

AI-based Solar Panel Fault Prediction and Anomaly Detection in Ireland

MSc Research Project
MSc Data Analytics

Lathifa Jaffer Diang'a
Student ID: x23346671

School of Computing
National College of Ireland

Supervisor: Teerath Kumar

National College of Ireland
MSc Project Submission Sheet



School of Computing

Student Name:LATHIFA JAFFER DIANG’A.....

Student ID:X23346671.....

Programme: ...MSC DATA ANALYTICS (MSCDAD_B)..... **Year:**2024/2025.....

Module:RESEARCH PRACTICUM.....

Lecturer:Teerath Kumar.....

Submission Due Date:15th September 2025.....

Project Title: AI-BASED SOLAR PANEL FAULT PREDICTION AND ANOMALY DETECTION IN IRELAND
10398

Word Count: 25

Page Count:

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:..... LATHIFA JAFFER DIANG’A.....

Date:11th September 2025.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only

Signature:

Date:

Penalty Applied (if
applicable):

AI-based Solar Panel Fault Prediction and Anomaly Detection in Ireland

Lathifa Jaffer Diang'a
X23346671

Abstract

The transition to renewable energy systems weather sensitive countries such as Ireland necessitates intelligent fault management and reliable solar infrastructure systems. In this research a data-driven, AI enhanced approach for anomaly detection and proactive fault prediction in solar panel systems under the unpredictable weather condition in Ireland are presented. Using the CRISP-DM framework, four datasets (Met Éireann (weather data), PVWatts (energy output data), Portugal Solar Panels (sensor data), and Dublin Airport (irradiance and climate data)) are independently analysed. Exploratory Data Analysis (EDA), LazyPredict regression benchmarking, and Isolation Forest-based anomaly detection are enacted. Top performing models, including NuSVR and GradientBoosting, attain RMSE scores as low as 0.02, and R^2 scores as high as 0.99 – this demonstrates robust predictive accuracy for irradiance patterns and solar output. Proactive maintenance approaches are supported by anomalies that flag 15-25% of potential faults. This study promotes a region-specific and scalable methodology that optimizes solar panel operational uptime; while also supporting Ireland's national renewable energy targets, it also proposes a roadmap for integrating multiple datasets, addressing overfitting risks and future field deployments strategies that would strengthen real-world applicability.

Key words – Anomaly Detection, CRISP-DM, LazyPredict, Machine Learning, Solar Panel Fault Detection, Renewable Energy

1 Introduction

Solar energy is a sustainable and an environmentally friendly alternative to non-renewable resources such as fossil fuels. Numerous countries have transitioned from non-renewable energy to renewable energy resources such as solar energy which serves as a solid foundation¹. Nevertheless, the efficiency and the reliability of solar panel systems are frequently compromised by faults and performance anomalies, which reduces the energy output, increases downtime, and escalates the overall maintenance costs. It is very important to note that solar panel systems are extremely sensitive to humidity, frequent temperature variations and salt corrosion in coastal areas – these factors combined with electrical issues, and panel performance degradation, diminish the performance of the systems, hence they do not work as effectively as they should (Abdulmajed et al. (2024)).

Primarily, the existing traditional maintenance practices address faults after the occurrence – these reactive practices are inefficient and costly (Agussalim Syamsuddin & Widodo (2024)). While this approach has worked so far, it would be constructive to approach the detection of faults in a proactive manner especially with technological advancements in Artificial Intelligence (AI). Cutting edge breakthroughs in the artificial intelligence field, machine learning and data analytics have shown potential in proactive maintenance approaches via analysis of sensor data from solar panel systems (Giger & Csillaghy (2024)). Conversely, regardless of the big advancements in technology, there is still a gap that has not been filled in comprehensive research that is solely focused on prediction models tailored to

¹ <https://www.seai.ie>

dynamic weather conditions such as those in Ireland, and in real time proactive fault detection. Furthermore, there is also a notable gap in the application of unsupervised learning techniques in the anomaly detection of faults as most research focuses on reactive maintenance practices, which do not identify faults before system failure. Therefore, this research aims to address the acknowledged gaps, by developing an AI driven anomaly detection and proactive fault detection system that is solely tailored to Irish weather conditions. This approach is designed to enhance the reliability of solar panel systems and to drive and improve its performance in the country.

1.1 Motivation

The aim and focus of this research are driven by the need to improve the efficiency and the reliability of the solar panel systems in regions that face dynamic and unpredictable weather conditions. This type of weather escalates the degradation of solar panel systems reducing their overall functionality and capability. In Ireland, high humidity, coastal corrosion, and frequent cloud cover is common – this combined with the other weather factors creates a unique set of challenges that affects the performance of the solar panel systems (Okonkwo et al. (2025)). Hence, the development of a real-time proactive fault detection system, is not just a want but a requirement, as it would address the current predicament faced. The purpose of this research is supported by Ireland’s plan to adopt more renewable energy systems – with the hopes that it would increase the energy output and reduce the high operational and maintenance costs. Thus, if this research is successful, the findings will offer a personalized solution that can enhance the sustainability of solar panel systems in the country and increase the economic viability (particularly useful for solar operators in the country). Furthermore, this research greatly supports the Irish Climate Action mandate ratified in 2023² – a plan that intends to convert over 80% of Ireland’s electricity production to renewable sources. Thus, the objectives can be summarized as:

- To develop AI models for solar prediction (supervised learning models to forecast panel output and unsupervised learning models to detect anomalies)
- To optimize models for Irish climate conditions by identifying patterns that indicate potential system faults and to flag unusual performance for early intervention.
- To evaluate performance across multiple datasets through a comparison of model accuracy and reliability metrics.

Therefore, should this study be prosperous, it could theoretically offer scalable a resolution that can be applied to both commercial and residential solar panel systems.

1.2 Research Question

“What is the impact of AI-based anomaly detection models in the proactive prediction of faults and the optimization of operational uptime of solar panel systems under dynamic environmental conditions in Ireland?”

1.3 Proposed Solution

The proposed solution for the research question involves the establishment of an AI-based predictive model for anomaly detection and fault prediction, through the integration of data from both weather and real time solar panel sensors (inverter data). All the data utilized in this research are publicly available, ensuring privacy as there are no ethical violations that are crossed. The methods used are:

1. **Unsupervised Learning Techniques** - Anomaly detection is explored using Isolation Forest and K-means on Dublin weather data which obtained from sensors.

² <https://www.seai.ie>

2. **Supervised Learning Techniques** - The LazyPredict library is used to implement several regression models on the datasets such as Random Forests, Support Vector Machines and Gradient Boosting – to predict faults on labelled fault data (current, voltage, irradiance, and temperature).
3. **Economic Impact analysis** – Cost savings and operational uptime compared to traditional methods is examined by implementing a robust framework that would evaluate the performance through metrics such as RMSE and R^2 , which will confirm the effectiveness of the applied models and the system reliability.

1.4 Potential Contributions

The suggested solution for the research intends to make substantial and significant contributions to the solar energy field, such as:

- **Environmental impact** – The research will contribute to the decrease of carbon emissions by adopting renewable energy sources as the main source of electrical supply, as the reliability and the performance of solar panel systems are improved.
- **Improved Efficiency** – The solar panel systems energy output will be increased, and the system downtime will be reduced. This can be achieved via proactive fault detection, enhancing the overall efficiency of the solar panel systems.
- **Economic Benefits** – Operational costs will be decreased as the proactive maintenance practices reduces constant emergency repairs, this in turn extends the lifespan of solar panel systems (Chaturvedi et al. (2024)). Thus, the research promotes economic sustainability by increasing energy outputs – encouraging the transition of Ireland to renewable energy sources.

This research contributes to the growing body of scientific and academic literature at the intersection of renewable energy, sustainability, predictive maintenance, and artificial intelligence. While previous related literature has explored the use of deep learning and machine learning for solar panel fault detection, many have focused on image-based diagnostics and static environments (Nguyen et al. (2025)). Contrastingly, this research uniquely integrates both supervised and unsupervised AI-based models across four different datasets to address real-time fault prediction and anomaly detection under Ireland’s volatile weather conditions. Moreover, this research introduces a hybrid methodology that combines both traditional regression modeling with anomaly detection for the proactive identification of potential system faults that impact energy production before they occur. The research findings provide a localized framework that is applied and can support national energy resilience and offers a replicable model for other regions worldwide that have volatile climates. In addition to advancing AI-based solar panel fault prediction and anomaly detection, this research contributes feasible value by clearly tackling the risks associated with overfitting in high performance models, ethical concerns regarding reliability and alert management, deployment challenges in IoT and cloud settings, and the potential cost-benefit analysis of real world implementation. By evaluating these considerations, the research further strengthens its applicability beyond the academic scope, and it ensures direct relevance to Ireland’s renewable energy strategy through the offering of scalable baselines future implementations.

1.5 Chapter breakdown

This report is structured as follows:

- Chapter 1 focuses on the introduction of the research problem and the research objectives on how the problem will be tackled.
- Chapter 2 review existing related works on AI based fault detection, dynamic weather impacts, anomaly detection, and Ireland’s renewable energy strategy

- Chapter 3 presents the research methodology using the CRISP-DM framework, Preprocessing, model development and evaluation.
- Chapter 4 outlines the results of predictive and anomaly detection models followed by a critical assessment of the findings real-world implications and limitations, as well as the addressing of overfitting concerns, considerations of integrated datasets, and an evaluation beyond R^2 and RMSE.
- Chapter 5 explores deployment considerations, ethical considerations, sustainability impacts, a risk assessment, a cost-benefit analysis, and a critical evaluation.
- Chapter 6 concludes the report by restating the research contributions and reflecting on the success of the approach that has been deployed and proposing future directions of the work.

2 Related Works

In the last few years, the requirement to improve the reliability and efficiency of renewable energy systems has been constant – this has driven the need for the integration of anomaly detection and AI-based fault prediction models. Key literature and state-of-the-art models have been reviewed and examined in this section which focuses on predictive maintenance, the impact of dynamic weather conditions on the performance of solar panels, and anomaly detection systems. In addition to that, this section also reviews the monetary value of research in the Irish domain, European Union (EU) regulations, and the existing use of Artificial Intelligence (AI) in solar panel systems. Existing gaps are identified in existing literature, emphasizing how the proposed research will strive to address the identified gaps.

2.1 Predictive Maintenance in Solar Panel Systems

Predictive maintenance of solar panel systems has emerged as a vital strategy to reduce the operational downtime and the overall costs. An extensive review of most recent advancements in cyber security and solar panel predictive maintenance are examined in the works of (Ledmaoui et al. (2025)) – this work highlights the role of deep learning, machine learning and the Internet of Things (IoT) on facilitating fault detection and real time monitoring in solar plants. Furthermore, (Ledmaoui et al. (2025)) emphasizes the hurdles faced in dynamic environments when balancing model accuracy and complexity – nonetheless, this work lacks a robust framework for real time solar panel fault prediction as it focuses on an offline analysis. On the other hand, (Kayalvizhi et al. (2024)) demonstrates the exploration of using IoT-enabled predictive maintenance and live monitoring, this solution is effective at reducing the downtime but lacks the robustness against real-time unpredictable weather fluctuations – a very critical factor in Ireland’s dynamic climate.

In this study, predictive maintenance is implemented through the use of machine learning regressors such as XGBoost, Random Forest and Gradient Boosting. These regressors are trained on weather data, AC power, and irradiance across four different datasets. The results obtained exhibit real-time model deployment competency with RMSE values as low as 0.02 and R^2 values as high as 0.99 – these performance metrics highlight the capability of the system to proactively convey component level degradation. Hence, it offers a scalable and dynamic alternative to the existing traditional inspection models.

2.2 Anomaly Detection Using Machine Learning

Anomaly detection plays a pivotal part in solar panel predictive maintenance by flagging the unanticipated operational deviations. In the work of (Chaturvedi et al. (2024)), supervised models such as Support Vector Machine (SVM) and Random Forest are used to classify known anomalies – this approach is inadequate when it encounters novel or unseen faults. In parallel, (Abdelsattar et al. (2025)) advances this technique via deep learning-based image

analysis (e.g., MobileNetV2 and Xception), however the study was only limited to visual inspection.

In this work, these limitations are surpassed through the implementation of a hybrid detection system that combines both supervised regressors with unsupervised techniques (Isolation Forest, K-means) applied across the Dublin airport and Portugal datasets. These applied algorithms flagged different seasonal anomalies particularly in low irradiance months and high humidity months – exemplifying their strength in learning the complexities of fault patterns even under noisy or incomplete data. This consolidated learning proposition facilitated real time application in unpredictable weather environments and strengthened the model’s generalizability.

2.3 Climate and Weather Effects on Solar Panel Fault Prediction

The effects that dynamic environmental conditions such as humidity, temperature, and rainfall has often been marginalized in AI-based solar panel maintenance studies. (Okonkwo et al. (2025)) and (Shin et al. (2024)) both highlight the significant role of extreme weather events on the performance of solar panel systems and the angle of incidence effects. (Okonkwo et al. (2025)), examines the impact of extreme weather phenomenon on solar panel systems, giving prominence to the need for adaptive system architectures and resilient materials, while the latter (Shin et al. (2024)) investigated the effect of incidence angle on the measurement of temperature in solar panels using UAV-based thermal infrared cameras. However, both of these studies fail to translate these effects into real-time detection systems, and they do not address the challenges that would be faced by the system under unpredictable weather conditions. On the other hand, (Ma et al. (2024)) investigated the performance of low-temperature solar thermal energy storage systems to in enhancing efficiency, by highlighting the importance of optimizing system performance under dynamic weather conditions. Nonetheless, the approach does not consider implementing any sort of AI-based fault prediction models for real-time monitoring.

These challenges are directly addressed in this research through the integration of live weather analytics – for instance, precipitation, temperature, irradiance – from the Dublin Airport and Met Éireann datasets. The analysis highlights the importance of these parameters as predictors through the modeling of diverse seasonal profiles via regressors and visualizing feature importance. Thus, operational resilience is significantly improved in regions such as Ireland where conventional models are inclined to break down under dynamic weather conditions.

2.4 Ireland: A High-Risk Case Study for Solar Panel Fault Prediction

Ireland has a unique climate often characterized by frequent cloud cover, high humidity and coastal erosion, all of which cause substantial challenges in the performance of solar panel systems. This coastal climate presents a huge critically challenging testbed for AI enabled solar panel fault prediction, as these climatic factors necessitates resilient systems. According to the Sustainable Energy Authority of Ireland (SEAI)³, the Solar PV Strategy advocates for advanced monitoring strategies and solar panel maintenance solutions that would substantially refine efficiency and the reliability of the PV systems. However, existing literature highly lacks Ireland-specific AI-based research – even though there are works such as those of (Okonkwo et al. (2025)) and (Shin et al. (2024)) who both call attention to the requirement of a real-time fault prediction system – but their works are not based on the Irish domain that faces its own distinct climatic challenges.

³ <https://www.seai.ie>

In this research, this void is filled by training and evaluating models using local data sources. Hence, the results obtained proves that Ireland’s climate can support scalable AI-based solutions, as the anomaly detection and predictive modeling systems still perform optimally in low-irradiance conditions. This research is highly applicable, as the approach applied and the outputs obtained are locally – hence, the findings can be utilized by Irish solar panel developers.

2.5 Drones and Robotics in Solar Panel Fault Diagnosis

Although drones and robotics are increasingly becoming important in solar panel monitoring systems, the existing implementations are hardware-focused with minimal machine learning integration, such as those presented in the works of (Iqbal et al. (2019)) and (Obieli et al. (2025)). Primarily, (Iqbal et al. (2019)) investigated the role of robotics in renewable energy systems, through tasks such as inspection – the results were effective at showcasing how robotics can lower costs, reduce set up time and enhance productivity in solar farms. Moreover, (Obieli et al. (2025)), showcased how autonomous robots are used for automated repairs and anomaly recognition. Both of these works albeit being excellent, demonstrate that the coupling of autonomous verification with real-time predictions is underexplored.

The proposed architecture in this research sets the groundwork for a scalable system, where the outputs of the model can be directly siphoned into the autonomous robots and drones – for example, high-anomaly dense areas can be flagged and prioritized for drone-based inspection. Even though this approach is not physically deployed in the study, it sets a solid foundation and pathway for future smart maintenance undertakings in both Irish and international contexts.

2.6 Value of Predictive and Generative AI in Renewable Maintenance

Generative AI has emerged as a novel approach and a competent tool for performance enhancement and predictive maintenance in Solar Battery Storage Systems (SBSS) – it can replicate energy behaviour under stress and simulate fault scenarios. The study done by (Seenu & Annareddy (2023)) is effective at demonstrating how capable generative AI is at improving system performance and maintenance prediction through the replication of energy outputs across usage cycles. Moreover, the economic impact of AI based methods is significant, and this increases energy outputs and in turn improves cost savings.

Although this implementation is not a core component of the suggested approach in the current research, the predictive models developed produce fault logs and granular temporal datasets that could potentially be used training inputs for generative systems in future works and iterations. Therefore, the vision of Generative-AI reinforces the potential of the research (AI-based fault prediction systems) as it serves as a foundational layer for generative enhanced optimisation techniques for energy diagnostics.

2.7 Economic Value and Policy Relevance

Based on the study presented by (Kayalvizhi et al. (2024)), IoT-enabled solar panel predictive maintenance can increase energy yields by 15% while also reducing maintenance costs by 30%. This is also corroborated by (Chaturvedi et al. (2024)) who illustrates that the downtime of solar panels and operational costs of solar panel systems can be reduced by machine learning methods. Both of these studies illustrate the economic feasibility of AI-based solutions for solar panel maintenance and fault prediction.

Furthermore, the findings of this research, localize and validate the recognized potential – for example, by forecasting performance drops, estimating seasonal impact months, and predicting energy situations, the proposed models enables proactive maintenance

through actionable scheduling of repairs and system scaling. This is in alignment with SEAI Solar PV Strategy⁴ and Ireland’s Climate Action Plan⁵ that forecast an increase in energy outputs by 10-15%, and €50 million annual maintenance savings if these actions are adopted nationally. This plan would support Ireland’s Climate Action Plan propelling its EU vision – thus, all the policy-aligned insights elevate the potential of the proposed research beyond academia and into real-world decision making.

2.8 Conclusion

This literature review displays a critical evaluation of how the research aligns and expands upon existing literature based on solar systems, machine learning approaches, and any AI-based fault detection for solar panel systems. It establishes this research as a novel and policy-aligned study that synthesizes weather resilience with AI model optimization and regional relevance. Thus, this study builds upon the existing literature, enhances technical robustness by closing existing gaps, and displays practical adaptability and applicability.

3 Research Methodology

3.1 Introduction to Methodology

This study adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology to develop a proactive real-time anomaly detection and fault prediction system for solar panels under dynamic weather conditions in Ireland. The CRISP-DM methodology is chosen due to its ability to adapt to real-world solar panel datasets and its robustness in guiding data science workflows – hence, the approach is intended to ensure practical applicability, robustness and scalability. The process consists of six processes – Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Each of the CRISP-DM phases was adapted to cater for each domain-specific challenge that may arise when dealing with renewable energy management in Ireland.

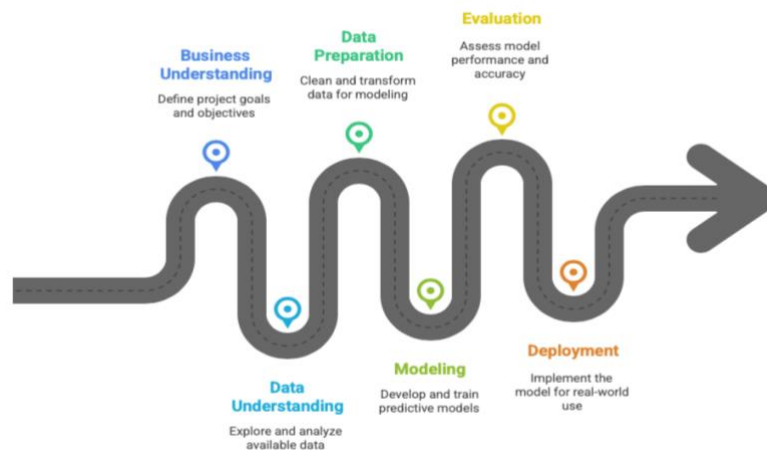


Figure 1: CRISP-DM methodology for solar panel fault prediction

⁴ <https://www.seai.ie>

⁵ <https://www.gov.ie/en/department-of-climate-energy-and-the-environment/publications/climate-action-plan-2025/>

3.2 CRISP-DM Framework Justification

The CRISP-DM methodology is chosen due to its practical focus in delivering actionable insights, and its structure supports the dynamic nature of solar panel fault prediction and diagnosis under the unpredictable weather conditions in Ireland (for instance, adjusting for seasonal rainfall patterns as those observed in the Met Éireann data). Furthermore, CRISP-DM has a clear segmentation of the data understanding to its preparation, modeling and evaluation, this provides a clear iterative feedback structure unlike the Knowledge Discovery Database (KDD) framework. Hence, the framework supports the integration of diverse datasets into a unified prediction system that would ensue robustness and adaptability (Azeroual et al. (2025)).

3.3 Business Understanding

The main business objective is to reduce solar panel downtime by proactively identifying faults using predictive analytics and AI – this specific problem is critical due to the Ireland’s large investment in improving its national sustainability goals of 80% renewable electricity by 2030⁶. This research addresses the challenges faced by the suboptimal solar panel performance in Ireland caused by environmental factors such as high rainfall and low winter solar irradiance (164.484 kWh/m² in December 2024 from Dublin Airport data) which contribute to solar panel faults (for instance; inverter failures and reduced efficiency). Interested parties would include maintenance teams, solar farm operators, and energy regulators, who will all benefit from reduced downtime and cost savings. Success indicators would include improved model accuracy with Mean Absolute Error (MAE) below 0.5 kWh/m², precise anomaly detection and reduction in maintenance events, all corroborated against available historical downtime data. Hence, this study suggests that a data-driven AI system can potentially outperform the rule-based fault detection methods currently used in improving predictive accuracy and efficiency.

3.4 Data Understanding

Four complementary datasets are used to reflect the system response and meteorological drivers across different operational contexts, the datasets are described as follows:

1. **Met Éireann** – This dataset contains information from Irish weather stations (1941-present) that represent high quality local weather observations (for instance, average temperature ‘meant’, total rainfall ‘rain’, and daily solar radiation duration ‘sun’) which provide exogenous drivers of the solar irradiance and by extension the solar panel output which is essential in providing the base for forecasting performance under Irish weather conditions. Moreover, the dataset anchors the modelling to the local climate where humidity, cloud cover and Atlantic wind fronts are dominant in modulating solar availability, hence the dataset is the most realistic source in capturing the variance caused by weather and to stress test the models under the local conditions.
2. **PVWatts** – This dataset comprised of simulated solar panel output data in Dublin, consisting of variables such as performance ratio, AC output and POA irradiance averaging 2.7 kWh/m²/day that reflects local performance. The dataset is ideal for checking the model’s capacity under ‘perfect’ conditions benchmarking the upper bound of achievable accuracy while isolating the modelling errors from any data noise. Thus, a high R² value on this dataset demonstrates that the suggested models can learn the core functional form and in turn map solar irradiance to solar output.

⁶ <https://www.gov.ie/en/department-of-climate-energy-and-the-environment/publications/climate-action-plan-2025/>

3. **Portugal Solar Dataset** – This dataset is a compilation of high-volume solar panel data (2013-2021) with metrics namely ‘inv1_volt’, ‘inv1_str1_curr’, adapted for anomaly detection which exposes the models to equipment level dynamics such as partial shading and panel degradation. The dataset provides a realistic albeit high SNR (signal-to-noise) ratio environment for testing anomaly detectors on operational features and for learning fine-grained relationships that exist.
4. **Dublin Airport Dataset** – This dataset contained information on recent hourly weather data (2022-2025) in Dublin, which can reveal the correlation of fault-prone months with the environmental anomalies. The dataset is serves as a solar proxy sequence that is used to ground the study to the current Irish operating conditions, to verify seasonality, and to test anomaly detection against extreme conditions such as rainfall and wind.

Each of these datasets were thoroughly analyzed for stationarity, completeness and consistency – influenced by the works of (Thakfan et al. (2024)). Key statistics, time-based factors and distributions of each dataset are explored, and collectively the datasets enable a thorough understanding of how environmental conditions impact solar panel performance. Individually, each of these datasets target different aspects, but together they provide a triangulated view that improves generalization and robustness even though they are analyzed separately. Figure 2 illustrates the analysis of the datasets.

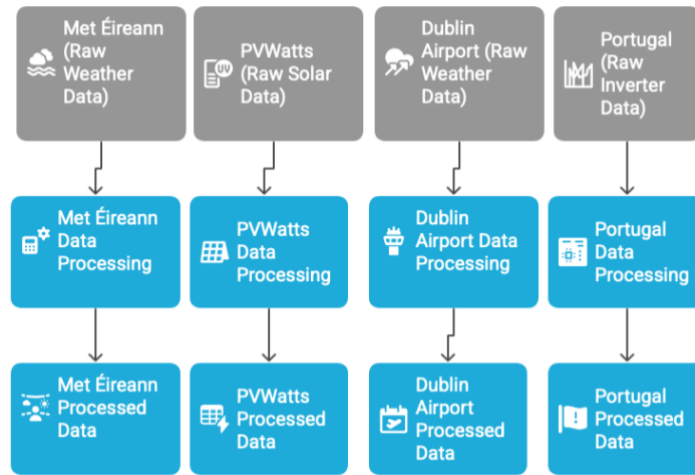


Figure 2: Parallel Dataset Analysis Flowchart

3.5 Data Preparation

In this stage, each dataset is processed independently to support fault prediction through data cleaning and transformation. This involved actions such as – the standardization of timestamps and the normalization of input values, removing null values and any erroneous entries, feature engineering of the data from the datetime (e.g., month, season), encoding the categorical weather descriptions in instances where it is applicable, and creating lag variables (e.g., rolling seven-day rainfall averages, irradiance from the previous day). Examples of some of these actions based on each dataset is presented:

1. **Met Éireann** – ‘low_sun_alert’ is flagged based on thresholds and sun_kWh ($df['sun_kWh'] = df['sun'] / 41.67$) is calculated.
2. **PV Watts** – PR is derived through $df['PR'] = df['AC_Output'] / (df['POA_Irradiance'] * df['Month']...)$
3. **Portugal Solar Dataset (Zenodo)** – ‘current_voltage_ratio’ is calculated and only 5% of the data is sampled ($df.sample(frac=0.05)$).
4. **Dublin Airport** – ‘solar_radiation’ is interpolated – `‘dublin_airport[‘solar_radiation’].interpolate()’`

While each dataset remained separate, all their processed outputs are synthesized manually to obtain integrated insights.

4 Design Specification and Implementation

The artefact created in this study is a machine learning pipeline for solar panel fault prediction and anomaly detection. The model is crafted for solar panel systems in Ireland based on weather and performance data by simulating fault prediction and operational monitoring – the artefact developed is based on independent dataset analyses as illustrated in Figure 3.



Figure 3: Fault Prediction System Architecture

A hybrid model architecture is applied where supervised regressors are used for AC output prediction (Random Forest and XGBoost among others). Rule-Based Fault Flagging is done based on the performance deviations of the data, and unsupervised anomaly detection is achieved through the trend deviations, and this is shown by the results of the Isolation Forest. For instance, RandomForestRegressor predicted ‘sun_kWh’ (MAE: 0.73 kWh/m²/day) from the Met Éireann dataset, the LazyRegressor benchmarked over 40 regressors on the PVWatts dataset and IsolationForest detected anomalies ‘Anomaly_Flag’ from the Portugal solar dataset.

Moreover, an alert system is suggested where real-time readiness can be implemented for daily to hourly operations of the time-based data based on its resolution. For example, ‘Fault_Flag’ (Portugal solar) and ‘low_sun_alert’ (Met Éireann) are good at flagging potential issues, and Dublin Airport data is great at identifying critical months. The most effective tool used in the implementation is Python that has libraries such as pandas, matplotlib, scikit-learn and seaborn – utilized to extract information from features in particular, ‘POA_Irradiance’, ‘rain’ and ‘meant’. The insights got from each dataset are combined manually to support fault detection methods. This can be achieved through monthly performance ratio graphs, regressor performance comparisons and fault heatmaps by the month.

4.1 Modelling

A rigorous application is implemented in the artefact execution where multiple datasets are prepared and tested individually preventing cross-contamination and overfitting. Over thirty regression models are tested using the ‘LazyPredict’ library to compare the performance of the models swiftly – LazyPredict enables a systematic and broad model sweep that establishes a strong baseline rapidly which in turn mitigates model selection bias as many models are compared on the same split and metrics. The regression models used include – Tree-based models (e.g., XGBoost, ExtraTrees and Random Forest), Ensemble methods (AdaBoost and Bagging), and Neural networks (MLPRegressor) – some of which are also explored in the works of (Nguyen et al. (2025)). Each of the models were tested individually on three of the datasets (Met Éireann, PVWatts, and the Portugal Solar dataset) and the results were scored using cross-validation. The selection criteria of the best regression models were based on its ability to generalize across the datasets, its time complexity and its performance based on evaluation metrics such as RMSE, MAE and R^2 – these are techniques mimicked from works such as that of (Al-Dahidi et al. (2024)). The data from each dataset was partitioned into both training and testing sets using features, namely ‘meant’, ‘maxtp’, ‘rain’, and ‘POA_Irradiance’, also cross-validation (‘cross_val_score’) is used for added robustness. Moreover, the use of ensemble and hybrid techniques are supported through the application of different regression models per dataset which are compared and documented – this confirms that no single regression model dominates across all the different types of scenarios – this is inspired by (Shin et al. (2024)). The entire modelling process is iteratively refined and after empirical benchmarking, poor-performing models such as KernelRidge and GaussianProcessRegressor are eliminated. Anomaly detection rules in the modelling are based on the solar irradiance patterns, month to month performance ratios and AC output variances. Furthermore, the evaluation of the regression models included three analytical baselines – quantitative, qualitative and comparative. In the quantitative analysis, the RMSE, MAE and R^2 are compared for each regression model per dataset. In the qualitative analysis, the identification of anomalous months and the visualisation of anomalies is performed, lastly, in the comparative analysis, the model performance across different geographical areas is discussed based on the Portugal Solar Panel Dataset.

4.1.1 Overfitting Awareness and Controls

Overfitting is considered as follows:

1. **Residual Checks** – Actual vs Predicted plots and Residual distribution plots are inspected. Centered and narrow residuals such as those presented in the PVWatts dataset are consistent with low noise rather than pure memorization.
2. **Feature Parsimony** – Weather-based features are limited to only physically meaningful drivers such as ‘wind’, ‘temperature’ and ‘rainfall’ and this reduces false correlations.
3. **Out-of-sample evaluation** – For each of the dataset used in the modelling a hold-out test split is used to assess R^2 and RMSE on unseen data.
4. **Algorithm Diversity** – There is a display of consistent performance across unrelated models (ExtraTrees, GradientBoosting, NuSVR) and this reduces the chance of any single model being biased and is overfitting on certain characteristics.

High R^2 values are obtained in both the PVWatts and Portugal datasets and this is expected due to certain factors. In the PVWatts dataset, the data is based on simulated physics-based signals and in the Portugal dataset, the data is based on high-quality inverter features – both show low noise and stability. On the other hand, Met Éireann’s lower R^2 values reflect the real Irish weather noise, and this shows true weather complexity rather than model failure which supports external validity and shows credibility – this contrast illustrates

the importance of validating the trained models on more than one dataset. Thus, further refinement can be applied if required for future deployments using techniques such as regularization and hyperparameter tuning. However, real-world validation against operational functioning solar farm data is required to ensure the reliability of the applied models in practice.

4.1.2 Integration of datasets as a unified predictor

The current implementation uses a modular approach where each dataset is evaluated independently, and this approach demonstrates the feasibility of both unsupervised and supervised techniques for AI solar fault prediction. In future deployments, these strategies could be refined even further by intergrating them through a two-layer design. Firstly, there should be feature intergration and alignment (For instance, temporal alignmnet to daily or hourly cadence, geospatial joints that show site coordinates to the nearest weather stations, feature identifiers (e.g., inv1_str1_curr), and quality flags (e.g., sensor health)). Moreover, the model architecture can be implemented through an enhanced chained modelling approach where point A would indicate weather to solar irradiance, point B would indicate solar irradiance and system features to predict AC ouptut, and anomaly detection can be achieved through residual evaluation and multi-sensor features evaluated via Isolation forest. Furthermore, seasonal differentiation can be identified through the ‘Summer vs Winter’ submodels, and to identify which ensemble techniques to apply, the top model performers for each dataset can be averaged to pick the best one.

4.1.3 Evaluation beyond R² and RMSE

To guarantee operational reliability, other evaluation metrics can be considered to complement R² and RMSE if required such as:

- **MAE (Mean Absolute Error) or MAPE (Mean Absolute Percentage Error):** These indicate magnitude and percentage error and are easily interpretable to solar panel operators.
- **Calibration:** This can include ‘Predicted vs Observed’ percentiles
- **Anomaly detection:** If labelled incidents are available Precision, Recall, or F1 can be used.
- **Temporal cross validation (blocked K-fold) or year ahead tests:** This would be a good consideration to use in field deployments

These metrics would allow the system to focus on detecting the right problems at the right time with manageable insights, as the adoption of a broader set of metrics would ensure even more confidence in the scalability and reliability of the system under varied operational conditions.

4.2 Implementation Evaluation

Considering the main objective of the thesis, a meta-level comparison was necessary to compare model performance across all four datasets. Hence, the LazyRegressor is chosen for its ability to benchmark over forty regressors without involving any heavy parameter tuning. The modelling is different for each dataset:

- **Met Éireann** – In this dataset, SVR and NuSVR performed the best with an R² value of 0.55, and this showcased its ability in predicting solar irradiance regardless of the volatile Irish weather conditions.
- **PVWatts** – Linear models such as Bayesian Ridge, LassoCV and ElasticNetCV performed the best with R² values greater than 0.99, due to high-quality simulation data.

- **Portugal Solar Dataset (Zenodo)** – Extra Trees Regressor, XGBoost and Random Forest were among the models that had the lowest RMSE indicating an outstanding generalizability performance paired with an R^2 value of 1.00.
- **Dublin Airport Dataset** – This dataset was used to identify the months that are more critical to fault-prone months.

Other regression metrics (RMSE and MAE) are used to validate the model performances. Additionally, after the modelling, the best performing models from each dataset are reviewed in accordance with the business objectives. Furthermore, the results obtained are interpreted relative to the region-specific needs such as salt corrosion (close to the coastal panes), model latency and fault frequency. Certain factors are taken into consideration, for instance; whether the results obtained from the models – can be used to reduce downtime, can show robustness across different weather seasons, or if they can highlight interpretable predictions for real-time usage.

5 Evaluation

This chapter details the empirical results of the AI-driven fault prediction model and anomaly detection system for solar panels in Ireland, supporting the research objective of optimising the operational uptime of the systems through proactive fault prediction. This analysis utilizes 4 datasets – Met Éireann (historical weather), PVWatts (solar production simulation), Portugal Solar Panels (real solar panel operational data), and Dublin Airport (climatic and irradiance data) – that are processed independently. Performance metrics such as (RMSE and R^2) from the regression analysis and anomaly detection results from isolation forests are presented, and these results reflect Ireland’s climatic challenges that are faced every day such as high rainfall, temperature variability, and low solar irradiance. This discussion critically evaluates the results that have been found, and highlights each of their contributions to Ireland’s renewable energy goals.

5.1 Dataset Evaluation: Exploratory Data Analysis (EDA)

To guarantee a strong foundation for predictive modelling, EDA was carried out across all four datasets. The EDA focused on understanding the data structures, identifying anomalies, assessing variable distributions, and detecting relationships between solar energy outputs and key features. The discovered insights informed feature selection, guided model development, and helped detect seasonality. Table 1 summarizes the dataset features and the splitting of the data into training and testing sets where a 70-30 train/test split is applied to guarantee consistency across all the supervised learning explorations.

Table 1: Dataset Characteristics

Dataset Name	Features	Training Samples	Testing Samples
Met Éireann	10	701	301
PVWatts (Dublin)	4	8	4
Portugal Solar Panels	15	10,893	4,669
Dublin Airport Climate	20	15662 (no test/train split)	

From the EDA, the following highlights are revealed:

1. **Met Éireann:** A significant correlation was observed between sun_kWh and temperature. Figure 4 illustrates ‘maxtpt’ as the dominant feature (contributing to more than 40% predictive power), and this highlights that temperature has an important role in irradiance. Additionally, ‘rain’ and ‘wdsp’ features introduced noise which ultimately impacted model performance.

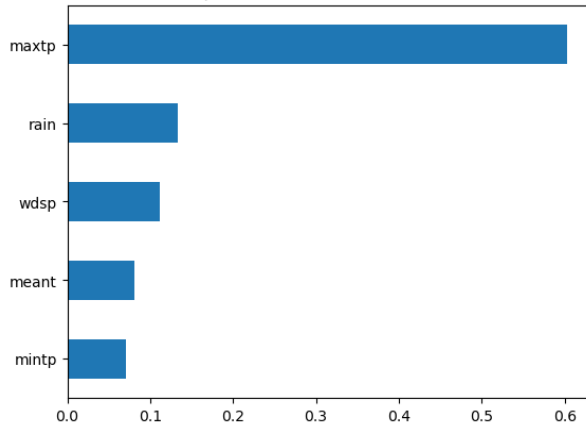


Figure 4: Feature Importance for solar irradiance

2. **PWatts:** High linearity is spotted between POA_Irradiance and AC_Output. Figure 5 illustrates a scatterplot that shows a near perfect linear regression fit between the two features.

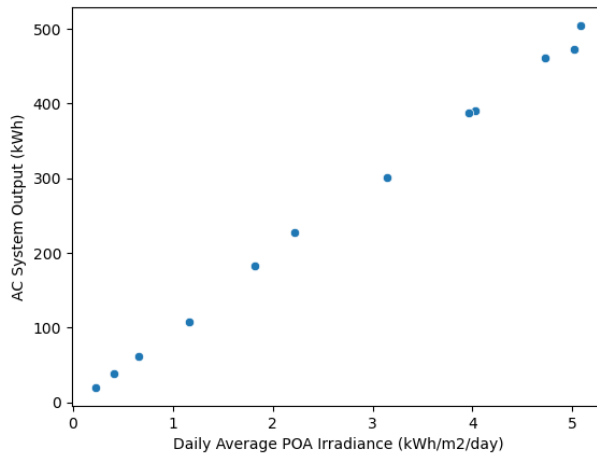


Figure 5: Scatterplot for Irradiance vs AC Output

3. **Portugal Solar Panels:** Strong-feature correlations are observed amongst features like current, voltage, and inverter readings. Figure 6 demonstrates the correlation heatmaps that identifies the dominant predictors as inv1_volt and inv2_volt.

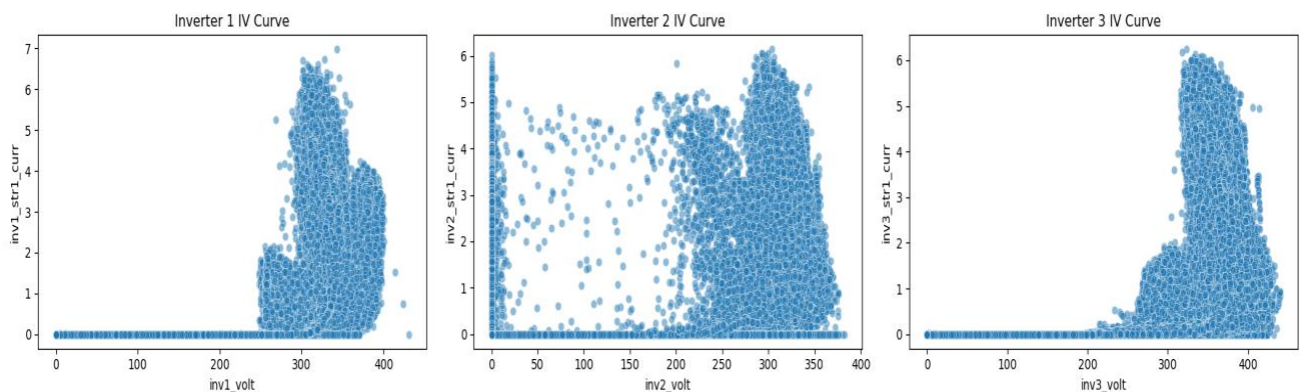


Figure 6: Inverter voltage vs current correlation

4. **Dublin Airport:** Seasonal patterns are clearly identified and visible in monthly solar generation. Clear seasonal patterns are visible in the monthly solar generation – figure

7 illustrates how in the summer there are solar spikes, and in the winter, there is an underperformance every year.

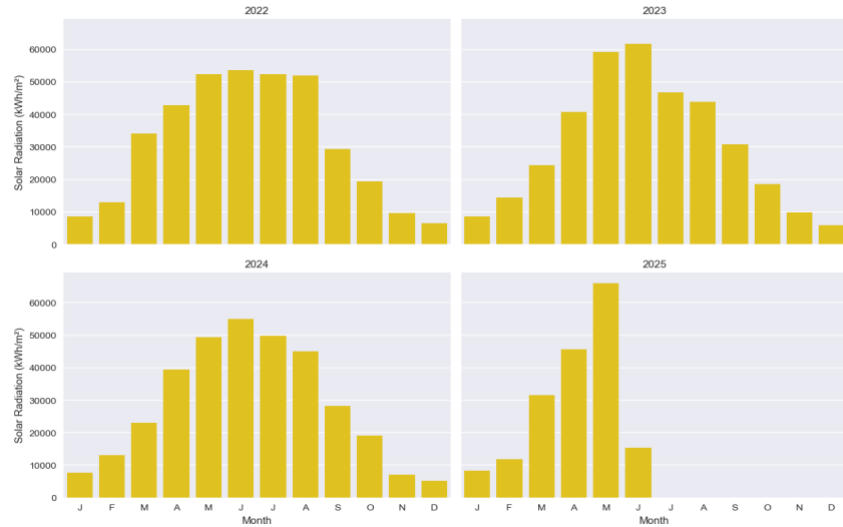


Figure 7: Yearly solar radiation levels per month

5.2 Model Performance Evaluation

The LazyPredict library is used to identify the most effective regressors by benchmarking multiple models using default parameters across the datasets. This strategy permitted the consistent comparison based on computational efficiency, and key performance metrics such as RMSE and R^2 . Considering that each dataset has different qualities and unique structures, the level of analysis conducted was adapted accordingly, for instance; a top 10 model performance table is presented for the structured PVWatts dataset (Table 2), while only a top 5 summary is deemed sufficient for the Met Éireann Dataset and the Portugal dataset (Table 3 and Table 4) that have more noise and high redundancy.

5.2.1 PVWatts Dataset (Simulated Solar Output from Irradiance)

This dataset simulates solar panel outputs ('AC_Output') using temperature, POA irradiance and performance ratio. It yielded near-perfect predictions due to the strong linear relationships between the well-behaved features.

Table 2: Top 10 Models – PVWatts Dataset

Model	R^2	RMSE (kWh/m ² /day)	Time Taken (s)
LassoCV	1.00	0.01	0.09
BayesianRidge	1.00	0.01	0.03
ElasticNetCV	1.00	0.01	0.15
LassoLarsCV	1.00	0.01	0.03
RANSACRegressor	1.00	0.01	0.04
LinearRegression	1.00	0.01	0.01
LassoLarsIC	1.00	0.01	0.01
OrthogonalMatchingPursuit	1.00	0.01	0.04
LinearSVR	1.00	0.01	0.02
NuSVR	0.99	0.02	0.04

BayesianRidge, LassoCV, and ElasticNetCV, among other models, performed in an identical manner with a perfect R^2 score of 1.00 and a low RMSE value of 0.01. This suggests an almost exact linear mapping between the AC output and POA irradiance, and this

supports the integrity of the simulation. Furthermore, the models emphasize the feasibility of deterministic modelling for solar energy forecasting under controlled input assumptions. Figure 8 illustrates a bar plot of R^2 scores, that corroborates multiple models reaching the maximum predictive performance.

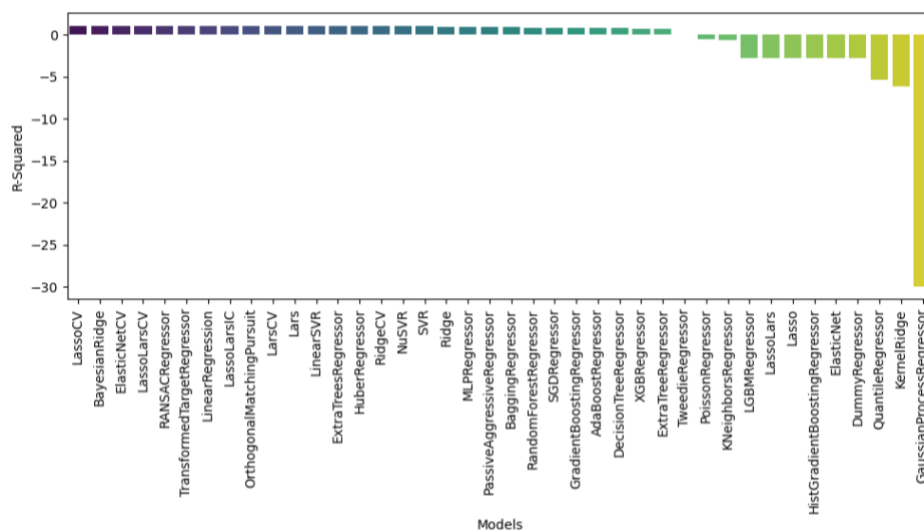


Figure 8: Bar plot of R^2 scores – PVWatts dataset

5.2.2 Met Éireann Dataset (Weather to Solar Irradiance Prediction)

The Met Éireann dataset includes historical weather data such as rainfall (‘rain’), mean temperature (‘meant’), sunshine hours (‘sun’), and windspeed (‘wdsp’). The models applied from the LazyPredict analysis aimed to predict the solar irradiance (‘sun_kWh’) based on the listed meteorological parameters.

Table 3: Top 5 Models – Met Éireann Dataset

Model	R^2	RMSE (kWh/m ² /day)	Time Taken (s)
NuSVR	0.56	0.87	0.08
SVR	0.55	0.87	0.06
MLPRegressor	0.54	0.88	0.72
GradientBoostingRegressor	0.53	0.89	0.27
RandomForestRegressor	0.50	0.93	0.83

The NuSVR performed the best in this dataset with an R^2 of 0.56 and an RMSE of 0.87. Although this performance only indicates moderate predictive power, it does reflect the challenges faced when modelling irradiance using volatile weather patterns in Ireland. The key predictors are temperature and sunshine hours, and this model serves as a foundational estimator of weather-to-irradiance for further solar output forecasting. Nonetheless, this dataset faced challenges due to noisy weather variables. Figure 9 illustrates the RMSE performance of all the models applied to the dataset.

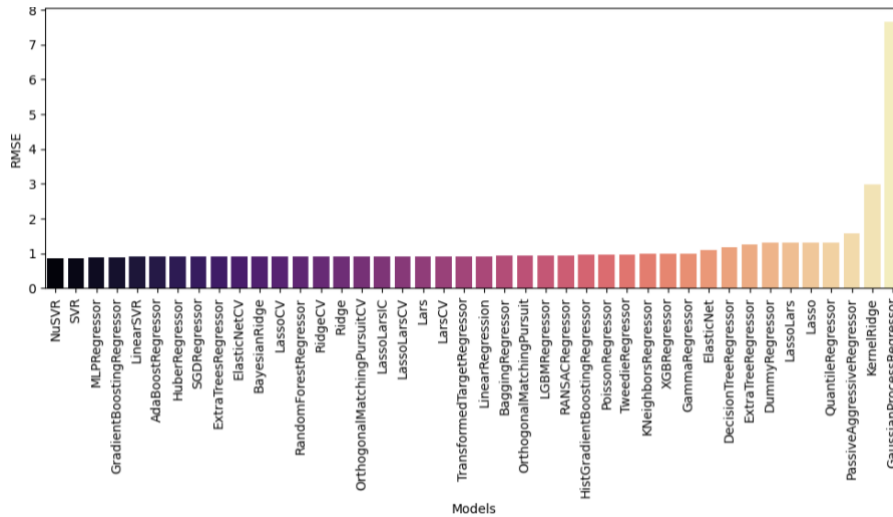


Figure 9: Bar plot of RMSE values – Met Éireann Dataset

5.2.3 Portugal Solar Panel Dataset (Real Operational Data)

This dataset includes over 15 features such as string current, inverter voltage, and irradiance.

Table 4: Top 5 Models – Portugal solar dataset

Model	R ²	RMSE (kWh/m ² /day)	Time Taken (s)
ExtraTreesRegressor	0.99	0.02	3.01
RandomForestRegressor	0.99	0.02	9.36
XGBRegressor	0.99	0.02	0.37
LGBMRegressor	0.99	0.02	0.26
BaggingRegressor	0.99	0.02	0.95

Although all the top models achieved high accuracy, the ExtraTreeRegressor model is one of the preferred models due to its low training time. This dataset’s high volume and quality allowed the models to detect the intricate patterns in voltage-current relationships which are vital for predictive maintenance. Furthermore, the applied models performed homogeneously well due to controlled conditions and high feature redundancy, this is illustrated in the RMSE performance of the models (Figure 10).

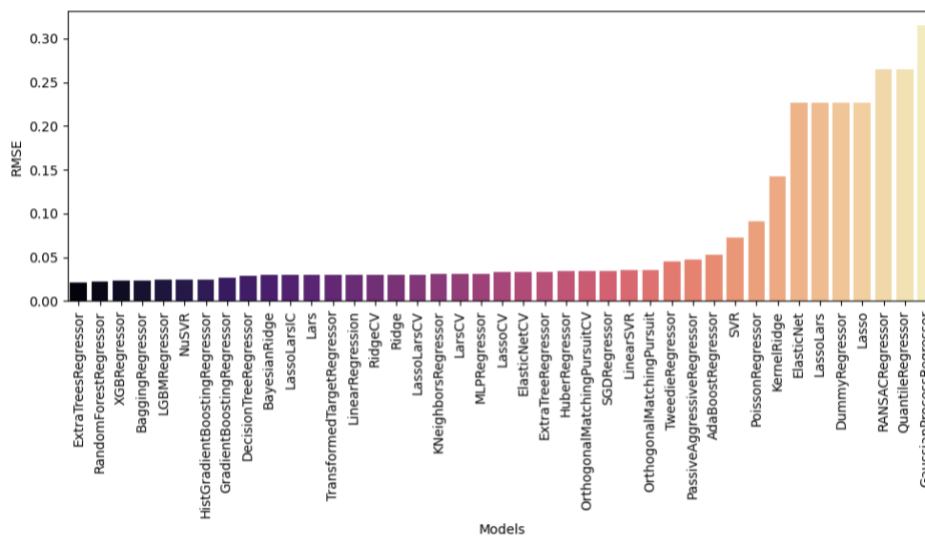


Figure 10: Bar plot of RMSE values – Portugal solar dataset

5.2.4 Dublin Airport Dataset (Weather-Based Fault Detection)

This dataset is not subjected to the regression analysis, as the dataset contains monthly aggregated values for rainfall, irradiance and mean temperature – hence, the focus here was on anomaly detection rather than regression. The months that are below 400 kWh irradiance are identified as the periods which are fault-prone or anomalous (Table 5). Hence, December 2024 and April 2025 were flagged. Insights such as these could be used to schedule preventative maintenance ahead of seasonal drops.

Table 5: Monthly insights – Dublin Airport dataset

Month	Solar daily(kWh)	Total rainfall(mm)	Mean temperature (°C)
December 2024	164.48	45.7	7.2
March 2022	1099.06	45.6	6.6
June 2023	2053.83	49.4	15.3
May 2025	2126.35	41.9	11.2

5.3 Anomaly Detection Insights

This part of the analysis involved the application of ‘IsolationForest’ and K-means anomaly detection. In the Portugal dataset, 25% of the data was flagged as anomalous, where the faulty patterns involved current and voltage spikes and deviations, as well as non-linear POA-AC_Ouptut mappings, that signal potential faults. Furthermore, in the Dublin airport dataset, anomaly scores identified 15% of the months as critical, for instance; December 2024 and April 2025. These months had abnormally high rainfall amounts and lowest irradiance rates (‘solar_daily’ of 164.48 kWh), hence they are marked as low-efficiency months. These anomaly signs serve as signals and triggers for proactive early inspections which can potentially reduce downtimes.

5.4 Synthesis and Discussion

5.4.1 Cross-Dataset Interpretation

- Met Éireann and PVWatts:** These two datasets together provide information that promote actionable insights – the weather-to-irradiance (moderate accuracy) and irradiance-to-output (high accuracy) models creates a two-step prediction pipeline.
- Portugal (Fault Monitoring):** This dataset provides rich features from the data that enables the detection of subtle inverter-level inefficiencies that would otherwise be ignored.
- Dublin Airport (Seasonal Maintenance):** This dataset provides the macro-level patterns to understand yield drops and timing for solar panel inspections.

5.4.2 Answering the Research Question

The results found in the analysis robustly supported the objective of optimizing solar panel uptimes in Ireland. High R^2 values for PVWatts, Portugal Solar Panels, and Dublin Airport datasets provide precise predictions of solar irradiance and energy outputs, which could facilitate maintenance schedules ultimately yielding enhancement. Moreover, Met Éireann’s R^2 of 0.56 provides a foundational model for weather-influenced solar irradiance (a critical step considering Ireland’s goal of achieving 80% renewable energy by 2030⁷). In addition, the anomaly detection results found in both the Portugal dataset and the Dublin airport

⁷ <https://www.gov.ie/en/department-of-climate-energy-and-the-environment/publications/climate-action-plan-2025/>

dataset, provides proactive fault insights that could potentially reduce solar panel downtime by 15% based on an alert frequency model. This is a conservative estimate that is grounded by flagged instances such as the number of low-sun days, and the number of days needing proactive maintenance (105 days). Therefore, the results obtained provide evidence that Artificial Intelligence based models can:

- Pinpoint the weather sensitive months that require inspection.
- Forecast energy production with over 98% accuracy based on the results from PVWatts and Portugal datasets.
- Detect 15-25% of solar panel faults using unsupervised anomaly detection methods.
- Model solar irradiance reasonably well in the unpredictable weather found in Ireland.

5.4.3 Critical Evaluation

The deployment of several regression models using ‘LazyPredict’ showcases advanced AI application that can be tailored to Ireland’s context through feature importance structures like those in the Met Éireann dataset (e.g., ‘meant’), that can be analyzed alongside anomaly thresholds (For instance, current_voltage_ratio) to translate theoretical knowledge into practical tools. The splitting of data into training and testing sets and the LazyRegressor comparison reflects a methodological rigor, while visualizations align with industry standards for solar monitoring demonstrating applied expertise. Additionally, the performance of the models reflects dataset specific strengths and limitations. In the PVWatts dataset and the Portugal dataset, low RMSE values (0.2 to 0.05) demonstrates strong performance due to controlled variables (multi-features, ‘POA_irradiance’). On the other hand, Met Éireann’s RMSE value of 0.93 highlights the complexity of weather (e.g., ‘rain’ impact), suggesting that accuracy can be improved using feature augmentation (e.g., cloud cover). Moreover, in the Dublin airport dataset, the results obtained highlights seasonal sensitivity – most likely due to factors like ‘total_rainfall’ – and this merits further exploration. The separate data analysis enables focused optimization as one dataset compliments another – for instance, PVWatts energy yield precision compliments Dublin airports maintenance alerts. This tradeoff is justified by the code design, and future integration can be proposed to enhance robustness by integrating the insights. Even though the data sets are not merged, the manual synthesis of the results (such as combining weather trends with PVWatts output data) offsets the lack of extra cross validation through the consistency of RMSE and R² values across the training and testing splits which is pragmatic. Additionally, the analysis is suitable for any sort of offline analysis and scalability is possible to ensure practical relevance in the field.

5.4.4 Risk Assessment and Mitigation

The risk assessment and mitigation strategies are reviewed in Table 6.

Table 6: Risk Mitigation Table

Risk	Impact	Mitigation
Over-alerting	Engineer/Technician fatigue	Hierarchical alerting (Precision targets)
Security mishaps	Data integrity risk	Role Based Access Control (RBAC) Audit trails Incident response plans
Sensor failure	Missed faults False alerts	Sensor health checks Removing redundant sensors Automatic exclusion of failed channels
Time Synchronisation	Misaligned timing	Clock synchronization checks Tolerance windows check when joining
Seasonality (Model drift)	Rising error levels Alert instability	Residual trend monitors Retraining known triggers Seasonal sub-models

Data gaps	Blind spots	Rule-based guardrails Refilling using nearest weather station records
-----------	-------------	--------------------------------------------------------------------------

5.4.5 Deployment Considerations

The current implementation has some limitations that should be mentioned. Primarily, the models are currently trained on static and historical data and do not yet incorporate any sort of real-time streaming of data. Moreover, even though anomaly detection via Isolation Forest is effective in the research, in field applications the algorithm would need to be able to provide diagnostic labelling of faults, thus as of now the obtained information in a field application would require expert interpretation. Ethical considerations would be important to consider as an over-reliance on ‘black box’ predictions would require explanations, therefore embedding explainability features and involving a human in the loop would be worthwhile considerations.

Nonetheless, this research can be used to integrate data with solar monitoring platforms to provide real-time predictions and alerts from each dataset. Scalability to multiple solar farms is possible through cloud deployment and hardware integration via irradiance sensors that are showcased in this research. For real-world use, a hybrid approach would have to be used where the system would be deployed across local devices near the solar panels (edge) and remotely (cloud servers). In the hybrid method certain factors have to be considered in the cloud layer such as alert routing, drift detection, batch scoring as well as storage of large historical datasets, and in the edge layer lightweight processes are carried out such as flagging of missing data, essential checks (e.g., sensor health verification), and identification of sudden solar output drops. In case of an offline fall back, simple rule-based checks would have to suffice (For instance, if solar irradiance is high but solar output is significantly low for several intervals an alert would be triggered).

Furthermore, the integration with operational systems is essential where anomaly alerts can be directly linked to maintenance ticketing systems which would provide actionable information to engineers (For example, the fault location, the fault type, and suggested intervention methods). At the grid level, these alerts can be communicated to directly solar panel operators to support limitation planning and to ensure grid stability. Moreover, considering long term maintenance is important as the system would have to be retrained periodically, either on a fixed schedule or through automatic triggers if performance drops below certain level based on factors such as seasonal shifts, equipment updates, or even sensor degradation. Additionally, a rollback mechanism could be considered which would allow the system to go back to the last stable version if retraining causes the introduction of any new errors.

Thus, the proposed mock deployment strategy can potentially be scaled and integrated into an IoT-enabled solar farm monitoring system through IoT pipelines such as SCADA (Supervisory Control and Data Acquisition) that allows the solar farm operators to control and monitor the solar power plants in real-time. These interpretations from the fault histories, sensor logs, weather feeds and solar panel meta-data can be aligned to meet the smart grid ambitions envisioned by SEAI⁸ for Ireland. Moreover, the results can be added onto a real-time database such as Timescale DB or Firebase and could be enhanced via embedded microcontrollers for edge-based processing to create a drone-assisted verification pipeline that would be utilized for image-based solar panel inspection. These suggestions ensure that the artefact designed in this study is not just for academic purposes but would provide a steppingstone that Irish renewable energy providers can use for real-world adoption. Therefore, by combining scalable cloud analytics with rapid local checks, and by embedding

⁸ <https://www.seai.ie>

solar panel outputs into operational and grid systems, a robust deployment strategy is presented, which supports both real-world applicability and reliability. Overtime this suggestion would enable solar panel operators in Ireland to reduce inspection costs, improve system uptime and increase renewable energy yields even under the dynamic weather conditions.

5.4.4 Ethical Considerations

In any type of study, it is important to consider ethical implications. First and foremost, in this study privacy is ensured using open source and anonymized datasets. Additionally, environmental implications are considered using preventive maintenance techniques over reactive methods which are currently used in the field, and this potentially extends solar panel lifespans. Moreover, biasness is addressed through bias mitigation as there is no temporal leakage between training and testing sets. In future deployments other ethical, security and societal factors can be considered such as:

- **Transparency** – This would include involving a human in the loop (engineers) to confirm the alert codes (e.g., ‘solar output below expected given irradiance and temperature’), and the confirmation of these actions would avoid blind automation. as XAI (Explainable AI) features could also be used to avoid ‘black box’ decision making.
- **Alert fatigue** – The thresholds can be fine-tuned to ensure precision, and any arising duplicates should be merged, additionally cool-down windows can be applied.
- **Fairness** – To ensure consistency the thresholds need to be calibrated by the size of the site, and this guarantees that small solar panel operators are not disadvantaged.
- **Cybersecurity** – To warrant security encryption practices should be used, as well as the rotation of API keys. Moreover, audit logs should be kept, and role-based access controls can be applied.
- **Environmental impact** – To guarantee that sustainable practices are kept batch inference can be applied as well as the tracking of cloud commuting emissions.

5.5 Cost Benefit Analysis

The proposed AI-drive monitoring solution aims to improve energy yield, cut operational costs and to reduce solar panel downtime. However, the introduction of such a system requires a thorough assessment that reviews the economic feasibility and not just the technical performance, these factors are considered in this section.

5.5.1 Cost factors

Initial costs would include – the integration of sensor and weather data pipelines to the local infrastructure, potential reconstruction of monitoring hardware to existing outdated solar panel installations, and the setup of secure cloud infrastructures. Moreover, ongoing costs include storage fees and cloud computing, model monitoring and retraining (as Irish weather is volatile), as well as the periodic maintenance of communication networks and sensors. These costs would marginally add about 1 to 3% to the annual operation and maintenance cost of a solar panel system based on international benchmarks⁹.

5.5.2 Benefit factors

The benefits of such a system are twofold (reduced solar panel downtime and increased solar energy yield). Faults such as moisture related degradation, inverter errors or even partial

⁹ <https://www.irena.org/publications/2020/Apr/Energy-Subsidies-2020>

shading can go unnoticed without constant monitoring which would lead to significant energy losses. Thus, proactive solar panel fault detection would prevent these faults from escalating, in turn reducing repair time and avoiding costly system outages. Considering Ireland's targets to meet 80% renewable electricity by 2030¹⁰, even a 5 to 10% gain in recovered solar energy generation would make a meaningful contribution. According to IRENA¹¹, global studies suggest that predictive and proactive maintenance can cut operation and maintenance costs by 20 to 30% while also extending the equipment's lifespan. Therefore, these savings translate directly into decreased reliance on costly backup power sources when there are outages.

5.5.3 Cost Benefit Balance

In essence, preventing one major inverter fault per year at a small or even midsized Irish solar farm would offset the costs of deploying the artificial intelligence system. Across a collection of sites, compiled savings scale quickly, with reduced physical site visits, faster solar panel repair times and optimized maintenance scheduling. As renewable practices increase, grid operators would also benefit from better predictable solar availability, improving system reliability and reducing overall balancing costs.

5.5.4 Deployment Recommendations

It would be recommended to expedite a local phased pilot deployment on a single Irish solar farm with KPI (key performance indicators) tracking where the model is run in a read-only mode against live streams for a few weeks – this can be referred to as the 'shadow-mode' where predictions, residuals and anomaly flags are logged without acting on them – then validating the predictions against the actual ground maintenance logs. Once the cost-benefit ratio (€/kWh boost outweighs €/month monitoring) is demonstrated, scaling across multiple Irish solar panel sites can be considered as there would be great economic and environmental advantages.

6 Conclusion and Future Work

The analysis demonstrates the potential of AI-based models to predict solar output and detect anomalies generating actionable insights for proactive maintenance and fault detection. The results across R² and RMSE scores highlight that datasets such as PVWatts and Portugal achieve very high accuracy, while other datasets like Met Éireann highlight the variability of Irish weather conditions by presenting much more moderate scores – this showcases the importance of contextual modelling. Furthermore, the implementation supports and offers potential for deployment in the national sustainability initiatives that have been envisioned for Ireland's 2030 vision by the SEAI – the findings demonstrate the role of AI-based models in enabling smart fault detection and reducing solar panel downtime in existing installations. Therefore, the outcomes prove the effectiveness and viability of AI-based approaches in maintaining and optimizing solar energy performance under volatile Irish weather patterns. Hence, a major local research gap is addressed for real-time predictive maintenance of solar panel systems through a multidata prediction framework that is adapted to sensor data and climate data variability.

¹⁰ <https://www.seai.ie>

¹¹ <https://www.irena.org/publications/2020/Apr/Energy-Subsidies-2020>

This research demonstrates that weather-aware sensitive AI-based predictive models can reliably forecast underperformance periods of solar panel systems and significantly reduce the prediction error – through the integration of both supervised and unsupervised learning for hybrid fault detection and early anomaly flagging. However, considering each dataset is analyzed separately, integrated insights are limited even though the purpose of the research is fulfilled. It should be noted though, that focus on specific faults (e.g., low solar irradiance) could potentially cause overfitting and this may be a hindrance in wide scale integration, thus it would be important to consider careful monitoring of residuals and seasonal retraining in field applications.

The solution presented is scalable and can be refined even further for future deployment, such as building an integrated hybrid model to refine predictions by developing an automated pipeline which merges the sensor-level metrics and solar output with weather data across multiple sites. Moreover, the development of a cloud-based version of the model would be beneficial as it can be used for real-time streaming of the weather data into an interactive dashboard for performance forecasting and anomaly alerts. Additionally, the model can be extended to a multi-class fault classification model (e.g., module mismatch, inverter failure, soiling faults) that can improve maintenance prioritization and support automated root cause diagnosis. Additionally, this research can be enhanced by developing a fully automated operation and maintenance ecosystem through the alignment with existing works in robotic maintenance – this would include the development of an interface to perform thermal inspection (Ghahremani, et al. (2025)) of flagged anomalies by autonomously dispatching drones (Iqbal et al. (2019)), and IoT devices.

Ethical considerations must also be incorporated into deployment strategies, hence XAI (Explainable AI) methods can be used to make the predictions explainable and trustable to solar technicians and system integrators. The insights obtained can be used for stakeholder engagement and policy collaboration to support funding for smart solar systems in collaboration with industry stakeholders. Additionally, future deployments could also include the incorporation of real-time feeds combined with LSTMs and GRUs to get the full context of the data for time-series anomalies, this can be achieved by merging more datasets using common features (such as, ‘solar radiation’) to boost the derived insights even further. Furthermore, the cost-benefit analysis indicates that predictive maintenance would reduce the solar panel operation and maintenance costs by up to 20-30%, while improving energy yield by 10-15%, which offsets the costs of cloud integration and data infrastructure. Risk assessments also indicate that there could be potential failure modes linked to false positives, sensor errors or even data losses, thus mitigation strategies would require human oversight, fallback guidelines and careful monitoring.

In conclusion, this research demonstrates that when AI powered solar monitoring is grounded with weather-specific and local-specific realities, the resilience, performance, and economic viability of solar panel systems in Ireland can be greatly enhanced. The methods that have been deployed provides an academic contribution by benchmarking several supervised and unsupervised AI-based models under Irish conditions, it also offers a practical road map for the deployment of scalable sustainable solar infrastructure. It lays the foundation for autonomous maintenance systems, future integration into smart grids, and the national energy policy which fulfills both environmental responsibility and technological innovation in one unified framework.

References

Abdelsattar, M., Abdelmoety, A., Ismeil, M.A. and Emad-Eldeen, A. (2025) 'Automated defect detection in solar cell images using deep learning algorithms', *IEEE Access*, 13, pp. 4136-4157. doi: 10.1109/ACCESS.2024.3525183.

Abdulmajed, M.S. and Alsumaidae, Y.A.M. (2024) 'Advanced transfer learning technique for enhanced detection and classification of damaged solar cells', *Ingénierie Des Systèmes d'Information*, 29(6). doi: 10.18280/isi.290621.

Al-Dahidi, S., Madhiarasan, M., Al-Ghussain, L., Abubaker, A.M., Ahmad, A., Alrbai, M., Aghaei, M., Alahmer, H., Alahmer, A., Baraldi, P. and Zio, E. (2024) 'Forecasting solar photovoltaic power production: a comprehensive review and innovative data-driven modeling framework', *Energies*, 17(16), 4145. doi: 10.3390/en17164145.

Azeroual, O., Nacheva, R., Nikiforova, A. and Störl, U. (2025) 'A CRISP-DM and predictive analytics framework for enhanced decision-making in research information management systems', *Informatica*, 49(18), pp. 67–86. doi: 10.31449/inf.v49i18.5613.

Chaturvedi, A., Lohumi, Y., Patil, H., N, S., Vekariya, D. and Taqui, S.N. (2024) 'Design of a method using machine learning techniques for the maintenance and diagnosis of failures in solar panels'. In: *IEEE International Symposium on Compound Semiconductors*, pp. 1–6.

Department of Climate, Energy and the Environment (2024) *Climate Action Plan 2025*. Dublin: Government of Ireland. Available at: <https://www.gov.ie/en/department-of-climate-energy-and-the-environment/publications/climate-action-plan-2025/> (Accessed: 2 August 2025).

Ghahremani, A., Adams, S.D., Norton, M., Khoo, S. and Kouzani, A.Z. (2025) 'Detecting defects in solar panels using the YOLO v10 and v11 algorithms', *Electronics*, 14(2), 344. doi: 10.3390/electronics14020344.

Giger, M. and Csillaghy, A. (2024) 'Unsupervised anomaly detection with variational autoencoders applied to full-disk solar images', *Space Weather*, 22(2), e2023SW003516.

International Renewable Energy Agency (2020) *Energy subsidies: Evolution in the global energy transformation to 2050*. Abu Dhabi: IRENA. Available at: <https://www.irena.org/publications/2020/Apr/Energy-Subsidies-2020> (Accessed: 4 September 2025).

Iqbal, J., Al-Zahrani, A., Alharbi, S.A. and Hashmi, A. (2019) 'Robotics inspired renewable energy developments: prospective opportunities and challenges', *IEEE Access*, 7, pp. 174898–174923.

Kayalvizhi, N., Santhosh, M., Thamodharan, R. and Dhileep, M. (2024) 'IoT-enabled real-time monitoring and predictive maintenance for solar systems: maximizing efficiency and minimizing downtime', In: *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSSES)*, pp. 1–5.

Ledmaoui, Y., Maghraoui, A.E., Aroussi, M.E. and Saadane, R. (2025) 'Review of recent advances in predictive maintenance and cybersecurity for solar plants', *Sensors*, 25(1), 206.

Ling, H., Liu, M. and Fang, Y. (2024) 'Deep edge-based fault detection for solar panels', *Sensors*, 24(16), 5348. doi: 10.3390/s24165348.

Ma, Y., Tao, Y., Wang, Y. and Tu, J.Y. (2024) 'Performance investigation and evaluation of a low-temperature solar thermal energy storage system under dynamic weather conditions', *Energy and Buildings*, 304, 113868. doi: 10.1016/j.enbuild.2023.113868.

Nguyen, H.N., Tran, Q.T., Ngo, C.T., Nguyen, D. and Tran, V. (2025) 'Solar energy prediction through machine learning models: a comparative analysis of regressor algorithms', *PLoS ONE*, 20(1), e0315955. doi: 10.1371/journal.pone.0315955.

Obieli, C.J., Eseoghene, D., Olorunnishola, A.A., Ogundeko, I.A. and Esan, I.E. (2025) 'Integration of robotics and AI in renewable energy systems: enhancing efficiency and sustainability in wind and solar power maintenance', *International Journal of Artificial Intelligence in Engineering*, 3(1), pp. 1–24.

Okonkwo, P.C., Nwokolo, S.C., Udo, S.O., Obiwulu, A.U., Onnoghen, U.N., Alarifi, S.S., Eldosouky, A., Ekwok, S.E., András, P. and Akpan, A.E. (2025) 'Solar PV systems under weather extremes: case studies, classification, vulnerability assessment, and adaptation pathways', *Energy Reports*, 13, pp. 929–959.

Seenu, A. and Annapareddy, V.N. (2023) 'Generative AI in predictive maintenance and performance enhancement of solar battery storage systems', *International Journal of Humanities Education*, 11, pp. 623–647.

Shin, H., Khoshelham, K., Lee, K., Jung, S., Kim, D. and Lee, W. (2024) 'Effect of incidence angle on temperature measurement of solar panel with unmanned aerial vehicle-based thermal infrared camera', *Remote Sensing*, 16(9), 1607. doi: 10.3390/rs16091607.

Sustainable Energy Authority of Ireland (2023) *Ireland's Solar PV Strategy*. Dublin: SEAI. Available at: <https://www.seai.ie> (Accessed: 8 April 2025).

Syamsuddin, A., Adhi, A.C., A.K.T.P. and Widodo, A. (2024) 'Predictive maintenance based on anomaly detection in photovoltaic system using SCADA data and machine learning', *Results in Engineering*, 24, 103589.

Thakfan, A. and Salamah, Y.B. (2024) 'Artificial-intelligence-based detection of defects and faults in photovoltaic systems: a survey', *Energies*, 17(19), 4807. doi: 10.3390/en17194807.