Time-To-Recovery Modelling for Photovoltaic Cells

Elia Abu-Manneh | Rutvij Rana

Abstract

This study presents a framework for modeling the time-to-recovery of utility-scale photovoltaic (PV) systems following adverse weather events. While prior research has focused on system degradation and fault detection, the recovery phase has remained underexplored despite its importance for operational forecasting and resilience planning. Leveraging a dataset of over 800 PV plants across the U.S., we applied survival analysis and machine learning techniques to estimate the duration required for a system's performance ratio (PR) to return to baseline after disruption. Feature engineering involved clustering, time labeling, and the creation of recovery episodes. Multiple models were evaluated—including XGBoost Cox, DeepSurv, Random Forest Survival, Logistic Hazard, and DeepHit—using the concordance index (C-index) as the primary metric. XGBoost Cox achieved the highest performance (C-index = 0.8426), demonstrating the best ability to rank recovery durations. The study also evaluated linear regression, logistic regression, and multi-layer perceptron models as comparative baselines. Our findings highlight the value of survival-based approaches in solar operations and suggest that tree-based models offer the most practical balance between performance and interpretability for predicting PV system recovery.

Contents

1 Introduction

As the global transition toward clean energy accelerates, photovoltaic (PV) systems have become a cornerstone of sustainable electricity generation. These systems offer scalable, low-emission solutions, but they remain vulnerable to environmental stressors. Adverse weather events, such as snow, hurricanes, storms, and floods, can significantly disrupt the performance of PV systems, reducing efficiency and power output for days or weeks at a time [1].

While extensive research has been conducted on PV system design, degradation rates, and failure mechanisms [2][3][4], little attention has been paid to the recovery phase that follows a weather-induced disruption. Understanding how long it takes for a PV system's performance ratio (PR) to return to its pre-event baseline can help improve energy forecasting, grid stability, and operational planning.

Climate change further amplifies this concern. The frequency and severity of extreme weather events are expected to increase globally, posing new challenges to renewable infrastructure [6]. Simultaneously, modern grids are relying more on intermittent power sources, demanding precise forecasting and robust system resilience [5]. Despite these needs, recovery modeling has remained largely absent in academic and industry tools.

1.1 Problem Statement

Current PV analytics focus on fault detection, degradation analysis, or long-term performance forecasting [2][3]. However, they do not estimate the time-to-recovery after a short-term performance drop caused by an adverse weather event. This leaves a critical gap in the tools available to PV operators and grid managers, who lack the ability to anticipate when systems will return to normal operation after disruptions.

1.2 Project Objective

This project aims to develop a data-driven framework to model the time-to-recovery of utility-scale photovoltaic systems following adverse weather events. Specifically, it seeks to:

- Estimate the number of days required for a PV plant's performance ratio (PR) to return to its pre-event baseline.
- Apply survival analysis techniques to treat recovery as a time-to-event problem.

- Apply probabilistic estimation techniques to predict the likelihood of an extreme weather event causing a power outage.
- Generate actionable insights for improving resilience, forecasting, and maintenance scheduling in solar energy operations.

This work contributes to both academic literature and practical utility operations by modeling PV system recovery in a way that accounts for environmental context and performance behavior over time.

2 Data

The dataset used in this study consists of daily-level operational and environmental data collected from over 800 utility-scale photovoltaic (PV) plants distributed across the United States. These data were compiled from internal PV operator performance records and integrated with external meteorological datasets such as the Global Historical Climatology Network (GHCN) and PRISM Climate Group. To ensure reliability and consistency in analysis, the study focuses on a nine-month window from August 2018 to April 2019. This period was selected based on completeness of reporting, consistency across sites, and observed seasonal variability. The final dataset offers a diverse set of observations across multiple climate zones, enabling a robust analysis of performance recovery following adverse weather events.

2.1 Key Variables and Features

The core of the data set is the performance ratio (PR), which serves as the primary outcome variable. Defined according to IEC 61724 standards, the PR measures the ratio of actual to theoretical energy output under ideal conditions, thereby indicating system efficiency. Alongside PR, the dataset includes indicators of adverse weather events such as snow, hurricanes, storms, floods, rain, and lightning. These are recorded as both binary flags and continuous measures, depending on the event type. Additional contextual information includes plant-level metadata such as capacity class, system age, and site-specific classifications (e.g., NOAA climate regions, thermal zones, and humidity zones). Daily weather features, including precipitation, temperature, and humidity, further enrich the data set help capture the environmental conditions that influence system performance.

2.2 Feature Engineering

To enable effective modeling of recovery times, several engineered features were developed.

- A key derived metric, *PR_change*, measures the daily change in performance ratio, allowing for the identification of sudden performance drops or rebounds.
- An *event onset flag* was created to mark the start of each adverse weather event, which serves as the baseline reference point for survival analysis.

Unsupervised learning techniques were also applied to enhance data representation. Principal Component Analysis (PCA) was used for dimensionality reduction, and K-means clustering was subsequently applied to the principal components to group PV plants with similar characteristics. These clusters were used as categorical features in downstream modeling to capture latent patterns in plant performance and environmental behavior.

Further feature engineering techniques will be discussed in the Methodology section where they are framed in the context of model input preparation and survival labeling.

2.3 Exploratory Data Analysis (EDA)

The EDA phase aimed to improve our understanding of the structure and behavior of photovoltaic (PV) system performance under adverse weather conditions, identify potential data quality issues, and inform downstream modeling.

2.3.1 Regional Distribution of Weather Events

Initial exploration of weather event frequency revealed stark discrepancies in reported event counts across NOAA climate regions. A bar chart visualizing the number of events per region (see Figure 1) showed that the West, Southeast, and Upper Midwest experienced the highest volume of weather-related disruptions, while regions such as Hawaii and the South reported no events. Given the improbability of zero weather events over a full year, these regions were excluded from further analysis due to likely data collection issues.

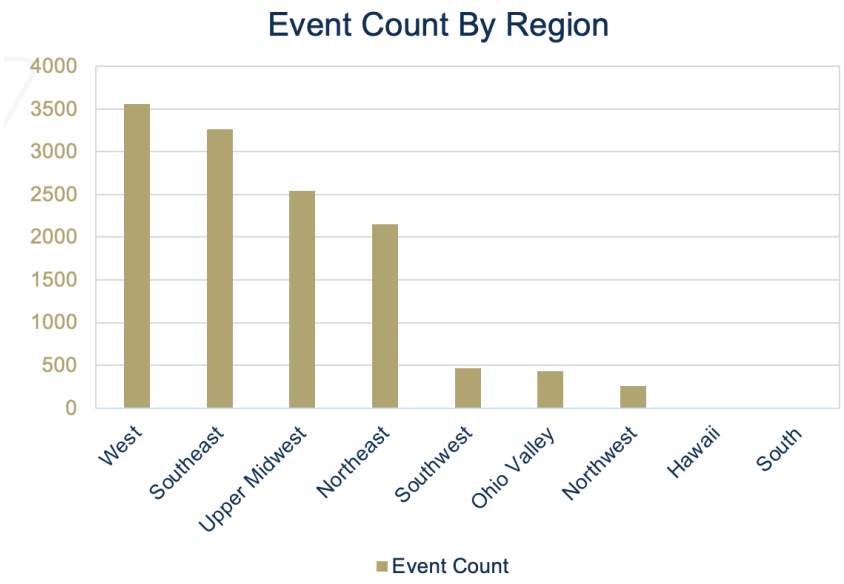

Figure 1: Hawaii and the South were excluded due to missing or incomplete data.

2.3.2 Performance Ratio Trends Over Time

To examine performance patterns, a heatmap of the monthly average performance ratio (PR) was generated across all regions (Figure 2). This visualization confirmed regional variability in PV system efficiency and supported the decision to focus the analysis on a consistent time window from August 2018 to April 2019. This period showed sufficient data density and seasonal coverage, making it well-suited for survival modeling.

2.3.3 Impact of Weather on Performance

As mentioned earlier, *PR_change*, was developed to capture the day-to-day difference in performance ratio. This metric helps identify abrupt performance shifts and provides a basis for defining recovery trajectories. A time-series plot of PR_change in the Ohio Valley region (Figure 3) shows a strong alignment between adverse weather events and large changes in PR. This finding empirically supports the hypothesis that weather events materially affect PV system performance and justifies the use of survival analysis to model time-to-recovery.

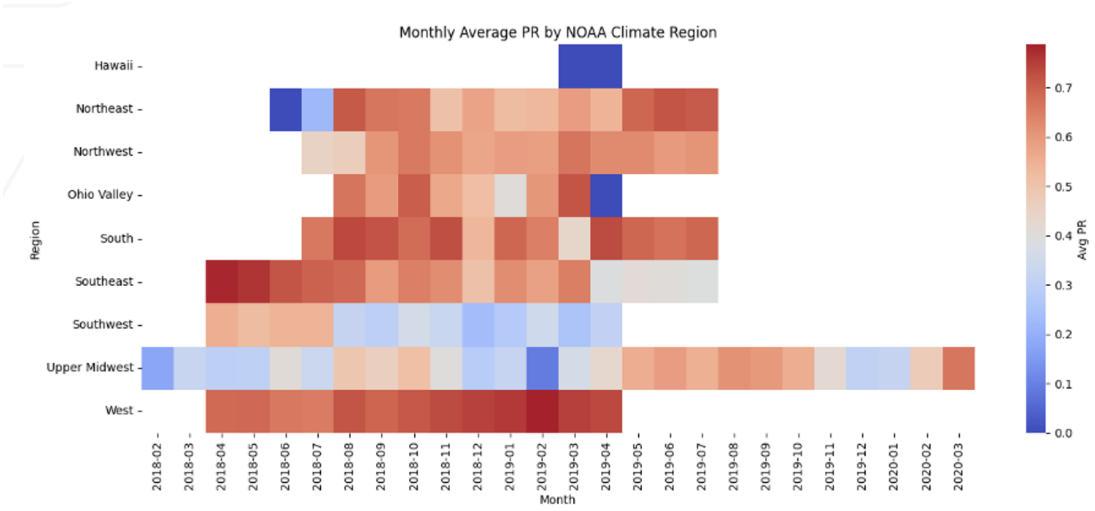

Figure 2: Monthly average performance ratio by region.

2.3.4 Clustering of Solar Plant Characteristics

To explore underlying structure in the dataset and mitigate the effects of multicollinearity, unsupervised learning methods were applied. Raw features were normalized to the range [-1, 1], and dimensionality reduction was performed using Principal Component Analysis (PCA).

K-means clustering models were created once using the normalised features and once using the PC components. The second approach produced more compact and meaningful clusters, as it effectively reduced the influence of correlated variables.

Table 1 lists the top features contributing to each principal component. For example, rainfall-related features such as `rain_value_mm`, `rain`, and `nearest_rain` were compressed into a single dimension, capturing shared variability and reducing the number of features needed for downstream analysis.

Table 1: Top features contributing to each principal component

PC1	**PC2**	**PC3**
rain_value_mm	snow_value_mm	flood
rain	total_daily_snow_mm	nearest_flood
nearest_rain	low_irradiation	cumulative_snow_mm

These engineered cluster features were later integrated into model inputs as categorical vari-

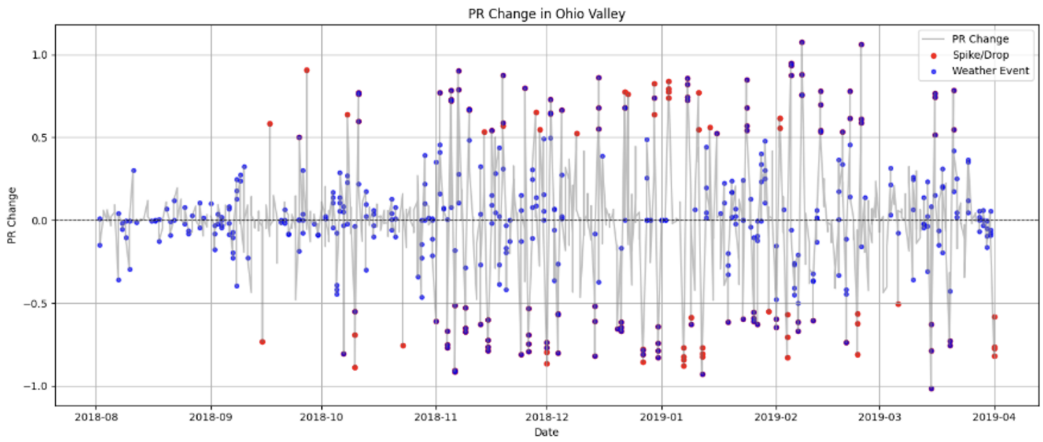

Figure 3: PR_change time series in the Ohio Valley.

ables, helping to account for regional and climatic variability.

3 Methodology

3.1 Model Selection and Rationale

To model the time-to-recovery of PV systems, we evaluated several survival analysis techniques that balance predictive power, flexibility and interpretability. The choice of models was based on their suitability for high-dimensional, time-to-event data, and their capacity to handle censoring and nonlinear relationships. For linear and logistic regression a feature engineering pipeline was created to adapt to regression relationships. For the censor handling models they were implemented using custom feature engineering pipelines and validated using the concordance index (C-index) as the primary performance metric. The models selected for this study are:

Table 2: Model Descriptions

Model	Description
Linear Regression	Linear Regression models the linear relationship between a dependent variable and independent variables by fitting a linear equation to data. It's used for predicting continuous outcomes and identifying the strength of relationships.
Logistic Regression	Logistic Regression is a statistical model primarily used for binary classification. It estimates the probability of an event by fitting data to a logistic function, transforming linear combinations into probabilities for classification.
Multi-Layer Perceptron Neural Network	A Multi-Layer Perceptron (MLP) Neural Network is a feedforward network with input, hidden, and output layers. Using non-linear activation functions, MLPs learn complex non-linear relationships, suitable for classification, regression, and pattern recognition.
Random Survival Forest (RSF)	RSF is a non-parametric ensemble learning method that extends random forests to right-censored survival data. It builds multiple decision trees where splits are chosen to maximize survival differences between nodes. The model does not assume proportional hazards or specific functional forms, making it well-suited for capturing complex, nonlinear interactions between features.
XGBoost Cox Model	XGBoost was adapted using a Cox partial likelihood loss function to model the hazard function. This approach preserves the efficiency and scalability of XGBoost while enabling time-to-event modeling under the proportional hazards assumption.
Logistic Hazard Model	The logistic hazard model, implemented using the PyCox library, estimates discrete hazards over pre-specified time intervals using a neural network architecture. Unlike Cox models, it allows for time-varying effects and does not rely on proportional hazards assumptions.
DeepSurv	DeepSurv is a deep learning-based extension of the Cox proportional hazards model that uses neural networks to learn a non-linear representation of the covariates. It maintains interpretability through its connection to the Cox partial likelihood while improving flexibility.
DeepHit	DeepHit directly estimates the probability mass function of event times using a multi-task neural network. It models both survival probability and time-specific risk in a unified architecture, enabling the capture of competing risks and multimodal recovery distributions.

Each approach is slightly different in modeling strategy for estimating recovery time distributions and was chosen to explore both classical and deep learning paradigms in survival analysis.

3.1.1 Feature Engineering and Preprocessing for Censor Handling Models

Before model training, a preprocessing pipeline was developed to convert raw time-series data into a structured format suitable for survival analysis. A key step in this process was the creation of the episodes_df, a tabular dataframe in which each row represents a distinct recovery episode following a detected adverse weather event. This transformation was important, as it aligned each sample with a start time (event onset), a recovery duration or censoring point, and corresponding covariates, making it compatible with models that expect time-to-event format input.

All features were scaled and encoded appropriately, with discrete time labels generated for models requiring interval-based training (e.g., DeepHit, Logistic Hazard). This standardized dataset enabled appropriate comparison across models while preserving the temporal and environmental context necessary to predict recovery time accurately.

4 Experiments and Results

4.1 Evaluation Metrics

4.1.1 Accuracy, Precision, Recall, F1-Score, RMSE, MAE

- **Accuracy:** Measures the proportion of correctly classified instances out of the total.
- **Precision:** Measures the proportion of true positive predictions among all positive predictions, indicating avoidance of false positives.
- **Recall:** Measures the proportion of true positive predictions among all actual positive instances, indicating the ability to find all positive cases.
- **F1-Score:** The harmonic mean of Precision and Recall, providing a balanced measure, especially for uneven class distributions.
- **RMSE (Root Mean Squared Error):** Measures the square root of the average of squared differences between predicted and actual values; a common regression error metric.

- **MAE (Mean Absolute Error):** Measures the average of the absolute differences between predicted and actual values; less sensitive to outliers than RMSE.

4.1.2 Concordance Index (C-index)

Model performance was assessed using the concordance index (C-index), a standard metric in survival analysis that quantifies how well the predicted ordering of recovery times aligns with the true order. The C-index ranges from 0.5 (random prediction) to 1.0 (perfect concordance), and reflects the model's ability to correctly rank the recovery times of censored and uncensored pairs. This metric was used across all models to ensure comparability and to guide hyperparameter tuning. In particular, changes to dropout, learning rate, number of time bins, and input dimensionality were evaluated based on their impact on the validation C-index.

4.1.3 Comparing the Evaluation Metrics

While traditional metrics such as accuracy, precision, recall, F1-score, MAE, and RMSE are valuable for evaluating classification and regression models in this study, particularly when predicting outage occurrence or recovery duration, they are not appropriate for survival models due to the presence of censored observations and the need to model time-to-event outcomes. In contrast, the Concordance Index (C-index) is specifically designed for survival analysis, measuring how well a model ranks recovery times without requiring uncensored ground truth for every instance. Unlike MAE or RMSE, which assess the magnitude of prediction error, the C-index evaluates the relative ordering of predicted durations, making it more robust in scenarios with partial event information. For this reason, the C-index was used as the primary evaluation metric for comparing survival models, ensuring consistent and meaningful performance assessment across varying architectures and assumptions. However, both types of metrics were essential for a holistic evaluation of model behavior and data characteristics across the different modelling approaches used in this study.

4.2 Result Summary

Table 3: Summary of Model Performances

Model	Goal	Metric	Score	Notes
XGBoost Survival Cox	Survival	C-index	**0.8426**	Best performer. High concordance with actual durations. Risk scores easily scaled to duration.
DeepSurv	Survival	C-index	0.7720	Lower than XGBoost. More variance in predictions; requires more data to generalize.
Random Forest Survival	Survival	C-index	0.7633	Moderate performance. Rank-based predictions noisy but directionally useful.
Logistic Hazard	Survival	C-index	0.7409	Slightly lower performance. Discrete bins make interpretation less intuitive.
DeepHit	Survival	C-index	0.6267	Poorest performer. Survival curves unstable; weak generalization.
Linear Regression (1)	Duration Regression	R^2 = 0.5993, MAE = 26.1	Moderate	Explains 60% of variance. Tends to underpredict long durations.
Linear Regression (2)	Classification	Accuracy = 0.8532	High	High precision, low recall. Misses many true outages.
Logistic Regression	Classification	Accuracy = 0.8585, AUC = 0.8241	High	Best classifier. Still suffers from high false negative rate.
MLP (1c)	Duration Regression	3-Day Accuracy = 58.0%	Moderate	Performs better than linear regression. Gains from more epochs. Less interpretable.

4.3 Model Outputs

This sub-section summarizes the output and behavior of each predictive model developed during the study. Models were trained for distinct objectives: estimating recovery duration (regression),

predicting outage occurrence (classification), and modeling time-to-event distributions (survival). Performance metrics were tailored to each modeling paradigm to ensure fair comparison.

4.3.1 Linear Regression Model 1:

This model was set up to predict the duration of a low-power/power outage event as a function of input features (weather and geographic).

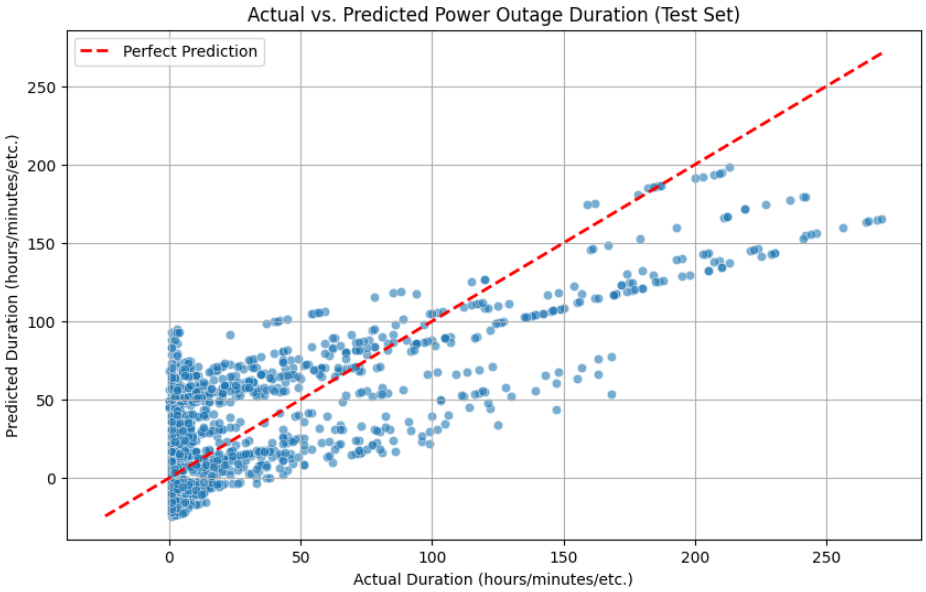

Figure 4: Linear Regression Model 1 - Actual vs Predicted Power Outage Duration

Evaluation Metric	Result
RMSE	33.5154
MAE	26.1156
R2	0.5993

Figure 4 shows that the linear regression model appears to predict shorter duration power outages more accurately than longer duration. This might be explained by the existence of more data points with shorter durations. The distribution of residuals in Figure 5 is not symmetric and shows a slight bias towards a positive residual. This means this model is more likely to underpredict the duration period. In a real-world application, this would be considered a liberal

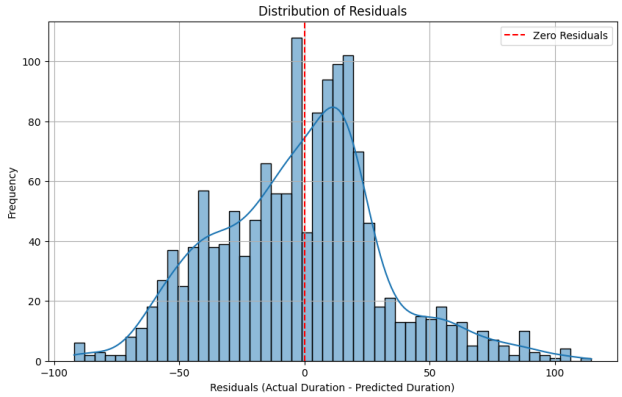

Figure 5: Linear Regression Model 1 - Distribution of Residuals

model and users would have to expect a great likelihood that a predicted outage duration will last longer than predicted.

The results table shows a decent R2 score of 0.5993 meaning the model can roughly explain 60 percent of the variance. However, the MAE of 26 days is rather high.

4.3.2 Linear Regression Model 2:

This model was set up to predict the likelihood of a low-power/power outage event occuring as a function of input features (weather and geographic) of a specific day.

Evaluation Metric	**Value**
Accuracy	0.8532
Precision	0.8100
Recall	0.1984
F1-Score	0.3187
Confusion Matrix	[[5797 57] [982 243]]

As a classifier, this model achieved an accuracy of 0.8532 meaning that it was able to accurately predict 85 percent of the time when a power outage was going to occur, which is quite useful for real-life applications.

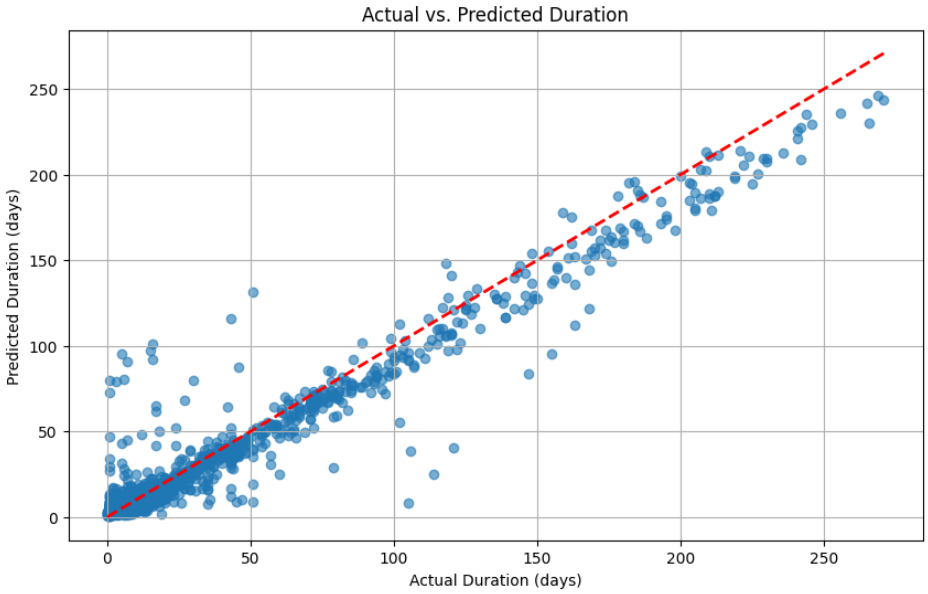

Figure 6: MLP Model 1c: Actual vs. Predicted Durations (days)

The high-precision score indicates a lower rate of false positives. However, the low recall score indicates a high rate of false negatives. This means that when the model does mispredict, it much more likely to be predicting no power outage when a power outage will occur.

4.3.3 Logistic Regression Model:

This model was set up to predict the likelihood of a lower-power/power outage event occuring as a function of input features (Weather and geographic) of a specific day.

Evaluation Metric	**Test Set**
Accuracy	0.8585
Precision	0.7636
Recall	0.2637
F1-Score	0.3920
ROC-AUC	0.8241

As a classifier, this model achieved an accuracy of 0.8585 on the test set, meaning that it was

able to accurately predict 85.85 percent of the time when a power outage was going to occur, which is quite useful for real-life applications.

The high precision score (0.7636) indicates a lower rate of false positives. However, the low recall score (0.2637) indicates a high rate of false negatives. This means that when the model does mispredict, it much more likely to be predicting no power outage when a power outage will occur. Overall, the logistic regression model performed very similarly to the second linear regression model.

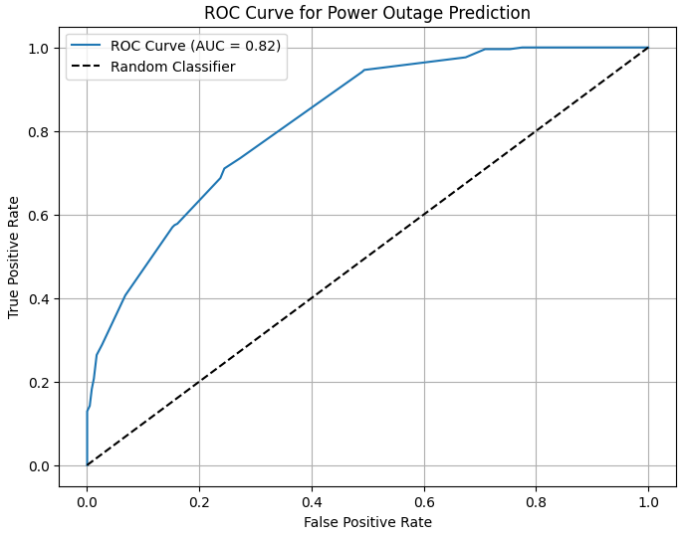

Figure 7: ROC Curve for Power Outage Prediction - Logistic Regression Model

The ROC (Receiver Operating Characteristic) curve is a common tool for evaluating the performance of binary classification models. It plots the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity) at various threshold settings. The curve rises from the bottom-left to the top-right, with a steeper curve towards the top-left indicating a better-performing model. The Area Under the Curve (AUC), given as 0.82, quantifies the overall performance, where a higher AUC (closer to 1) suggests that the model is better at distinguishing between the two classes (power outage vs. no power outage).

Figure 7 shows the distribution of predictions against the actual result. The logistic regression model uses p=0.5 as the threshold. The True Negatives (TN) can be interpreted as the blue bars left of the 0.5 threshold and False Positives are to the right of the 0.5 threshold. We can observer that there is a very low rate of false positive. Meaning that if this model does predict an outage, users have a confidence that it is correct.

The True Positives (TP) are the red bars to the right of the 0.5 threshold. The False Negatives

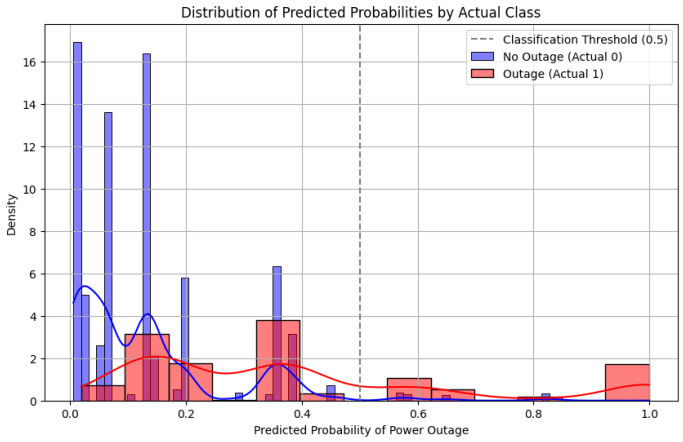

Figure 8: Distribution of Predictions - Logistic Regression Model

are the red bars to the left of the 0.5 threshold. We can observe that there is a relatively high rate of false negatives occurring. This means that when using the model, if the model predicts no power outage, there is a decent chance that there will be indeed a power outage.

4.3.4 Multi-Layer Perceptron Neural Net:

This model was set up to predict the duration of a power outage using all the normalized and OHE features as inputs. 3 different model architectures were created. Each model was trained using 4 different number of epochs. The purpose was to examine the effects of different number of hidden layers, the size of the hidden layers and the number of training epochs on the model accuracy.

The models were evaluated based the percentage of duration predictions made that fall within n-Days of the correct duration. These are referred to as the 1-day Accuracy, 2-Day Accuracy, and the 3-Day Accuracy.

Model Name	Num of Hidden Layers	Largest Layer as Ratio of Input Layer	Epochs	1-Day Accuracy	2-Day Accuracy	3-Day Accuracy
1a	3	4	32	22.53%	38.97%	50.86%
1b	3	4	42	27.40%	45.70%	57.70%
1c	3	4	52	28.40%	46.50%	58.00%
1d	3	4	80	27.70%	46.40%	57.70%
2a	3	2	8	14.70%	28.60%	41.10%
2b	3	2	32	23.80%	40.10%	50.00%
2c	3	2	42	26.70%	44.30%	56.20%
2d	3	2	52	26.50%	45.50%	57.30%
3a	4	4	8	14.90%	30.00%	41.20%
3b	4	4	32	26.20%	41.10%	52.30%
3c	4	4	42	27.90%	45.40%	57.00%
3d	4	4	52	22.10%	41.30%	53.00%

The results from the MLP Models shows that Model 1c with 3 hidden layers, with the largest layer being 4 times the size of the input layer, trained over 52 epochs achieves the best result. We can observed a clear pattern where more epochs of training makes the model more accurate. However, in some cases adding more epochs did not yield a major improvement and any differences in accuracy can be attributed to randomness introduced during the train-test data split. This can be seen for example when comparing models 1b, 1c and 1d.

4.3.5 XGBoost Survival (Cox)

This model was used to predict the time required for the performance ratio (PR) of a utility-scale photovoltaic system to return to its baseline after a disruptive weather event. A Cox proportional hazards framework was implemented using XGBoost to estimate recovery risk based on engineered environmental and operational features.

Figure 9 shows the predicted (scaled from risk score) versus actual recovery duration, where actual values have been log-transformed to improve interpretability. The scatterplot reveals that the model captures the expected inverse relationship: lower predicted risk scores correspond to longer recovery times. The clustering of samples around the lower risk–higher duration region supports this interpretation and reflects the model's ability to prioritize or rank sites by recovery urgency. Despite the Cox model's inherent limitation in predicting exact durations, the XGBoost

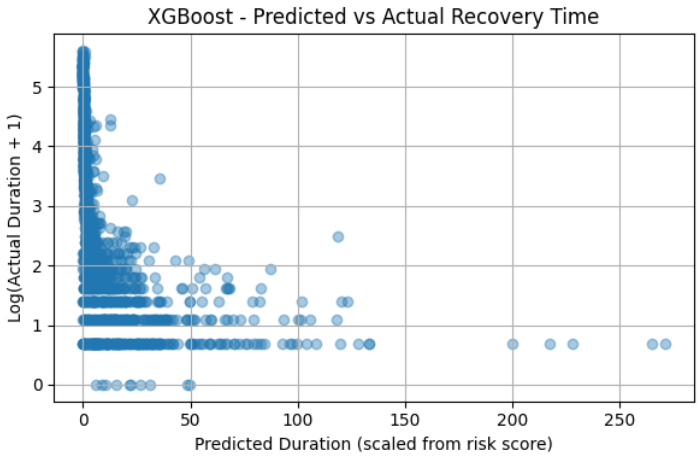

Figure 9: XGBoost - Predicted vs Actual Recovery Time (log scale)

implementation excelled at identifying relative differences in recovery risk. This makes it particularly well-suited for triaging operational responses post-weather events or prioritizing site inspections.The XGBoost Survival Cox model outperformed all other tested methods.

Evaluation Metric	**Result**
C-index	0.8426

Table 4: XGBoost Survival Model Performance

4.3.6 DeepSurv (Neural Cox PH)

The DeepSurv model uses a neural network to estimate individualized hazard rates under the Cox proportional hazards framework. Unlike traditional Cox models, DeepSurv captures nonlinear relationships between input features and recovery risk. This makes it a strong candidate for modeling photovoltaic recovery, where environmental and operational interactions may not follow simple trends.

Figure 10 shows the predicted survival curves for five test samples. The vertical axis represents the probability that the system will have not yet recovered by a given time. Curves that decay more slowly represent samples expected to recover later, while sharper drops indicate faster recovery. This individual-level modeling capability can support risk stratification across a solar portfolio. Despite strong theoretical advantages, the DeepSurv model performed slightly below the XGBoost Cox model in terms of concordance. It still demonstrated meaningful stratification

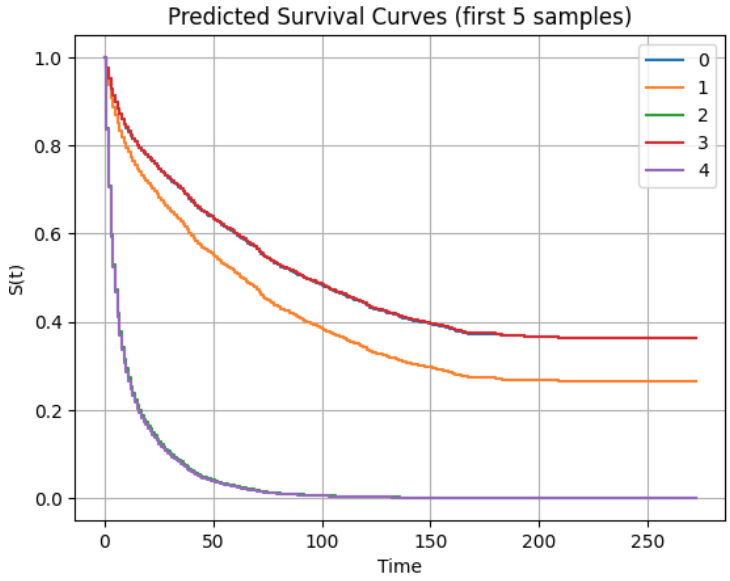

Figure 10: DeepSurv - Predicted Survival Curves for First 5 Samples

of recovery risk, but may have been constrained by data volume or overfitting given its complexity. In operational settings, DeepSurv could be useful when personalized recovery risk curves are needed for high-value or critical assets.

Evaluation Metric	**Result**
C-index	0.7772

Table 5: DeepSurv Evaluation Results

4.3.7 Random Forest Survival

The Random Forest model was trained using the scikit-survival package, leveraging a non-parametric ensemble of decision trees to rank the relative risk of delayed performance ratio (PR) recovery following disruptive weather events. Rather than estimating exact recovery times, the model provides relative risk scores that were converted into ranks for visualization purposes.

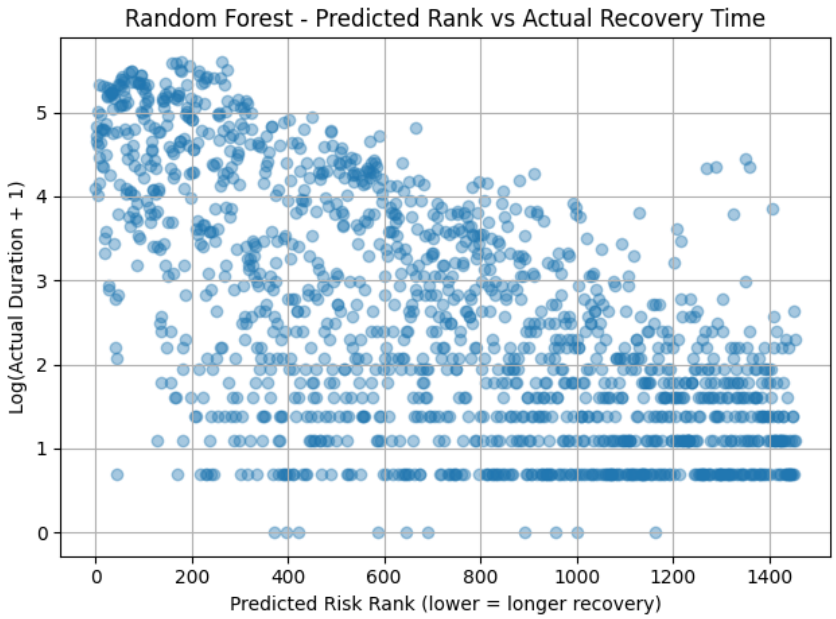

Figure 11: Random Forest - Predicted Rank vs Actual Recovery Time

Figure 11 plots the predicted rank (derived from risk scores, where a lower rank indicates a longer expected recovery) against the log-transformed actual recovery duration. A downward trend is evident, samples with lower risk ranks generally exhibit longer recovery times, which indicate that the model is effectively ordering cases in the correct relative direction. However, a wide vertical spread of actual durations is seen within most ranks, particularly at the lower (i.e., higher-risk) end. This suggests that while the Random Forest model captures relative ordering well, it struggles to discriminate finely between samples within similar risk bands. The dense clustering of short-duration cases near the bottom of the plot also reveals class imbalance, which may have biased the model towards overestimating the risk of quick recovery. The model achieved a C-index of 0.7633, supporting its utility for survival ranking tasks, but its limitations in handling noisy and skewed duration distributions reduce its suitability for applications requiring precise time predictions. Nevertheless, it remains an interpretable and robust baseline for risk-based stratification.

Evaluation Metric	**Result**
C-index	0.7633

Table 6: Random Forest Survival Evaluation Results

4.3.8 Logistic Hazard

This model was built using a discretized version of time-to-event data to predict the probability of photovoltaic system recovery within specific time intervals. It segments the time horizon into discrete bins and estimates the hazard (failure) probability in each bin, making it suitable for irregularly distributed recovery events across time.

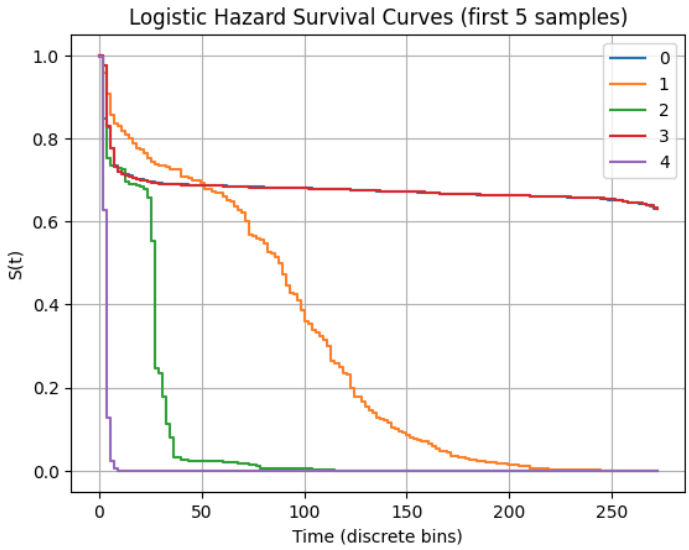

Figure 12: Logistic Hazard Model - Predicted Survival Curves (First 5 Samples)

The survival curves in Figure 12 show varying shapes for different test samples, indicating the model is responsive to distinct underlying patterns in the data. For example, some samples predict rapid recovery (sharp drop), while others expect prolonged recovery periods (gradual decline). However, the model often underestimated the complexity of long recovery times, with most curves flattening prematurely near zero probability. This may be due to over-discretization or insufficient gradient updates in rare-duration bins. Additionally, although the model achieved a moderate C-index of 0.7460, it struggled with calibration across high-recovery durations. In future iterations, tuning the number of discretized time intervals or using a hybrid model with continuous outputs may help resolve this. Ultimately, the Logistic Hazard model performed best at capturing short-to-moderate recovery events but lacked the temporal resolution needed for precise long-duration modeling.

Evaluation Metric	Result
C-index	0.7460

Table 7: Logistic Hazard Model Evaluation Results

4.3.9 DeepHit

The DeepHit model, a deep neural network approach that estimates the probability of survival over discrete time intervals. The model was trained using engineered features related to environmental and operational conditions of photovoltaic (PV) systems, with the goal of predicting recovery duration following adverse weather events.

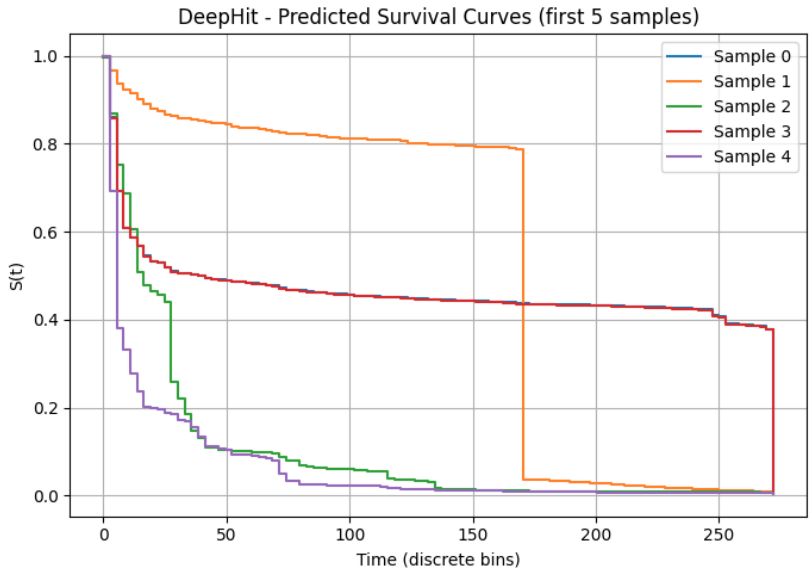

Figure 13: DeepHit - Predicted Survival Curves (first 5 samples)

Figure 13 shows the predicted survival probabilities over time for five test samples. These curves reflect how the model estimates the likelihood of system recovery persisting beyond each time step. The diversity in survival patterns—such as a sharp drop for Sample 4 compared to the gradual decline in Sample 1—illustrates the model's ability to account for heterogeneous recovery profiles across different weather-affected PV systems. Despite the model's advanced design, the DeepHit implementation struggled to surpass traditional survival models, achieving a C-index of approximately 0.63. This lower performance may be attributed to the model's sensitivity to hyperparameters, the discrete time binning used, and a relatively small dataset size for a deep

learning model. Additionally, some predicted survival curves flatten too early or late, potentially reflecting over- or underestimation of certain recovery durations. Nonetheless, DeepHit offers high flexibility and may become more competitive with further tuning or with access to a larger training corpus. Its survival function outputs could also be integrated into broader maintenance planning tools, offering probabilistic forecasts of recovery timelines.

Evaluation Metric	**Result**
C-index	0.6379

Table 8: DeepHit Evaluation Results

4.4 Model Limitations

The use of linear regression and logistic regression models for time-to-recovery analysis presents challenges. Linear regression is designed for continuous outcome variables but struggles with censoring (when an event has not occurred). This was overcame with feature engineering. Linear regression also assumes a normally distributed residuals which, as observed in the analysis above, did not hold true. Logistic regression can predict on the probability of an event occurring but provides a binary output. The prediction is heavily influence by the choice of probability threshold used.

The survival models used—XGBoost Cox, DeepSurv, Random Forest, Logistic Hazard, and DeepHit—each have limitations. XGBoost Cox assumes proportional hazards, which may not always hold. DeepSurv and DeepHit are prone to overfitting and require large datasets. Random Forest provides only relative risk rankings, not true probabilities. Logistic Hazard requires time discretization, which can reduce accuracy. DeepHit also showed unstable survival curves and poor calibration in this setting. These factors limit interpretability and real-world deployment.

4.5 Insights and Implications

During the process of feature selection, interested patterns emerged. The use of one-hot encoding on the 'climate region' feature meant each region was seperated as its own feature. RFE feature selection determined that 'NOAAClimRegion_Upper Midwest' and 'NOAAClimRegion_West' were among the top 10 out of 40 best feature to predict a power outage. In Layman's term, weather a solar power plant is located in the Upper Midwest or West region is one of the strongest indicators of the likelihood of a power outage.

5 Conclusion

This study set out to evaluate the most effective modeling approach for predicting recovery duration following power disruptions in photovoltaic systems, an essential step toward enabling preemptive maintenance, optimized resource allocation, and enhanced operational resilience. The modeling objective specifically required an approach capable of handling right-censored data, accommodating time-to-event characteristics, and adapting to nonlinear environmental influences.

While traditional models such as linear regression and logistic regression provided interpretable baselines, they were fundamentally limited by their assumptions. Linear regression does not account for censoring and struggled to capture the heavy skew in residuals, particularly for longer recovery times. Logistic regression, though useful for classifying outage likelihood, was constrained by its binary output and low recall, limiting its ability to reliably flag potential high-risk recovery periods. The multi-layer perceptron (MLP) improved flexibility and accuracy across short prediction horizons, but lacked the inherent structure to properly model survival functions or uncertainty over time. Several survival models were also explored, including DeepSurv, DeepHit, Logistic Hazard, and Random Survival Forests. While each showed varying levels of promise—particularly DeepSurv and Logistic Hazard—none outperformed **XGBoost Survival Cox**. XGBoost produced the highest C-index of 0.8426, demonstrating the best concordance between predicted and actual recovery orderings. Its ability to handle nonlinearity, manage censoring, and provide interpretable feature importance makes it the most robust and scalable option for real-world implementation. As such, tree-based and neural-network-based survival models, especially XGBoos, can offer practical and theoretically grounded solutions for recovery prediction, and should be prioritized in future deployment pipelines.

6 Future Work

The various models and techniques developed in this project can be used in conjunction with other tools and techniques to derive greater utility.

- GUI: A user-interactive tool can be develop that utilities one or multiple models to provide users a prediction of time-to-recovery duration based on user input features. The input can also be connected to weather prediction data sources to automate the input data.
- Solar Power Generation Modeling: Using the time-to-recovery models, the expected amount of solar energy produced per year for existing and proposed solar power projects can be es-

timated. This can help governments and private companies choose the ideal geographic location to build solar power plants to optimise ROI. The tool can also help financial institutions derive a more accurate cash flow discount rate for future cash flows from solar power sold when a solar power plant project is expected to repay debt using generated power.

7 Division of Responsibilities

Task	**Contributors**
Data Wrangling	Rutvij, Elia
Exploratory Data Analysis (EDA)	Rutvij, Elia
Models (Linear Regression, Logistic Regression, MLP)	Elia
Models (Random Forest Survival, XGBoost Cox, Logistic Hazard, DeepSurv, DeepHit)	Rutvij
Report Writing	Rutvij, Elia

References

[1] Mekhilef, S., Saidur, R., & Safari, A. (2011). A review of photovoltaic systems: Design, operation and maintenance. *Renewable and Sustainable Energy Reviews*, 15(9), 4470–4485. https://www.sciencedirect.com/science/article/abs/pii/S0038092X19305912

[2] Jordan, D. C., & Kurtz, S. R. (2012). *Photovoltaic Degradation Rates—An Analytical Review*. National Renewable Energy Laboratory (NREL). https://docs.nrel.gov/docs/fy12osti/51664.pdf

[3] Abate, A., Martinez, A., & Kamble, A. (2024). Current Challenges in Operation, Performance, and Maintenance of Photovoltaic Panels. *Energies*, 17(6), 1306. https://www.mdpi.com/1996-1073/17/6/1306

[4] IEA PVPS Task 13. (2014). *Review of Failures of Photovoltaic Modules*. International Energy Agency. https://iea-pvps.org/wp-content/uploads/2020/01/IEA-PVPS_T13-01_2014_Review_of_Failures_of_Photovoltaic_Modules_Final.pdf

[5] Reinders, A., & Veldhuis, A. (2021). Grid balancing challenges illustrated by two European examples: Interactions of electric grids, photovoltaic power generation, energy storage and

power generation forecasting. *Energy Reports*, 7, 1603–1614. https://www.sciencedirect.com/science/article/pii/S2352484721003644

[6] Jerez, S., Tobin, I., Vautard, R., Montávez, J. P., de la Cuadra, J., Manzanas, R., Herrera, S., & Gutiérrez, J. M. (2015). The impact of climate change on photovoltaic power generation in Europe. *Nature Communications*, 6, 10014. https://doi.org/10.1038/ncomms10014

[7] Kumar, P., Singh, A., & Zhao, L. (2025). Enhanced deep-learning-based forecasting of solar photovoltaic generation for critical weather conditions. *Clean Energy*, (Forthcoming). https://doi.org/10.1093/ce/zkad002

[8] Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1), 24. https://doi.org/10.1186/s12874-018-0482-1