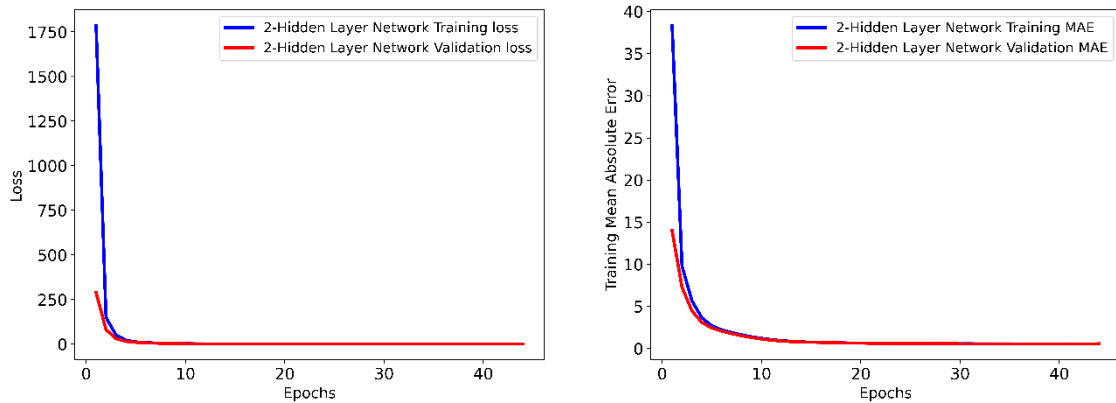
Figure 4. 5: (a) Shallow network configuration, (b) Two hidden layers DNN, and (c) Deep neural network configuration (multiple hidden layers with multiple neurons in each layer).

Table 4. 3 Optimal hyperparameters for RF and DNN architecture.

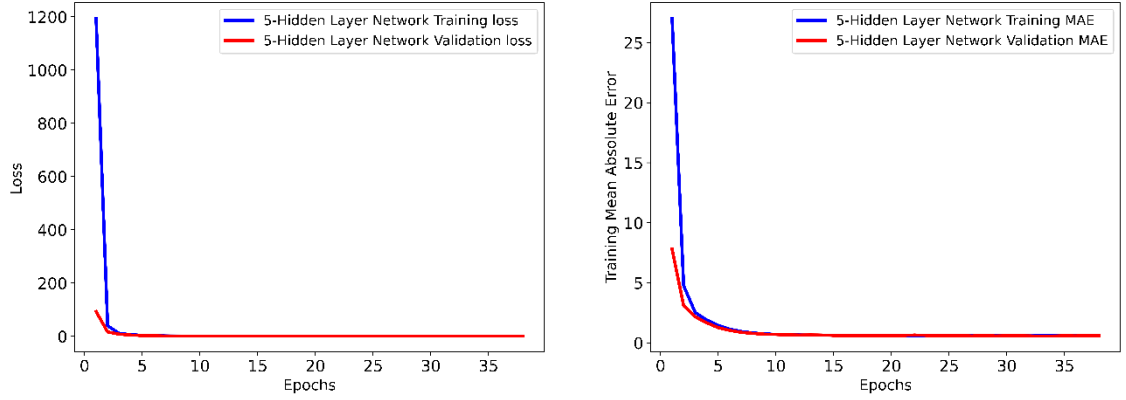
RF Model		DNN	
Hyperparameter	Optimal Hyperparameter	Hyperparameter	Optimal Hyperparameter
Number of Estimators	100	Dense layer	5
Maximum depth	10	No. of Neurons	[50,35,20,15,5]
Minimum sample leaf	5	Optimizer	Adam
Minimum sample split	3	Validation metric	MAE



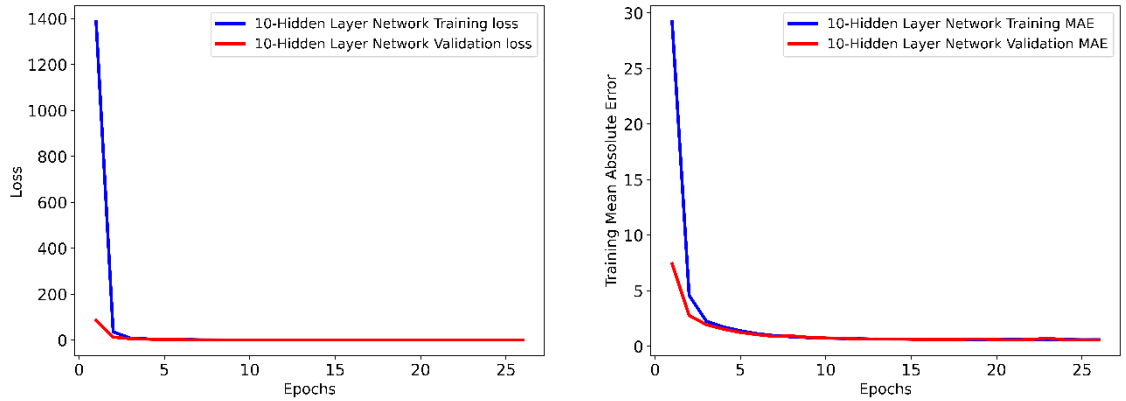
(a)



(b)



(c)



(d)

Figure 4.6: Training error and validation loss curve for (a) Shallow neural network, (b) 2 - hidden layer network, (c) 5 - hidden layer network, (d) 10 - hidden layer network

The performance evaluation metrics result for the DNN is illustrated in Figure 4.7. Also, the performance evaluation metrics result for both the standalone and ensemble NBM prediction is presented in Figure 4.8. The prediction results from the stand-alone NBM are illustrated in Figures (4.9 - 4.11). The plot describes the prediction accuracy for each model on the test dataset for the first 100th, 1000th, and the whole test dataset.

Table 4. 4 NBM performance metric comparison

Normal Behaviour Model	R^2	WIA	MSE	MAE	RMSE
MLR	0.9693	0.9921	1.0407	0.7463	1.0201
RF	0.9831	0.9957	0.5710	0.5058	0.7556
DNN	0.9811	0.9952	0.6225	0.5532	0.7999
Ensemble	0.9858	0.9963	0.3492	0.4196	0.5909

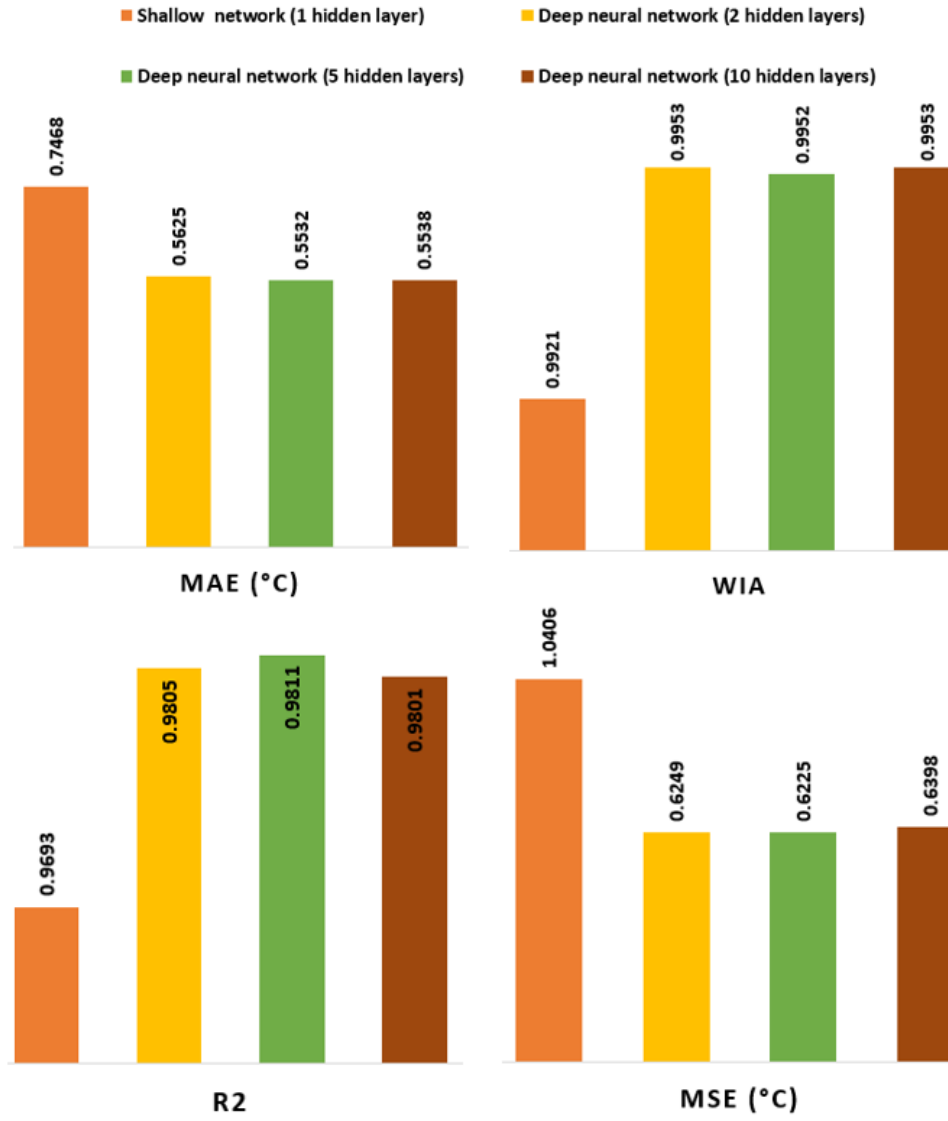


Figure 4. 7: DNN evaluation result for MAE, WIA, R², and MSE

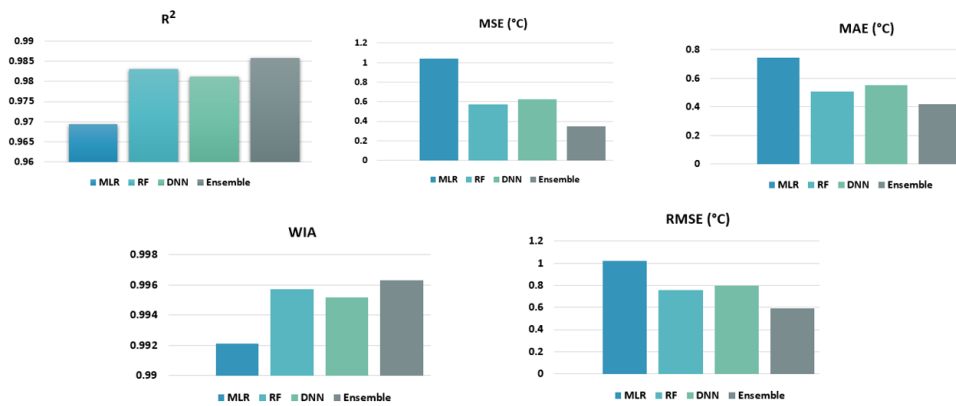
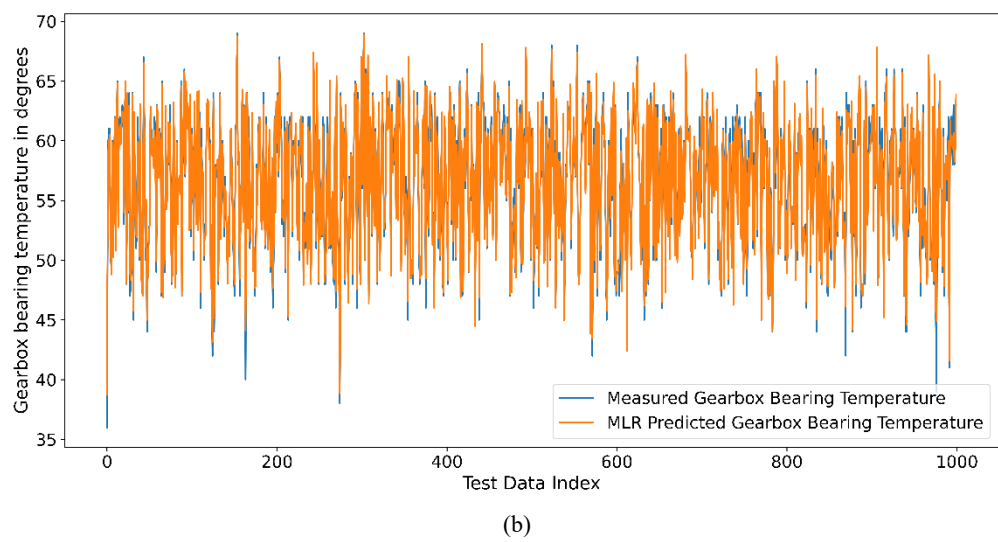
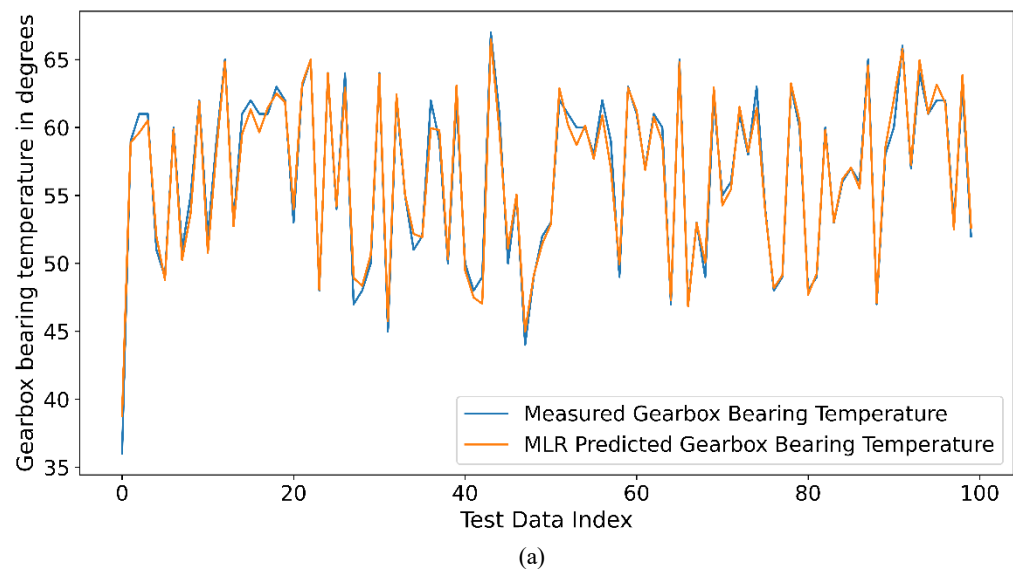
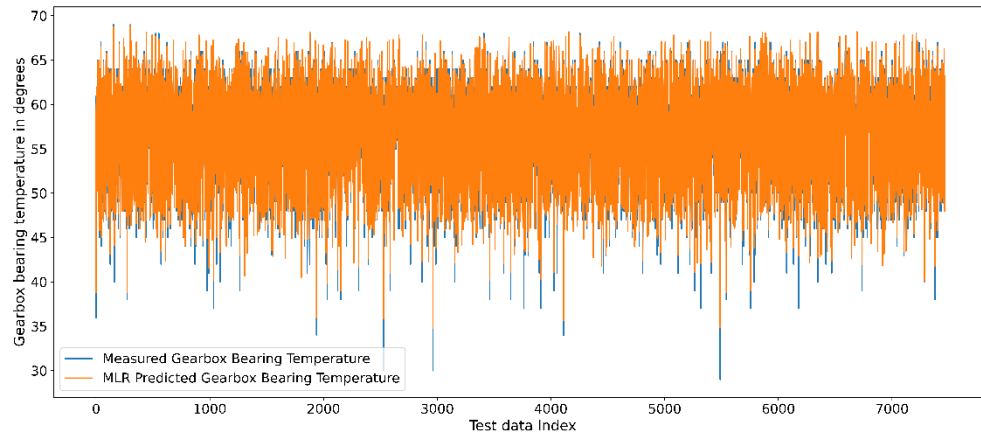


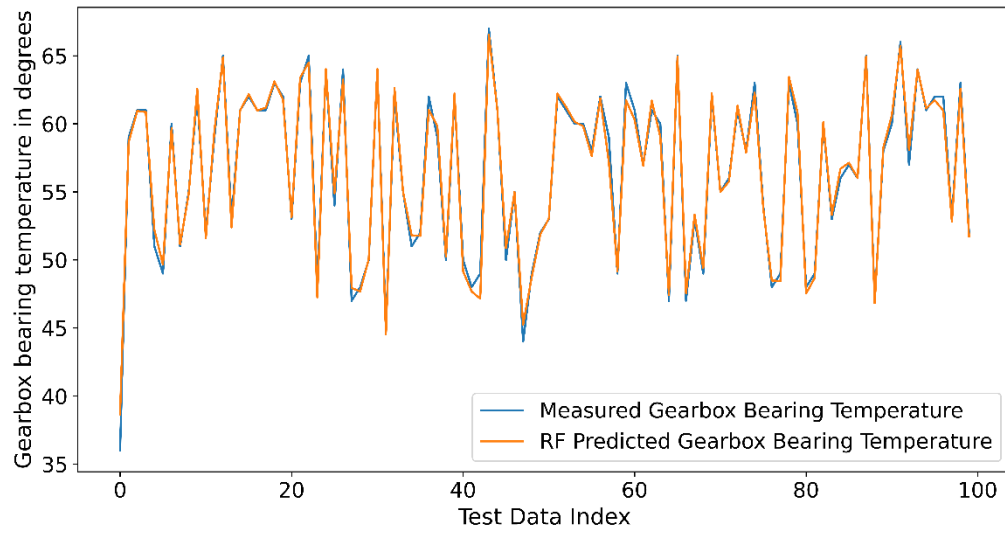
Figure 4. 8: Standalone and ensemble NBM evaluation result



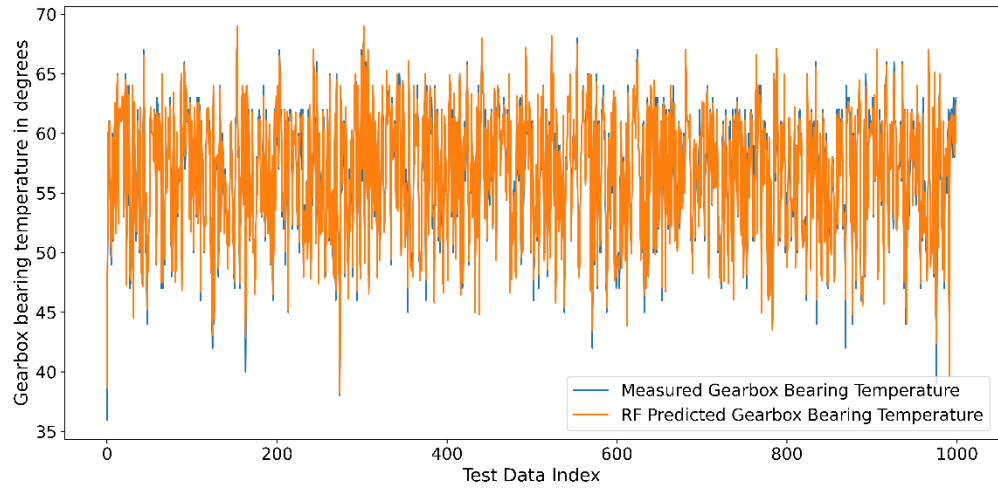


(c)

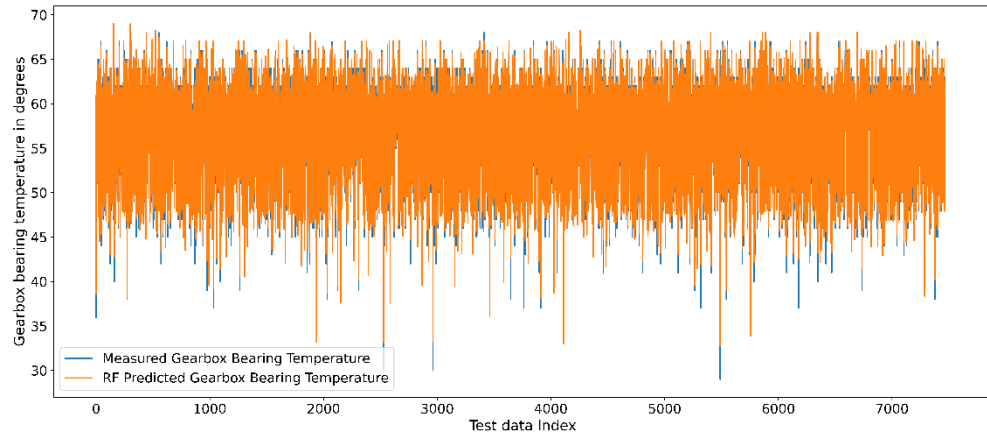
Figure 4. 9: Prediction of Gearbox bearing actual and predicted temperature for (a) MLR first 100th data point, (b) MLR first 1000th data point, (c) MLR Test set.



(a)

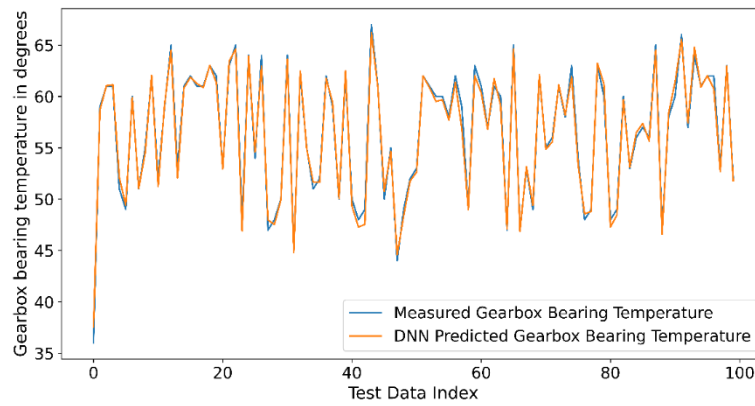


(b)

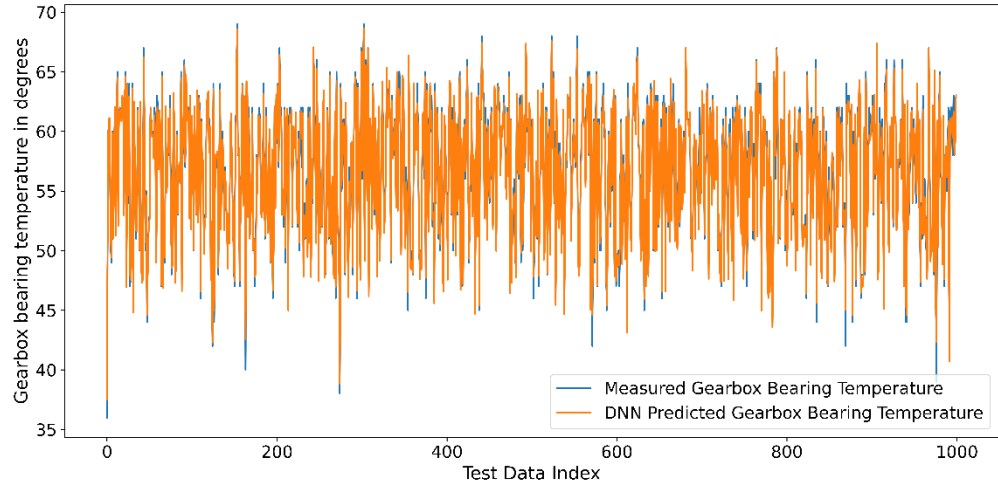


(c)

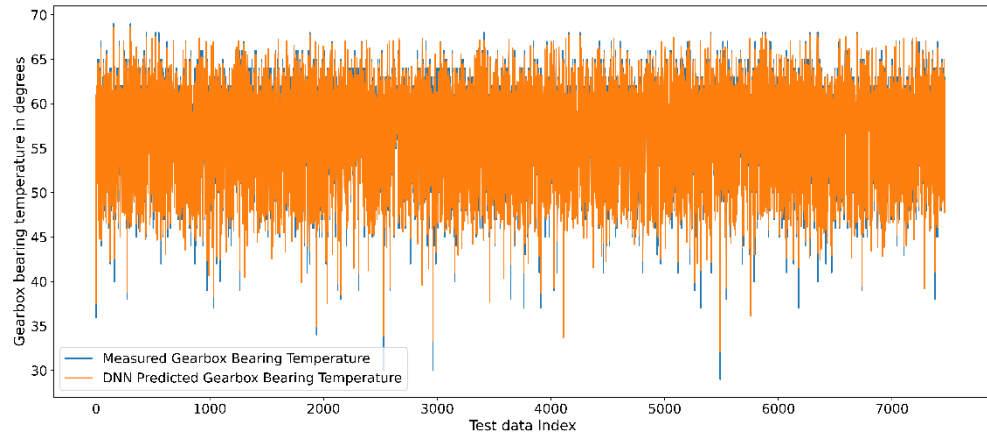
Figure 4. 10: Prediction of Gearbox bearing actual and predicted temperature for (a) RF first 100th data point, (b) RF first 1000th data point, (c) RF Test set.



(a)



(b)



(c)

Figure 4. 11: Prediction of Gearbox bearing actual and predicted temperature for (a) DNN first 100th data point, (b) DNN first 1000th data point, (c) DNN Test set.

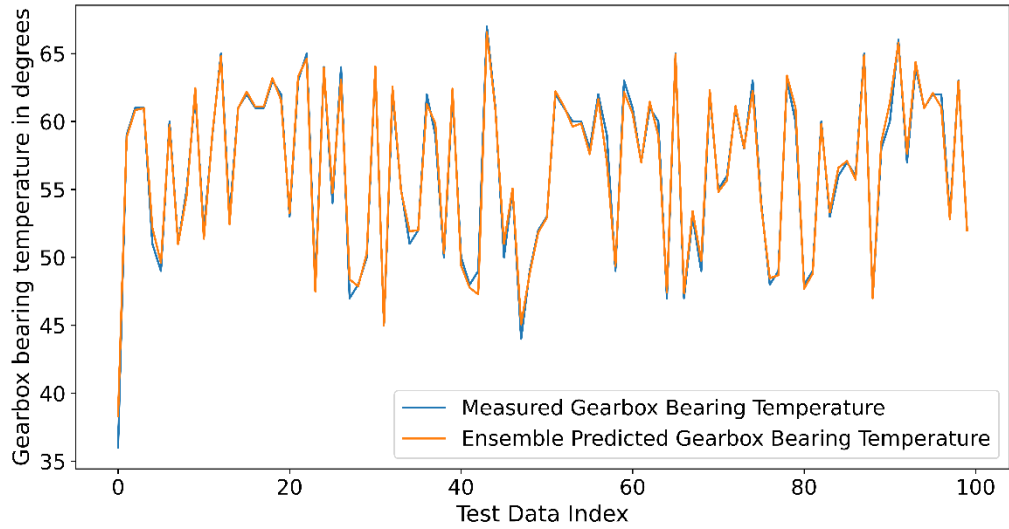
Following the performance evaluation of three standalone NBM, the initial weight was set based on the model prediction accuracy. The defined raw weights are presented in (4.1), with w_{RF} , w_{MLR} and w_{DNN} as gearbox bearing temperature prediction weights and their corresponding prediction as RF_{pred} , MLR_{pred} and DNN_{pred} from RF, MLR, and DNN, respectively.

$$\begin{bmatrix} w_{RF} \\ w_{MLR} \\ w_{DNN} \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.01 \\ 0.9 \end{bmatrix} \quad (4.1)$$

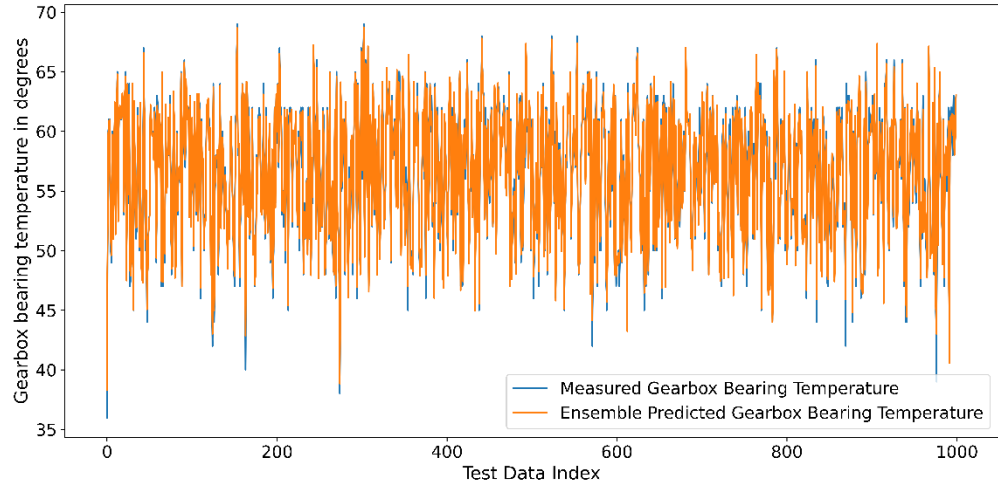
The softmax function transforms the set weight by scaling between zero and one ([0,1]) to ensure that the contribution of each model is proportional but normalized. Finally, the ensemble prediction weight is given by (4.2), and the ensemble prediction as in (4.3), having Sw_{RF} , Sw_{MLR} , and Sw_{DNN} as the transformed softmax weight for RF, MLR, and DNN, the results from the ensemble NBM are illustrated in Figure 4.10. The plot describes the prediction accuracy for each model on the test dataset for the first 100th, 1000th, and the whole test dataset.

$$\begin{bmatrix} Sw_{RF} \\ Sw_{MLR} \\ Sw_{DNN} \end{bmatrix} = \begin{bmatrix} 0.4148 \\ 0.1703 \\ 0.4148 \end{bmatrix} \quad (4.2)$$

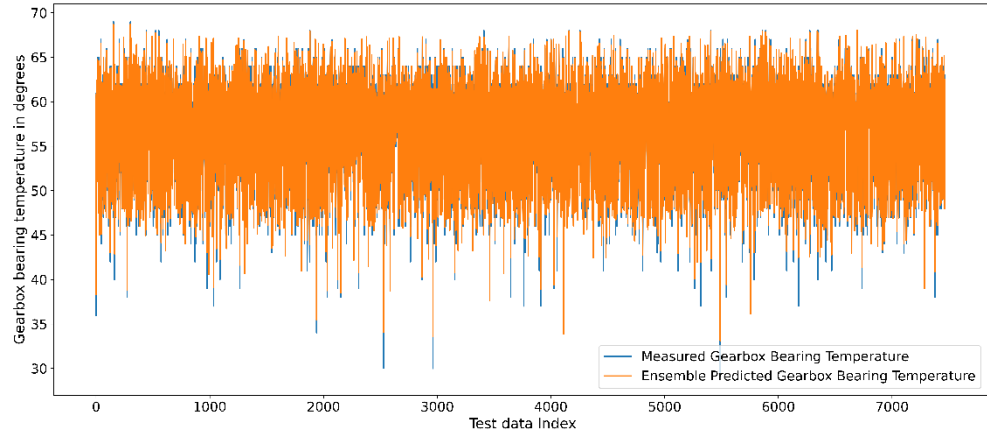
$$Ensemble Prediction = [RF_{pred} \quad MLR_{pred} \quad DNN_{pred}] \begin{bmatrix} Sw_{RF} \\ Sw_{MLR} \\ Sw_{DNN} \end{bmatrix} \quad (4.3)$$



(a)



(b)



(c)

Figure 4. 12: Ensemble NBM prediction of Gearbox bearing actual and predicted temperature for (a) first 100th data point, (b) first 1000th data point, and (c) Ensemble NBM Test set.

4.4 Solar Defect Prediction Model

This section will implement the step-by-step methodology defined for solar PV cell defect classification in Chapter 3.

4.4.1 EL Images Preparation

This section involves all data preparation steps. A total of 1500 EL images were employed for the solar defect classification. Figure 4.11 describes the classes for the EL cell images [1].

4.4.1.1 Normalization

The electroluminescence images are 8-bit grayscale images. Therefore, normalization is applied to both the training and test images to scale their pixel values to a fixed range of $[0, 1]$, as highlighted in Chapter 3.

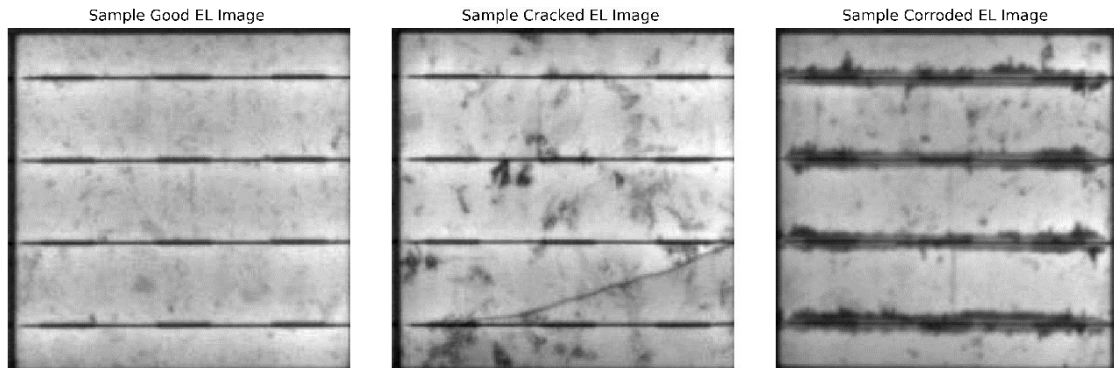


Figure 4. 13: Samples of the EL cell image classes [1]

4.4.1.2 Augmentation

The cracked and corroded cell EL image was of limited quantity, therefore they both under a geometrical transformation during augmentation by rotation of about 180 degrees, and flipping vertically and horizontally to have data balance by avoiding data bias within the EL image dataset. The quantity was increased to reach 500 EL cell images for each class. According to [1], Figure 4.12 illustrates the augmentation transformations.

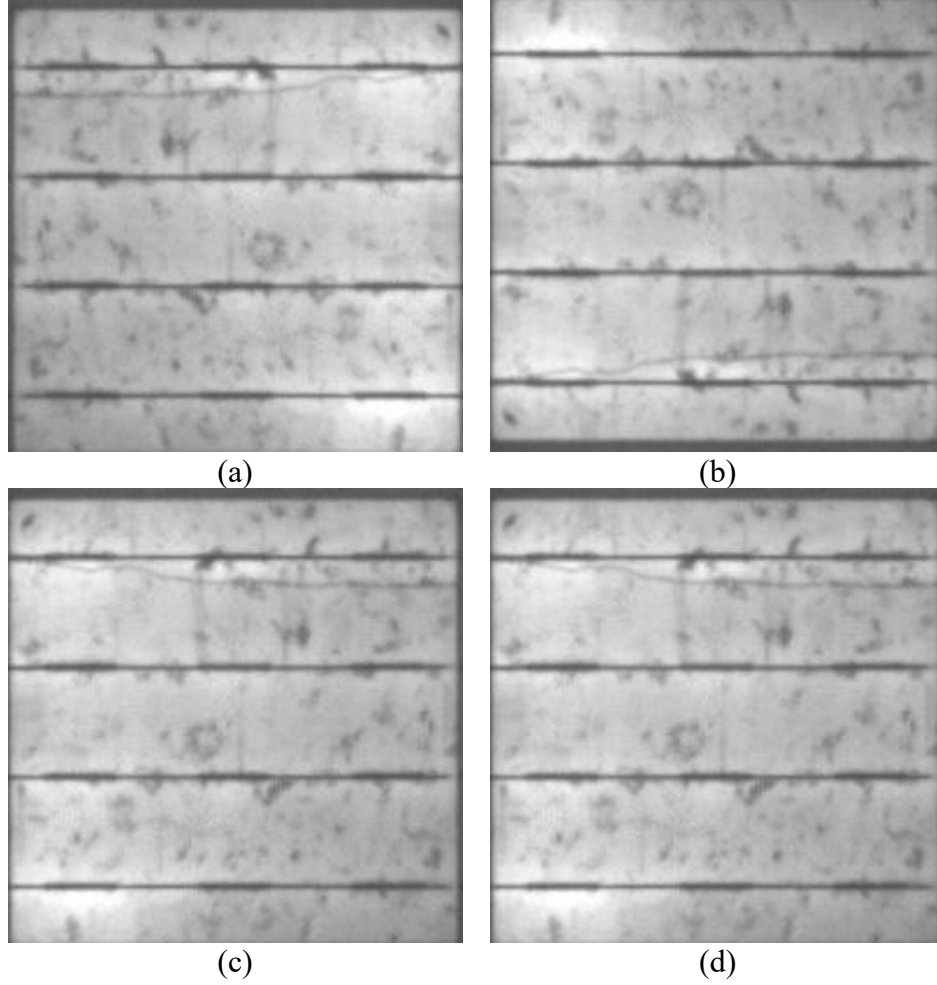
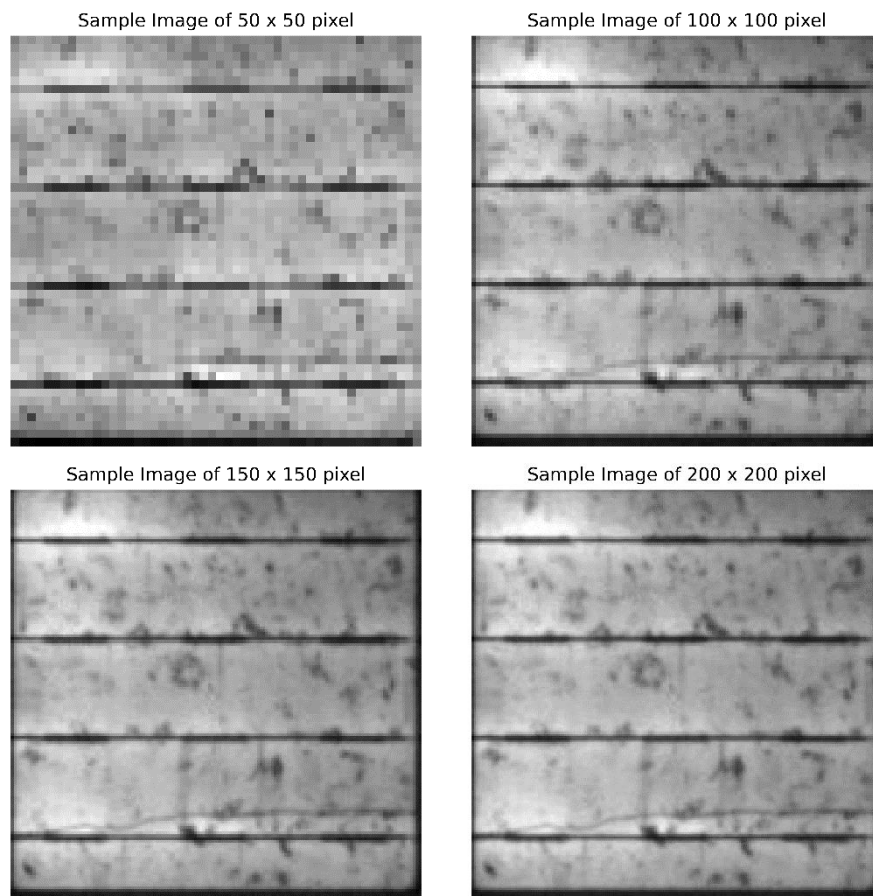


Figure 4. 14: (a) Original, (b) 180 degrees flip, (c) Horizontal flip, and (d) Vertical flip.

4.4.2 CNN-Based Defect Classification Using EL Images

This segment comprehensively describes the proposed convolutional neural network-based approach for defect detection in photovoltaic modules using electroluminescence images. After the CNN-based model was implemented, the resulting output was flattened to serve as the input to a fully connected layer, following the methodologies adopted by [77], [82]. In this work, a softmax function is utilized as the final layer to align the extracted features with their corresponding defect types, following the recommendations of Ref. [129].

In developing the defect classification model, pixel sizes ranging from 50×50 to 250×250 were utilized to evaluate the model's performance based on the resolution of electroluminescence cell images. This approach aimed to achieve higher classification accuracy by reducing noise and optimizing computational efficiency. The samples for the low to high pixel cell image resolution are illustrated in Figure 4.13.



Sample Image of 250 x 250 pixel

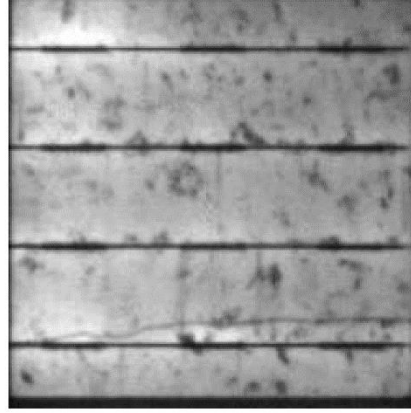


Figure 4. 15: Samples of EL cell images based on resolution according to their pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200, and (2) 250 x 250.

4.4.2.1 Experimental Settings and Training

In this thesis, the electroluminescence image dataset was sourced from a publicly available database [77]. The dataset comprises 852 electroluminescence (EL) images, with 149 images representing the corroded class, 10 corresponding to the cracked class, and 693 images classified as good. To address the class imbalance, the cracked and corroded defect classes were augmented to include 490 and 358 additional EL images, respectively, while the good class was adjusted to contain 402 EL images. In total, 1,409 EL images were utilized for constructing the CNN model, with 80% (1127) of the dataset allocated for training and the remaining 20% (282) reserved for testing. In these experiments, the parameters for the CNN models are provided in Table 4.5 and Table 4.6. The CNN-based model was implemented in Python using the TensorFlow framework within a computing environment running Windows 10 OS, equipped with an Intel Core i9-13900HX processor, 32 GB of RAM, and an NVIDIA GEFORCE RTX graphics card.

During the experimental training, we demonstrated the accuracy and loss of the classification model based on the varying pixel size, such as 50 x 50, 100 x 100, 150 x 150,

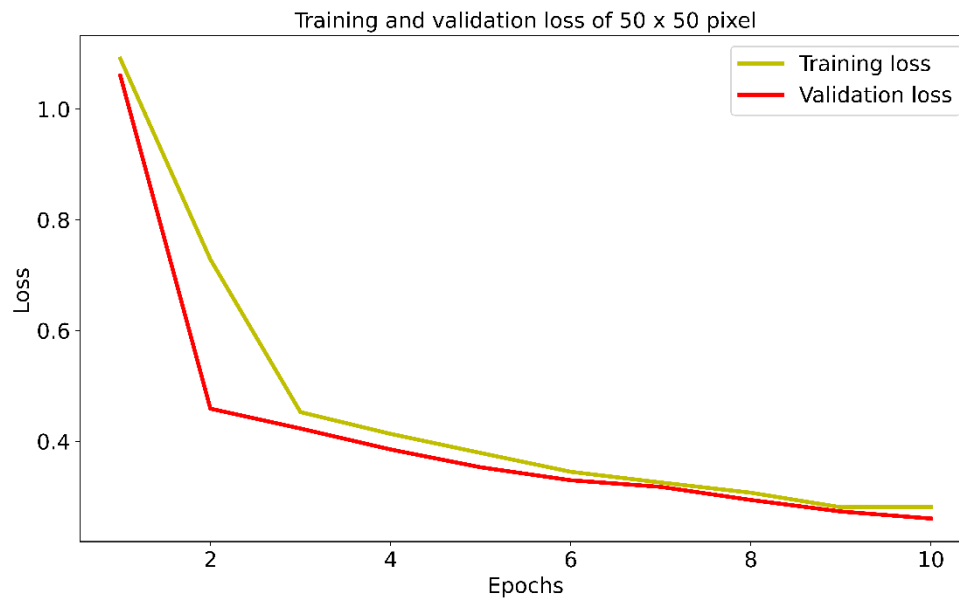
200 x 200, and 250 x 250 in Figure 4.14-15. The test accuracy of each model based on the pixel values is presented in Figure 4.16. Similarly, each model confusion matrix is illustrated in Figure 4.17.

Table 4. 5 CNN-based model architecture for defect classification.

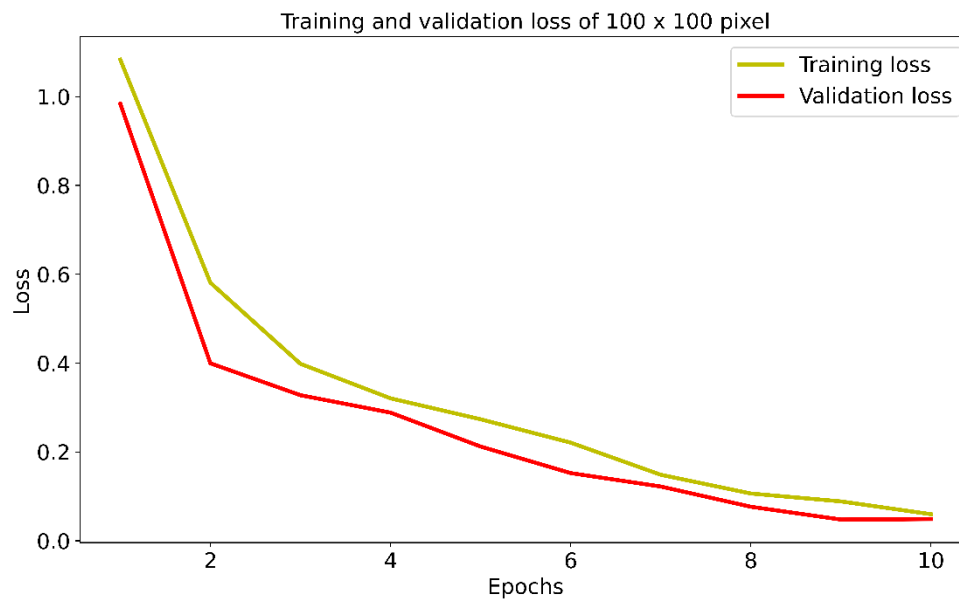
CNN	Output Shape	Parameters
Input	250 x 250 x 1	0
Convolution layer 1	248, 248, 32	320
MaxPooling2D	124, 124, 32	0
Convolution layer 2	122, 122, 16	4624
MaxPooling2D	61, 61, 16	0
Convolution layer 3	59, 59, 16	2320
MaxPooling2D	29, 29, 16	0
Flatten	13456	0
Dropout	16	0
Dense	3	51

Table 4. 6 Training parameters of the defect detection CNN model.

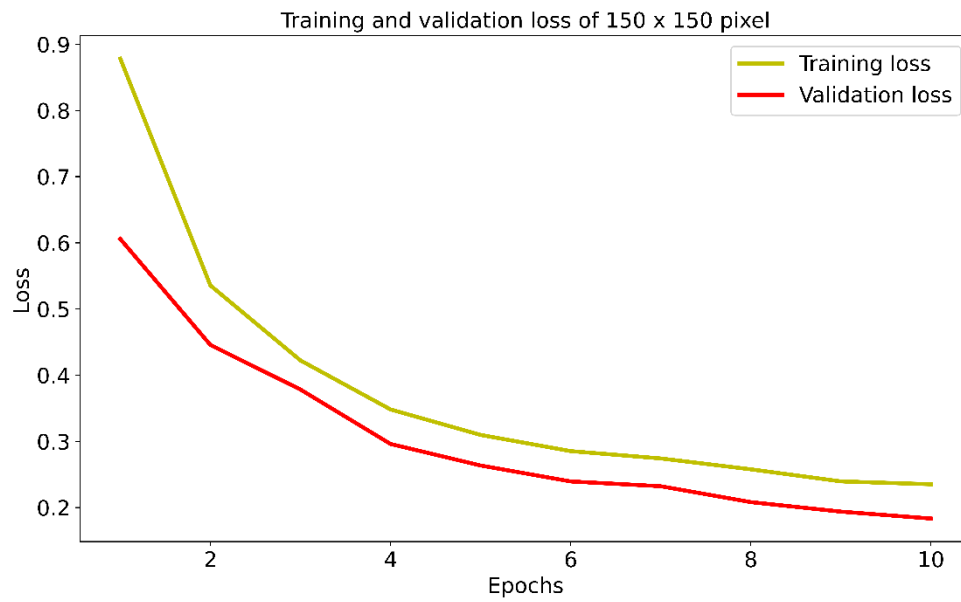
Description	Value
Loss function	Categorical Cross Entropy
optimizer	Adam
metrics	Accuracy
Batch size	16
Iteration times	10



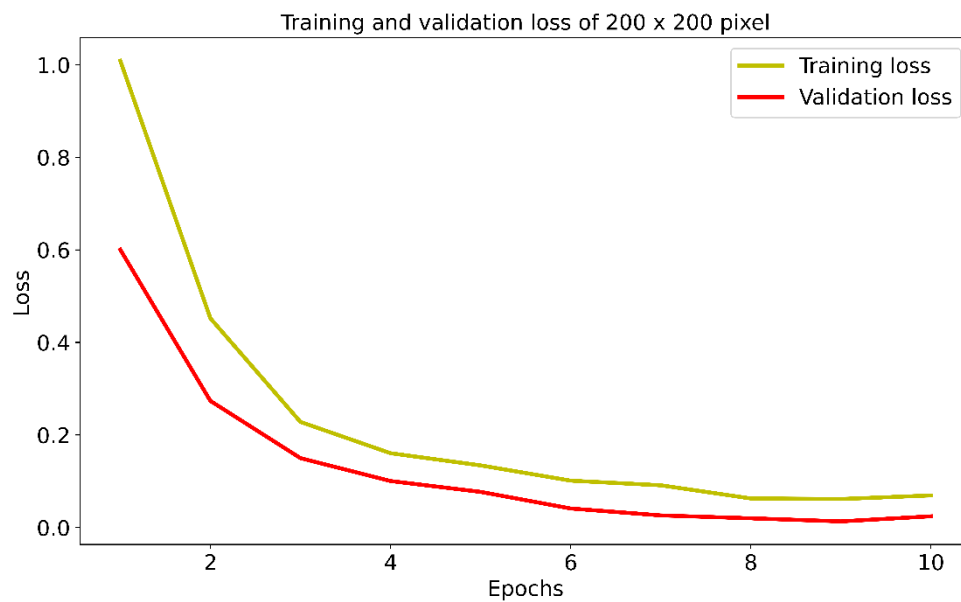
(a)



(b)



(c)



(d)

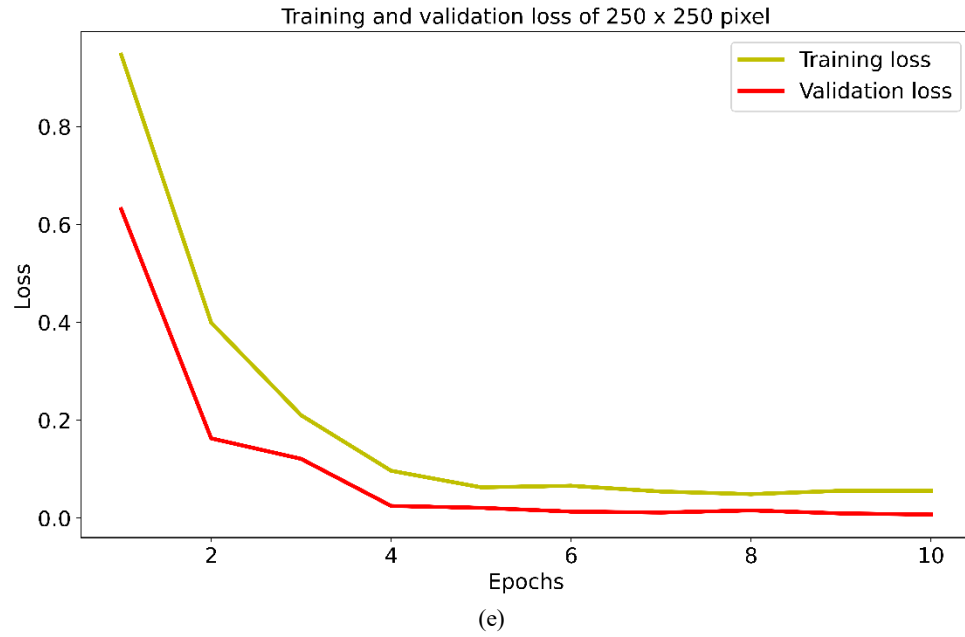
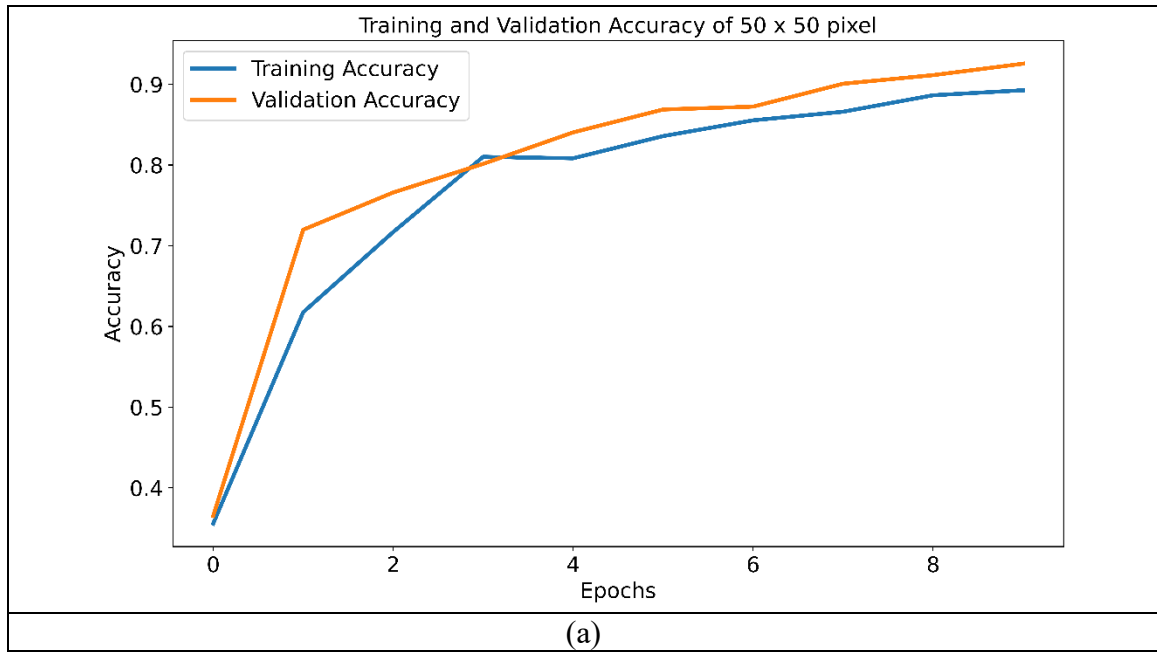
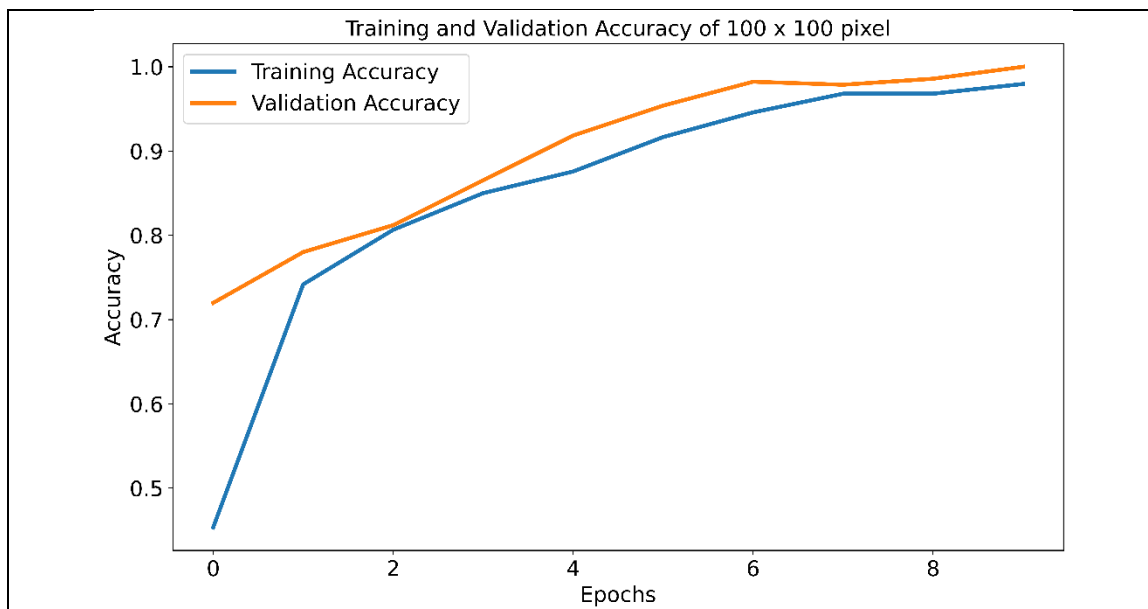
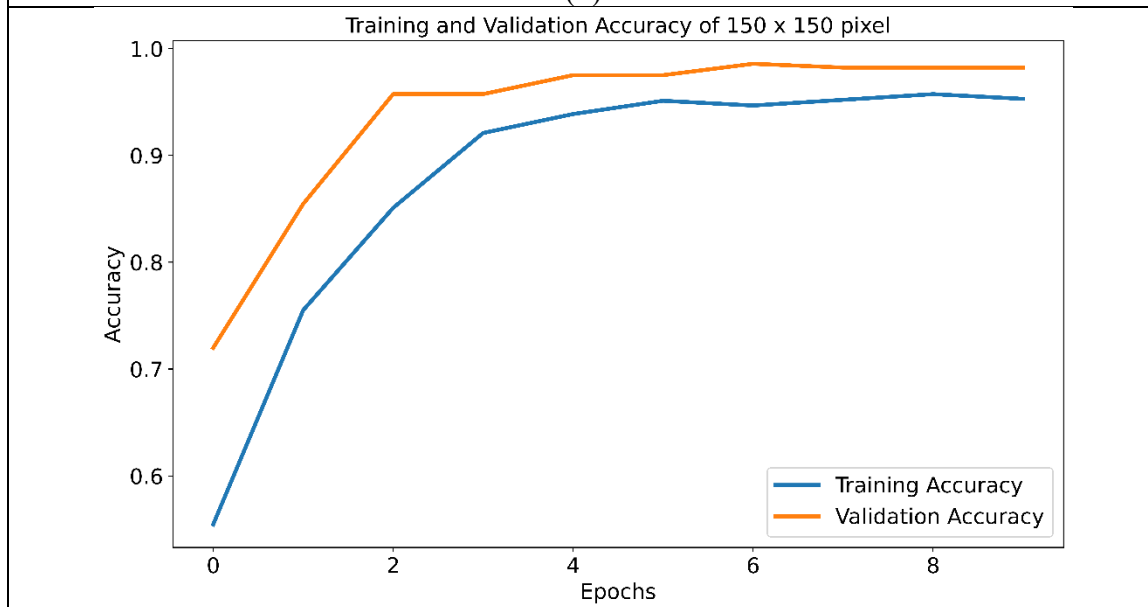


Figure 4. 16: Presents the training and validation loss based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (2) 250 x 250.

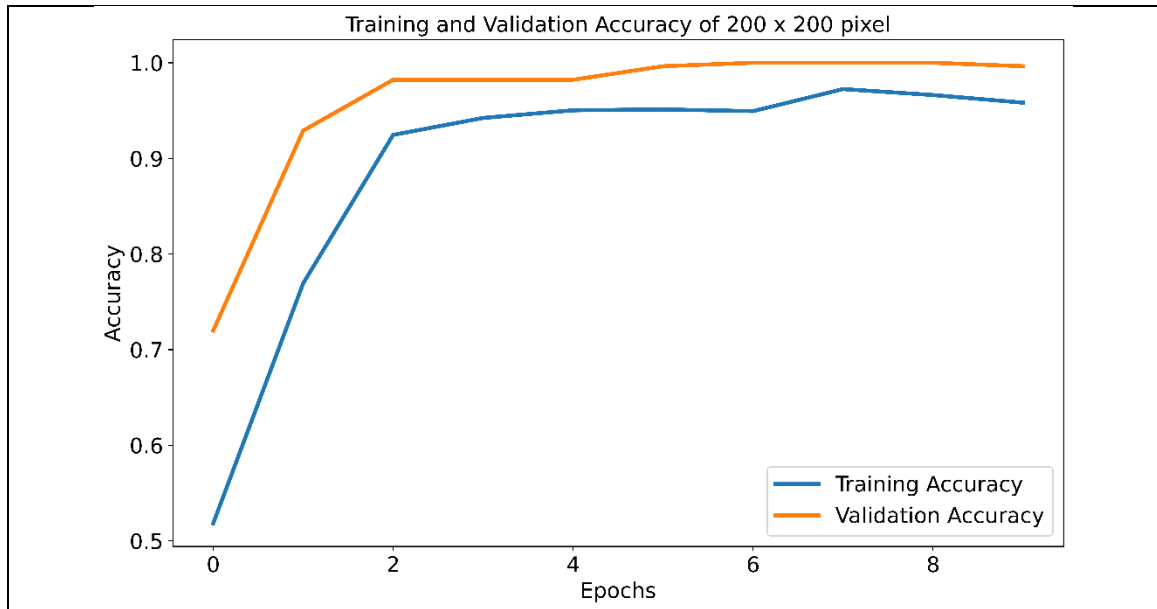




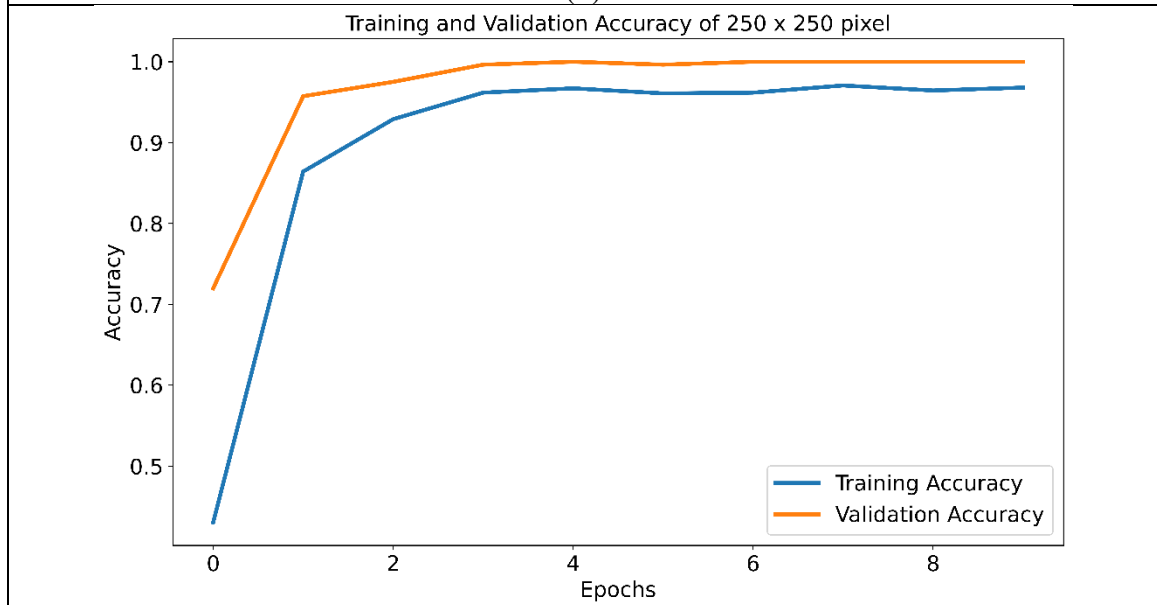
(b)



(c)

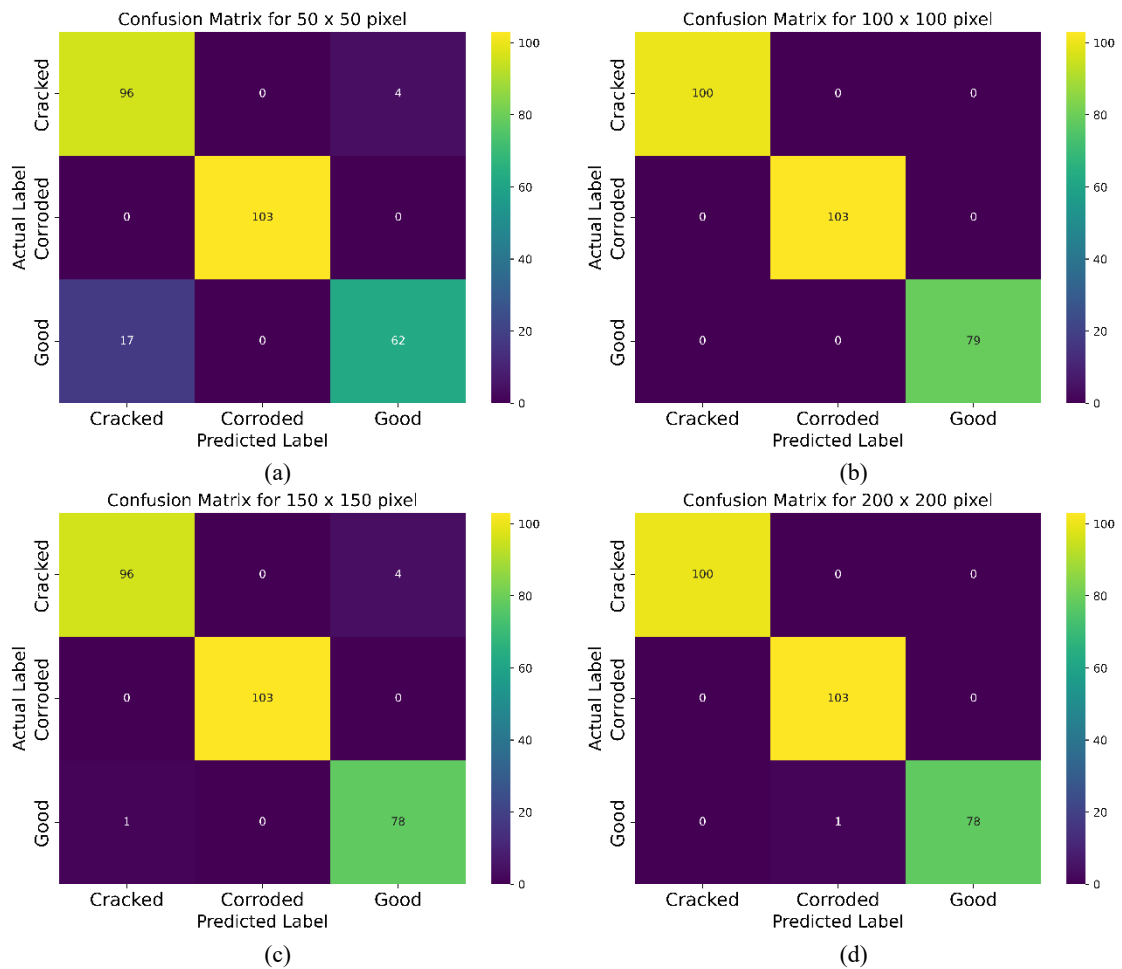
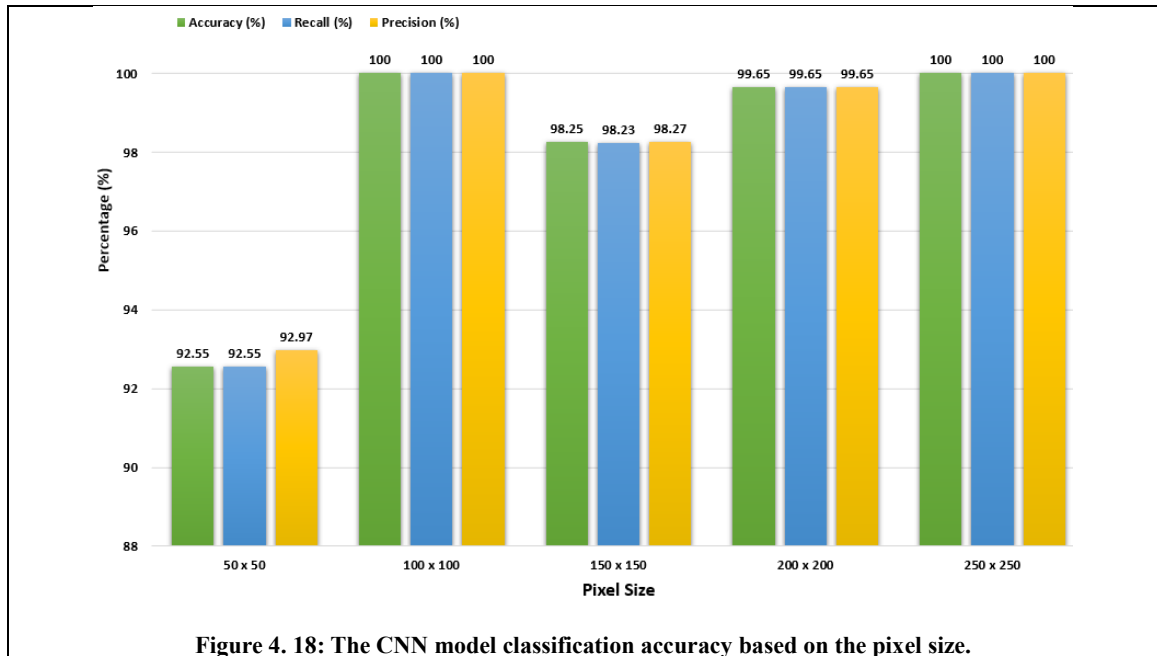


(d)



(e)

Figure 4. 17: Presents the training and validation accuracy based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (e) 250 x 250.



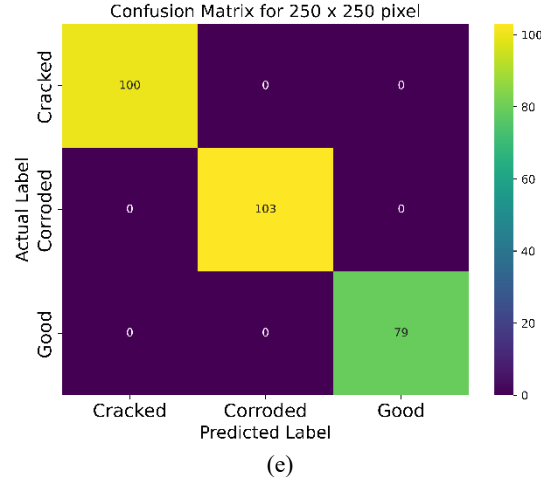


Figure 4. 19: Presents the confusion matrix based on the pixel values for (a) 50 x 50, (b) 100 x 100, (c) 150 x 150, (d) 200 x 200 and (e) 250 x 250.

4.5 Solar Defect Classification Result Discussion

The solar PV defect classification model was developed using the convolutional neural network based on electroluminescence cell image pixel values, ranging from lower to higher resolutions. The training accuracy of the model is illustrated in Figure 4.15, while the training loss is depicted in Figure 4.14. The model's efficacy in classifying test data and the evaluation of accuracy is presented alongside its corresponding confusion matrix in Figures 4.16 and 4.17.

The confusion matrices revealed misclassified electroluminescence cell images at pixel resolutions of 50×50 , 150×150 , and 200×200 . Figure 4.18 illustrates examples of the misclassified classes identified by the CNN model.

The results indicated that a medium pixel size of 100×100 and a high pixel size of 250×250 achieved the highest classification accuracy, with no misclassified electroluminescence images as in Figure 4.17 (b) and (e). While the highest pixel size of

250×250 provided higher-resolution EL cell images, offering more detailed information to the model and enhancing classification accuracy, it posed potential challenges. These include an increased risk of overfitting and a significant rise in computational requirements for training and inference, as each pixel contributes to the calculations. Additionally, high-resolution images are more sensitive to noise, as small artifacts or noise within the data can be amplified, potentially confusing the model.

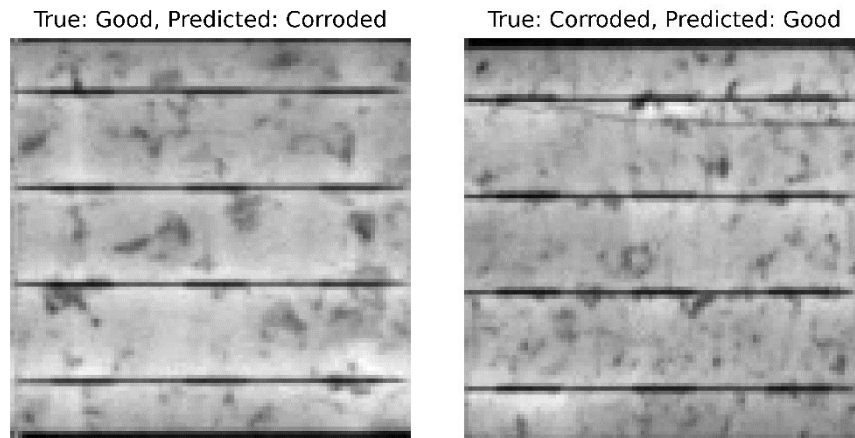


Figure 4. 20: Samples of misclassified classes.

Therefore, a medium pixel size of 100×100 is more suitable due to its lower pixel values, which reduce the size of the input electroluminescence (EL) cell images during training and testing/deployment. This requires fewer computations than higher pixel value EL cell images, improving computational efficiency. Additionally, the medium pixel size (low-resolution images) can mitigate noise, as it contains fewer details for the model to process, making it less sensitive to noisy pixels. However, this pixel size may compromise classification accuracy, particularly for smaller image sizes where the model has fewer features to learn from, potentially overlooking finer details such as microcracks in EL cell images.

4.6 Summary

This chapter presented the development of the NBM along with an evaluation of its performance metrics. The simulation results indicate that the ensemble model outperforms the standalone NBMs, establishing it as the most robust model for fault prediction in gearbox bearings. Similarly, for the solar PV defect classification model, the pixel sizes ranging from 50×50 to 250×250 were utilized to evaluate the model's performance based on the resolution of electroluminescence cell images. This approach aimed to achieve higher classification accuracy by reducing noise and optimizing computational efficiency.

CHAPTER 5

STANDALONE AND ENSEMBLE MODEL FAULT PREDICTION/ALARM GENERATION SYSTEM

5.1 Introduction

This section presents the development of a fault alarm system as a post-processing step following the construction of the standalone and ensemble NBM. Statistical process control, specifically the Shewhart control chart, was implemented to trigger an alarm in response to significant deviations between the ensemble NBM predictions and the measured SCADA data. An experimental setup was also developed to demonstrate the alarm system on hardware using an embedded system.

5.2 Fault Alarm System

Following data pre-processing, feature selection, and constructing an ensemble prediction model using healthy SCADA data from the T06 wind turbine, a Shewhart Control Chart was developed to monitor process control. The SCC provides insights into the condition of the bearing. Figures 5.1 illustrate the MLR and RF NBM prediction with the measured gearbox bearing temperature from the 13th to the 18th of October and the corresponding prediction error. Similarly, Figure 5.2 illustrates DNN standalone and ensemble model bearing temperature predictions alongside the corresponding SCC response for fault prediction within the gearbox bearing during October, when bearing damage was reported

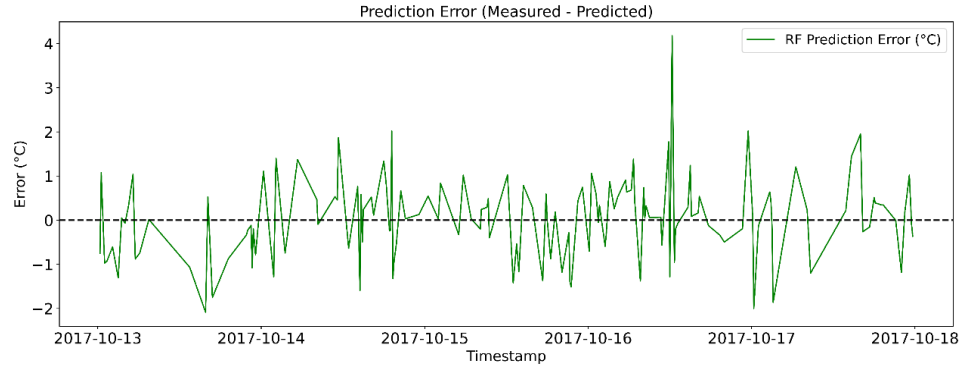
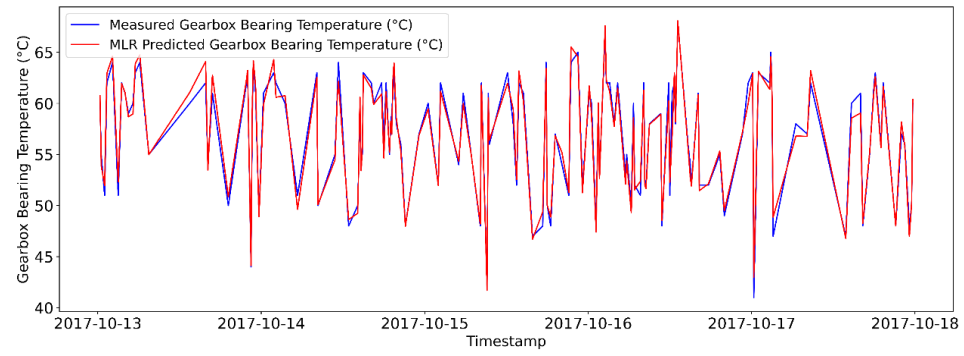
in the maintenance log. Table 5.1 summarizes the standalone and ensemble SCC threshold and fault signal alarms.

5.3 Experimental Setup

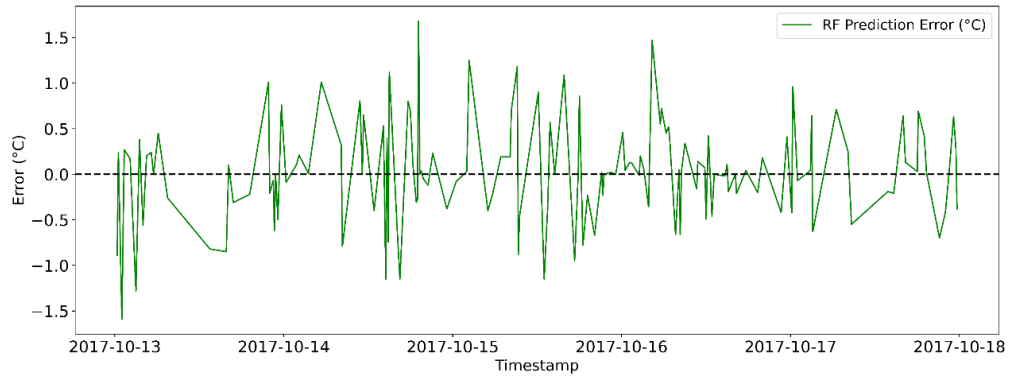
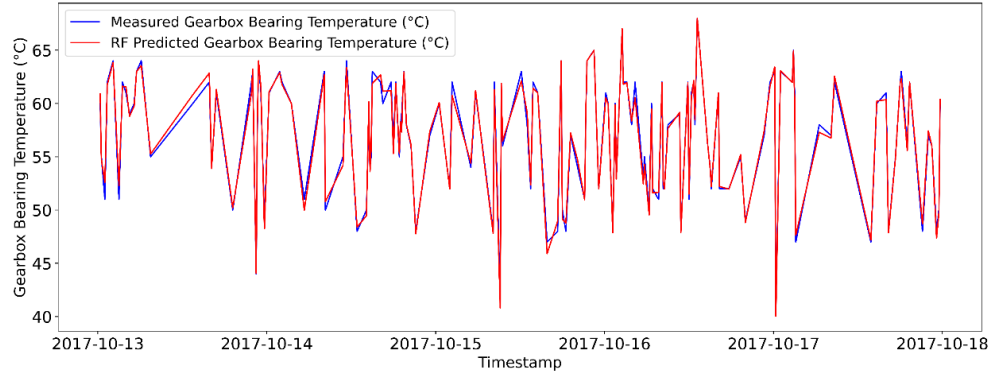
This demonstration simulates the online condition monitoring of the developed fault prediction model (Ensemble NBM) within a computing environment running Windows 10 OS on an Intel Core i9-13900HX processor with 32 GB RAM and an NVIDIA GEFORCE RTX graphics card. Additionally, two Integrated Development Environments (IDEs), Spyder (for Python programming) and Arduino IDE (for Raspberry Pi Pico), were employed to facilitate serial communication using the pyserial module and scikit-learn modules.

Table 5. 1 Standalone and ensemble SCC threshold and fault signal alarm

Prediction Model	SCC Threshold (°C)	Fault Signal
Shallow network	± 1.6054	29
DNN (2 Hidden layer)	± 1.4245	29
DNN (5 Hidden layer)	± 1.4204	22
DNN (10 Hidden layer)	± 1.3769	23
Ensemble	± 1.2962	22

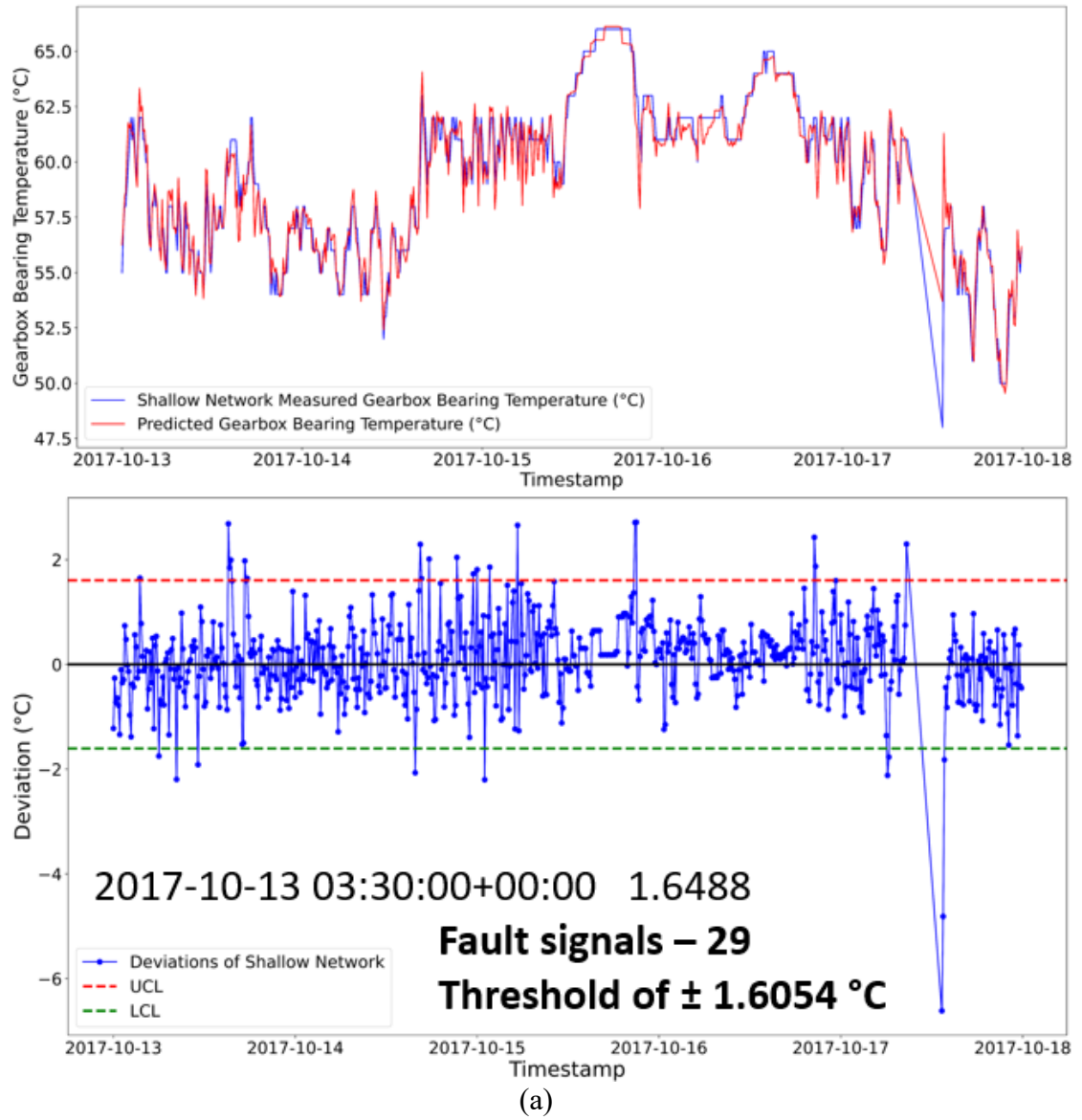


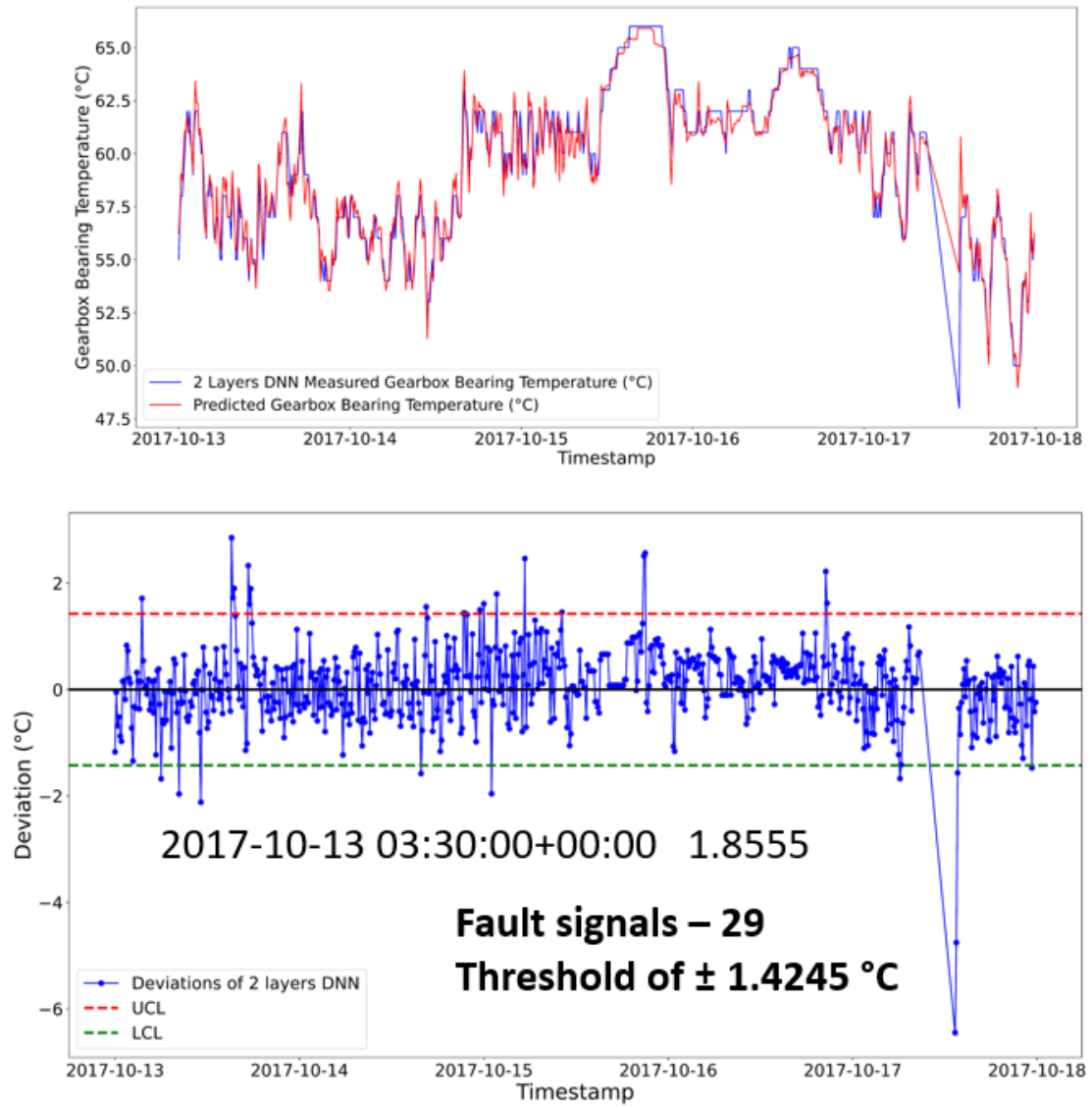
(a)

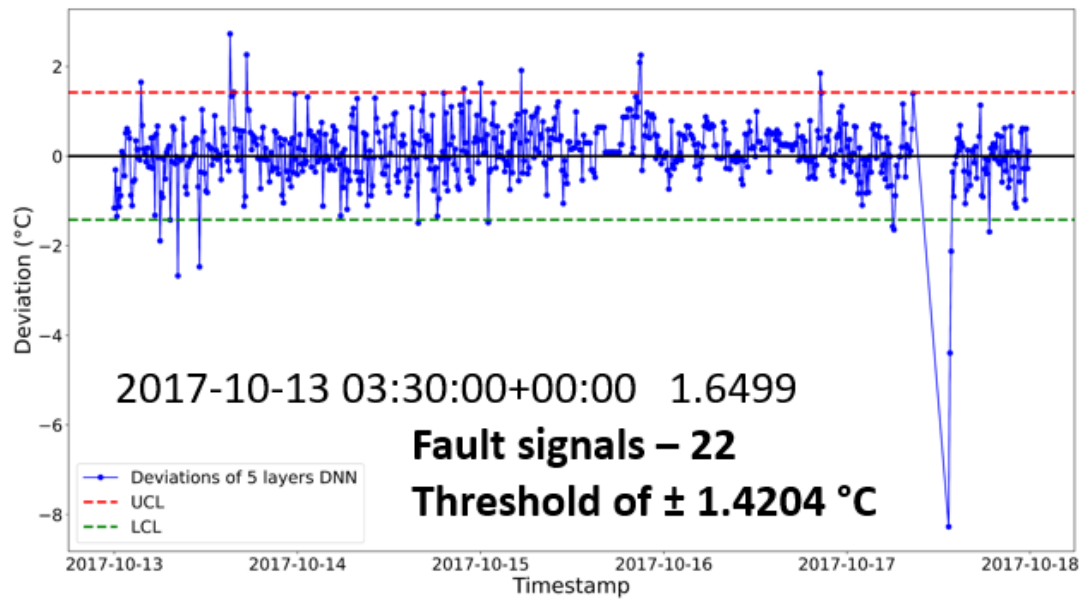
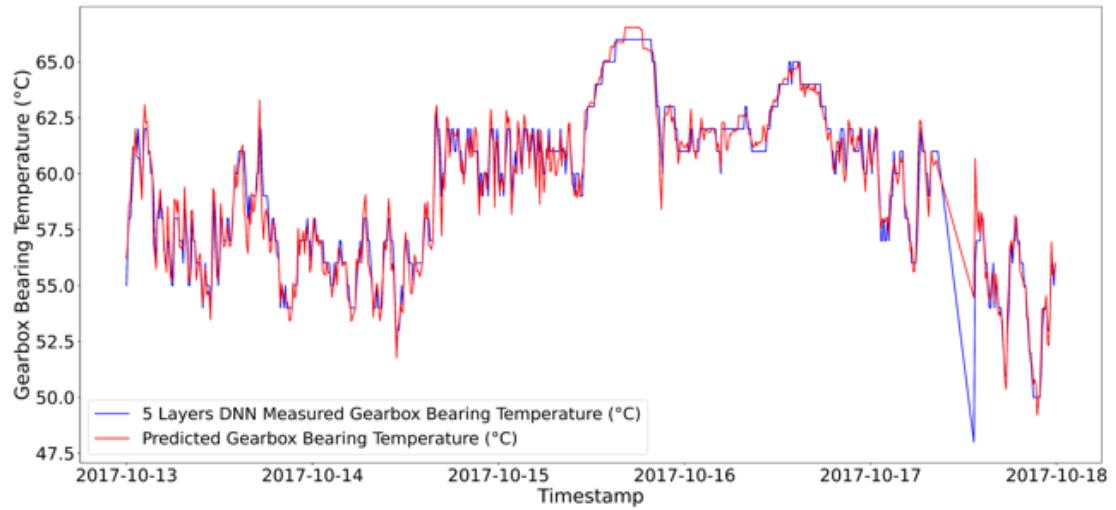


(b)

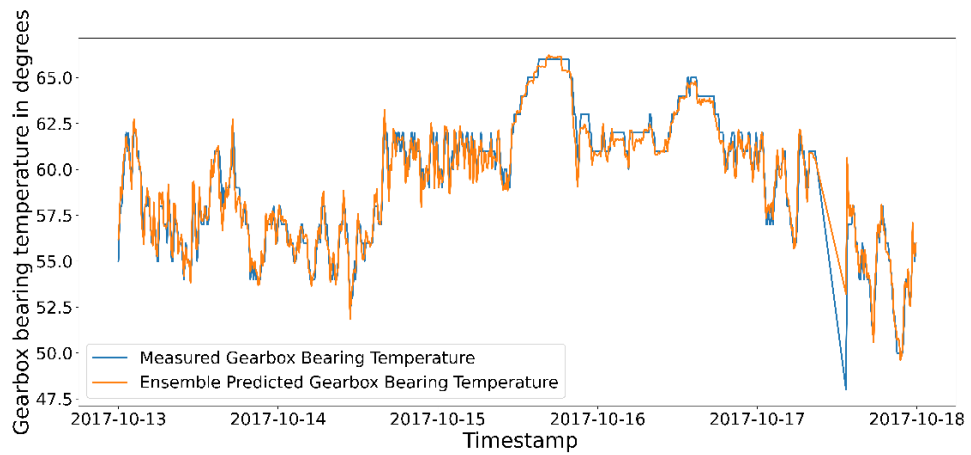
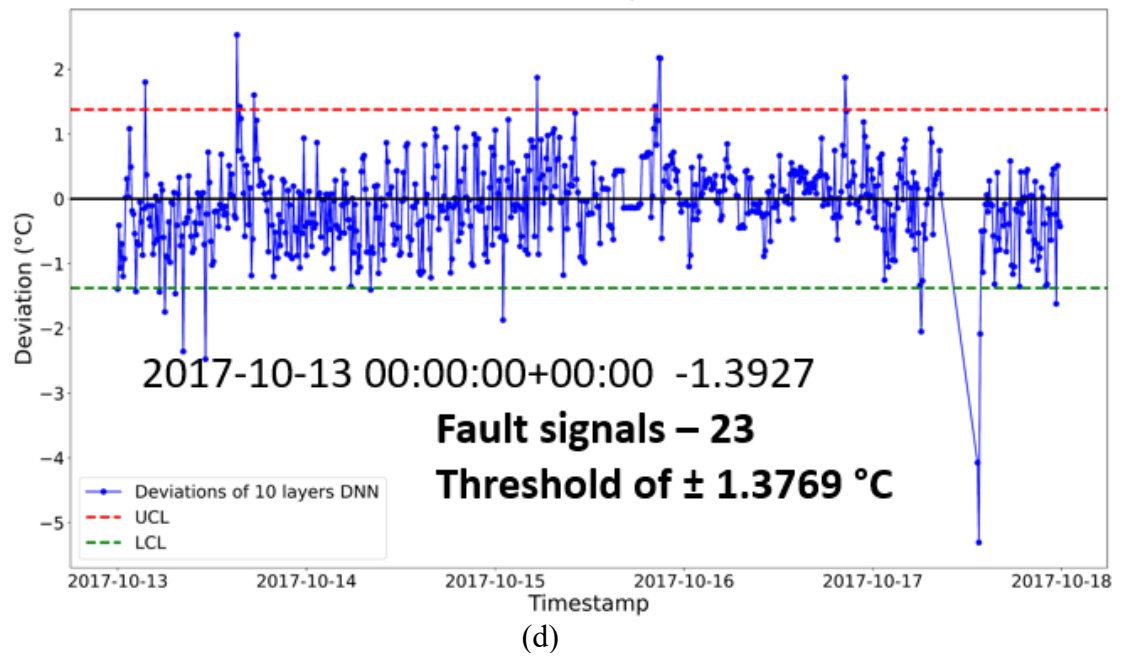
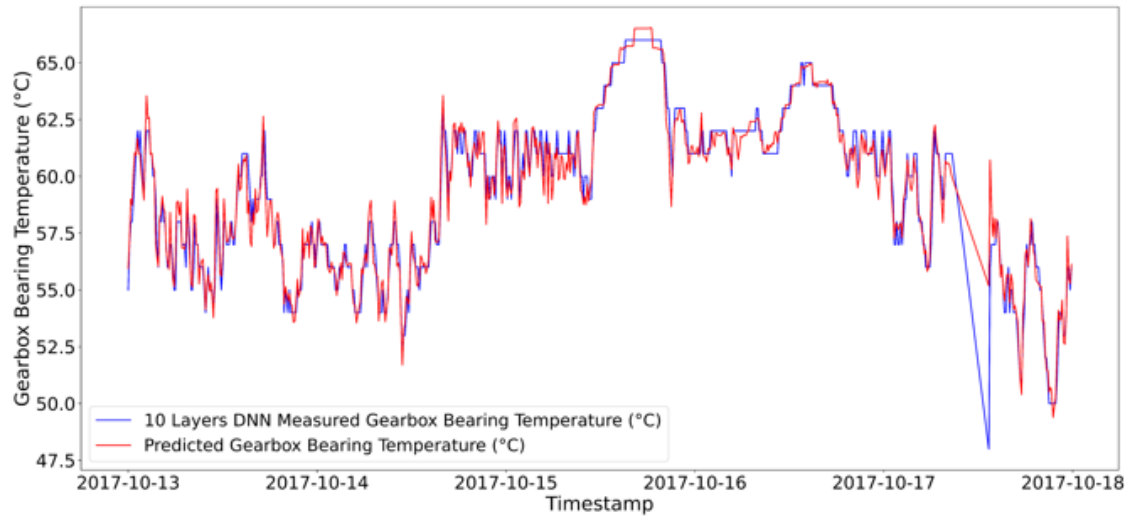
Figure 5. 1: NBM prediction with the measured gearbox bearing temperature from 13th to 18th October and the corresponding prediction error for (a) MLR and (b) RF







(c)



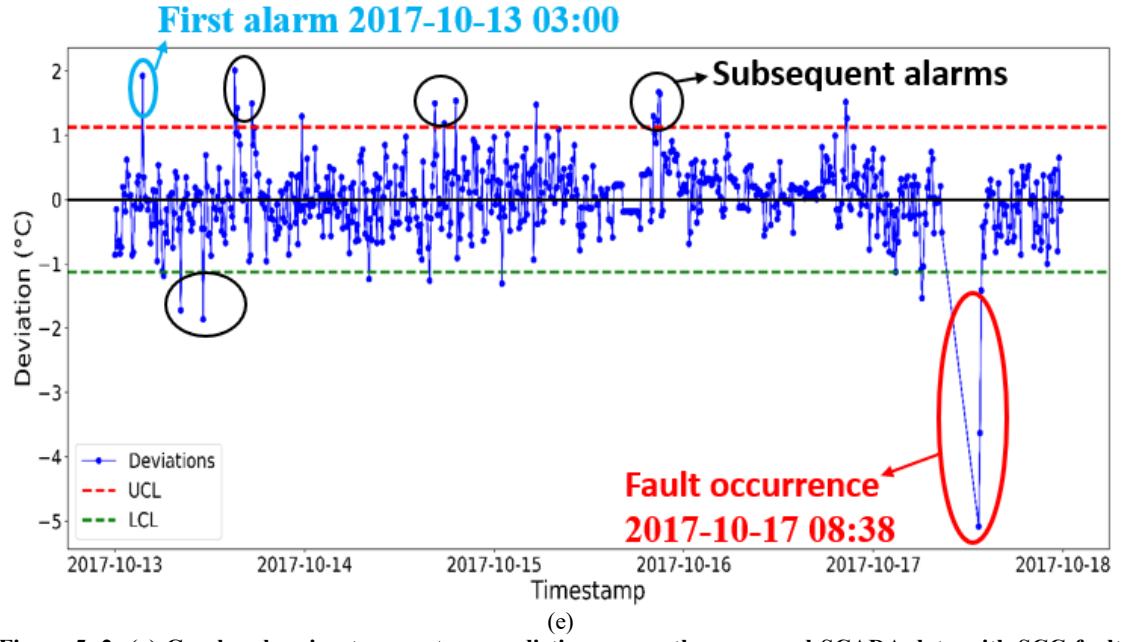


Figure 5. 2: (a) Gearbox bearing temperature prediction versus the measured SCADA data with SCC fault alarm generation for (a) Shallow network, (b) 2 hidden layer network, (c) 5 hidden layer network, (d) 10 hidden layer network and (e) ensemble model.

5.4 Online Condition Monitoring Simulation Components

This experimental setup requires both software and hardware components. The software components include the data acquisition system integrated with the prediction model, while the hardware components consist of a computing device, a microcontroller system (Raspberry Pi Pico) for prediction signal processing, and a buzzer for alarm generation. The experimental setup framework description is illustrated in Figure 5.2.

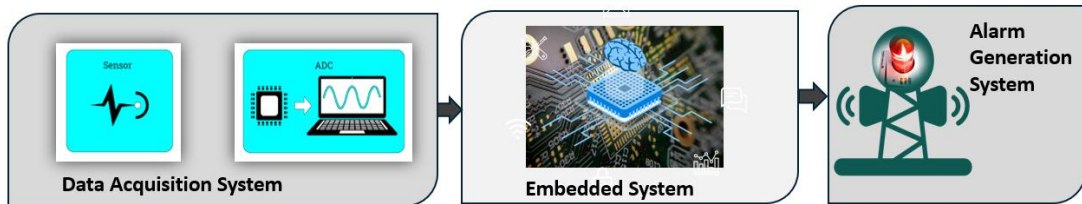


Figure 5. 3: Experimental setup framework

5.4.1 Data Acquisition System

The data utilized for this work was described in Chapter 3; this data is SCADA data that is suitable for machine learning tasks.

5.4.2 Buzzer - PKM22EPP – 40

This piezoelectric sounder is designed for external drive applications and is commonly used on devices like digital watches and electronic calculators. It operates by receiving a signal, typically at 2048 Hz or 4096 Hz, from a Large-scale integration circuit to produce sound [131]. Figure 5.3 presents the piezoelectric buzzer.

5.4.3 Raspberry Pi Pico

Raspberry Pi Pico is a microcontroller board developed by the Raspberry Pi Foundation, featuring the RP2040 microcontroller chip explicitly designed for this board. This makes it ideal for embedded applications and IoT projects [130]. Figure 5.4 is the description of the Raspberry Pi Pico ports.



Figure 5. 4: Piezoelectric buzzer [131]

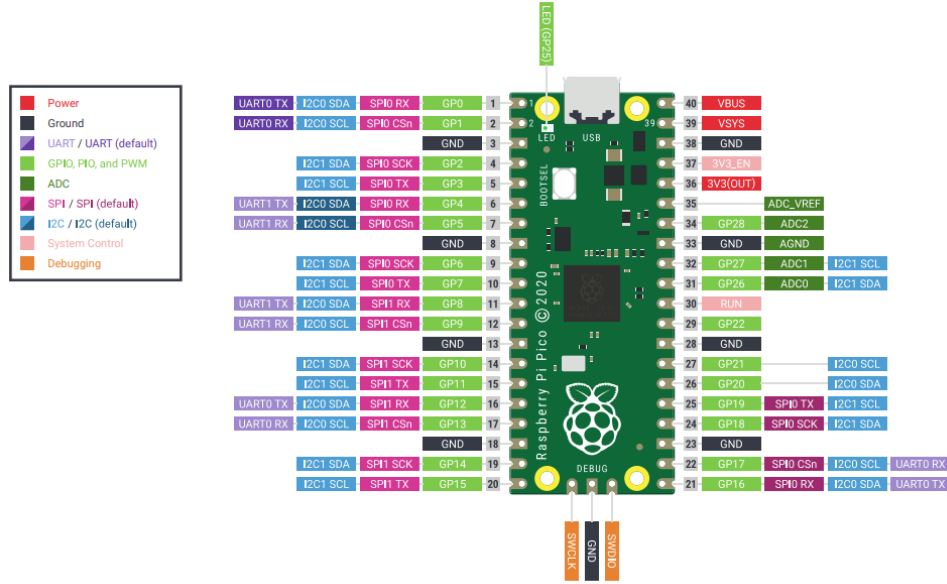


Figure 5. 5: Raspberry Pi Pico ports [130]

5.5 Result Discussion

This study developed three standalone NBMs with their ensemble multi-model approach for gearbox fault prediction, and a Statistical process control technique was incorporated to generate fault alarms. The methodology begins with data acquisition and pre-processing. SCADA data was obtained from an open-source EDP database, and anomalies were identified and removed using the quartile method and International Electrotechnical Commission (IEC) regulations on wind turbine power curve characteristics. Following the data pre-processing, model input feature selection was conducted using three feature selection algorithms: Mutual information selector, Principal component analysis, and Lasso regression. Based on the feature selection results in Table 4.2, the six most relevant features out of 21 average features were selected for training the three NBM models.

Subsequently, an ensemble fault prediction model was developed using a soft-weighted average voting technique. The weights assigned to each base model were based on the performance metrics presented in Table 4.4 for the three individual NBMs: multiple linear regression, random forest, and deep neural network. A softmax function was used to scale the weights between 0 and 1 and to ensure they sum to 1, RF and DNN were assigned the highest weights, as they demonstrated superior performance compared to MLR, with lower error metrics and WIA and R^2 values are almost unity. Additionally, time series plots comparing the predicted gearbox bearing temperature from the ensemble and individual NBMs with the measured gearbox bearing temperature from the SCADA system are shown in Figures (4.7– 4.10). These time series plots corroborate with the performance metrics, demonstrating that the ensemble NBM achieved the highest prediction accuracy among the models.

The multi-model approach enhances the prediction accuracy of the ensemble model, outperforming individual models, mainly due to the advantages of the weighted averaging technique. Furthermore, a fault alarm generation system was developed using statistical process control, specifically through the SCC, to determine the deviations of the predicted values from the actual SCADA measurements through control limits to generate the alarms before the actual occurrence of the failure. According to the maintenance log, the gearbox bearing of the T06 wind turbine system sustained damage on October 17, 2017, at 08:38. The SCC of the ensemble fault prediction model successfully triggered multiple fault-bearing alarms before this failure with a control limit/threshold of $\pm 1.1261^\circ\text{C}$. The first alarm occurred on October 13, 2017, at 03:30, as illustrated in Figure 5.1. Thus, it can be concluded that the proposed data-driven ensemble model strategy effectively generates

predictive alarms before the actual bearing damage, demonstrating the robustness of the developed model. Furthermore, the predictive maintenance framework exhibited high accuracy, signalling potential failures at least 72 hours in advance.

5.6 Summary

This section has discussed the experimental implementation of the ensemble model fault prediction with alarm generation using the Python IDE, Arduino IDE, Raspberry Pi Pico microcontroller, and piezoelectric buzzer. It also discussed the overall NBM result analysis.

CHAPTER 6

CONCLUSION AND RECOMMENDATION

This section presents the summary of the research undertaken in the thesis. It highlights the significant findings from the study and draws relevant conclusions. In addition, the limitations encountered during this research are outlined, and the potential directions for future research in these areas are suggested.

6.1 Summary of the Work

Among renewable energy sources, solar and wind energy are crucial in meeting the increasing global demand for clean energy. The wind turbine gearbox system is one of the critical components of WTs and has been associated with the highest failure rates and downtime in the wind energy system. Therefore, a data-driven condition-monitoring approach was employed to predict faults before they occur, reducing huge operation and maintenance costs and allowing the wind energy system to remain more competitive with traditional energy sources.

Similarly, cracks or corrosion in the solar PV module will degrade solar cell performance by interrupting the power supply in the connected solar cells. These challenges necessitate the development of an intelligent solar PV module defect detection model that will monitor and optimize power generation because manual labor alone is inadequate as the number of modules grows.

The objective of this thesis is to develop an efficacious data-driven ensemble machine-learning model for the predictive maintenance of wind turbine gearbox bearings in wind farms and a solar PV module defect prediction model using CNN. To achieve this, these specific objectives are established: developing an NBM using multiple linear regression, random forest, and deep neural network for wind turbine gearbox bearing temperature prediction; and creating an ensemble NBM for gearbox bearing fault prediction in wind turbines, incorporating an alarm generation system based on statistical process control techniques. The solar PV defect prediction model objectives involve the construction of a deep neural network architecture using a convolutional neural network to predict crack and corrosion within the PV model cells using EL images.

Chapter 2 provides a comprehensive overview of wind turbine predictive maintenance techniques relevant to the thesis. This chapter segments the predictive maintenance techniques into three major categories: gearbox temperature prediction NBMs, ensemble gearbox temperature prediction NBMs, and ensemble NBMs with corresponding statistical process control for fault indication. Several challenges must be addressed to fully exploit predictive maintenance's capabilities, such as predictive model accuracy, model deployment capability, and false fault alarm generation. On solar PV defect prediction, this chapter provides comprehensive past studies of solar PV module defect detection techniques relevant to the thesis. Also, a summary of the recent studies on solar farm PV defect classification was presented.

Chapter 3 defined the data-driven ensemble predictive maintenance framework methodology for the WT system's gearbox bearing. The framework was constructed through four primary steps: data acquisition and pre-processing, input feature selection, ensemble model development and evaluation, and fault alarm generation. A step-by-step methodological framework is presented in the solar farm defect prediction segment, encompassing data acquisition and preprocessing and developing and evaluating the deep learning model.

Chapter 4 discussed the step-by-step NBM construction and evaluation. From the acquired historical SCADA data, outliers, and abnormal data points were filtered out during the pre-processing step using the quartile method and International Electrotechnical Commission regulations on wind turbine power curve characteristics. Following data pre-processing, model input feature selection was conducted using three feature selection algorithms: Mutual Information selector, Principal Component Analysis, and Lasso Regression. Based on the feature selection results, the six most relevant features out of 21 average features were selected for training the three NBMs. Then, the cleaned and filtered SCADA data with the selected features were split into training and testing to validate the developed data-driven ensemble and the individual normal behaviour machine learning models through key statistical performance measures, e.g., R^2 , WIA, MSE, MAE, and RMSE. The model evaluation result shows that error metrics values were relatively low, and the R^2 and WIA values were almost unity for the ensemble fault prediction model, indicating their effectiveness and strengths due to the weighted average formation based on the performance of the individual NBM.

In the context of solar farm defect prediction, the implementation of the step-by-step methodological approach outlined in Chapter 3 is presented. Preprocessing tasks were conducted after acquiring electroluminescence images of solar photovoltaic modules. First, the EL images were normalized to enhance the training performance of the deep learning model, and second, data augmentation was applied to balance the classes of EL cell images before training. A convolutional neural network-based classification model was employed to classify the EL cell images, with pixel resolutions ranging from 50×50 to 250×250 , into good, cracked, and corroded classes. Finally, model performance was evaluated using accuracy, precision, and recall metrics. A discussion of the results is provided, analyzing the performance of each classification model

Chapter 5 presents the development of a fault alarm system as a post-processing step following the construction of the ensemble NBM. A fault alarm generation system was incorporated using a Shewhart control chart to evaluate the models further and to identify deviations between the predicted values from the ensemble model and the actual SCADA measurements through control limits, enabling the generation of alarms prior to the occurrence of failures. The SCC analysis indicated that the developed framework could issue multiple alarms before the occurrence of actual faults, as reported in the wind turbine system maintenance logbook.

In contribution to the existing work, this proposes a data-driven ensemble model strategy that effectively generates predictive alarms before the actual bearing damage, demonstrating the robustness of the developed model. Furthermore, the predictive maintenance framework exhibited high accuracy, successfully signaling potential failures at least 72 hours in advance. This predictive maintenance approach has the potential to

generate substantial cost savings by enhancing the economic viability of wind turbine system operations. Early fault detection gives wind farm operators adequate time to implement corrective actions, allowing for timely repair or replacement of faulty gearboxes. This proactive maintenance strategy has the potential to minimize unnecessary costs and reduce downtime effectively.

6.2 Limitations and Recommendations for Future Work

This study has contributed to a comprehensive understanding and practical application of predictive maintenance for wind turbine gearboxes and solar farm photovoltaic modules. However, certain factors may have constrained the scope and depth of the findings. This section highlights the limitations and provides recommendations for future thesis research.

6.2.1 Wind Turbine Gearbox Fault Prediction

- The NBMs developed for the bearing temperature prediction for this thesis rely on labeled data used for the supervised machine learning model, while access to labeled data is generally limited. Future research could explore methods to build models that perform effectively with unlabeled data, such as unsupervised learning approaches.
- In terms of generalizability of the developed predictive model, this study relied on SCADA data from a single wind turbine (T06) within the EDP wind turbine system. Consequently, the model's effectiveness and generalizability to other wind turbines or wind farms remain uncertain. Therefore, there is a need further to test the model deployment on edge devices for faster local inference

- The accuracy of the ensemble prediction model is essential for reliable fault prediction. A low-accuracy model may cause false or missed alarms, thus affecting the operational efficiency and economic viability of wind farms. Future work should focus on enhancing model accuracy to minimize false and missed alarms, ensuring the practical effectiveness of predictive maintenance systems.
- In a data-driven machine learning application, data preprocessing is critical in effective data cleaning. This study applied specific preprocessing methods; however, future research could explore additional preprocessing techniques for outlier elimination, including integrating physics-based models for data quality and model robustness.

6.2.2 Solar Farm Photovoltaic Modules Defect Prediction

- The acquisition of photovoltaic module data is crucial for developing an efficient defect prediction model. Future research could focus on expanding the dataset, particularly by including PV modules subjected to damp heat exposure ranging from 500 hours to 3,000 hours, especially the defect samples, to enhance model robustness and generalizability.
- In terms of the preprocessing task, more techniques, such as generative adversarial networks, could be explored to synthesize complex defect EL images such as micro-crack.
- This thesis investigated the capability of convolutional neural networks as feature extractors for the electroluminescence cell images. Future research could explore

the potential of pre-trained models, such as VGG and ResNet, for feature extraction from EL cell images to further enhance the defect prediction framework.

REFERENCES

- [1] J. S. Fada, M. A. Hossain, J. L. Braid, S. Yang, T. J. Peshek, and R. H. French, “Electroluminescent Image Processing and Cell Degradation Type Classification via Computer Vision and Statistical Learning Methodologies,” in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, IEEE, Jun. 2017, pp. 3456–3461. doi: 10.1109/PVSC.2017.8366291.
- [2] H. Jie, I. Khan, M. Alharthi, M. W. Zafar, and A. Saeed, “Sustainable energy policy, socio-economic development, and ecological footprint: The economic significance of natural resources, population growth, and industrial development,” *Util Policy*, vol. 81, p. 101490, Apr. 2023, doi: 10.1016/J.JUP.2023.101490.
- [3] T. X. Wiesli and T. Hammer, “How can Quality of Life be Achieved in a Sustainable Way? Perceptions of Swiss Rural Inhabitants,” *Discover Sustainability*, vol. 3, no. 1, pp. 1–17, Dec. 2022, doi: 10.1007/S43621-022-00114-6/TABLES/7.
- [4] K. Abbass, M. Z. Qasim, H. Song, M. Murshed, H. Mahmood, and I. Younis, “A review of the global climate change impacts, adaptation, and sustainable mitigation measures,” *Environmental Science and Pollution Research* 2022 29:28, vol. 29, no. 28, pp. 42539–42559, Apr. 2022, doi: 10.1007/S11356-022-19718-6.
- [5] M. Shafiullah, S. D. Ahmed, and F. A. Al-Sulaiman, “Grid Integration Challenges and Solution Strategies for Solar PV Systems: A Review,” *IEEE Access*, vol. 10, pp. 52233–52257, 2022, doi: 10.1109/ACCESS.2022.3174555.

- [6] S. D. Ahmed, F. S. M. Al-Ismail, M. Shafiullah, F. A. Al-Sulaiman, and I. M. El-Amin, “Grid Integration Challenges of Wind Energy: A Review,” *IEEE Access*, vol. 8, pp. 10857–10878, 2020, doi: 10.1109/ACCESS.2020.2964896.
- [7] “REN21. 2024. Renewables 2024 Global Status Report Collection: Energy Supply (Paris: REN21 Secretariat).,” (Paris: REN21 Secretariat)., 2024. Accessed: Jul. 04, 2024. [Online]. Available: <http://www.ren21.net/gsr-2024>
- [8] J. Carroll, A. McDonald, and D. McMillan, “Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines,” *Wind Energy*, vol. 19, no. 6, pp. 1107–1119, Jun. 2016, doi: 10.1002/we.1887.
- [9] Y. Vidal, B. Puruncajas, F. Castellani, and C. Tutivén, “Predictive maintenance of wind turbine’s main bearing using wind farm SCADA data and LSTM neural networks,” in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2507/1/012024.
- [10] Andrea Devenish, “Data Analytics for predictive maintenance of Wind Turbines.,” Master of Engineering (Mechatronic) , Stellenbosch University, 2021. Accessed: Jun. 05, 2024. [Online]. Available: <https://scholar.sun.ac.za/items/3619871c-c925-436c-a332-3a33c938243c>
- [11] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, and L. Rambault, “Gearbox condition monitoring in wind turbines: A review,” *Mech Syst Signal Process*, vol. 111, pp. 251–264, Oct. 2018, doi: 10.1016/j.ymssp.2018.03.052.

- [12] J. Keller, S. Sheng, J. Cotrell, and A. Greco, “Wind Turbine Drivetrain Reliability Collaborative Workshop: A Recap,” Golden, CO (United States), Aug. 2016. doi: 10.2172/1314863.
- [13] Feng Yanhui, Qiu Yingning, D. Infield, Li Jiawei, and Yang Wenxian, “Study on Order Analysis for Condition Monitoring Wind Turbine Gearbox,” in *3rd Renewable Power Generation Conference (RPG 2014)*, Institution of Engineering and Technology, 2014, pp. 8.19-8.19. doi: 10.1049/cp.2014.0904.
- [14] E. Hart, A. Turnbull, J. Feuchtwang, D. McMillan, E. Golysheva, and R. Elliott, “Wind turbine main-bearing loading and wind field characteristics,” *Wind Energy*, vol. 22, no. 11, pp. 1534–1547, Nov. 2019, doi: 10.1002/we.2386.
- [15] O. I. Owolabi, N. Madushele, P. A. Adedeji, and O. O. Olatunji, “FEM and ANN approaches to wind turbine gearbox monitoring and diagnosis: a mini review,” *J Reliab Intell Environ*, vol. 9, no. 4, pp. 399–419, Dec. 2023, doi: 10.1007/s40860-022-00183-4.
- [16] Y. and P. A. and S. R. Malik Hasmat and Pandya, “Feature Extraction Using EMD and Classifier Through Artificial Neural Networks for Gearbox Fault Diagnosis,” in *Applications of Artificial Intelligence Techniques in Engineering*, S. and S. Y. R. and A. A. Malik Hasmat and Srivastava, Ed., Singapore: Springer Singapore, 2019, pp. 309–317.
- [17] E. Artigao, S. Martín-Martínez, A. Honrubia-Escribano, and E. Gómez-Lázaro, “Wind turbine reliability: A comprehensive review towards effective condition

- monitoring development,” *Appl Energy*, vol. 228, pp. 1569–1583, Oct. 2018, doi: 10.1016/j.apenergy.2018.07.037.
- [18] S. Sheng *et al.*, “Wind Turbine Drivetrain Condition Monitoring During GRC Phase 1 and Phase 2 Testing,” Golden, CO (United States), Oct. 2011. doi: 10.2172/1027157.
- [19] I. Renewable Energy Agency, *RENEWABLE POWER GENERATION COSTS IN 2023*. 2024. [Online]. Available: www.irena.org
- [20] T. Stehly, P. Duffy, and D. Mulas Hernando, “2022 Cost of Wind Energy Review,” Golden, CO (United States), Dec. 2023. doi: 10.2172/2278805.
- [21] International Renewable Energy Agency, *FUTURE OF SOLAR PHOTOVOLTAIC Deployment, investment, technology, grid integration, and socio-economic aspects, A Global Energy Transformation paper About IRENA*, 2019. Accessed: Apr. 30, 2024. [Online]. Available: www.irena.org/publications
- [22] REN21, “Renewables 2023 Global Status Report Energy Supply Collection,” 2023.
- [23] A. Herez, M. Ramadan, B. Abdulhay, and M. Khaled, “Short review on solar energy systems,” 2016, p. 030041. doi: 10.1063/1.4959437.
- [24] REN21, “Renewables 2021 Global Status Report (Paris: REN21 Secretariat),,” 2021.
- [25] L. Pratt, D. Govender, and R. Klein, “Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic

- segmentation,” *Renew Energy*, vol. 178, pp. 1211–1222, Nov. 2021, doi: 10.1016/j.renene.2021.06.086.
- [26] IRENA, *Renewable capacity statistics 2024*, International Renewable Energy Agency, Abu Dhabi. 2024. [Online]. Available: www.irena.org
- [27] N. Renewable Energy Laboratory, S. National Laboratory, S. Alliance, and S. National Laboratory Multiyear Partnership, “Best Practices for Operation and Maintenance of Photovoltaic and Energy Storage Systems; 3rd Edition,” 2018. [Online]. Available: www.nrel.gov/publications.
- [28] H. Wang, W. Zhang, D. Yang, and Y. Xiang, “Deep-Learning-Enabled Predictive Maintenance in Industrial Internet of Things: Methods, Applications, and Challenges,” *IEEE Syst J*, vol. 17, no. 2, pp. 2602–2615, Jun. 2023, doi: 10.1109/JSYST.2022.3193200.
- [29] J. Ramirez-Vergara, L. B. Bosman, E. Wollega, and W. D. Leon-Salas, “Review of forecasting methods to support photovoltaic predictive maintenance,” *Clean Eng Technol*, vol. 8, p. 100460, Jun. 2022, doi: 10.1016/j.clet.2022.100460.
- [30] C.-Y. Lin, Y.-M. Hsieh, F.-T. Cheng, H.-C. Huang, and M. Adnan, “Time Series Prediction Algorithm for Intelligent Predictive Maintenance,” *IEEE Robot Autom Lett*, vol. 4, no. 3, pp. 2807–2814, Jul. 2019, doi: 10.1109/LRA.2019.2918684.
- [31] H. F. Mateo Romero *et al.*, “Optimized estimator of the output power of PV cells using EL images and I–V curves,” *Solar Energy*, vol. 265, Nov. 2023, doi: 10.1016/j.solener.2023.112089.

- [32] A. K. Acharya, P. K. Sahu, and S. R. Jena, “Deep neural network based approach for detection of defective solar cell,” in *Materials Today: Proceedings*, Elsevier Ltd, 2019, pp. 2009–2014. doi: 10.1016/j.matpr.2020.09.048.
- [33] R. Fu, D. Feldman, R. Margolis, M. Woodhouse, and K. Ardani, “U.S. Solar Photovoltaic System Cost Benchmark: Q1 2017,” Sep. 2017. doi: 10.2172/1395932.
- [34] L. Hernández-Callejo, S. Gallardo-Saavedra, and V. Alonso-Gómez, “A review of photovoltaic systems: Design, operation and maintenance,” *Solar Energy*, vol. 188, pp. 426–440, Aug. 2019, doi: 10.1016/j.solener.2019.06.017.
- [35] N. G. Dhere, “Reliability of PV modules and balance-of-system components,” in *Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005.*, IEEE, pp. 1570–1576. doi: 10.1109/PVSC.2005.1488445.
- [36] D. C. Jordan and S. R. Kurtz, “Photovoltaic Degradation Rates -- An Analytical Review: Preprint,” 2012. [Online]. Available: <http://www.osti.gov/bridge>
- [37] R. Pandit, D. Astolfi, J. Hong, D. Infield, and M. Santos, “SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends,” Apr. 01, 2023, *SAGE Publications Inc.* doi: 10.1177/0309524X221124031.
- [38] P. W. Khan and Y.-C. Byun, “A Review of machine learning techniques for wind turbine’s fault detection, diagnosis, and prognosis,” *Int J Green Energy*, vol. 21, no. 4, pp. 771–786, Mar. 2024, doi: 10.1080/15435075.2023.2217901.

- [39] Z. Liu and L. Zhang, “A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings,” *Measurement*, vol. 149, p. 107002, Jan. 2020, doi: 10.1016/j.measurement.2019.107002.
- [40] M. Molęda, B. Małysiak-Mrozek, W. Ding, V. Sunderam, and D. Mrozek, “From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry,” *Sensors 2023, Vol. 23, Page 5970*, vol. 23, no. 13, p. 5970, Jun. 2023, doi: 10.3390/S23135970.
- [41] P. Nunes, J. Santos, and E. Rocha, “Challenges in predictive maintenance – A review,” *CIRP J Manuf Sci Technol*, vol. 40, pp. 53–67, Feb. 2023, doi: 10.1016/J.CIRPJ.2022.11.004.
- [42] A. D. Scaife, “Improve predictive maintenance through the application of artificial intelligence: A systematic review,” *Results in Engineering*, vol. 21, p. 101645, Mar. 2024, doi: 10.1016/J.RINENG.2023.101645.
- [43] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, “Short-Term Wind-Power Prediction Based on Wavelet Transform–Support Vector Machine and Statistic-Characteristics Analysis,” *IEEE Trans Ind Appl*, vol. 48, no. 4, pp. 1136–1141, Jul. 2012, doi: 10.1109/TIA.2012.2199449.
- [44] K. Lee, S. Im, and B. Lee, “Prediction of renewable energy hosting capacity using multiple linear regression in KEPCO system,” *Energy Reports*, vol. 9, pp. 343–347, Nov. 2023, doi: 10.1016/j.egy.2023.09.121.

- [45] M. Garan, K. Tidriri, and I. Kovalenko, “A Data-Centric Machine Learning Methodology: Application on Predictive Maintenance of Wind Turbines,” *Energies (Basel)*, vol. 15, no. 3, p. 826, Jan. 2022, doi: 10.3390/en15030826.
- [46] T. Yuan, Z. Sun, and S. Ma, “Gearbox Fault Prediction of Wind Turbines Based on a Stacking Model and Change-Point Detection,” *Energies (Basel)*, vol. 12, no. 22, p. 4224, Nov. 2019, doi: 10.3390/en12224224.
- [47] F. Elasha, S. Shanbr, X. Li, and D. Mba, “Prognosis of a wind turbine gearbox bearing using supervised machine learning,” *Sensors (Switzerland)*, vol. 19, no. 14, Jul. 2019, doi: 10.3390/s19143092.
- [48] L.-K. Song, C.-W. Fei, G.-C. Bai, and L.-C. Yu, “Dynamic neural network method-based improved PSO and BR algorithms for transient probabilistic analysis of flexible mechanism,” *Advanced Engineering Informatics*, vol. 33, pp. 144–153, Aug. 2017, doi: 10.1016/j.aei.2017.05.005.
- [49] A. A. Chojaczyk, A. P. Teixeira, L. C. Neves, J. B. Cardoso, and C. Guedes Soares, “Review and application of Artificial Neural Networks models in reliability analysis of steel structures,” *Structural Safety*, vol. 52, pp. 78–89, Jan. 2015, doi: 10.1016/j.strusafe.2014.09.002.
- [50] X. Xue, Y. Li, X. Yang, X. Chen, and J. Xiang, “Prediction of Slope Stability Based on GA-BP Hybrid Algorithm,” *Neural Network World*, vol. 25, no. 2, pp. 189–202, 2015, doi: 10.14311/NNW.2015.25.010.

- [51] P. A. Adedeji, O. O. Olatunji, N. Madushele, Z. Z. Rasmeni, and N. Janse van Rensburg, “Effectiveness of evolutionary-tuned neurofuzzy inference system in predicting wind turbine gearbox oil temperature,” *Mater Today Proc*, Aug. 2023, doi: 10.1016/j.matpr.2023.08.034.
- [52] X. Bai *et al.*, “Wind turbine gearbox oil temperature feature extraction and condition monitoring based on energy flow,” *Appl Energy*, vol. 371, p. 123687, Oct. 2024, doi: 10.1016/j.apenergy.2024.123687.
- [53] A. Verma, D. Zappalá, S. Sheng, and S. J. Watson, “Wind turbine gearbox fault prognosis using high-frequency SCADA data,” in *Journal of Physics: Conference Series*, Institute of Physics, Jun. 2022. doi: 10.1088/1742-6596/2265/3/032067.
- [54] F. Castellani, F. Natili, D. Astolfi, and Y. Vidal, “Wind turbine gearbox condition monitoring through the sequential analysis of industrial SCADA and vibration data,” *Energy Reports*, vol. 12, pp. 750–761, Dec. 2024, doi: 10.1016/j.egyr.2024.06.041.
- [55] H. R. F. Fotso, C. V. A. Kazé, and G. D. Kenmoé, “Real-time rolling bearing power loss in wind turbine gearbox modeling and prediction based on calculations and artificial neural network,” *Tribol Int*, vol. 163, p. 107171, Nov. 2021, doi: 10.1016/j.triboint.2021.107171.
- [56] G. Yan, C. Yu, and Y. Bai, “Wind Turbine Bearing Temperature Forecasting Using a New Data-Driven Ensemble Approach,” *Machines*, vol. 9, no. 11, p. 248, Oct. 2021, doi: 10.3390/machines9110248.

- [57] Q. Zhao, K. Bao, J. Wang, Y. Han, and J. Wang, “An Online Hybrid Model for Temperature Prediction of Wind Turbine Gearbox Components,” *Energies (Basel)*, vol. 12, no. 20, p. 3920, Oct. 2019, doi: 10.3390/en12203920.
- [58] Z. Cheng, Y. Deng, X. Wang, and Z. Xie, “A case study of industrial data analysis: Gearbox temperature prediction of wind turbines using ensemble deep learning regression,” in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, IEEE, Mar. 2021, pp. 1006–1009. doi: 10.1109/ICBAIE52039.2021.9390053.
- [59] K. Zhuang, C. Ma, H.-F. Lam, L. Zou, and J. Hu, “Pre-trained 1DCNN-BiLSTM Hybrid Network for Temperature Prediction of Wind Turbine Gearboxes,” *Processes*, vol. 11, no. 12, p. 3324, Nov. 2023, doi: 10.3390/pr11123324.
- [60] L. Campoverde, C. Tutivén, Y. Vidal, and C. Benaláazar-Parra, “SCADA Data-Driven Wind Turbine Main Bearing Fault Prognosis Based on Principal Component Analysis,” in *Journal of Physics: Conference Series*, Institute of Physics, Jun. 2022. doi: 10.1088/1742-6596/2265/3/032107.
- [61] D. Menezes, M. Mendes, J. A. Almeida, and T. Farinha, “Wind Farm and Resource Datasets: A Comprehensive Survey and Overview,” *Energies (Basel)*, vol. 13, no. 18, p. 4702, Sep. 2020, doi: 10.3390/en13184702.
- [62] P. Sun, J. Li, C. Wang, and X. Lei, “A generalized model for wind turbine anomaly identification based on SCADA data,” *Appl Energy*, vol. 168, pp. 550–567, Apr. 2016, doi: 10.1016/j.apenergy.2016.01.133.

- [63] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, “Wind Turbine Gearbox Failure Identification With Deep Neural Networks,” *IEEE Trans Industr Inform*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017, doi: 10.1109/TII.2016.2607179.
- [64] Y. Wang, Z. Zhu, H. Song, and K. Shi, “Wind turbine gearbox condition monitoring based on extreme gradient boosting,” in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, IEEE, Oct. 2017, pp. 6017–6023. doi: 10.1109/IECON.2017.8217044.
- [65] W. Udo and Y. Muhammad, “Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data,” *IEEE Access*, vol. 9, pp. 162370–162388, 2021, doi: 10.1109/ACCESS.2021.3132684.
- [66] S.-H. Lee and K.-P. Park, “Development of a Prediction Model for the Gear Whine Noise of Transmission Using Machine Learning,” *International Journal of Precision Engineering and Manufacturing*, vol. 24, no. 10, pp. 1793–1803, Oct. 2023, doi: 10.1007/s12541-023-00845-0.
- [67] Y. Wang, T. Chen, R. Zou, D. Song, F. Zhang, and L. Zhang, “Ensemble probabilistic wind power forecasting with multi-scale features,” *Renew Energy*, vol. 201, pp. 734–751, Dec. 2022, doi: 10.1016/j.renene.2022.10.122.
- [68] M. Beretta, Y. Vidal, J. Sepulveda, O. Porro, and J. Cusidó, “Improved Ensemble Learning for Wind Turbine Main Bearing Fault Diagnosis,” *Applied Sciences*, vol. 11, no. 16, p. 7523, Aug. 2021, doi: 10.3390/app11167523.

- [69] Y. Zhu, C. Zhu, J. Tan, Y. Tan, and L. Rao, “Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning,” *Renew Energy*, vol. 189, pp. 90–103, Apr. 2022, doi: 10.1016/j.renene.2022.02.061.
- [70] C. F. de Lima Munguba *et al.*, “Ensemble learning framework for fleet-based anomaly detection using wind turbine drivetrain components vibration data.,” *Eng Appl Artif Intell*, vol. 133, p. 108363, Jul. 2024, doi: 10.1016/j.engappai.2024.108363.
- [71] H. Zhang, H. Niu, Z. Ma, and S. Zhang, “Wind Turbine Condition Monitoring Based on Bagging Ensemble Strategy and KNN Algorithm,” *IEEE Access*, vol. 10, pp. 93412–93420, 2022, doi: 10.1109/ACCESS.2022.3164717.
- [72] D. Wu *et al.*, “A novel online framework for gear machining quality prediction based on ensemble deep regression,” *Measurement*, vol. 201, p. 111716, Sep. 2022, doi: 10.1016/j.measurement.2022.111716.
- [73] Z. Wang *et al.*, “Early anomaly detection of wind turbine gearbox based on SLFormer neural network,” *Ocean Engineering*, vol. 311, p. 118925, Nov. 2024, doi: 10.1016/j.oceaneng.2024.118925.
- [74] L. Campoverde-Vilela, M. C. Feijóo, Y. Vidal, J. Sampietro, and C. Tutivén, “Anomaly-based fault detection in wind turbine main bearings,” *Wind Energy Science*, vol. 8, no. 4, pp. 557–574, 2023, doi: 10.5194/wes-8-557-2023.
- [75] M. Elforjani, “Diagnosis and prognosis of real world wind turbine gears,” *Renew Energy*, vol. 147, pp. 1676–1693, Mar. 2020, doi: 10.1016/j.renene.2019.09.109.

- [76] J. Carroll, S. Koukoura, A. McDonald, A. Charalambous, S. Weiss, and S. McArthur, “Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques,” *Wind Energy*, vol. 22, no. 3, pp. 360–375, Mar. 2019, doi: 10.1002/we.2290.
- [77] A. Maroof Karimi *et al.*, “Automated Pipeline for Photovoltaic Module Electroluminescence Image Processing and Degradation Feature Classification,” *IEEE J Photovolt*, vol. 9, no. 5, 2019, doi: 10.17605/OSF.IO/QT5HJ.
- [78] E. Sovetkin and A. Steland, “Automatic processing and solar cell detection in photovoltaic electroluminescence images,” *Integr Comput Aided Eng*, vol. 26, no. 2, pp. 123–137, Feb. 2019, doi: 10.3233/ICA-180588.
- [79] C. Mantel *et al.*, “Machine learning prediction of defect types for electroluminescence images of photovoltaic panels,” in *Applications of Machine Learning*, M. E. Zelinski, T. M. Taha, J. Howe, A. A. Awwal, and K. M. Iftekharruddin, Eds., SPIE, Sep. 2019, p. 1. doi: 10.1117/12.2528440.
- [80] S. Deitsch *et al.*, “Automatic classification of defective photovoltaic module cells in electroluminescence images,” *Solar Energy*, vol. 185, pp. 455–468, Jun. 2019, doi: 10.1016/j.solener.2019.02.067.
- [81] K. G. Bedrich, M. Bliss, T. R. Betts, and R. Gottschalg, “Electroluminescence imaging of PV devices: Camera calibration and image correction,” in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, IEEE, Jun. 2016, pp. 1532–1537. doi: 10.1109/PVSC.2016.7749875.

- [82] W. Tang, Q. Yang, K. Xiong, and W. Yan, “Deep learning based automatic defect identification of photovoltaic module using electroluminescence images,” *Solar Energy*, vol. 201, pp. 453–460, May 2020, doi: 10.1016/j.solener.2020.03.049.
- [83] D.-M. Tsai, S.-C. Wu, and W.-Y. Chiu, “Defect Detection in Solar Modules Using ICA Basis Images,” *IEEE Trans Industr Inform*, vol. 9, no. 1, pp. 122–131, Feb. 2013, doi: 10.1109/TII.2012.2209663.
- [84] X. Chen, T. Karin, and A. Jain, “Automated defect identification in electroluminescence images of solar modules,” *Solar Energy*, vol. 242, pp. 20–29, Aug. 2022, doi: 10.1016/j.solener.2022.06.031.
- [85] Y. Jiang and C. Zhao, “Attention classification-and-segmentation network for micro-crack anomaly detection of photovoltaic module cells,” *Solar Energy*, vol. 238, pp. 291–304, May 2022, doi: 10.1016/j.solener.2022.04.012.
- [86] I. A. Zulfauzi, N. Y. Dahlan, H. Sintuya, and W. Setthapun, “Anomaly detection using K-Means and long-short term memory for predictive maintenance of large-scale solar (LSS) photovoltaic plant,” *Energy Reports*, vol. 9, pp. 154–158, Nov. 2023, doi: 10.1016/j.egyr.2023.09.159.
- [87] J. Maldonado-Correa, S. Martín-Martínez, E. Artigao, and E. Gómez-Lázaro, “Using SCADA Data for Wind Turbine Condition Monitoring: A Systematic Literature Review,” *Energies (Basel)*, vol. 13, no. 12, p. 3132, Jun. 2020, doi: 10.3390/en13123132.

- [88] “Wind Turbine SCADA signals from year 2017,”
<https://www.edp.com/en/innovation/open-data/wind-turbine-scada-signals-2017>.
- [89] Z. Lin, X. Liu, and M. Collu, “Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks,” *International Journal of Electrical Power & Energy Systems*, vol. 118, p. 105835, Jun. 2020, doi: 10.1016/j.ijepes.2020.105835.
- [90] D. C. Montgomery, *Introduction to Statistical Quality Control*, 6th ed. Wiley, 2008.
Accessed: Oct. 22, 2024. [Online]. Available:
<https://books.google.com.sa/books?id=53JlPgAACAAJ>
- [91] L. Ying, Z. Yisu, H. Xiangsheng, and M. Youlin, “Interpretation of IEC 61400-12-1 Update,” in *2019 4th International Conference on Power and Renewable Energy (ICPRE)*, IEEE, Sep. 2019, pp. 102–105. doi: 10.1109/ICPRE48497.2019.9034853.
- [92] G. Sahin, G. Isik, and W. G. J. H. M. van Sark, “Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression,” *Energy Reports*, vol. 10, pp. 2837–2849, Nov. 2023, doi: 10.1016/j.egyr.2023.09.097.
- [93] F. Castellani, D. Astolfi, and F. Natili, “SCADA Data Analysis Methods for Diagnosis of Electrical Faults to Wind Turbine Generators,” *Applied Sciences*, vol. 11, no. 8, p. 3307, Apr. 2021, doi: 10.3390/app11083307.
- [94] S. Chen, Y. Ma, L. Ma, F. Qiao, and H. Yang, “Early warning of abnormal state of wind turbine based on principal component analysis and RBF neural network,” in

- 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), IEEE, Apr. 2021, pp. 547–551. doi: 10.1109/ACPEE51499.2021.9437063.
- [95] R. Bartkowiak Anna M. and Zimroz, “Importance of Variables in Gearbox Diagnostics Using Random Forests and Ensemble Credits,” in *Computer Information Systems and Industrial Management*, J. Saeed Khalid and Dvorský, Ed., Cham: Springer International Publishing, 2021, pp. 3–13.
- [96] K. Shukla, W. Holderbaum, T. Theodoridis, and G. Wei, “Enhancing Gearbox Fault Diagnosis through Advanced Feature Engineering and Data Segmentation Techniques,” *Machines*, vol. 12, no. 4, p. 261, Apr. 2024, doi: 10.3390/machines12040261.
- [97] C. Yang, Z. Qian, Y. Pei, and L. Wei, “A data-driven approach for condition monitoring of wind turbine pitch systems,” *Energies (Basel)*, vol. 11, no. 8, Aug. 2018, doi: 10.3390/en11082142.
- [98] M. A. Sulaiman and J. Labadin, “Feature selection with mutual information for regression problems,” in *2015 9th International Conference on IT in Asia (CITA)*, IEEE, Aug. 2015, pp. 1–6. doi: 10.1109/CITA.2015.7349826.
- [99] T. Pan and M. Li, “Feature Extraction From Spectroscopy Using LASSO and Net Analyte Signal,” *IEEE Sens J*, vol. 22, no. 13, pp. 12997–13004, Jul. 2022, doi: 10.1109/JSEN.2022.3180734.
- [100] Yi-Ren Yeh, Ting-Chu Lin, Yung-Yu Chung, and Y.-C. F. Wang, “A Novel Multiple Kernel Learning Framework for Heterogeneous Feature Fusion and

- Variable Selection,” *IEEE Trans Multimedia*, vol. 14, no. 3, pp. 563–574, Jun. 2012, doi: 10.1109/TMM.2012.2188783.
- [101] A. Mohammed and R. Kora, “A comprehensive review on ensemble deep learning: Opportunities and challenges,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014.
- [102] Z. Wang, H. Huang, and Y. Wang, “Fault diagnosis of planetary gearbox using multi-criteria feature selection and heterogeneous ensemble learning classification,” *Measurement*, vol. 173, p. 108654, Mar. 2021, doi: 10.1016/j.measurement.2020.108654.
- [103] A. Gupta, I. Vijay, S. Garg, and J. Yadav, “Ensemble ML Model for Automobile Price Prediction Using Image Dataset,” in *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, IEEE, May 2024, pp. 1–6. doi: 10.1109/ISCS61804.2024.10581309.
- [104] V. C. S. Rao, M. Sujatha, J. Navitha, C. Karthikeya, B. Goutham, and P. Jairam, “Development of an Ensemble Model for Disease Prediction using Laboratory Tests,” in *2024 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Apr. 2024, pp. 907–913. doi: 10.1109/ICICT60155.2024.10544626.
- [105] T. R. Mahesh, V. Vinoth Kumar, V. Vivek, K. M. Karthick Raghunath, and G. Sindhu Madhuri, “Early predictive model for breast cancer classification using blended ensemble learning,” *International Journal of System Assurance*

- Engineering and Management*, vol. 15, no. 1, pp. 188–197, Jan. 2024, doi: 10.1007/s13198-022-01696-0.
- [106] N. Mohd Khairudin, N. Mustapha, T. N. Mohd Aris, and M. Zolkepli, “Ensemble neural networks with input optimization for flood forecasting,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3686–3694, Oct. 2024, doi: 10.11591/eei.v13i5.6863.
- [107] T. Surapunt and S. Wang, “Ensemble Modeling with a Bayesian Maximal Information Coefficient-Based Model of Bayesian Predictions on Uncertainty Data,” *Information*, vol. 15, no. 4, p. 228, Apr. 2024, doi: 10.3390/info15040228.
- [108] P. Umamaheswari and V. Ramaswamy, “An integrated framework for rainfall prediction and analysis using a Stacked Heterogeneous Ensemble Model (SHEM),” *Expert Syst Appl*, vol. 256, p. 124831, Dec. 2024, doi: 10.1016/j.eswa.2024.124831.
- [109] C. Zhang *et al.*, “Research and application of a novel weight-based evolutionary ensemble model using principal component analysis for wind power prediction,” *Renew Energy*, vol. 232, p. 121085, Oct. 2024, doi: 10.1016/j.renene.2024.121085.
- [110] J. Heinermann and O. Kramer, “Machine learning ensembles for wind power prediction,” *Renew Energy*, vol. 89, pp. 671–679, Apr. 2016, doi: 10.1016/j.renene.2015.11.073.
- [111] G. Wang, R. Jia, J. Liu, and H. Zhang, “A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning,” *Renew Energy*, vol. 145, pp. 2426–2434, Jan. 2020, doi: 10.1016/j.renene.2019.07.166.

- [112] Y. Wang, T. Chen, R. Zou, D. Song, F. Zhang, and L. Zhang, “Ensemble probabilistic wind power forecasting with multi-scale features,” *Renew Energy*, vol. 201, pp. 734–751, Dec. 2022, doi: 10.1016/j.renene.2022.10.122.
- [113] M. Ali *et al.*, “Software Defect Prediction Using an Intelligent Ensemble-Based Model,” *IEEE Access*, vol. 12, pp. 20376–20395, 2024, doi: 10.1109/ACCESS.2024.3358201.
- [114] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Y. Li, “Soft-Weighted-Average Ensemble Vehicle Detection Method Based on Single-Stage and Two-Stage Deep Learning Models,” *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 100–109, Mar. 2021, doi: 10.1109/TIV.2020.3010832.
- [115] Md. A. I. Nelooy, N. Nahar, M. S. Hossain, and K. Andersson, “A Weighted Average Ensemble Technique to Predict Heart Disease,” 2022, pp. 17–29. doi: 10.1007/978-981-16-7597-3_2.
- [116] C. Valle, F. Saravia, H. Allende, R. Monge, and C. Fernández, “Parallel Approach for Ensemble Learning with Locally Coupled Neural Networks,” *Neural Process Lett*, vol. 32, no. 3, pp. 277–291, Dec. 2010, doi: 10.1007/s11063-010-9157-6.
- [117] K. B. Abdusamad, D. W. Gao, and E. Muljadi, “A condition monitoring system for wind turbine generator temperature by applying multiple linear regression model,” in *2013 North American Power Symposium (NAPS)*, IEEE, Sep. 2013, pp. 1–8. doi: 10.1109/NAPS.2013.6666910.

- [118] L. Breiman, “Random Forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [119] S. V. V. S. N. Pichika, R. Yadav, S. Geetha Rajasekharan, H. M. Praveen, and V. Inturi, “Optimal sensor placement for identifying multi-component failures in a wind turbine gearbox using integrated condition monitoring scheme,” *Applied Acoustics*, vol. 187, p. 108505, Feb. 2022, doi: 10.1016/j.apacoust.2021.108505.
- [120] J. A. Bellido-Jiménez, J. Estévez Gualda, and A. P. García-Marín, “Assessing new intra-daily temperature-based machine learning models to outperform solar radiation predictions in different conditions,” *Appl Energy*, vol. 298, p. 117211, Sep. 2021, doi: 10.1016/j.apenergy.2021.117211.
- [121] I. Lillo-Bravo, J. Vera-Medina, C. Fernandez-Peruchena, E. Perez-Aparicio, J. A. Lopez-Alvarez, and J. M. Delgado-Sanchez, “Random Forest model to predict solar water heating system performance,” *Renew Energy*, vol. 216, p. 119086, Nov. 2023, doi: 10.1016/j.renene.2023.119086.
- [122] A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, and D. Ruiz-Hernández, “A survey of artificial neural network in wind energy systems,” *Appl Energy*, vol. 228, no. April, pp. 1822–1836, 2018, doi: 10.1016/j.apenergy.2018.07.084.
- [123] A. Santolamazza, D. Dadi, and V. Introna, “A Data-Mining Approach for Wind Turbine Fault Detection Based on SCADA Data Analysis Using Artificial Neural Networks,” *Energies (Basel)*, vol. 14, no. 7, p. 1845, Mar. 2021, doi: 10.3390/en14071845.

- [124] M. Shafiullah, M. A. Abido, and T. Abdel-Fattah, "Distribution Grids Fault Location employing ST based Optimized Machine Learning Approach," *Energies (Basel)*, vol. 11, no. 9, p. 2328, Sep. 2018, doi: 10.3390/en11092328.
- [125] M. Shafiullah and M. A. Abido, "S-Transform Based FFNN Approach for Distribution Grids Fault Detection and Classification," *IEEE Access*, vol. 6, pp. 8080–8088, 2018, doi: 10.1109/ACCESS.2018.2809045.
- [126] N. G. Şengöz, "Control Charts to Enhance Quality," in *Quality Management Systems - a Selective Presentation of Case-studies Showcasing Its Evolution*, InTech, 2018. doi: 10.5772/intechopen.73237.
- [127] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017, doi: 10.1016/j.measurement.2017.07.017.
- [128] H. Mewada, "Electroluminescence image-based defective photovoltaic (solar) cell detection using a modified deep convolutional neural network," Aug. 2024, pp. 13–20. doi: 10.21741/9781644903216-2.
- [129] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-Margin Softmax Loss for Convolutional Neural Networks," Dec. 2016, [Online]. Available: <http://arxiv.org/abs/1612.02295>
- [130] "Technical And Reliability Data for Raspberry Pi Products (Including Datasheets)," 2023.

[131] "Specification of Piezoelectric Sounder." [Online]. Available:
[http://www.murata.com/en-us/products/productdetail?partno=PKM22EPPH4001-](http://www.murata.com/en-us/products/productdetail?partno=PKM22EPPH4001-B0URL:https://www.murata.com/)
[B0URL:https://www.murata.com/](https://www.murata.com/)

VITAE

Name	Abdulbasit Hassan
Nationality	Nigerian
Date of Birth	18 th September, 1993
Email	abdulbasithassan5@gmail.com
Address	No. 33 Albarkawa, Zaria City, Kaduna State, Nigeria. Postal code: 810241
Academic Background	M.Sc. in Systems and Control Engineering (2024)
	<i>King Fahd University of Petroleum & Minerals, KSA.</i>
	B.Eng. Mechatronics Engineering
	<i>Bayero University Kano, Kano, Nigeria.</i>
Publication	Hassan, A., Shafiullah, M., Islam, A., Gulzar, M. M., & Hamanah, W. M. (2024). Data-driven intelligent prediction model for gearbox temperature in wind turbines. In 2024 IEEE 21st International Conference on Smart Communities: Improving Quality of Life using AI, Robotics, and IoT (HONET). (Accepted).
	Shafiullah, M., Panda, S. K., Sahani, M., & Hassan, A. (2024). Deep neural networks for predictive maintenance of wind turbine system. In 12th Saudi Arabia Smart Grid Conference 2024. (Accepted).

	Abdulbasit Hassan, Ghali Ahmad, Md Shafiullah, Asif Islam, Md Shafiul Alam. Review of the intelligent frameworks for pitch angle control in wind turbines (2024). IEEE Access. (Submitted).
	Ghali Ahmad, Asif Islam, Abdulbasit Hassan, Md Shafiullah, Mohammad A. Abido, and Mujahed Al-Dhaifallah. Distributed Control Strategies for Microgrids: A Critical Review of Technologies and Challenges (2024). IEEE Access. (Submitted).
	Hassan, A., Al-Ahdal, A. A., Abdul-wahid, A. S., & Yahaya, J. U. (2024, April). Analysis of Robust Control Method for Single Link Flexible Manipulator," 2024 21st International Multi-Conference on Systems, Signals & Devices (SSD), Erbil, Iraq, 2024, pp. 465-473, doi: 10.1109/SSD61670.2024.10548073.
	Mohammed F. Al-mashdali, Asif Islam, Abdulbasit Hassan, Md Shafiullah, Mujahed Al-Dhaifallah. Signal Processing-free Intelligent Model for Power Quality Disturbances Identification (2024). IEEE Access. (Submitted).
	K. K. Dandago, A. Mohammed, J. Y. Umar, A. Hassan and M. S. Zango, "Trajectory Tracking of a Ground Agricultural Robot Using Sliding Mode Control," 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), 2021, pp. 1-5, doi: 10.1109/ICMEAS52683.2021.9692360.
Experience	RESEARCH/ LABORATORY ASSISTANT (Aug/2022 to Current)

	<i>King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia (15 hours per week)</i> • Utilized IT-5200 (Process Control Trainer), IT 5927 (Transducer and Instrumentation Trainer), IT 5928 (Pressure Transducer Trainer), and IT 5931 (Strain Gauge Trainer) for the Instrumentation Engineering Laboratory course to train Undergraduate students on Force and strain, Temperature sensor, Light sensor, and Variable Resistor measurements. • Utilized LABVIEW software and Data Acquisition (DAQ) System for the Computer Control Laboratory course to train undergraduate students on Robotic Arm PID control and the SCADA system data generation. • Developed and Improved the laboratory manual for the Instrumentation Engineering Laboratory. • Development of modern laboratory experiments based on IOT devices. • Cleaned and maintained equipment to achieve optimum performance. • Implemented health and safety guidelines to minimize risks in the laboratory GRADUATE ASSISTANT (Apr/2020 to Aug/2022) <i>Air Force Institute of Technology (AFIT), Kaduna State, Nigeria.</i> • Teaching assistantship on two (2) undergraduate courses (MCT 307-Electronic I and EE-205 Student Work Experience Scheme (SWEP).
--	--

	• Supervised examination assessments and managed departmental examination records while serving as Examination Officer (EO). • Coordinated with faculty members to develop learning materials and course resources while serving on the Curriculum committee to enhance educational delivery. • Conducted literature reviews and database searches to support the ongoing Institute's research projects while serving in the Research committee. INTERN <i>Ashaka/LafargeHolcim Cement Plc, Gombe State, Nigeria</i> (Sept 2016 - Jan 2017) • Supported daily production performance meetings by preparing presentation slides and daily performance data and taking notes while holding a Secretaryship position. • Performed a Root Cause Analysis (RCA) on kiln thruster bearing failure. • Worked with Programmable Logic Controller (PLCs) and Supervisory Control and Data Acquisition (SCADA) Systems of the Engineers workstation to manage the industrial processes. • Carried out calibration of Instrumentation Systems. • Conducted Preventive Maintenance on Various types of Sensors and Transducers with the Maintenance crew.
--	---

	• Conducted Vibration Test on Electrical Drives, Gearbox, and bearings • Worked on Kiln Gas Analyzer System ULTRA MAT 23 Nigerian <i>College of Aviation Technology (NCAT), Zaria, Nigeria</i> (Sept 2016 - Jan 2017) • Conducted Annual Inspection, 100 hours Inspection, Non-destructive Tests on Aircraft • Replaced faulty avionics components such as the artificial horizon and altimeter in the aircraft cockpit. • Servicing of Landing Gears Braking System. <i>African Textile Manufacturers Ltd (ATM), Kano, Nigeria</i> (Jan 2016 – March 2016) • Conducted Maintenance on Pneumatic and Hydraulic Systems with the Maintenance crew on the warping and sewing machines. • Replaced some electronic components on the Electronic Card (ECG) of a Sewing machine. • Involved in Fire Fighting Demonstrations.
--	---